

A GENOMIC INVESTIGATION OF BONOBO (*PAN PANISCUS*) AND
CHIMPANZEE (*PAN TROGLODYTES*) DIVERGENCE

by

COLIN M. BRAND

A DISSERTATION

Presented to the Department of Anthropology
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2021

DISSERTATION APPROVAL PAGE

Student: Colin M. Brand

Title: A Genomic Investigation of Bonobo (*Pan paniscus*) and Chimpanzee (*Pan troglodytes*) Divergence

The dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Anthropology by:

Frances J. White	Chair
Nelson Ting	Core Member
Larry R. Ulibarri	Core Member
Timothy H. Webster	Core Member
Andrew D. Kern	Institutional Representative

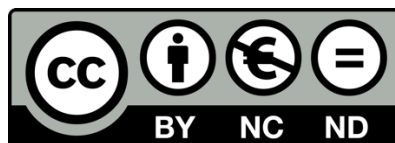
and

Andrew Karduna	Interim Vice Provost for Graduate Studies
----------------	---

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2021.

© 2021 Colin M. Brand
This work is listed under a Creative Commons
**Attribution-NonCommerical-NoDerivatives 4.0 International (CC BY-NC-ND
4.0) License.**



DISSERTATION ABSTRACT

Colin M. Brand

Doctor of Philosophy

Department of Anthropology

June 2021

Title: A Genomic Investigation of Bonobo (*Pan paniscus*) and Chimpanzee (*Pan troglodytes*) Divergence

Our closest living relatives are two species in the genus *Pan*: bonobos and chimpanzees. Chimpanzees are further divided into four subspecies. While there are a number of phenotypic similarities between bonobos and chimpanzees, there are also a number of differences, particularly in social behavior. Additionally, some phenotypes are highly variable among chimpanzees and within each of the five lineages. The absence of an extensive bonobo and chimpanzee fossil record means that genomic data provide the best window into their evolutionary past. This dissertation uses reassembled and remapped autosomal genomic data from all five *Pan* lineages to answer questions about adaptation and demography in the time following lineage divergence, ~ 1.88 Ma. We find evidence for positive selection in deep time within genes related to the brain, immune system, musculature, reproduction, and skeletal system. Most of these patterns are lineage specific and only one candidate gene was shared across all chimpanzee subspecies and another two were shared across all five taxa. We also observe that recent positive selection is largely the result of variable environmental conditions acting on standing genetic variation rather than *de novo* mutation in the four *Pan* lineages we could analyze. Finally, we consider previous models for the demographic history of these taxa. The best fit model includes a single introgression event from bonobos and central chimpanzees. We also find that the

common ancestor of chimpanzees is older than previously estimated. Our results collectively broaden our understanding of the complex evolutionary history of the *Pan* genus. The identification of positively selected genes both recently and earlier during lineage divergence as well as understanding the processes that drove recent positive selection in these taxa contributes to better estimating the timing of lineage-specific adaptations, reconstructing the behavior and genetics of the *Pan* common ancestor, and recognizing potential selective pressures for these adaptations during key time periods in chimpanzee evolution. Estimates of demographic parameters can also offer further insight into adaptation and other evolutionary processes in these species and more broadly. This dissertation includes previously unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Colin M. Brand

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
Miami University, Oxford, OH, USA

DEGREES AWARDED:

Doctor of Philosophy, Anthropology, 2021, University of Oregon
Master of Science, Anthropology, 2015, University of Oregon
Bachelor of Arts, Anthropology, Botany, Environmental Science, Zoology,
2014, Miami University

AREAS OF SPECIAL INTEREST:

Behavioral Ecology
Biological Anthropology
Evolutionary Anthropology
Molecular Anthropology
Population Genetics

PROFESSIONAL EXPERIENCE:

Sample Collector, COVID Monitoring and Assessment Program, University of
Oregon, Eugene, OR, 2020-2021

Tutor, Lane Community College, Eugene, OR, 2019-2021

Instructor of Record, Department of Anthropology, University of Oregon,
Eugene, OR, 2016-2021

Graduate Teaching Fellow, Department of Anthropology, University of
Oregon, Eugene, OR, 2014-2021

Animal Care Intern, Cincinnati Zoo and Botanical Garden, Cincinnati, OH,
2014

GRANTS, AWARDS, AND HONORS:

William S. Pollitzer Student Travel Award, American Association of Physical
Anthropologists, 2021

Health Education Award, Anthropology, University of Oregon, 2020

Risa Palm Graduate Fellowship, University of Oregon, 2018

Young Explorer's Grant, National Geographic Society, 2017

Pauline Wollenberg Juda Memorial Endowment Fund Award, Anthropology,
University of Oregon, 2017

Malcom McFee Award, Anthropology, University of Oregon, 2016

International Research Award, University of Oregon Global Studies Institute,
2016

Undergraduate Presentation Award, Office for the Advancement of
Scholarship and Research, Miami University, 2014

Senior Service Leadership Award, Miami University, 2014

Senior Service Award, Anthropology, Miami University, 2014

President's Distinguished Service Award, Miami University, 2014

Employee Service Leadership Award, Miami University, 2014

Best Paper in Archaeology Award, Anthropology, Miami University, 2014

Rebecca Jeanne Andrew Memorial Award, Miami University Department of
Anthropology, 2013.

Provost's Student Academic Achievement Award, Miami University, 2013

Dean's Scholar Award, Miami University College of Arts and Science, 2013.

Cambridge Junior Visiting Fellows Fund, College of Arts and Science, Miami
University, 2013

Undergraduate Presentation Award, Office for the Advancement of
Scholarship and Research, Miami University, 2012

Rebecca Jeanne Andrew Memorial Award, Miami University Department of
Anthropology, 2012.

Employee Service Leadership Award, Miami University, 2012

PUBLICATIONS:

- White FJ, Brand CM, Hickmott AJ, Minton IR. 2020. Sex differences in bonobo (*Pan paniscus*) terrestriality: implications for human evolution. *J Anthropol Sci.* 98:5-14.
- Gartland KN, Brand CM, Ulibarri LR, White FJ. 2020. Variation in adult male-juvenile affiliative behavior in Japanese macaques (*Macaca fuscata*). *Fol Primatol.* 91:610-621.
- Brand CM, Johnson MB, Parker LD, Maldonado JE, Korte L, Vanthomme H, Alonso A, Ruiz-Lopez MJ, Wells CP, Ting N. 2020. Abundance, density, and social structure of African forest elephants (*Loxodonta cyclotis*) in a human-modified landscape in southwestern Gabon. *PLoS ONE.* 15:e0231832.
- Brand CM, Marchant LF. 2019. Social hair plucking is a grooming convention in a group of captive bonobos (*Pan paniscus*). *Primates.* 60:487-491.
- Wakefield ML, Hickmott AJ, Brand CM, Takaoka IY, Meador LM, Waller MT, White FJ. 2019. New observations of meat eating and sharing in wild bonobos (*Pan paniscus*) at Iyema, Lomako Forest Reserve, Democratic Republic of Congo. *Fol Primatol.* 90:179-189.
- Boose KJ, White FJ, Brand CM, Meinelt A, Snodgrass JJ. 2018. Infant handling in bonobos (*Pan paniscus*): exploring functional hypotheses and the relationship to oxytocin. *Phys Behav.* 193:154-166.
- Brand CM, Marchant LF. 2018. Prevalence and characteristics of hair plucking in captive bonobos (*Pan paniscus*) in North American zoos. *Am J Primatol.* 80:e22751.
- Brand CM, Marchant LF, Boose KJ, Rood TM, White FJ, Meinelt A. 2017. Laterality of grooming and tool use in a group of captive bonobos (*Pan paniscus*). *Fol Primatol.* 88:210-222.
- Brand CM, Boose KJ, Squires EC, Marchant LF, White FJ, Meinelt A, Snodgrass JJ. 2016. Hair plucking, stress, and urinary cortisol among captive bonobos (*Pan paniscus*). *Zoo Biol.* 35:415-422.
- Brand CM, White FJ, Wakefield ML, Waller MT, Ruiz-Lopez MJ, Ting N. 2016. Initiation of genetic demographic monitoring of bonobos (*Pan paniscus*) at Iyema, Lomako Forest, DRC. *Primate Conservation.* 30:103-111.
- Brand CM, Marchant LF. 2015. Hair plucking in captive bonobos (*Pan paniscus*). *App Anim Behav Sci.* 171:192-196.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the incredible love and support of a large social network. First, I thank my dissertation committee: Frances White, Nelson Ting, Larry Ulibarri, Tim Webster, and Andy Kern. I am so grateful for my incredible colleagues and chosen family including Monya Anderson, Klaree Boose, Diana Christie, Elisabeth Goldman, Alex Hickmott, Cam Johnson, Josh Schrock, Evan Simons, Noah Simons, Jessica Stone, Nicky Ulrich, and Hannah Wellman. I thank the UO Anthropology faculty who have taught me so much over the past seven years both in and out of the classroom including Diana Baxter, Alison Carter, Steve Frost, Ana Lara, Madonna Moss, Carol Silverman, Uli Streicher, Michelle Sugiyama, and Larry Sugiyama. I am proud to call myself an anthropologist and am appreciative of the Miami University faculty and staff who taught me so much: Jeb Card, Kathy Erbaugh, Cameron Hay-Rollins, Neringa Klumbyte, Leighton Peterson, Mark Peterson, Scott Suarez. The amazing four years I spent in Oxford also resulted in many friendships for which I am so grateful including Andrea Blackburn, Alex Cowper, Amanda Friend, Alex Intorcio, Jacob Negrey, Jordan Martin, Rob O'Malley, and Ashley Skolits. My experiences in the Division of Biological Anthropology at the University of Cambridge were fundamental to my undergraduate training and created lifelong memories. I thank Bill McGrew, Lia Betti, Jake Dunn, Leslie Knapp, Frank Marlowe, Alex Piel, Fiona Stewart, and Peter Walsh. I am also very grateful for Andrea Blackburn, Julianne Joswiak, Tina Lasisi, Dan Schofield, and many others for their friendship and helping make Cambridge feel like home. I feel so lucky to have an incredible network of colleagues beyond UO who make my life so joyful. These include the PEGL crew, especially Hazel Byrne, Tina Lasisi, Liz Tapanes, Andrew Zamora as well as Alexander Baxter, Melanie Beasley, Joel Brown,

Morgan Chaney, Ashley Edes, Kelsey Ellis, Drew Enigk, Stephanie Fox, Brett Frye, Katie Gerstner, Luke Larter, Kaedan O'Brien, Ian Takaoka, Monica Wakefield, and Shasta Webb. Also thank you to Alan Rogers for his incredible mentorship and support. Many of these friendships stem from the various professional societies and committees, of which I have been fortunate to be a part including the ASP Student Committee, the UO Graduate Student Association, the UO Committee on Courses. I am very thankful for mentorship and support over the years from Tori Byington. Many of the ideas presented here stem from my wonderful time in the field. I thank our amazing field staff: Abdulay, Augustin, Beken, Bellevie, Christian, Dipon, Gedeon, Isaac, Mathieu, Papa Siri, and Teddy, as well as Alfred Simba, Hugues Akpona, Jef Dupain, Christelle Ilanga, Moïse Amisi Luenga, the African Wildlife Foundation, the Institut Congolais pour la Conservation de la Nature, and many others. This research was generously funded by National Geographic and the UO Global Studies Institute. The work presented in this dissertation would not be possible with the assistance of Mark Allen, Mike Coleman, and Rob Yelle at UO RACS as well as Talapas itself and the Utah CHPC. I thank my incredible work family in the TRiO department at Lane Community College including Lynn, Gwen, Jane, Rose, Shijo, Giulia, Alex, Bailey, Dru, Dustin, Aimee, and countless students for their friendship and support. Thank you also to my work family at the UO COVID Monitoring and Assessment Program including Hannah, Jaimyn, Katelyn, Josh, Matthew, Sam, Tanner, Katie, Clara, and Shuhao. All of you helped make this difficult year so much more bearable. I am privileged have long lasting friendships and am especially grateful for Megan Jackson, Katie Tela, and many others. Finally, I thank my amazing family who has supported my every endeavor from the very beginning.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
Brief Overview of Bonobos and Chimpanzees.....	1
Genetic and Genomic Perspectives on <i>Pan</i> Evolutionary History	15
Project Overview	22
II. ADAPTATION DURING DIVERGENCE IN BONOBO (<i>PAN PANISCUS</i>) AND CHIMPANZEES (<i>PAN TROGLODYTES</i>)	25
Introduction.....	25
Methods.....	30
Results.....	36
Discussion.....	43
III. SOFT SWEEPS PREDOMINATE RECENT POSITIVE SELECTION IN BONOBO (<i>PAN PANISCUS</i>) AND CHIMPANZEES (<i>PAN TROGLODYTES</i>)	52
Introduction.....	52
Methods.....	57
Results.....	66
Discussion.....	70
IV. ESTIMATION OF <i>PAN</i> DEMOGRAPHY FROM SITE PATTERNS.....	80
Introduction.....	80
Methods.....	84

Chapter	Page
Results.....	92
Discussion.....	96
V. CONCLUSION.....	100
REFERENCES CITED.....	104

LIST OF FIGURES

Figure	Page
1. Unique and shared candidate genes for positive selection.....	40
2. Unique and shared hard selective sweep windows	69
3. Unique and shared soft selective sweep windows	69
4. Demographic model and introgression events considered.....	89
5. Observed site patterns	93
6. Parameter estimate bias.....	95

LIST OF TABLES

Table	Page
1. Number of MK candidate genes under selection and tested.....	38
2. Candidate genes under positive selection	39
3. Candidate exons under positive selection.....	42
4. Candidate genes under positive selection via SnIPRE	42
5. Number and proportion of sweep, linked, and neutral windows	68
6. Demography parameter estimates.....	94

CHAPTER I

INTRODUCTION

Brief Overview of Bonobos and Chimpanzees

The genus *Pan* consists of two species: bonobos and chimpanzees. Additionally, there are four subspecies of chimpanzee: central (*Pan troglodytes troglodytes*), eastern (*P. t. schweinfurthii*), Nigeria-Cameroon (*P. t. ellioti*, formerly *P. t. vellerosus*), and western (*P. t. verus*). These four subspecies are supported by both morphological (Pilbrow 2006) and genetic evidence although Groves (2005) has argued for a fifth chimpanzee subspecies (*P. t. marungensis*) based on skull morphology. Such an additional subspecies would include chimpanzees living in the southern portion of *P. t. schweinfurthii* range. No subspecies for bonobos have been formerly recognized, although variation in craniodental morphology has been described (Pilbrow and Groves 2013) and there is some genetic evidence for population structure (Kawamoto et al. 2013, but see Eriksson et al. 2004). All currently recognized lineages are allopatric and all are presently separated by rivers, except for *P. t. ellioti* and *P. t. verus*, which are separated by the Dahomey Gap. Bonobo and chimpanzee populations are currently trending downward and both species are currently listed as endangered by the IUCN (Fruth et al. 2016; Humle et al. 2016). The largest recognized threats include habitat loss, disease transmission, and hunting (Fruth et al. 2016; Humle et al. 2016).

Scientists have recognized chimpanzees as a species for centuries. The scientific name for chimpanzees, *Pan troglodytes*, was coined by Johann Friedrich Blumenbach, which is variably cited as Blumenbach 1775, Blumenbach 1779, and Blumenbach 1799. Bonobos were first classified as a subspecies of chimpanzee by

Ernst Schwarz based on a skin and skull collected near Befale, DRC in 1927 (Schwarz 1929). He coined the new subspecies, *Pan satyrus paniscus*, believing it to be a dwarf of the right bank apes (chimpanzees), *Pan satyrus satyrus* (Schwarz 1929). Indeed, the subspecific name “paniscus” translates to “a little Pan”, in reference to the Greek god, reflecting the perceived size difference between bonobos and chimpanzees. Two specimens were sent to the American Museum of Natural History in December 1930 (Thompson 2001) and in 1933 American anatomist Harold Coolidge elevated the taxon to species status, *Pan paniscus*, based on his analysis of an adult female (Coolidge 1933). Coolidge had, in fact, compared this specimen to a central chimpanzee (Thompson 2001), which is the largest subspecies of chimpanzee (Smith and Jungers 1997) (see below). Later, Jungers and Susman (1984) would describe Coolidge’s specimen as the smallest adult bonobo they had ever encountered. This belief in a species size difference resulted in the use of the common names “dwarf chimpanzee” and “pygmy chimpanzee”, which were routinely used until the 1980s and 1990s.

The uniqueness of bonobos and the differences between bonobos and chimpanzees were described before the taxonomic difference was officially designated. Anton Portielje, a Dutch naturalist, wondered if a popular ape housed at the Amsterdam Zoo, named Mafuca, was a new species of ape (de Waal and Lanting 1997). Years later, the individual’s stuffed remains would be recognized as a bonobo. In August 1923, Robert Yerkes purchased two young apes from a dealer in New York (Susman 1984). He named the male “Prince Chim” and the female “Panzee” and would later reflect on the differences between the two in his book *Almost Human* (Yerkes 1925). Photographs of Prince Chim and Panzee would later confirm that they were a bonobo and chimpanzee, respectively. Despite early speculation of species’

differences, it would be decades until the first systematic study comparing the two was carried out.

In the aftermath of World War II, Tratz and Heck (1954) published their findings collected before the war at the Hellabrunn Zoo in Munich. This publication also represents the first use of the term “bonobo” although this term was proposed as a new genus distinct from chimpanzees. This study identified eight differences between bonobos and chimpanzees and included comparisons of these taxa beyond morphology. Not long after this study, the first scientific studies of wild chimpanzees began in Tanzania. In 1960 Jane Goodall began research at Gombe Stream Reserve and Toshisada Nishida began research ~ 200 km south of Gombe at Mahale Mountains National Park (Goodall 1986; Nishida 2011). Fieldwork on bonobos followed just over a decade later. Takayoshi Kano surveyed bonobos across the Democratic Republic of Congo in 1973 establishing two sites: Wamba and Yalosidi (Kano 1992). That same year Noel and Alison Badrian established a field site at Lomako (Susman 1984). The late 1970s saw the start of research on wild Western chimpanzees at Bossou, Guinea (Matsuzawa and Humle 2011) and Tai, Ivory Coast (Boesch and Boesch-Achermann 2000). In the following decades, many other long-term and more recent research sites have been established for chimpanzees, totaling > 43 to date (van Leeuwen et al. 2020). Yet, our understanding of chimpanzees is heavily biased toward eastern and western chimpanzees, with only a few sites at which central chimpanzees are studied and no long-term data are available for Nigeria-Cameroon chimpanzees. Conflict in the 1990s and 2000s in central Africa impacted studies of wild bonobos for over a decade (Furuichi et al. 2012; Waller and White 2016). In the time since the conflict’s end, research has resumed at Wamba and intermittently at Lomako. Additionally, new sites have been established or a long-

term research presence has resumed: Iyondji (Sakamaki et al. 2012), Lonoa/Kokolopori (Surbeck, Coxe, et al. 2017), LuiKotale (Hohmann and Fruth 2003a), and Lake Tumba/Malebo (Inogwabini et al. 2007; Serckx et al. 2014). While some have argued that bonobos should no longer be considered “scientifically endangered” (Hare and Yamamoto 2015), there likely remains a gulf in our understanding of bonobos and chimpanzees.

Central to understanding the evolution of *Pan* is assessing similarities and differences in their behavior and biology. There are several reviews and volumes that explicitly cover such similarities and differences (Stanford 1998; Boesch et al. 2001; Stumpf 2011; Gruber and Clay 2016) and other volumes that focus on either bonobos or chimpanzees. Recapitulating an exhaustive review is beyond the scope of this introduction. Therefore, I describe the more relevant similarities and differences in the paragraphs below, focusing on morphology, ecology, social behavior, and reproduction.

The common names for the bonobo clearly reflect that earlier analyses were performed on relatively small adult individuals and the view that bonobos are specialized dwarf chimpanzees (Johnson et al. 1981) or exhibit paedomorphic characteristics (Shea 1983). Such species dimorphism is not reflected in mean body mass where bonobos and chimpanzees overlap, with male weight averaging between 42.7 to 45 kg and 42.7 and 59.7 kg, respectively, and female weight averaging 33.2 to 34.3 kg and 33.7 to 45.8 kg, respectively (Smith and Jungers 1997; Zihlman and Bolter 2015). The larger range of body mass in chimpanzees highlights the considerable variation across the subspecies in which both eastern and western chimpanzee females and males are much smaller than central chimpanzees (Smith and

Jungers 1997). Such data are lacking for Nigeria-Cameroon chimpanzees although it appears males weigh < 70 kg (Hof and Sommer 2010).

Multiple post-cranial skeletal differences have been described. These include differences in clavicles, scapulae, and pelvis as well as the humerus/femur and femoral head/length ratios (Zihlman and Cramer 1978). However, long bone lengths and talar breadths are similar between species (Zihlman and Cramer 1978). More recently, Turley and Frost (2014) described that the appositional articular morphology of the talo-crural joint in adult bonobos was different to that of chimpanzees. Bonobos are more similar to highly arboreal hylobatids whereas chimpanzees exhibited a more terrestrial pattern. This morphology is consistent with locomotor patterns for these taxa. While females across all African apes exhibit a higher degree of terrestriality, bonobos are more arboreal than chimpanzees or gorillas (Doran 1993).

Other aspects of morphology are more difficult to assess. For example, there remains considerable debate regarding whether or not the bonobo cranium is paeodomorphic relative to chimpanzees. While some have maintained that heterochrony explains skull shape differences (Shea 1983), other analyses have yielded partial support (Lieberman et al. 2007) or rejected this hypothesis altogether (Mitteroecker et al. 2005; Simons and Frost 2020).

In addition to osteological differences, bonobos and chimpanzees differ in their overall appearance. Bonobos have hair that parts down the middle of the head whereas chimpanzees do not (Stumpf 2011). The hair around their cheeks is also quite long (Kano 1992). The lips of bonobos are depigmented and pink and they are born with dark faces whereas most chimpanzees are born with lighter faces (Kano 1992; Stumpf 2011). While both bonobos and chimpanzees are born with a white tail tuft only bonobos maintain this trait into adulthood (Kano 1992; Stumpf 2011). In bonobo

females, the vulva is more anterior than in chimpanzees, which may be related to female-female sexual behavior (see below) (de Waal and Lanting 1997).

Both species live in multi-male/multi-female groups known as communities and exhibit a fission-fusion social structure (Goodall 1986; Kano 1992). Community sizes vary greatly across both species, averaging between 20 and 40 individuals in most bonobo communities (Stumpf 2011) and ranging from ~ 20 at Bossou (Sugiyama and Fujita 2011) to > 200 individuals at Ngogo, Uganda (Sandel and Watts 2021) in chimpanzees. While there is considerable intraspecific variation and complications with how to quantify party size, bonobo parties appear to be larger, on average, than those of chimpanzees (Furuichi 2009). Unlike many other primates, apes do not exhibit female residence and both bonobos and chimpanzees are typically male philopatric where females emigrate upon sexual maturation (Goodall 1986; Kano 1992). There are some exceptions, such that only ~ 50% of females at Gombe emigrate (Pusey et al. 1997) and both males and females are thought to emigrate at Bossou (Sugiyama 1999). Additionally, male transfer has been documented in the Eyengo community of bonobos at Lomako (Hohmann 2001). Immigration itself appears to be different between bonobos and chimpanzees. Among chimpanzees, newly immigrant females form bonds with males as resident females are largely intolerant of new females (Kahlenberg, Thompson, et al. 2008; Kahlenberg, Emery Thompson, et al. 2008). This starkly contrasts with bonobos, where new females form bonds with resident females (Furuichi 1989; Idani 1991).

Some of the most striking differences between the two *Pan* species are those related to intracommunity and intercommunity social relationships. Chimpanzees exhibit a primate-typical pattern of male dominance (Goodall 1986; Boesch and Boesch-Achermann 2000). Additionally, the strongest relationships in chimpanzee

communities are between adult males. Given that both *Pan* species are male philopatric, this observation is consistent with kin selection theory (Hamilton 1964). Despite this, affiliation between kin among males is complex. For example, Nishida (1983) noted that the dynamic nature of alliances between chimpanzee males contradicts the prediction from kin selection theory. Further, male chimpanzees at Ngogo, Uganda preferentially associate with maternal kin rather than paternal kin (Langergraber et al. 2007; Mitani 2009).

In contrast, bonobos are not male dominant and the strongest bonds do not occur between adult males. While some authors have described bonobos as female-dominant (Parish 1994; Parish 1996; Parish et al. 2000), others have described their dominance patterns as more equivocal (Furuichi 1989; Kano 1992; White 1996) because while the highest ranking individual in bonobo communities can be female, not all adult females outrank all adult males. It is possible that captive conditions facilitate full female dominance in bonobos as those studies that describe female dominance were of zoo-housed individuals. However, other zoo-housed bonobo groups do not show full female dominance (Paoli et al. 2006; Stevens et al. 2010; Brand and Marchant 2019). Another important perspective is the consideration of female feeding priority rather than only social dominance. For example, adult and subadult males at Lomako were socially dominant to females; however, females had feeding priority (White and Wood 2007). Unlike chimpanzees, the strongest bonds in bonobos occur among females and between females and males (White 1988; White and Burgman 1990; Kano 1992; Parish 1996). Female-female bonds are dynamic such that grooming relationships are stable over time, while proximity and genito-genital (GG) rubbing, a female-female sexual behavior, preferences are more flexible

(Moscovice et al. 2017). This behavioral coordination may function to enable cooperation with a wider range of individuals (Moscovice et al. 2017).

Aggression is common in primates and other animals; however, lethal aggression is rare. When lethal aggression occurs, it is most often in the form of male infanticide although female infanticide does occur in primates, including chimpanzees (Townsend et al. 2007; Pusey et al. 2008). Lethal aggression directed towards non-infants does occur in a number of primate species in addition to chimpanzees including capuchins (Scarry and Tujague 2012), spider monkeys (Campbell 2006; Valero et al. 2006), miqui (Talebi et al. 2009), and orangutans (Marzec et al. 2016). Intracommunity lethal aggression has been documented in multiple eastern chimpanzee communities (Fawcett and Muhumuza 2000; Watts 2004; Kaburu et al. 2013; Sandel and Watts 2021), at least one central chimpanzee community (Boesch et al. 2007) and one western chimpanzee community (Pruetz et al. 2017). This behavior may function to reduce mating competition (Watts 2004). Lethal intercommunity aggression also occurs (Wilson and Wrangham 2003) and may be explained by the imbalance of power hypothesis (Wrangham 1999). In contrast, neither lethal aggression nor infanticide have ever been directly observed in bonobos (Stanford 1998; Wrangham 1999), although one potential case has been reported (Hohmann and Fruth 2011, but see White 2012). Lethal aggression in *Pan* appears to be adaptive and is not the result of anthropogenic effects (M.L. Wilson et al. 2014). While the presence/absence of lethal aggression (and infanticide) differentiates bonobos and chimpanzees, there is considerable variation in the rate of this behavior not only across subspecies but within subspecies as well (M.L. Wilson et al. 2014). The absence of lethal aggression and infanticide in bonobos has been suggested to stem from the frequency of copulation and an alleged extended period of receptivity (de

Waal and Lanting 1997) or the high status of females and initiative in various social behaviors (Furuichi 2011).

Bonobos and chimpanzees exhibit such notable differences in sexual behavior that some of these were described as early as the 1950s (Tratz and Heck 1954). The females of both species exhibit a sexual swelling (Stumpf 2011). Chimpanzees have sexual cycle that averages 35 days, estrus lasts 10 - 15 days, and ovulation occurs near the end of maximal swelling (Stumpf and Boesch 2005). Bonobos have a slightly longer sexual cycle (~ 40 days) than chimpanzees and ovulation does not always occur during maximum tumescence (Reichert et al. 2002). Both species copulate dorsoventrally; however, ventroventral mating is common in bonobos (Thompson-Handler et al. 1984; Kano 1992; de Waal and Lanting 1997). Bonobos are also well known for socio-sexual behavior, such that sex may serve non-reproductive functions (Hashimoto 1997; de Waal and Lanting 1997). Males may engage in rump-rubbing (Kano 1992) whereas GG rubbing is common among females where two individuals rub their sexual swellings together (Thompson-Handler et al. 1984; Kano 1992; Hohmann and Fruth 2000). Chimpanzees sometimes engage in socio-sexual behavior and GG rubbing; however, it is much less common than what is observed in bonobos (Anestis 2004; Sandel and Reddy 2021).

Sexual strategies may also differ between species. Tutin (1979) described four types of mating among chimpanzees: opportunistic, consortship, possessive, and extragroup. Both consortship and possessive mating seem rare in bonobos (Kano 1992). While dominance rank is often elicited as a predictor for mating success there is mixed evidence in primates (Fedigan 1983). This holds true in chimpanzees. Such a relationship is sometimes absent (Tutin 1979; Boesch and Boesch-Achermann 2000) or present (Goodall 1986). High rank in chimpanzees can also result in higher

reproductive success (Vigilant et al. 2001). Dominance is also positively related to both mating and reproductive success in bonobos (Gerloff et al. 1999). One notable species difference lies in the relationship between dominant males and females such that high ranking male bonobos tend to be the sons of high ranking females (Kano 1992; Surbeck et al. 2011). This mother-son relationship only affects reproductive success in bonobos and not chimpanzees (Surbeck et al. 2019). Finally, it is worth noting that aggression in the context of mating in both bonobos and chimpanzees may constrain female choice. Among eastern chimpanzees, sexual coercion is a well described male reproductive strategy (Muller et al. 2007; Muller et al. 2011; Feldblum et al. 2014) and can result in higher reproductive success (Feldblum et al. 2014). Male aggression also occurs in bonobo mating (Kano 1992; Surbeck et al. 2011), of which some events are coercive (White and Wood 2007).

Despite the complexity of differences between bonobos and chimpanzees that are further complicated by intraspecific variation in both species, multiple models have been suggested to explain species' differences. Below, I describe some of these models and consider evidence that supports and is counter to each model.

One early ecological model proposed to explain differences in female sociality between bonobos and chimpanzees has been coined the terrestrial herbaceous vegetation or THV hypothesis (Wrangham 1986). THV is both ubiquitous and non-seasonal in both bonobo and chimpanzee habitat (Yamakoshi 2004). However, chimpanzees can occur sympatrically with gorillas whereas bonobos do not. Thus, the THV hypothesis posits that the feeding competition female chimpanzees experience from gorillas, who consume large amounts of THV, prevents the formation of larger parties. As bonobos are not subjected to such competition, THV is used to compensate during periods of fruit scarcity and maintain larger parties (Wrangham

1986). In this paper, Wrangham (1986) reports THV represents 7% of the monthly food intake for chimpanzees at Gombe and 33% of monthly food intake for bonobos at Wamba. However, the difference is less pronounced when considering other *Pan* sites. For example, the Lomako bonobos spend 2% of total feeding time consuming THV (White 1992). Yamakoshi (2004) notes that while there are no data from chimpanzees sympatric with gorillas, chimpanzees at sites outside the range of gorillas consume THV that ranges from 3% at Taï (Boesch 1996) to 17% at Kibale (Wrangham et al. 1996). Averaging across months may obscure seasonal patterns; however, which is a prediction from the model. Chimpanzees at Kahuzi-Biega, Lopé, and Ndoki are sympatric with gorillas and fibrous content in their feces is greater during non-fruiting periods (Tutin et al. 1991; Kuroda et al. 1996; Basabose 2002). At non-gorilla sites, there is mixed evidence for seasonality. Fibrous content in Kibale chimpanzee feces was higher during fruit scarcity (Wrangham et al. 1991) but party size also decreased during these periods (Wrangham et al. 1992). At Bossou, party size is stable regardless of fruit scarcity and THV is consumed consistently (Yamakoshi 1998). Furuichi et al. (2001) also reported that THV consumption was unrelated to fruit scarcity at Kalinzu. Among the Lomako bonobos, THV consumption is unrelated to fruit scarcity nor does it exhibit a seasonal pattern (Malenky and Wrangham 1994; White 1998). Rates of THV consumption could be driven simply by its density. Malenky et al. (1994) compared THV density at Kibale, Lomako, and Ndoki. While there was a significant difference between Kibale and Lomako, neither of the other pairs of sites were significantly different in THV density. A further complication lies in possible nutritional differences across sites. THV at Kibale was reported to have lower protein compared to Lomako, suggesting THV may act as a fruit substitute in the Kibale chimpanzees, which is consistent with

its increased consumption during periods of low fruit availability (Malenky and Wrangham 1994). Collectively, these results do not support the THV hypothesis. Indeed, Wrangham et al. (1996) proposed a revised hypothesis in which they divide THV into low and high quality, L-THV and H-THV respectively. These authors argue that H-THV is protein-rich, has relatively high nutritional value, is more preferred than “typical fig fruits”, and occurs at low density prompting consumption upon encounter. H-THV are said to occur at Kahuzi, Lomako, Wamba while L-THV occurs at Kibale (Wrangham et al. 1996). As such, the occurrence of H-THV at Kahuzi should result in increased gregariousness among those female chimpanzees compared to Kibale, however, there is no present evidence for such a difference. In addition to the issues with this revised hypothesis, there are few data presently available to evaluate its predictions (Yamakoshi 2004).

Another potential model for *Pan* differences considers the observations of female coalitionary behavior reported in both captive and wild bonobos (Kano 1992; White and Wood 2007; Furuichi 2011; Tokuyama and Furuichi 2016). While female coalitions are generally thought to function in the context of female-female competition (Sterck et al. 1997), they may also deter males from aggressing against females due to the threat of female coalitionary counteraggression. However, female coalitions have also been reported in chimpanzees at Budongo and Tai (Boesch and Boesch-Achermann 2000). Thus, Tokuyama and Furuichi (2016) suggest that this may be a shared *Pan* trait. However, the context of coalitionary behavior may be important. Given the male deference is evident in feeding contexts, female coalitionary action may function to gain feeding priority through male deference (White and Wood 2007).

Recently, similarity between some bonobo phenotypes and those of other domesticated mammals, largely canids, prompted the introduction of the self-domestication hypothesis (SDH) (Wrangham and Pilbeam 2001; Hare et al. 2012). This model invokes sexual selection theory and argues that female bonobos selected for less aggressive males, producing the bonobo phenotype. Despite the simplicity of the argument, the model is complex and contains a multitude of predictions for various phenotypes including behavior, morphology, and psychology. While some bonobo morphological traits, such as depigmented lips and white tail tufts (Stumpf 2011), support the SDH, other morphological predictions are less well supported. The most recent studies of bonobo crania find that they are, at best, only partially paedomorphic or not paedomorphic at all (Mitteroecker et al. 2005; Lieberman et al. 2007; Simons and Frost 2020). This prediction from the model also supposes dogs are paedomorphic wolves, which is not supported by three-dimensional geometric morphometric analyses of dog and wolf crania (Drake 2011). Additionally, few morphological characteristics are even shared across domesticated mammals, aside from canids (Sánchez-Villagra et al. 2016).

Similar to morphology, the behavioral and psychological data from *Pan* offers mixed support for the SDH. Bonobos may exhibit some delay in psychological development (Wobber et al. 2010). Yet, there is conflicting evidence based on behavioral experiments for the core of the SDH: tolerance. In one set of experiments, bonobos were more tolerant and more cooperative when food was monopolizable (Hare et al. 2007). Bonobos have also been observed to share food, even with unfamiliar conspecifics (Hare and Kwtuenda 2010; Tan and Hare 2013). Yet, one attempt to replicate Hare and Kwtuenda's finding in bonobos found no evidence of this behavior (Bullinger et al. 2013). These authors speculate that this discrepancy

may be related to the rearing of each study's subjects as Hare and Kwetuenda (2010) studied sanctuary-housed bonobos whereas Bullinger et al. (2013) studied zoo-housed bonobos. Bonobos have also been described as equally or less tolerant than chimpanzees (Jaeggi et al. 2010). In one study, chimpanzees not only shared more frequently but also more actively (Jaeggi et al. 2010). A follow up to this study highlighted how bonobos received more aggression and were less successful at acquiring food from conspecifics than chimpanzees (Jaeggi et al. 2013). Further, social tolerance, as measured by the proportion of a group at a resource (an artificial termite mound) or inside a resource zone (scattered food), was found to be lower in bonobos when compared to chimpanzees (Cronin et al. 2015) and both chimpanzees and gorillas (Boose et al. 2013).

There are notable species differences with respect to play such that adult bonobos play more frequently than adult chimpanzees and rough play is common in bonobos, which may reflect higher tolerance (Palagi 2006). Bonobo socio-sexual behavior lends support to the SDH; however, the assertion that males compete less intensely than chimpanzees is not immediately obvious given the high reproductive skew in bonobos (Surbeck, Langergraber, et al. 2017; Ishizuka et al. 2018), aggression in mating contexts (White and Wood 2007; Surbeck et al. 2011), and mate defense behavior during intergroup encounters or IGEs (Tokuyama et al. 2019).

Further evaluation of these models requires considerable behavioral, ecological, morphological, and physiological data that is both cross-sectional and longitudinal. Further insight could be gained from data on fossil panins as is widely used for many other taxonomic groups. However, to date, the *Pan* fossil record is limited to two central incisors and a first molar, recovered from the Kapthurin Formation, Kenya (McBrearty and Jablonski 2005). A second molar was reported but

is not discussed in detail. These fossils are likely from the same individual and are estimated to be near 545 ka in age (McBrearty and Jablonski 2005). Overlap in the dental variation in bonobos and chimpanzees does not permit any insight on whether this individual was more bonobo or chimpanzee-like. However, these fossils do highlight that *Pan* lived beyond its current range around 545 ka.

In the absence of such a fossil record, we must turn to genetic and genomic data to gain additional insight on the evolutionary history of the genus *Pan*. The section below provides a review of the previous research on this topic.

Genetic and Genomic Perspectives on *Pan* Evolutionary History

The phylogenetic proximity of *Pan* to humans meant that these taxa were the focus of some of the earliest studies on non-human primate genetics. The first draft of the chimpanzee genome became available in 2005 (The Chimpanzee Sequencing and Analysis Consortium 2005) and the bonobo genome in 2012 (Prüfer et al. 2012). Following the publication of these genomes, additional genomes were sequenced across all great ape species as part of the Great Ape Genome Project (GAGP) (Prado-Martinez et al. 2013). To date, this remains the largest genomic dataset for non-human hominids. While some additional data have been subsequently generated, these data form the foundation from which the majority of our understanding of great ape genomes stems, including this dissertation. In the following paragraphs, I briefly review this body of literature focusing solely on *Pan*.

The first analysis for admixture using whole genome sequences calculated D statistics from two western, seven eastern, and seven central chimpanzees and three bonobos and found no evidence of interspecies gene flow (Prüfer et al. 2012). However, de Manuel et al. (2016) used a larger sample size from the GAGP and

found evidence of gene flow within chimpanzee lineages and at least two episodes of introgression from bonobos into central chimpanzees. Further evidence of these events come from an analysis examining the potential adaptiveness of the putatively introgressed regions (Nye et al. 2018). More recently, introgression from an extinct *Pan* species into bonobos was reported (Kuhlwilm et al. 2019).

Central to understanding the evolution of bonobos and chimpanzees, is assessing the potential nature of mutations that have occurred in these lineages over time following their divergence. This range of possible effects is captured by the distribution of fitness effects or DFE (Eyre-Walker and Keightley 2007). Two important parameters for DFE are the shape parameter (b) and the mean selection coefficient for deleterious mutations (S_d). One recent analysis of great ape genomes, including bonobos and chimpanzees, found that the model with a shared b across all species and a lineage-specific S_d fit the genomic data better than other models (Castellano et al. 2019). This suggests a strong effect of effective population size, or N_e , on purifying selection, which is consistent with nearly neutral theory (Ohta 1992). Another analysis found that lineages with the smallest historical N_e had low levels of genetic diversity, larger numbers of deleterious homozygous alleles, and an increased proportion of deleterious variants at low frequency (Han et al. 2019). However, the efficacy of purifying may be less constrained given higher deleteriousness. Analysis of loss of function variants indicated that the number of variants was related to N_e but the number of variants was more equal across lineages with different N_e for variants that had drastic phenotypic effects (de Valles-Ibáñez et al. 2016).

Genomic data from bonobos, chimpanzees, gorillas, and orangutans reveal a more complex relationship between N_e and adaptation in these taxa. Both the proportion of nonsynonymous substitutions and the ratio of adaptive to neutral

divergence were positively correlated to long-term N_e (Cagan et al. 2016). Assuming that the targets for most selective sweeps are near or in genes, Nam et al. (2017) found that the relative amount of genetic diversity in great apes was more reduced in species with higher N_e . Simulations suggested that background selection alone could not explain this pattern. This reduction in diversity could be explained by either stronger sweeps or a higher frequency of selective sweeps in larger populations. The authors of this study suggest the latter, which is consistent with the theoretical prediction that larger populations wait “less” than smaller populations for beneficial mutations to occur, per a hard sweep model in which a beneficial allele arises *de novo* and rapidly sweeps to fixation (Maynard Smith and Haigh 1974). A recent application of machine learning to the GAGP data yielded partial supported this perspective. Nye et al. (2020) used a random forest algorithm and employed 15 different statistics to detect selective sweeps including a demographic model from Schmidt et al. (2019). While central chimpanzees had the most selective sweeps, the highest genomic proportion of putative sweeps, and total number of genes, this linear relationship was not upheld for the remaining three chimpanzee subspecies with smaller N_e . Further, Castellano et al. (2019) did not find a relationship between the proportion of beneficial alleles and N_e based on zerofold nonsynonymous and fourfold synonymous sites, although bonobos had a substantially high number of beneficial mutations, despite their low estimated N_e .

Studies of *Pan* genomics are useful to testing hypotheses and predictions from population genetic theory and these results inform the nature of evolutionary processes in these lineages. Of equal interest is the identification of genomic regions that exhibit various selection signatures that may inform the genomic underpinnings of phenotypes involved in lineage divergence.

Considerable attention has focused on host evolutionary responses to disease in apes, especially bonobos and chimpanzees. The ongoing COVID-19 pandemic highlights the need for understanding zoonoses for human health as well as anthroponoses given the endangered status of *Pan* and other primates, some of whom are susceptible to COVID-19 (Melin et al. 2020; Melin et al. 2021). One example infection that may have shaped *Pan* immune systems is simian immunodeficiency virus (SIV). SIV has been known to occur in chimpanzees (SIV_{cpz}) for over two decades (Gao et al. 1999) although it is curiously absent from bonobos (Inogwabini 2020). Indeed, HIV-1 is partially the result of a zoonotic event from an SIV infected chimpanzee (Sharp and Hahn 2011). Until recently, SIV_{cpz} was thought to be non-pathogenic in chimpanzees although this is no longer the case (Keele et al. 2009; Etienne et al. 2011; Terio et al. 2011). Genomic data point to several regions that may reflect the potential selective pressure. There is evidence for at least one selective sweep near the major histocompatibility complex in chimpanzees (de Groot et al. 2010; Prüfer et al. 2012). Cagan et al. (2016) used Fay and Wu's H statistic (Fay and Wu 2000) to identify *IDO2* as a candidate for recent positive selection in all four chimpanzee subspecies and bonobos. McDonald-Kreitman tests (McDonald and Kreitman 1991) were used by Cagan et al. (2016) and revealed *HIVEP1* as a positive selection candidate in bonobos and eastern chimpanzees. While SIV is absent from bonobos, a number of potentially zoonotic diseases were recently reported in this species (Medkour et al. 2021). HKA tests (Hudson et al. 1987) also identified genes related to the activation of the innate immune system (GO category: complement activation) to be significantly enriched in bonobos (Cagan et al. 2016). Schmidt et al. (2019) used a modification of population-branch statistics to examine recent adaptation in central and eastern chimpanzees. These authors did not find evidence for

enrichment in immune genes in central chimpanzees. However, multiple immune related GO categories as well as genes in three different sets of viral interacting proteins were significantly enriched. This signature was so strong that the removal of genes in these categories greatly reduces the selection signature (Schmidt et al. 2019).

Beyond the immune system, other *Pan* phenotypes show evidence of selection. Kovalaskas et al. (2020) reported two candidate SNPs subject to recent positive selection near *AMY2A* using XP-EHH, a test that detects recent selection. *AMY2A* codes for the production of pancreatic amylase and the authors suggest their findings offer support that bonobos are adapted to the consumption of starchy resources compared to chimpanzees (Kovalaskas et al. 2020).

Positive selection may have also shaped phenotypes related to the SDH. Kovalaskas et al. (2020) report a strong signal near *DIO2*. This gene provides the brain with triiodothyronine (T3). Interestingly, bonobos exhibit higher levels of circulating T3 than compared to chimpanzees and humans (Verena Behringer et al. 2014). Both *SOX5* and *SOX14* were identified as under recent positive selection (Kovalaskas et al. 2020). *SOX5* organizes the production of cartilage cells (Lefebvre et al. 2001) and is involved in nervous system development, which may have consequences on skeletal morphology, particularly in the cranium. *SOX14* has been associated with nervous system development and several disorders that impact the face (Arsic et al. 1998).

Genes that underlie social behavior may also been the targets of adaptation. Kovalaskas et al. (2020) identified variants in *CD38*, *DRD1*, *OT*, and *OXTR* as well as *AVPR1A* in bonobos. These genes are well characterized in modulating social behavior and genetic variation in *AVPR1A* has been previously linked to sociality in *Pan*. Staes et al. (2014) originally reported no polymorphism for a regulatory element

(RS3) of *AVPR1A*; however, both bonobos and chimpanzees appear to be polymorphic for this locus and this variation has been linked to differences in personality (Anestis et al. 2014; Staes et al. 2015; Staes et al. 2016). Yet, Staes et al. (2015) did not find an association between *OXTR* variation and sociality in chimpanzees. Finally, humans and bonobos shared a single amino-acid change in *TAAR8*, which encodes a G-coupled protein receptor that may provide social cues (Prüfer et al. 2012).

Positive selection may have also shaped *Pan* brains. Cagan et al. (2016) described an enrichment for genes under recent positive selection in the GO categories “dendrite” and “neuron spine” in central chimpanzees. *NRXN3* exhibited a strong signature of recent positive selection using Fay and Wu’s H statistic in central, eastern, and Nigeria-Cameroon chimpanzees. This gene is largely expressed in the brain and related to synaptic plasticity and transmission. This test also detected *CSMD1* in bonobos, eastern chimpanzees, and Nigeria-Cameroon chimpanzees, which is a gene with unknown function but is highly expressed in the nervous system.

Signals of balancing selection appear to be shared to a higher degree than adaptive signals in great apes, including *Pan* (Cagan et al. 2016). It is not surprising that many immunity-related genes were found to be under balancing selection in bonobos and chimpanzees (and the other apes) (Cagan et al. 2016). This study also found evidence for enrichment in genes involved in keratinocyte differentiation (*LCE3D*, *LCE3E*, *SCEL*, *SPRR2B*, *SPRR2G*) in western chimpanzees and cornified envelope development in central, Nigeria-Cameroon, and western chimpanzees. *CDSN* was also identified as a putative balancing selection candidate in bonobos and western chimpanzees. Cagan et al. (2016) note that balancing selection on these genes

may enable low levels of pathogen penetrance into a host potentially resulting in immunity to such pathogens.

Other immunity-related candidates for balancing selection have been identified. Cheng and DeGiorgio (2020) developed a suite of statistics, called *B* statistics, and applied these to both humans and bonobos. *MHC-DQ* and *MHC-DP* were identified as well as *KLRD1*, which encodes a cell-surface antigen, and *GPNMB*, which encodes osteoactivin (a transmembrane glycoprotein found on several cells) (Cheng and DeGiorgio 2020). Balancing selection may act on innate immune genes as an intergenic region between *BPIFA2* and *BPIFB4* exhibited a strong selection signal. Balancing selection may also act on non-immunity related phenotypes. Cheng and DeGiorgio (2020) describe potential selection on genes related to pain and neurodevelopment including *EPHA6*, *HPCAL1*, *SCN9A*, and *SUSD2*. This study also noted that such a signal may arise because of conflicting functions, which may explain the observed signatures in *CAMK4*, *GPNMB*, and *PDE1A*.

The studies above primarily focus on allelic variation; however, many other changes and/or interactions can occur that impact bonobo and chimpanzee phenotypes. Inversions can play an important role in disease but they are notoriously difficult to characterize. Porubsky et al. (2020) recently identified novel simple inversions and inverted duplications in the great apes, including bonobos and chimpanzees, which may contribute to differences in *Pan* phenotypes. Soto et al. (2020) described new structural variation in chimpanzee genomes, including variants in 56 genes that may underlie chimpanzee phenotypes. A recent high-quality bonobo genome assembly also revealed novel structural variants (Mao et al. 2021). This study identified gene family expansion in *EIF4A3*, a translation initiation factor subunit, that began ~2.9 Ma and resulted in six and five copies in bonobos and chimpanzees,

respectively. These authors also described 15,786 bonobo-specific insertions and 7,082 deletions. These deletions are enriched in membrane-associated genes with extracellular domains and two structural variants ablate *LYPD8* and *SAMD9* (Mao et al. 2021). As with allelic variation, structural variation may be related to N_e . Sudmant et al. (2013) found that western chimpanzees and bonobos (and Sumatran orangutans) exhibited an excess of segregating duplications > 30 kb. Further, western chimpanzees also exhibited an excess of segregating deletions > 30 kb. As these populations are estimated to have experienced recent bottlenecks, it appears that N_e may affect the extent of structural variation in great apes and other species. While the project described in this dissertation does not focus on structural variation, this area is a key future avenue for understanding the genomic architecture of *Pan* phenotypes.

Project Overview

This dissertation uses genomic data on all five *Pan* lineages to answer questions about their evolutionary history following divergence, specifically related to adaptation and demography. The second chapter of this dissertation, which includes unpublished but co-authored material with Frances White and Timothy Webster, focuses on signatures of positive selection that reflect adaptation in deeper time using two approaches. We find that most genes with sufficient statistical power to evaluate for selection have been subject to purifying selection. Candidates for positive selection are largely unique to each lineage and include genes related to the brain, immune system, musculature, reproduction and skeletal system. We did not find evidence of a shared pattern among chimpanzee lineages except for one gene, which may reflect the deep divergence and variation within the species.

The third chapter of this dissertation, which includes co-authored unpublished material with Frances White, Nelson Ting, and Timothy Webster, considers some of the most recent evolutionary processes in *Pan* evolution. We use supervised machine learning to identify genomic regions that are evolving neutrally, are linked to selective sweeps, or subject to a recent hard or soft sweep. In the four lineages we could analyze, we find that soft sweeps are overwhelmingly more common than hard although most of the genome is linked to these sweeps or is evolving neutrally. Most sweep windows are unique to each lineage although there are some shared windows, particularly for soft sweeps and especially between central and eastern chimpanzees. We find evidence of enrichment for genes related to the nervous system in central chimpanzees and identify candidates that may drive phenotypic differences in these taxa.

The fourth chapter of this dissertation addresses the evolutionary history of these lineages and includes unpublished but co-authored material with Frances White, Alan Rogers, and Timothy Webster. This topic has been the subject of many analyses resulting in some increased agreement of particular demographic parameter estimates whereas others remain less well known. Some currently used demographic methods produce biased parameters. We build and compare various demographic models by analyzing the site patterns of derived alleles and find that a simpler model than have been previous proposed best fits the data. This model includes an episode of introgression from bonobos into central chimpanzees and also points to a deeper divergence in the chimpanzee common ancestor than formerly estimated.

These results not only shed light on different facets of *Pan* evolutionary history at various points following speciation, they also offer important insight on evolutionary processes broadly and, more specifically, processes that occurred

specifically in western and central Africa during a critical time period for the evolution of other species in this region, including humans.

CHAPTER II

ADAPTATION DURING DIVERGENCE IN BONOBOS (*PAN PANISCUS*) AND CHIMPANZEES (*PAN TROGLODYTES*)

Frances White, Timothy Webster, and I conceived of this analysis. The assembly and mapping of the genomic data in this analysis was conducted by Timothy Webster and he provided some code for the preparation of the SnIPRE analysis. I performed the other data analyses and wrote the initial draft of the manuscript. Frances White, Timothy Webster, and I edited the manuscript.

Introduction

Genomic data can provide an important window into the evolutionary past of a population, particularly when paleontological and archaeological data are lacking. Considerable emphasis has been placed on positive selection and the identification of adaptive traits that may differentiate a lineage from others. However, methods for detecting positive selection are dependent on genetic variation, which can be impacted by a population's demographic history (Nielsen 2001; Przeworski 2002). Mitigating such effects require either models robust to demography or the specification of a demographic model, when such information is known or can be inferred. Further, different metrics are informative for specific timescales such that some selection tests are better suited for more recent events, whereas others speak to the distant past (Weigand and Leese 2018). In particular, two tests are especially useful for detecting older signatures of selection: Hudson-Kreitman-Aguadé (HKA) tests (Hudson et al. 1987) and McDonald-Kreitman (MK) tests (McDonald and Kreitman 1991).

However, HKA tests can result in false positives under certain migration rates (Nielsen 2001).

MK tests, on the other hand, are more robust to certain aspects of demography because the rate of polymorphism and divergence for two site categories are compared within a gene in a single lineage and an outgroup is used to determine divergence. These rates are compared because neutral theory predicts that the ratio of polymorphic to divergent sites should be equal for both synonymous and non-synonymous sites (McDonald and Kreitman 1991). A significant result from this test does not offer any information about the type of selection, only that a neutral model can be rejected (Nielsen 2001). Yet, a significant difference can arise in a population of constant size and under an additive model of selection in one of two ways. An excess of divergent nonsynonymous mutations would suggest that different amino acids are being selected for, and thus suggesting positive selection, whereas fewer than expected divergent nonsynonymous mutations indicates the locus is under negative or purifying selection, removing mutations that would alter the resulting amino acids. This also assumes that mutations are strongly deleterious. However, if weakly deleterious mutations are present at a locus, they are unlikely to become fixed and can inflate the number of polymorphic sites, reducing the power to detect positive selection. This issue may be ameliorated by excluding rare alleles (i.e., biallelic sites whose minor allele frequency (MAF) was < 0.1). The other major consideration for MK tests are changes in effective population size, hereafter referred to as N_e (McDonald and Kreitman 1991; Eyre-Walker 2002). When N_e increases, a larger number of mutations shift from nearly neutral to deleterious, thus increasing the constraint on a gene and decreasing the effectively neutral mutation rate (Wright and Andolfatto 2008). Therefore, sufficient differences in N_e between the time period that

is captured by polymorphisms vs the time period reflected in substitutions may result in different effectively neutral mutation rates. For example, slightly deleterious alleles may have been fixed during a population's divergence but may not affect polymorphisms following population size increase resulting in false signatures of adaptation (McDonald and Kreitman 1991). This suggests that, given particular population histories, MK tests may not be able to distinguish between positive selection or reduced constraint during divergence. Awareness of such caveats is key to applying this critical test and its extensions to detect positive selection.

The genus *Pan* provides an intriguing model for understanding positive selection at deeper time scales. The two extant species, bonobos (*Pan paniscus*) and chimpanzees (*P. troglodytes*), diverged ~ 1.88 Ma (de Manuel et al. 2016) and chimpanzees subsequently split into four subspecies (Stumpf 2011). While both species exhibit a considerable and often overlooked number of similarities, phenotypic differences, particularly evidenced in behavior, are well documented. Despite sharing a male philopatric fission-fusion social structure with similar community sizes, bonobos and chimpanzees, on the whole differ in patterns of power, adult sex-based bondedness, and gregariousness, such that adult male chimpanzees exhibit the strongest bonds with other males and are typically aggressively dominant to females (Goodall 1986; Wrangham 1986; Boesch and Boesch-Achermann 2000; Mitani 2009; Nishida 2011), whereas relationships among females and between males and females are strongest in bonobos, females can hold high ranking positions, and aggression is less intense and less frequent than in chimpanzees (Kano 1992; White 1996; White and Wood 2007; Furuichi 2011; Tokuyama and Furuichi 2016; Moscovice et al. 2017). Additionally, some chimpanzees engage in lethal aggression both within and between communities (Watts 2004; Kaburu et al. 2013; M.L. Wilson

et al. 2014). While lethal aggression is not a defining characteristic of chimpanzees because it is so variable and has not been seen in all communities, the behavior has never been observed in bonobos, with only one potential suspected case (Hohmann and Fruth 2011, but see White 2012). These patterns appear unlinked to anthropogenic influence and lethal aggression (or lack thereof) may be adaptive (M.L. Wilson et al. 2014). Indeed, the typical nature of intergroup encounters (IGEs) appears fundamentally different between the two *Pan* species (Kano 1992; Boesch et al. 2008; Mitani et al. 2010; Furuichi 2011; Fruth and Hohmann 2018; Sakamaki et al. 2018; Lucchesi et al. 2020).

A number of hypotheses have been proposed to explain these differences, either socio-ecological or behavioral. Socio-ecological hypotheses point to differences in available terrestrial herbaceous vegetation that would allow grouping by reducing competition (Wrangham 1986; Wrangham et al. 1996) or qualitative and/or quantitative differences in food patches that would actively select for female cooperation (White 1986). Behavioral hypotheses focus on such factors as the importance of mothers on male reproductive success (e.g., Kano 1992), the role of tension regulation in social contexts (de Waal 1989), and the impact of female coalitions (Tokuyama and Furuichi 2016). Additionally, sexual selection may drive phenotypic differences in *Pan* as suggested by the self-domestication hypothesis that posits female bonobos selected for less aggressive males resulting in some phenotypes that are similar to some domesticated mammals (Wrangham and Pilbeam 2001; Hare et al. 2012). Females may specifically select for less aggressive males to reduce infanticide rather than monopolize resources, the “baby dominance hypothesis” (Walker and Hare 2017).

Although testing of these hypotheses typically requires considerable behavioral and ecological data, a new, complementary approach uses genomic data to identify signatures of adaptation that address various explicit predictions of these hypotheses. These include, for example, looking at the potential impact of the thyroid on morphology and behavior, examining digestive enzymes to test hypotheses on the importance of different foods, and considering the proximate mechanisms for species differences in reproduction-related traits. Variation in the ontogenetic patterns of circulating thyroid hormone (triiodothyronine or T3) has been suggested as an explanation for differences in bonobos and chimpanzees (Verena Behringer et al. 2014). Recently, a single nucleotide polymorphism, or SNP, near *DIO2*, a gene that catalyzes the conversion of thyroxine (T4) to T3, has been reported to exhibit a signature of positive selection in bonobos (Kovalaskas et al. 2020). This study also described an adaptive signature near the *AMY2* locus in bonobos. As *AMY2* codes for pancreatic amylase, the authors interpreted this result as support for the THV hypothesis. Embedded in several of the hypotheses above are differences in female reproduction between bonobos and chimpanzees (Stumpf 2011). Han et al. (2019) previously noted enrichment for bonobo-specific nonsynonymous changes at loci associated with menarche in humans. Thus, we predicted genes related to reproduction to exhibit signatures of positive selection.

In addition to testing these and other candidate genes, genome-wide selection scans can also shed light on previously underappreciated unique or shared phenotypes between extant lineages (i.e., “reverse ecology” (Li et al. 2008)), that have not yet been built into a hypothesis.

The present study investigates adaptation in *Pan* that occurred in the distant past, closer to the speciation of the extant members of this genus using MK tests. We

build on a previous analysis (Cagan et al. 2016) using reassembled data and a different and improved chimpanzee reference genome, addressing contamination issues, and including all five, rather than four, *Pan* lineages. We also apply a Bayesian implementation of a generalized linear mixed model to identify putative candidates for positive selection that leverages genome-wide averages to increase statistical power.

Methods

Genomic Data

We retrieved raw short read data on bonobos and all four chimpanzee subspecies from the Great Ape Genome Project (GAGP) (Prado-Martinez et al. 2013). This dataset contained high coverage genomes (<https://github.com/brandcm/Dissertation>: File S0: Figures S1, S2) from 13 bonobos (*P. paniscus*), 18 central chimpanzees (*P. troglodytes troglodytes*), 19 eastern chimpanzees (*P. t. schweinfurthii*), 10 Nigerian chimpanzees (*P. t. ellioti*), and 11 western chimpanzees (*P. t. verus*). See <https://github.com/brandcm/Dissertation>: Files S0 and S1 for more information on these samples. MK tests require an outgroup to determine whether substitutions are unique or shared. We retrieved short read data on a high-coverage human female, HG00513, collected as part of the 1000 Genomes Project (Auton et al. 2015) to use as the outgroup sequence (Biosample ID: SAME123526).

Read Mapping and Variant Calling

Initial quality assessments in fastqc (Andrews 2010) and multiqc (Ewels et al. 2016) indicated a number of quality issues, including failed runs, problematic tiles,

and substantial variation in base quality. We removed adapters and trimmed all reads with BBduk (<https://sourceforge.net/projects/bbmap/>). For trimming, we used the parameters “ktrim=r k=21 mink=11 hdist=2 qtrim=rl trimq=15 minlen=50 maq=20” for all reads and added “tpo and tpe” for paired reads.

We used XYalign (Webster et al. 2019) to create versions of the chimpanzee reference genome, panTro6 (Kronenberg et al. 2018), for male- and female-specific mapping. Specifically, the version of the reference for female mapping has the Y chromosome completely masked, as its presence can lead to mismapping (Webster et al. 2019). We then mapped reads with BWA MEM (Li 2013) and used SAMtools (Li et al. 2009) to fix mate pairs, sort BAM files, merge BAM files per individual, and index BAM files. We use Picard (Broad Institute 2018) to mark duplicates with default parameters, before calculating BAM statistics with SAMtools. We next measured depth of coverage with mosdepth (Pedersen and Quinlan 2018), removing duplicates and reads with a mapping quality less than 30 for calculations.

We used GATK4 (Poplin et al. 2018) for joint variant calling across all samples. We used default settings for all steps—HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs—with three exceptions. First, we turned off physical phasing for computational efficiency and downstream VCF compatibility with filtering tools. Second, because multiple samples in this dataset suffer from contamination from other samples both within and across taxa (Prado-Martinez et al. 2013), we employed a contamination filter to randomly remove 10% of reads during variant calling. This should have the effect of reducing confidence in contaminant alleles. Finally, we output non-variant sites to allow equivalent filtering of all sites in the genome and more accurate assessments of callability.

The above quality control, assembly, and variant calling steps are all contained in an automated Snakemake (Köster and Rahmann 2012) available on GitHub (https://github.com/thw17/Pan_reassembly). The repository also contains a Conda environment with all software versions and origins, most of which are available through Bioconda (Grüning et al. 2018).

Variant Filtration

We considered only autosomes for this analysis as the X and Y chromosome violate many of the assumptions for the following methods (Webster and Wilson Sayres 2016). We also excluded unlocalized scaffolds (N = 4), unplaced contigs (N = 4,316), and the mitochondrial genome from any downstream analyses. Additional filtration steps were completed using bcftools (Li 2011) and command line inputs are provided in parentheses. MK tests rely on accurate assessments of whether a SNP is synonymous or nonsynonymous. We first normalized variants by joining biallelic sites and merging indels and SNPs into a single record (“norm -m +any”) using the panTro6 FASTA. Next, we filtered to retain only coding sequence (“-R CDS_autosomes.bed”) as designated by the panTro6 GFF (retrieved from: https://www.ncbi.nlm.nih.gov/genome/202?genome_assembly_id=380228). Further, we only included single nucleotide polymorphisms (SNPs) (“-v snps”) that were biallelic (“-m2 -M2”). On a per sample basis within each site, we marked genotypes where sample read depth was less than 10 and/or genotype quality was less than 30 as uncalled (“-S . -i FMT/DP \geq 10 && FMT/GT \geq 30”). To ensure that missing data did not bias our results, we further excluded any sites where less than ~ 80% of individuals (N = 56) were confidently genotyped (“AN \geq 112”). We also removed any positions that were monomorphic for either the reference or alternate allele (“AC > 0

&& $AC \neq AN$ ”). While lack of or low coverage at a locus is problematic, loci with excessive coverage are also of concern. These sites may yield false heterozygotes that are usually the result of copy number variation or paralogous sequences (Li 2014). As our data exhibit a high degree of inter-individual and inter-chromosomal variation in mean coverage (Brand et al. 2021), we applied Li’s (2014) recommendation for a maximum depth filter ($d + 4\sqrt{d}$) to the mean chromosomal coverage of the individual in our sample (*Pan* or *Homo*) with the highest coverage and excluded any loci that exceeded this value (“filter -e FMT/DP > d + $4\sqrt{d}$ ”) (<https://github.com/brandcm/Dissertation>: File S2). These filtration steps yielded 291,782 SNPs for our downstream analyses (<https://github.com/brandcm/Dissertation>: File S0: Table S1).

Analysis

We built a custom database in snpEff (Cingolani et al. 2012) using only the assembled autosomes from the panTro6 FASTA and the panTro6 GFF (“java -jar snpEff/snpEff.jar build -gff3 -v chimp”). We included only rows for coding sequences (‘CDS’) for assembled autosomes in the GFF and used one transcript per gene (N = 20,265). In cases where genes had multiple transcripts, we determined the longest transcript (based on CDS bp) using a custom R script (R Core Team 2020), `one_transcript_per_gene_filtered_gff.R`, and used that transcript. This database was then used to annotate VCFs for each autosome via snpEff. Allele frequencies per SNP per *Pan* population were calculated via VCFtools (Danecek et al. 2011).

For each autosome in each *Pan* lineage, we used a modified R script, `command_line_mk_script.R`, to run MK tests. This script was based on an existing

script (<https://github.com/thomasblankers/popgen/blob/master/MKTtest>) and uses the stringr, version 1.4.0 (Wickham 2019) and tidyverse, version 1.3.1 (Wickham et al. 2019) packages. Our script first filtered for SNPs identified by snpEff as either synonymous or missense (nonsynonymous) and subsequently categorized each SNP as 1) divergent (i.e., fixed for different alleles) and synonymous (D_s), 2) divergent and nonsynonymous (D_n), 3) polymorphic and synonymous (P_s), or 4) polymorphic and nonsynonymous (P_n) via the *Pan* allele frequencies calculated in VCFtools above (for all four categories: D_s , D_n , P_s , P_n) and using the human sample as the outgroup (for D_s and D_n). We summed the number of SNPs per category per gene and then ran Fisher's exact test on the contingency table for each gene using $\alpha < 0.05$.

Our script also calculated the neutrality index (Rand and Kann 1996), $NI =$

$$\frac{P_n/D_n}{P_s/D_s} = \frac{P_n D_s}{P_s D_n}, \text{ per gene. Values greater than one reflect more polymorphic}$$

nonsynonymous sites than expected, or an abundance of weakly deleterious alleles, whereas values less than one suggests more fixed nonsynonymous mutations than expected, i.e., adaptive mutations. This statistic is informative when P_s and D_n are defined; however, this is not always the case (Stoletzki and Eyre-Walker 2011). NI can also be biased when either or both P_s and D_n are small (Stoletzki and Eyre-Walker 2011). Therefore, we also calculated the direction of selection, DoS , statistic

$$\text{(Stoletzki and Eyre-Walker 2011), } DoS = \frac{D_n}{D_n + D_s} - \frac{P_n}{P_n + P_s} \text{ and used this metric to}$$

classify genes as subject to either positive ($DoS > 1$) or purifying ($DoS < 1$) selection.

We immediately discarded genes where the contingency table was incomplete and Fisher's test could not be performed (i.e., there were either no fixed SNPs or no polymorphic SNPs). We further removed genes ($N = 6,892$) for which $< 50\%$ of the coding sequence exhibited poor coverage across the entire *Pan* sample ($N = 71$)

(Brand et al. 2021) and retained 13,228 genes (<https://github.com/brandcm/Dissertation>: File S3). Fisher's exact test is underpowered, i.e., exhibits a high false positive rate, when the overall observations in a contingency table are low (Begun et al. 2007; Holloway et al. 2007; Andolfatto 2008; Darolti et al. 2018). We followed Holloway et al. (2007) and excluded genes for which the sum of each row and column in the 2 x 2 table was < 5 . We designated a gene as a candidate for being previously subjected to natural selection when both the p-value for Fisher's exact test was < 0.05 and the sum of each row and column in the contingency table was ≥ 5 . We further categorized these genes as subject to positive selection where DoS was > 1 and purifying selection where DoS was < 1 .

We repeated the analysis above two times. In our second analysis, we removed SNPs whose minor allele frequency (MAF) was < 0.1 in order to assess the effects of weakly deleterious mutations on our results. We also considered within gene heterogeneity, i.e., differences between exons of the same gene, by constructing contingency tables and running the aforementioned analyses per exon rather than per gene. To ensure that bias in CDS length did not affect our analyses, we visualized the distribution of SNPs/bp for all genes/exons per lineage that passed our initial filter (an incomplete contingency table) and the distribution of SNPs/bp for candidate genes/exon under positive selection.

Additionally, we performed another set of selection analyses implemented using SnIPRE (Eilertson et al. 2012). This non-parametric approach uses the same input data as MK tests (i.e., the contingency table) as well as genome-wide information on the average and variance in polymorphism to divergence, therefore, increasing power. Additionally, if the assumptions of a neutral demographic model are met, the resulting parameters can be used to estimate the strength, directionality,

and timing of selection. We applied the SnIPRE method to our data where the contingency table was complete and did not set a row/column filter for the number of SNPs. SnIPRE also requires the fraction of time a site can mutate synonymously or nonsynonymously. We generated these data per gene using the panTro6 FASTA and a custom script, `collect_snipre_data.py`. We used the Bayesian implementation of SnIPRE, using the default MCMC sampling settings by discarding the first 10,000 values, retaining every fourth value, and running 15,000 iterations per chain (“BSnIPRE.run(data, burnin = 10000, thin = 4, iter = 15000)”).

All scripts used in our analysis can be found at:

https://github.com/brandcm/Pan_MK. Figures were generated using `ggplot2`, version 3.3.3 (Wickham 2016). Additionally, some scripts to build figures used `gridExtra`, version 2.3 (Baptiste 2015). Unique candidate genes for positive selection and those shared between two or more lineages were visualized using Upset plots created with the `ComplexUpset` package, version 1.2.1 (Lex et al. 2014; Krassowski 2020).

Data Availability

The raw data underlying this article are previously published (Prado-Martinez et al. 2013; de Manuel et al. 2016) and are available from the Sequence Read Archive (PRJNA189439 and SRP018689) and the European Nucleotide Archive (PRJEB15086).

Results

The distribution for the number of annotations per SNP ranged from 1 to 5 (<https://github.com/brandcm/Dissertation>: File S0: Figure S3), with approximately 90% having one annotation. Therefore, disagreement between variant effects across

multiple annotations as determined by snpEff is unlikely to bias these results. We found that the number of SNPs per autosome available for use in our MK analysis was similar across all five populations after filtering (<https://github.com/brandcm/Dissertation>: File S4). After filtering for loci whose MAF was < 0.1 , the remaining number of SNPs was strongly related to estimated N_e and partially related to sample size (<https://github.com/brandcm/Dissertation>: File S4). Thus, $\sim 20\%$ of SNPs were excluded for central chimpanzees, $\sim 10\%$ for eastern chimpanzees, and $< \sim 6\%$ for the other lineages.

Gene Analysis

The results for all assessable genes can be found at <https://github.com/brandcm/Dissertation>: Files S5 and S6. Based on the DoS statistic, the majority of significant candidate genes for each population were found to be under purifying selection (Table 1). The number of candidate genes for both positive and purifying selection was variable across lineages and somewhat mirrored both N_e as well as the sample size used in the analysis. As predicted, the number of statistically significant genes with a positive selection signature changed when the MAF filter was applied. These lists were not only shorter but also included a number of genes not previously identified in the analysis without the MAF filter. This was particularly true for eastern chimpanzees where 14/23 (60.9%) of the list included different genes as well as central chimpanzees for which 11/21 (52.4%) were new. The presence of slightly deleterious mutations reduces the power to detect positive selection because such mutations will disproportionately affect polymorphic sites. Therefore, we combined the list of genes exhibiting a signature of positive selection generated both with and without the MAF filter per lineage and consider every gene in these collated

lists to be a candidate for positive selection. The results of these analyses with and without the MAF filter can be found in <https://github.com/brandcm/Dissertation>: Files S5 and S6 and the subset of positive selection candidates are provided in <https://github.com/brandcm/Dissertation>: Files S7.

Table 1. Number of candidate genes under positive selection, purifying selection, and the number of total genes/exons tested per lineage after all filtration steps.

Analysis	Type	<i>P. paniscus</i>	<i>P. t. ellioti</i>	<i>P. t. schweinfurthii</i>	<i>P. t. troglodytes</i>	<i>P. t. verus</i>
per gene	Positive	11	24	31	41	5
	Purifying	55	51	74	89	34
	Total Tested	1399	1679	2433	3180	906
per gene, without rare alleles	Positive	6	14	23	21	4
	Purifying	14	27	19	19	17
	Total Tested	477	1091	1101	1170	474
per exon	Positive	5	5	8	17	3
	Purifying	5	4	15	18	9
	Total Tested	378	482	705	1028	231

We found that the distributions of SNPs per bp for candidates under positive selection per lineage overlapped the distribution for all assessable genes/exons (<https://github.com/brandcm/Dissertation>: File S0: Figures S4-S6). The candidate genes/exons that fell within the right tail of these distributions were almost completely short sequences (< 1k bp). Only one gene, *RNF213*, exhibited a signature of positive selection in all five lineages (Figure 1, <https://github.com/brandcm/Dissertation>: File S8). Contrary to our prediction, we did not find any genes under positive selection that were unique to all four chimpanzee lineages (Figure 1). Consistent with phylogenetic expectations, the most shared selection signals were between eastern and central chimpanzees (N = 10) (Figure 1). Further, candidate genes for bonobos were

unique to their lineage except for one gene, *NPAPI*, which was also detected in eastern chimpanzees (Figure 1). Figures S7 and S8 (<https://github.com/brandcm/Dissertation: File S0>) show candidate gene overlap for our datasets including and excluding rare polymorphisms, respectively.

Table 2. Candidate genes under positive selection per lineage from the MK analyses. Results from the analyses with and without rare polymorphisms (MAF < 0.1) are combined here.

Lineage	Genes
<i>P. paniscus</i>	<i>ALPK2, C2AH2orf78, CC2D2B, EFCAB8, KCNU1, NPAPI, OR5J2, PIK3C2G, RNF213, SCAPER, TSHR, WDR49, ZNF135</i>
<i>P. t. ellioti</i>	<i>ABCC2, ALMS1, ANKRD30A, CXCR1, DNAH14, DNAH6, FAN1, FARP2, HASPIN, HEATR5A, IL1RL2, LOC100608661, LOC100613827, LOC466407, LOC739832, LOC741747, MIA3, OR51G1, PARP14, RNF213, SHPRH, SLC17A3, SLC26A3, SNTG1, SOHLH2, TDRD15, TGM3, XRN1, ZFAND4</i>
<i>P. t. schweinfurthii</i>	<i>ADAM2, ADGRV1, ANKS4B, BDPI, C7, CAGE1, CCPG1, CEACAM5, CX3CR1, DDX60L, DNAH14, DNAH6, DOCK8, EFCAB5, FAM111A, FAN1, FGA, FLT3, GCNT2, HAVCR1, HEATR5A, IFI16, JHY, KIAA1257, KIAA2026, LOC100608661, LOC107972003, LOC451494, LOC469634, MROH8, MYH2, NPAPI, NXPE2, OR6X1, RNF213, SAMD7, SLC17A3, SLC26A3, SLC6A16, TMC05A, TRIM5, UBR2, VWA8, XRN1, ZFAND4</i>
<i>P. t. troglodytes</i>	<i>ABC5, ADGRV1, ANKS4B, ANP32C, BUB1, CMYA5, COL24A1, CX3CR1, DHTKD1, DNAH6, FAM111A, FAM209A, FAM71A, FGA, FLT3, GEMIN4, HASPIN, HEATR5A, HERC5, HMCN1, IQCA1L, JHY, LOC100608047, LOC107966998, LOC452946, LOC456268, LOC466407, LOC468520, LOC469634, LRRC53, MIAP, MROH8, MYH2, NLRP11, PCDHB10, PKDREJ, PPP1R15A, RNF213, SCUBE2, SLC17A3, SLC26A3, TANC1, TGM3, TMPRSS2, TOGARAM1, TTC6, TTLL6, TULP2, XRN1, ZFAND4, ZGRF1, ZNF480, ZNF518A, ZNF649</i>
<i>P. t. verus</i>	<i>HASPIN, KIAA1257, PRAME, RNF213, SHPRH, SLC6A16, ZNF473</i>

The function of some candidate loci (N = 13) in our analysis is unknown (genes beginning with “LOC”), in addition to a few other genes that have been assigned an identifier but whose function remains unknown or poorly understood. However, some interesting patterns emerge for the remaining loci. We caution that the following counts should be treated with caution as the full function of many genes is unknown, which may underestimate this value. Conversely, some of these genes are only associated with particular phenotypes and the causality is not fully determined.

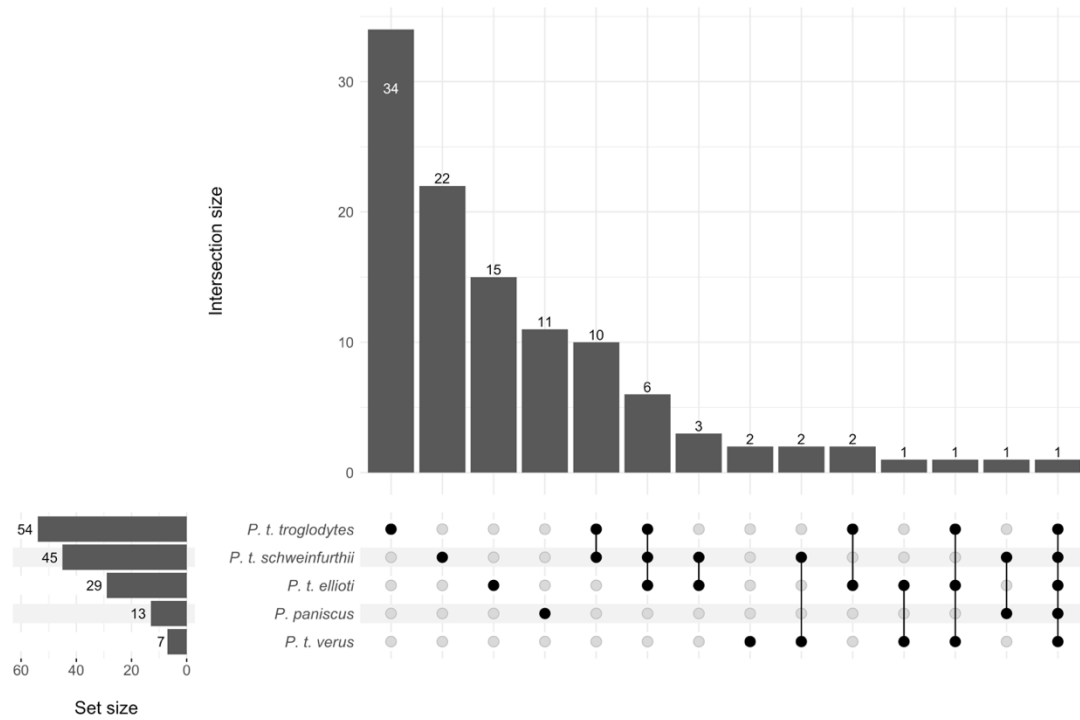


Figure 1. Upset plot of unique and shared candidate genes for positive selection.

We note a number of genes that are associated with the brain and the central nervous system were found to be positive selection candidates ($N = 18$), including *JHY* in eastern and central chimpanzees. Consistent with the hypothesis that disease has strongly shaped recent hominid evolution, many candidates ($N=12$), particularly for eastern and central chimpanzees, are known or are speculated to play a role in immune function. We identified a few other functional categories including genes related to sensory systems ($N = 10$), muscle development, function, and maintenance ($N = 4$), the skeletal system ($N = 4$), and reproduction ($N = 4$). As predicted, *TSHR* emerged as a candidate for positive selection in bonobos.

We found some overlap in genes exhibiting signatures of positive selection at the gene-level when compared to a previous MK analysis in four of the same lineages

(Cagan et al. 2016). This includes 4 genes in *P. paniscus*, 6 in *P. t. ellioti*, 10 in *P. t. schweinfurthii*, and 4 in *P. t. verus* (<https://github.com/brandcm/Dissertation>: File S9).

Exon Analysis

Despite reduced power to detect positive selection at the exon-level, we found a number of specific exons (N = 16) across all lineages that exhibited a significantly different ratio of polymorphisms to substitutions relative to the entire gene, when excluding genes with unknown function and those with only one exon (Table 3) (<https://github.com/brandcm/Dissertation>: Files S10 and S11). Six of the genes (*FSIP1*, *KIAA1755*, *LRRC63*, *MROH7*, *NLRP8*, *VPS13D*) were only discovered using this exon-level analysis as they were not significant when the entire gene was considered, both with and without rare alleles.

SnIPRE Analysis

The SnIPRE approach yielded an entirely new set of candidate genes for positive selection as compared to the MK tests (Table 4) (<https://github.com/brandcm/Dissertation>: Files S12-16). The majority of the genes are unknown or poorly characterized. One gene, *MUC17*, was detected in all five *Pan* lineages. This analysis did yield a shared gene among chimpanzees, *RFPL4B*, and a gene shared by all chimpanzees except for *P. t. verus*, *MS4A3*.

Table 3. Exons with a statistically significant signature of positive selection excluding genes whose function is unknown and genes with only one exon. ppn = *P. paniscus*, pte = *P. t. ellioti*, pts = *P. t. schweinfurthii*, ptt = *P. t. troglodytes*, ptv = *P. t. verus*.

Gene	Exon	Chromosome	Lineage	pN	pS	dN	dS	p-value	DoS
<i>ANKS4B</i>	exon 2	chr16	pts	1	4	7	1	0.03185703	0.675
<i>BDP1</i>	exon 24	chr5	ptt	13	7	14	0	0.02621228	0.35
<i>CMYA5</i>	exon 12	chr5	ptt	49	34	43	12	0.02650057	0.19145674
<i>FAM111A</i>	exon 2	chr11	pts	2	5	14	0	0.00103199	0.71428571
			ptt	5	6	13	0	0.00343249	0.54545455
<i>FGA</i>	exon 2	chr4	pts	2	4	24	1	0.00224235	0.62666667
			ptt	3	5	24	1	0.00128931	0.585
<i>FSIP1</i>	exon 2	chr15	ptt	1	6	5	0	0.01515152	0.85714286
<i>KIAA1755</i>	exon 12	chr20	ptt	1	4	7	1	0.03185703	0.675
<i>LRRC63</i>	exon 8	chr13	pts	2	5	5	0	0.02777778	0.71428571
<i>MROH7</i>	exon 1	chr1	pts	1	5	7	2	0.04055944	0.61111111
<i>NLRP8</i>	exon 3	chr19	ptt	6	12	8	2	0.0460717	0.46666667
<i>PPP1R15A</i>	exon 1	chr19	ptt	3	4	11	1	0.0379257	0.48809524
<i>SLC6A16</i>	exon 11	chr19	ptv	1	4	8	1	0.02297702	0.68888889
<i>TSHR</i>	exon 10	chr14	ppn	0	5	6	4	0.04395604	0.6
<i>VPS13D</i>	exon 18	chr1	pts	1	8	5	3	0.04977376	0.51388889
			ptv	0	6	5	3	0.03096903	0.625
<i>ZNF135</i>	exon 4	chr19	ppn	0	5	5	2	0.02777778	0.71428571
<i>ZNF480</i>	exon 4	chr19	ptt	0	6	5	2	0.02097902	0.71428571

Table 4. Candidate genes under positive selection per lineage from the SnIPRE analyses.

Lineage	Genes
<i>P. paniscus</i>	<i>LOC104001091, LOC107972003, MUC17</i>
<i>P. t. ellioti</i>	<i>C2AH2orf16, C6H6orf201, C9H9orf131, CD244, FGA, LOC101058953, LOC104001091, LOC107967308, LOC107969623, LOC107970484, LOC107972003, LOC470467, LOC739951, MS4A3, MUC17, NBPF7, RFPL4B</i>
<i>P. t. schweinfurthii</i>	<i>C6H6orf201, LOC100608661, LOC101058953, LOC104001091, LOC107970484, LOC107972003, LOC739951, MS4A3, MUC17, RFPL4B</i>
<i>P. t. troglodytes</i>	<i>C6H6orf201, FAM208B, LOC101057029, LOC104001091, LOC107972003, LOC470467, MS4A3, MUC17, RFPL4B, TXNDC2</i>
<i>P. t. verus</i>	<i>C2AH2orf16, LOC104001091, LOC107970484, LOC107972003, MUC17, RFPL4B</i>

Discussion

The limited number of methods appropriate for detecting adaptation at deeper time scales underscores the importance of their appropriate application to understanding the evolutionary history of a lineage. Here, we confirm and expand upon genomic insights into the past of the genus *Pan*. Using the direction of selection statistic, we found that the majority of candidates appear subjected to purifying selection. Curiously, controlling for rare alleles as a means to correct for slightly deleterious alleles was related to sample size and estimated N_e . As these variables are correlated in this dataset, additional samples are needed to tease apart whether this result could be driven by sample size, N_e , or another unknown factor.

Only one gene, *RNF213*, emerged as a candidate for positive selection across all lineages based on the traditional MK test. This gene is relatively large (CDS = 15,771 bp) and encodes a finger motif that functions as an E3 ubiquitin ligase (Wu et al. 2012). The locus has been associated with Moyamoya disease, an uncommon cerebrovascular progressive disease that results in the narrowing and blockage of blood vessels (Kamada et al. 2011; Liu et al. 2011; Wu et al. 2012). The phenotypic effects of this gene may partially explain some of the morphological and physiological differences in the brain and resulting cognitive differences between *Homo* and *Pan*. This shared gene is striking given the estimated divergence for the genus. Thus, convergence in bonobos and chimpanzees may be more plausible as an explanation for a shared adaptive signature at this locus.

As predicted, we identified one thyroid associated gene, a receptor, as a positive selection candidate in bonobos: *TSHR*. This is consistent with the hypothesis that thyroid-related differences may contribute to species differences between bonobos and chimpanzees. Additionally, our results align with the currently available

data on ontogenetic changes in circulating T3 in *Pan* and *Homo* (Verena Behringer et al. 2014). That study found that chimpanzees exhibited a decline in T3 at approximately ten years of age, falling within the variation observed in humans, whereas T3 in bonobos did not decline until about 20 years of age. While differences in thyroid hormone “rhythms” may facilitate speciation (Crockford 2003), it is unclear whether sequence variation would result in such ontogenetic differences. Future work at and near this locus and *DIO2*, as reported by Kovalaskas et al. (2020), may help shed light on how the thyroid contributes to differences in *Pan*.

We did not find any evidence supporting positive selection at *AMY2A* or *AMY2B* in bonobos. Based on our sample, these genes did not exhibit enough variants for robust statistical testing; however, the data suggest that both genes may be subject to purifying selection (<https://github.com/brandcm/Dissertation>: File S5). One caveat is that starch digestion differences in bonobos vs. chimpanzees could be shaped by only one or a few SNPs. MK tests would fail to detect such a signal because multiple genic changes increase power to identify selection signatures. This also does not preclude structural variation or differences in gene expression that may enable these loci to shape bonobo feeding ecology. For example, the SNPs identified by Kovalaskas et al. (2020) that occurred near *AMY2A* may be related to gene expression. Additional study of these loci, as well as *AMY1*, may yield support for the THV hypothesis. This is particularly needed as it has been proposed that, despite a gain in copy number relative to chimpanzees, *AMY1* may be non-functional in bonobos (Perry et al. 2007). However, the current behavioral and ecological data do not offer more than sparse support (Yamakoshi 2004). Further, if the SNPs identified by Kovalaskas et al. (2020) at *AMY2A* are responsible for such phenotypic effects,

this would imply that differences in *Pan* feeding ecology are ~ 100 ka old, which seems unlikely.

Differences in the reproductive and sexual cycles of female bonobos and chimpanzees have been well documented. Indeed, Han et al. (2019) described a number of nonsynonymous changes in bonobo genes that have been associated with menarche in humans. Additionally, sperm competition is not evenly distributed among hominoids and one would predict selection to act on genes related to this phenotype. While one study found little evidence for adaptation in these genes in *Pan* compared to *Gorilla* and *Homo* (Good et al. 2013), we found evidence of at least one reproduction related gene in all *Pan* lineages, including genes related to male reproduction. A few genes appear to impact reproduction broadly: *KIAA1257* in *P. t. verus*, which may impact gene expression of the gene *NR5A1* that is a transcriptional activator involved in sex determination (Sakai et al. 2008); and *SOHLH2* in *P. t. ellioti*, that affects both oogenesis and spermatogenesis (Toyoda et al. 2009). Some remaining candidates appear to primarily act on male gametes: *PKDREJ* in *P. t. troglodytes* (Hamm et al. 2007), and *PRAME* in *P. t. verus* (Chang et al. 2011). Two other candidates have been implicated in male murine reproduction but remain unknown in humans: *ADAM2* in *P. t. schweinfurthii* (Choi et al. 2016) and *KCNU1* in *P. paniscus* (Vyklicka and Lishko 2020).

Physical differences between bonobos and chimpanzees are well described. Additionally, variation and the distribution of such variation between different chimpanzee populations has also been reported (Groves 2001), as data are somewhat lacking for *P. t. ellioti* and the full extent of variation may not be yet realized in the other lineages. Four of the positive selection candidates are related to skeletal variation and another four seem to affect the development and maintenance of muscle

tissue. Collectively, these candidates covered all chimpanzee subspecies except for western chimpanzees and we did not find any such candidates in bonobos. For example, *FAM111A* was detected as a candidate in both eastern and central chimpanzees. While this gene appears to be highly pleiotropic, its link to skeletal development is well established (Unger et al. 2013). *MYH2* also appeared as a candidate in these lineages. This gene codes for a protein that is an essential component of myosin (Tajsharghi et al. 2005) and variants at this locus or the other skeletal/muscle related candidate genes may contribute to physical differences in *Pan*.

Infectious disease has been long thought to play an important role in human and hominid evolution. Concordant with this perspective, we found multiple genes involved in antiviral activity and immune function in the non-Western chimpanzees. These include *DDX60L*, *HERC5*, and *TMPRSS2*. We also note that balancing selection likely plays an equally, if not more important, role in the evolution of immune systems (Andrés et al. 2009). A number of important studies (Ferguson et al. 2012; Cagan et al. 2016; Cheng and DeGiorgio 2020) have provided valuable insight on such roles and work in this arena is very clearly just beginning.

Positive or purifying selection may not uniformly impact the entire coding sequence of a gene. Therefore, the consideration of selection per exon can identify previously unknown candidates and better pinpoint the region(s) under selection. We identified a number of specific exons that exhibited signatures of positive selection, some of which were detected in our gene-level analyses. The functions of three of the novel positive selection candidates are not well understood beyond some associations with cancer: *FSIP1*, *KIAA1755*, and *LRRC63*. One candidate gene revealed by our exon-level analysis, *MROH7*, was detected in eastern chimpanzees and may be related to reproduction (Kenigsberg et al. 2017). This class of genes, Maestro Heat-Like

Repeat Family Members, is generally not well understood; however, another family member *MROH8*, was detected in both eastern and central chimpanzees by our gene-level analysis and has been associated with Alzheimer's, suggesting brain-related functions for that locus and potentially other *MROH* genes (Potkin et al. 2009). We found another signature of positive selection at a reproduction related-gene (Tian et al. 2009) in the third exon of *NLRP8* in central chimpanzees. Finally, the eighteenth exon of *VPS13D* emerged as a positive selection candidate in eastern and western chimpanzees. Variants in this gene have been associated with neurological and movement effects and the locus appears involved in mitochondrial clearance (Anding et al. 2018; Gauthier et al. 2018; Seong et al. 2018).

Application of a complementary method, SnIPRE (Eilertson et al. 2012), to detect signatures from positive selection from MK data also yielded a few additional candidate genes, most of which do not even have a gene symbol identifier and are thus poorly characterized. *MUC17* was categorized as a candidate in all five lineages. This gene codes for mucins that protect epithelial cells (Moniaux et al. 2006) and variants at this locus were recently associated with endometriosis and infertility in Taiwanese women (Yang et al. 2015). Similar to *RNF213*, the deep divergence of *Pan* makes finding another gene with a shared selective signature across all descendent lineages quite puzzling. *RFPL4B* was identified across all four chimpanzee subspecies but the function of this gene and its associated phenotypes are not well understood.

While the absence of multiple shared genes across all four chimpanzee subspecies based on the traditional MK test and the identification of only one shared gene by SnIPRE is initially puzzling, this pattern can be explained by a number of factors. The neutral theory of molecular evolution posits that genetic diversity is a balance between mutation and drift (Kimura 1983). Accordingly, genetic drift is

inversely proportional to N_e . As such, the power to detect selection, particularly positive selection, seems to be particularly reduced for bonobos and western chimpanzees. Chimpanzees may share more genes under positive selection than reported here but low polymorphism in *P. t. verus* does allow for such patterns to be detected. This is also exacerbated by our strict filtering to reduce the number of false positives. Reducing our row/column sum to values less than five SNPs, immediately decreases the number of genes that can be evaluated in each lineage. Yet, we favor this more conservative approach used here and suggest that potential positive selection candidates present in other lineages but missing from *P. t. verus* be examined using a larger sample and with different methods.

These data may also or conversely suggest that the selection signatures described here are older than signatures from selective sweeps, yet far younger than the time period immediately following the estimated *Pan* divergence (~ 1.88 Ma). This may point to rapid and strong divergence in the chimpanzee subspecies. A recent analysis of selective sweeps in these lineages revealed that soft sweeps make up a substantial proportion of recent positive selection (Brand et al. 2021). It is quite possible that if these evolutionary processes operated similarly for the past ~ 2 Ma, then rapid adaptation could be more likely.

Along with other genome-wide analyses, MK tests may result in a higher number of false positives due to multiple hypothesis testing. That is, there is a 5% chance that non-significant data could be deviant enough due to random chance to produce a statistically significant result. This probability increases with the number of tests run. Therefore, the significant results from a set of MK tests may be composed of both true positives and false positives. The false discovery rate reflects the proportion of false positives among all significant results. Typically, one adjusts (or

corrects) α to address false positives; however, given the nature of the data used in MK tests, this is not typically done (e.g., Begun et al. 2007; Cagan et al. 2016). This is because such a correction might result in true positives (i.e., genes under positive selection) being missed based on an adjusted α . For example, application of a Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) to these data using a 25% false discovery rate results in zero significant results for both negative and positive selection across all lineages. A critical second step is to subsequently examine candidate gene sequences for those genes deemed statistically significant. In addition, because significant results are used in MK tests to only identify possible candidate genes and these are then further evaluated, this second step reduces potential impact of false discoveries while minimizing the damaging effect of excluding true positives through a conservative correction. Subsequent examination of gene sequences, therefore, allows for a confident assessment of true positives in the candidate gene list.

As described above, the MK test is generally robust to specific aspects of demography when compared other methods. However, differences in more recent N_e , reflected in the polymorphism data, vs. N_e during divergence, reflected in the substitution data, violates the premise of equivalent neutral mutation rates because the effective neutral mutation rate is related to N_e . Eyre-Walker (2002) examined the conditions under which population size differences yielded false signals of adaptation. When there is no selection on synonymous codon use and the population size changed recently a 3-fold increase can generate a false signature, whereas size differences further back in evolutionary time requires changes of even larger magnitude (Eyre-Walker 2002). Curiously, when there is selection on synonymous codon use, it becomes more difficult to generate a false signal of positive selection; however, this

does not appear to be the case in humans. In light of these findings and recent estimates for *Pan* population history, we conclude that differences in N_e were not sufficient to generate widespread false positives in this study.

Another MK test model assumption is that selection coefficients do not vary over time at a locus. The environmental variability of the Plio-Pleistocene in Africa (deMenocal 2004) would suggest that this assumption is unlikely for many taxa, including *Pan*. Yet, there does not appear to be a consensus on how fluctuating selection coefficients would shape MK tests. Huerta-Sanchez et al. (2008) found that variation in s increased the ratio of substitutions to polymorphisms; mimicking the signature of positive, directional selection. Gossmann et al. (2014) replicated this finding, but, they also reported that mutations that contribute to divergence and polymorphism tend to be net positive over their lifetimes. Those authors conclude that such adaptive signatures are genuine but rates of adaptation are likely underestimated when s fluctuates.

Finally, we note that a number of genes were not analyzed in the present study due to suboptimal read depth across the entire sample. As this issue is inherent to genome-wide analyses, complementary subsequent candidate gene analyses for loci of particular interest are warranted (e.g., *AVPRIA*). Additionally, genes located on the sex chromosomes could not be tested here without violating model assumptions. Different methods are needed to identify signatures of positive selection at deeper time scales on X and Y chromosome genes.

This analysis highlights candidate genes that may be involved in lineage divergence in *Pan*, including genes related to the brain, immunity, musculature, reproduction, and skeletal system. Further analysis of *TSHR* in bonobos is warranted, particularly given the developmental differences in T3 between bonobos and

chimpanzees. The absence of multiple genes unique to chimpanzees may point to deep divergence in common chimpanzees and may be supported by phenotypic variation observed within chimpanzees.

CHAPTER III

SOFT SWEEPS PREDOMINATE RECENT POSITIVE SELECTION IN BONOBOS (*PAN PANISCUS*) AND CHIMPANZEES (*PAN TROGLODYTES*)

Frances White, Nelson Ting, Timothy Webster, and I conceived of this analysis. The assembly and mapping of the genomic data in this analysis was conducted by Timothy Webster. I performed the data analyses and wrote the initial draft of the manuscript. Frances White, Nelson Ting, Timothy Webster, and I edited the manuscript.

Introduction

The identification of adaptive traits and their genetic basis is one of the central goals of evolutionary biology. Two approaches, top-down and bottom-up, have been used to accomplish this goal; the latter of which leverages population-level data to recognize the genomic signatures of positive selection (Barrett and Hoekstra 2011). At the genomic level, the process of adaptation results in a window of reduced variation that erodes over time. As these signatures do not persist, they can only be used to infer selection over a particular time scale in a population. In most species, this time frame is restricted to a few thousand generations, and roughly ~ 200,000 years in humans (Oleksyk et al. 2010). The classic model for positive selection for a given locus proposes that a single, novel mutation, that confers a fitness advantage (i.e., a beneficial allele) will rapidly spread in a population and eventually reach fixation (Maynard Smith and Haigh 1974). Neutral polymorphism adjacent to the novel allele will ‘hitchhike’, resulting in a distinct pattern of reduced genomic

diversity at the locus and surrounding sites. The term ‘hard sweep’ has been used to identify this pattern and process.

‘Soft sweeps’ describe the presence of two or more haplotypes that occur at intermediate frequencies (Hermisson and Pennings 2005). Thus, the signature of a soft sweep is intermediate to those of neutral or ‘background’ genomic variation and the signature of a hard sweep. This pattern can result from recurrent *de novo* mutations followed by positive selection. Alternatively, soft sweeps can also result from positive selection on standing genetic variation where alleles were already present in a population before selection. This variation may be the result of independent mutations (multiple origin soft sweep) or when an adaptive allele arose before selection, but multiple copies have subsequently swept through the population (single origin soft sweep). Soft sweeps are often incorrectly viewed synonymously with standing genetic variation; hard sweeps can emerge from standing genetic variation if a single copy of the beneficial allele was the ancestor of all beneficial alleles in a sample (Hermisson and Pennings 2017).

Hard and soft sweeps are locus-specific and, thus, not mutually exclusive across a genome. Unsurprisingly, soft sweeps are also much more difficult to recognize than hard sweeps because their genomic patterns are intermediate. Additionally, the identification of selective sweeps, hard or soft, is further complicated by the possibility that neutral loci linked to either soft or hard sweeps may produce a false signature similar to that of a sweep (Schridder et al. 2015; Kern and Schridder 2018).

With these challenges in mind, a considerable amount of work has been dedicated to both developing robust methods to identify selective sweeps and also understanding the evolutionary parameters that determine hard or soft sweeps.

Mutation-limited scenarios are expected to exclusively produce hard sweeps because beneficial alleles rarely occur (Hermisson and Pennings 2017). Thus, the most important parameter for estimating the likelihood of hard vs soft sweeps is the population-scaled mutation rate: $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate. However, this single parameter can vary widely depending on the advantage of the beneficial allele, the effective population size, the size of the mutational target, and the timescale for adaptation (Messer and Petrov 2013; Hermisson and Pennings 2017). Therefore, adaptation across the genome for a given population can be simultaneously mutation-limited and non-mutation-limited (B.A. Wilson et al. 2014). While it has become clear that most populations will likely exhibit a mosaic of hard and soft sweeps (Hermisson and Pennings 2017), additional data on sweep type frequencies in various species are sorely needed to better tease apart which parameters may determine each of those frequencies.

Both species of the *Pan* genus represent important evolutionary models due to their phylogenetic proximity to humans. *Homo* and *Pan* diverged ~ 5 to 7 Ma (Sarich and Wilson 1967; Bradley 2008; Scally et al. 2012; Besenbacher et al. 2019) and the most recent estimates for the divergence of bonobos and chimpanzees range between 1 and 2 Ma (Prüfer et al. 2012; de Manuel et al. 2016). Four extant chimpanzee subspecies evolved from a chimpanzee common ancestor that split ~ 600 ka with both subsequent lineages further splitting: one ~ 250 ka and the other ~ 160 ka (de Manuel et al. 2016). These two species exhibit stark differences in aspects of their morphology, physiology, behavior, and ecology (Susman 1984; Goodall 1986; Wrangham 1986; Kano 1992; White 1996; Furuichi 2011; Nishida 2011; Stumpf 2011; Verena Behringer et al. 2014; Turley and Frost 2014; M.L. Wilson et al. 2014). Many of these distinguishing traits are inferred to have occurred shortly after

divergence, while much less is known about recent evolutionary processes in these lineages.

Understanding recent positive selection in *Pan* is intriguing because of the dynamic physical and social environments in which they evolved. Climatic variation across Africa is well-documented for the Pleistocene (deMenocal 2004) and has been proposed to drive the evolution of early *Homo* (Potts 1998; Antón et al. 2014), and such variation probably impacted other taxa throughout the Pleistocene, including the genus *Pan*. Chimpanzee populations living in more stable environments that were closer to Pleistocene refugia were recently described to exhibit less behavioral diversity than chimpanzees living in more seasonal habitats that are more distant to forest refugia (Kalan et al. 2020). While the formation of these refugia may have resulted in periods of habitat stability for some bonobo and chimpanzee populations during glacial periods (Takemoto et al. 2017; Barratt et al. 2020), climatic fluctuations throughout the Pleistocene likely affected both the physical environment—via changes in habitat structure and type—and the social environment—via changes in the frequency of dispersal and intergroup encounters. Further, evidence of admixture within extant and between extant and extinct members of the *Pan* genus adds even more variation to the social environments in which these apes evolved (Hey 2010; Wegmann and Excoffier 2010; de Manuel et al. 2016; Kuhlwilm et al. 2019). A dynamic environment may result in selection for multiple existing alleles, resulting in a greater frequency of soft sweeps than in a more stable environment where one would expect a greater frequency of hard sweeps.

In this study, we apply a recently developed supervised machine-learning approach to population-level genomic data for bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*) to assess the extent of different completed sweep

types in these species. While a few studies have examined recent positive selection in bonobos and chimpanzees (e.g., Cagan et al. 2016; Han et al. 2019; Schmidt et al. 2019; Kovalaskas et al. 2020; Nye et al. 2020), the role of hard and soft sweeps in shaping their adaptations is currently unknown. We sought to categorize genomic regions as subject to recent hard or soft sweeps, as linked to recent hard or soft selective sweeps, or as evolving neutrally. Data from simulations have predicted that hard sweeps would be common in humans because of our overall low mutation rate (Hermisson and Pennings 2017). Under this “mutation limitation hypothesis” and given the similarity in mutation rate between *Homo* and *Pan*, one could predict that bonobos and chimpanzees should also exhibit a high degree of hard sweeps. However, hard sweeps have been thought and observed to be quite rare in recent human evolution (Hernandez et al. 2011; Schrider and Kern 2017), although this perspective is debated (Jensen 2014; Harris et al. 2018). This could be explained by several non-mutually exclusive alternatives including demographic effects. Larger populations can have more standing variation for selection to act on (Hermisson and Pennings 2005) which may result in more soft sweeps, whereas bottlenecks can result in drift and thus potentially more hard sweeps if intermediate frequency haplotypes are lost (B.A. Wilson et al. 2014). For example, some human populations experienced recent demographic changes (e.g., Schiffels and Durbin 2014), such as a bottleneck upon leaving Africa (e.g., Henn et al. 2012). Indeed, Schrider and Kern (2017) found that hard sweeps were more frequent in non-African than African populations. Chimpanzees and bonobos have also experienced recent demographic changes, including in effective population size, within the time frame (< 200 ka) for selective sweeps, based on PSMC analyses (Prado-Martinez et al. 2013; de Manuel et al. 2016). Three of the five lineages appear to have declined, whereas the other two have

increased and then decreased. Under such changes in population size, the strength of selection plays a strong role in the likelihood of soft sweeps (B.A. Wilson et al. 2014). We therefore predicted that we would observe a higher frequency of soft sweeps in *Pan*, but that lineage-specific population histories might affect the degree to which soft sweeps dominate.

Methods

Genomic Data

We retrieved raw short read data on bonobos and all four chimpanzee subspecies from the Great Ape Genome Project (GAGP) (Prado-Martinez et al. 2013). This dataset contained high coverage genomes (<https://github.com/brandcm/Dissertation>: File S0: Figures S1, S2) from 13 bonobos (*P. paniscus*), 18 central chimpanzees (*P. troglodytes troglodytes*), 19 eastern chimpanzees (*P. t. schweinfurthii*), 10 Nigeria-Cameroon chimpanzees (*P. t. ellioti*), and 11 western chimpanzees (*P. t. verus*).

Read Mapping and Variant Calling

Initial quality assessments in fastqc (Andrews 2010) and multiqc (Ewels et al. 2016) indicated a number of quality issues, including failed runs, problematic tiles, and substantial variation in base quality. We removed adapters and trimmed all reads for quality with BBduk (<https://sourceforge.net/projects/bbmap/>). For trimming, we used the parameters “ktrim=r k=21 mink=11 hdist=2 qtrim=rl trimq=15 minlen=50 maq=20” for all reads and added “tpe and tpo” for paired reads.

We used XYalign (Webster et al. 2019) to create versions of the chimpanzee reference genome, panTro6 (Kronenberg et al. 2018), for male- and female-specific

mapping. Specifically, the version of the reference for female mapping has the Y chromosome completely masked, as its presence can lead to mismapping (Webster et al. 2019). We then mapped reads with BWA MEM (Li 2013) and used SAMtools (Li et al. 2009) to fix mate pairs, sort BAM files, merge BAM files per individual, and index BAM files. We use Picard (Broad Institute 2018) to mark duplicates with default parameters, before calculating BAM statistics with SAMtools. We next measured depth of coverage with mosdepth (Pedersen and Quinlan 2018), removing duplicates and reads with a mapping quality less than 30 for calculations. Visualizations for coverage and demography (see Generation of Simulated Chromosomes below) were created in R, version 3.6.3 (R Core Team 2020), using ggplot2, version 3.3.3 (Wickham 2016).

We used GATK4 (Poplin et al., unpublished data) for joint variant calling across all samples. We used default settings for all steps—HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs—with three exceptions. First, we turned off physical phasing for computational efficiency and downstream VCF compatibility with filtering tools. Second, because multiple samples in this dataset suffer from contamination from other samples both within and across taxa (Prado-Martinez et al. 2013), we employed a contamination filter to randomly remove 10% of reads during variant calling. This should have the effect of reducing confidence in contaminant alleles. Finally, we output non-variant sites to allow equivalent filtering of all sites in the genome and more accurate assessments of callability.

The above quality control, assembly, and variant calling steps are all contained in an automated Snakemake (Köster and Rahmann 2012) available on GitHub (https://github.com/thw17/Pan_reassembly). The repository also contains a Conda

environment with all software versions and origins, most of which are available through Bioconda (Grüning et al. 2018).

Variant Filtration and Genome Accessibility

We considered only autosomes for this analysis as the X and Y chromosome violate many of the assumptions for the following methods (Webster and Wilson Sayres 2016). We also excluded unlocalized scaffolds (N = 4), unplaced contigs (N = 4,316), and the mitochondrial genome from any downstream analyses. Additional filtration steps were completed using bcftools (Li 2011); command line inputs are provided in parentheses. Given our focus on selective sweeps, we only included single nucleotide variants (SNPs) (“-v snps”) that were biallelic (“-m2 -M2”). On a per sample basis within each site, we marked genotypes where sample read depth was less than 10 and/or genotype quality was less than 30 as uncalled (“-S . -i FMT/DP \geq 10 && FMT/GT \geq 30”). To ensure that missing data did not bias our results, we further excluded any sites where less than ~ 80% of individuals (N = 56) were confidently genotyped (“AN \geq 112”). We also removed any positions that were monomorphic for either the reference or alternate allele (“AC > 0 && AC \neq AN”). These filtration steps yielded 41,869,892 SNPs for our downstream analyses (<https://github.com/brandcm/Dissertation>: File S0: Table S2).

We considered sites in our sample with low to no coverage to be ‘inaccessible’ in the reference genome. Using the output of mosdepth (see Read Mapping and Variant Calling above), we identified and filtered sites exhibiting low coverage as defined above. We used the ‘maskfasta’ function in bedtools (Quinlan and Hall 2010) to mark these sites (N) in the panTro6 FASTA, featuring only the autosomes, for use

in downstream analyses. This resulted in 86.3% of the assembled autosomes as accessible (<https://github.com/brandcm/Dissertation>: File S17).

Generation of Simulated Chromosomes

We used the software ‘discoal’ to generate simulated chromosomes on which we trained a classifier per lineage (Kern and Schrider 2016). We generated a matching number of simulated haploid chromosomes for the sample size of each *Pan* lineage (i.e., 26 chromosomes for 13 *P. paniscus*, 20 chromosomes for 10 *P. t. ellioti*, etc.). Simulated chromosomes were set to 1.1 Mb in length and divided into 0.1 Mb subwindows for a total of 11 subwindows. These simulations included a population-scaled mutation rate ($4N\mu L$), where N is the effective population size, μ is the per base pair per generation mutation rate, and L is the length of the simulated chromosome. We used the median of the previously reported effective population size range per lineage (Prado-Martinez et al. 2013). As estimates of genome-wide mutation rates vary considerably and are complicated in that mutation rates vary across individual genomes, we based our parameter on a mutation rate of 1.6×10^{-8} , which falls between estimates from genome-wide data and phylogenetic estimates (Narasimhan et al. 2017). We introduced some variation in this rate by setting a lower and upper-bound to 1.5 and 1.7×10^{-8} and sampled a new mutation rate per simulation drawing from this uniform prior. All simulations also included a population-scaled recombination rate ($4NrL$), where r is the recombination rate per base pair per generation, again calculated from the median effective population size for each lineage from Prado-Martinez et al. (2013) and a recombination rate drawn from a uniform prior of $1.1 - 1.3 \times 10^{-8}$, based on the mean genome-wide rate (1.2×10^{-8}) reported for bonobos, chimpanzees, and gorillas (Stevison et al. 2015). Recent results

from a different selective sweep classifier, Trendsetter, suggest that including a range of recombination rates is important to reducing misclassification (Mughal and DeGiorgio 2019). We note that while some of the estimated recombination rates in bonobos and chimpanzees are beyond the uniform distribution used in our simulations, many of these values are the high rates present in the telomeres, regions that generally exhibit lower or no coverage and thus will be largely if not entirely masked from this analysis (see Variant Filtration and Genome Accessibility above). We also included a demographic string reflecting approximate changes in population size for each lineage between ~ 0.05 and 2 Ma. Changes in population size were set in units of $4N_0$ generations, N_0 was set to the approximate median effective population size from Prado-Martinez et al. (2013) and we used a generation time of 25 years (Langergraber et al. 2012). Population size changes for this time period were drawn from a previous PSMC analysis (de Manuel et al. 2016) (Figure S3). While this is only one study from which to draw demographic information and reconstructions of *Pan* demography vary widely across studies, the downstream program used to classify genomic windows, diploS/HIC, is robust to demographic misspecification (Kern and Schrider 2018). We generated 2×10^3 simulations using these parameters as a set of simulations under neutral evolution per lineage.

Hard and soft selective sweeps were simulated with all of the aforementioned parameters and using a uniform prior of population-scaled selection coefficients ($\alpha = 2Ns$) derived from each lineage's median effective population size (Prado-Martinez et al. 2013) and moderately weak to moderately strong selection coefficients between 0.02 and 0.05. Sweeps also included a parameter (τ) for the time to fixation of the beneficial allele over a uniform range in units of $4N$ generations. This value ranged from 0 to 0.001 for all lineages. Linked-hard and linked-soft sweeps were generated

by placing the selected site at the center of each of the 10 subwindows flanking the center (6th) subwindow. Additionally, we included a uniform prior on the frequency at which a mutation is segregating at the time it becomes beneficial for soft and linked-soft sweeps, setting this range from 0 to 0.2. We generated 1×10^3 simulations per subwindow for linked-hard and linked-soft sweeps ($N = 10$) and 2×10^3 simulations for hard and soft sweeps. This resulted in a total of 2×10^3 hard, 1×10^4 hard-linked, 2×10^3 soft, and 1×10^4 soft-linked simulated sweeps. Parameters for these simulations can be found here: <https://github.com/brandcm/Dissertation>: File S18.

Calculation of Simulation Feature Vectors and Classifier Training

We calculated feature vectors from these simulated chromosomes using the ‘fvecSim’ function in the program diploS/HIC (Kern and Schrider 2018). Briefly, diploS/HIC calculates 12 summary statistics for all 11 subwindows: π , Watterson’s θ , Tajima’s D , the variance, skew, and kurtosis of genotype distance (g_{kl}), the number of multilocus genotypes, J_1 , J_{12} , J_2/J_1 , unphased Z_{ns} , and the maximum value of unphased ω . Collectively, these summary statistics capture information about the site frequency spectrum (SFS), haplotype structure, and linkage disequilibrium (LD). diploS/HIC uses a convolutional neural network (CNN) to capture essential aspects of a feature (the feature vector) by sliding a receptive field over the image to compute dot product between the original filter and the convolutional filter. In diploS/HIC, the CNN uses three branches of a CNN, of which each has two dimensional convolutional layers with ReLu activations followed by max pooling. This is followed by a dropout layer to control for model overfitting. Outputs from all three units are fed into two fully connected dense layers, which also use dropout layers, before arriving at a softmax activation that outputs the probability for each categorical class (hard, hard-

linked, neutral, soft-linked, or soft). Complete details for this procedure can be found in Kern and Schrider (2018).

When calculating feature vectors for the simulated chromosomes, we used the optional arguments for the 'fvecSim' function to mask each simulation with 110,000 bp segment randomly drawn from our masked FASTA where > 0.25 of SNPs in a subwindow were accessible (i.e., not marked by Ns). This enabled us to train our classifiers on simulated data featuring the same patterns of inaccessible genomic regions that the classifier would encounter in the empirical data.

We created a balanced set with equal representation (2×10^3) of all five classes via sampling without replacement in which to train the classifier using diploS/HIC's 'makeTrainingSets' function. These were divided into 8,000 training examples, 1,000 validation examples, and 1,000 testing examples to test the accuracy of the classifier via the 'train' function in diploS/HIC. We built ten classifiers per lineage and selected the one with the highest accuracy to apply to the empirical data (<https://github.com/brandcm/Dissertation>: File 19).

A second, independent set of simulated chromosomes was generated per lineage using the same parameters. We then calculated feature vectors and created another balanced training set with 2×10^3 chromosomes per class (hard, linked-hard, neutral, linked-soft, and soft). We used diploS/HIC's 'predict' function by applying each trained classifier to all five classes separately per lineage. In other words, we ran each classifier on 2,000 simulated hard sweeps, 2,000 simulated linked-hard sweeps, 2,000 simulated neutral regions, 2,000 simulated linked-soft sweeps, and 2,000 simulated soft sweeps and for each lineage. We used a binary classification scheme, where the identification of a sweep (hard or soft) was considered to be positive and linked or neutral regions were negative, to assess the true positive rate, false positive

rate, and obtain a second estimate of accuracy for each trained classifier (<https://github.com/brandcm/Dissertation>: File S0: Tables S3-S6). We also calculated class-specific accuracy, by summing the number of instances per lineage where the predicted class matched the simulated class divided by the total (1×10^4) (<https://github.com/brandcm/Dissertation>: File S0: Tables S3-S6).

Empirical Data Feature Vectors and Prediction

Upon achieving > 0.8 accuracy, each trained classifier was applied to its respective *Pan* lineage. Each autosome was analyzed separately and feature vectors calculated using diploS/HIC's 'fvecVcf' function. We supplied this function with the masked FASTA for that chromosome and discarded windows where any subwindow had < 0.25 unmasked sites following Schrider and Kern (2017) (<https://github.com/brandcm/Dissertation>: File S20). This step reduces the potential effect of the number of SNPs in a given window on sweep classification. Finally, the trained classifier was applied to the feature vector files using the 'predict' function.

Sweep Identification, Potential Target Genes, and Gene Ontology

As diploS/HIC outputs the probability for each sweep class, we first report the class inferred to be the most likely. However, as the difference between the most likely class and the next most likely may be small, we further report windows where the sweep class probability is > 0.5 , > 0.75 , and > 0.9 (<https://github.com/brandcm/Dissertation>: File S21). We also examined our data for spatial patterns. Windows classified as immediately abutting other windows with the same sweep type for hard and soft sweeps were considered to be a single sweep. Unique sweep windows and those shared between two or more lineages were

visualized using Upset plots created with the ComplexUpset package, version 1.2.1 (Lex et al. 2014; Krassowski 2020) in R, version 3.6.3 (R Core Team 2020).

We examined what genes lie in the windows identified as being subject to a recent selective sweep by extracting the genomic coordinates of all autosomal coding regions for the longest transcript per gene (N = 20,119 genes) in the panTro6 genome via the panTro6 GFF (retrieved from:

https://www.ncbi.nlm.nih.gov/genome/202?genome_assembly_id=380228). We used the bedtools ‘intersect’ function (Quinlan and Hall 2010) to identify overlap between coding regions and candidate sweep windows after converting both CDS and sweep window coordinates to 0-start, half-open format. As some coding sequences may have been masked (see Variant Filtration and Genome Accessibility above), we extracted FASTAs for each coding sequence using bedtools ‘getfasta’ function (Quinlan and Hall 2010) and used a custom R script to calculate the percent of each gene that was masked. Overall, 66.2% of all coding sequence was unmasked. We excluded listing genes for candidate sweep regions if > 50% of the total coding sequence per gene was masked. Thus, we considered 13,228 genes as potential targets for selective sweeps (<https://github.com/brandcm/Dissertation>: File S3).

We investigated the enrichment of particular pathways by performing a gene ontology analysis using the Functional Annotation Tool in DAVID (Huang et al. 2008; Huang et al. 2009). We used the custom background described above (genes whose total coding sequence was > 50% unmasked) rather than all panTro6 genes to ensure our analysis was not underpowered. DAVID does not allow for official gene symbols to be used in a background list, so we converted gene symbols to Entrez gene IDs. As not all gene symbols have a corresponding Entrez gene ID, we removed genes for which there was no Entrez gene ID (N = 98 in background list). We collated genes

for both hard and soft sweeps into a single input per lineage. We evaluated statistical significance for biological process gene ontology terms via p-values adjusted using the Benjamini-Hochberg method (Benjamini and Hochberg 1995).

Scripts for all data analyses are available on GitHub (https://github.com/brandcm/Pan_Selective_Sweeps).

Data Availability

The raw data underlying this article are previously published (Prado-Martinez et al. 2013; de Manuel et al. 2016) and are available from the Sequence Read Archive (PRJNA189439 and SRP018689) and the European Nucleotide Archive (PRJEB15086).

Results

We generated four classifiers that reached an acceptable level of accuracy for bonobos (*P. paniscus*), central chimpanzees (*P. t. troglodytes*), eastern chimpanzees (*P. t. schweinfurthii*), and Nigeria-Cameroon (*P. t. ellioti*) chimpanzees. These classifiers ranged in accuracy from 85.6% (Nigeria-Cameroon chimpanzees) to 93.9% (central chimpanzees) (<https://github.com/brandcm/Dissertation: File S19>). We could not produce a sufficiently accurate classifier using realistic parameters for western chimpanzees (*P. t. verus*); therefore, they were excluded from downstream analyses. Following Kern and Schrider (2018), we calculated false positive rates by testing our classifiers on a second, independent set of simulated chromosomes per lineage. We used a binary classification, considering the identification of either sweep type as a positive and identification of a linked or neutral region to be negative. Our trained classifiers had considerable statistical power (1 - false positives) ranging from 96.6 to

99.2% and a low false positive rate (false positives / false positives + true negatives) that ranged from 1.4 to 4.3% across all four classifiers (Tables S2 - S5). When considered separately—i.e., true positives only included one sweep type (hard or soft) rather than both—we had greater power to detect hard sweeps than soft sweeps, averaging 99% and 96.9% across lineages, respectively (Tables S2 - S5). Accuracy (true positives + true negatives / total) for identifying sweep regions vs non-sweep regions ranged from 94.1 to 98.3%. In addition to the initial class-specific accuracy generated during classifier training, a second estimate of class-specific accuracy ranged from 81.6 to 92.1% (Tables S2 - S5).

We classified ~ 91.6% of the assembled autosomes in each lineage (Table 5, <https://github.com/brandcm/Dissertation: File S0: Tables S3-S7>), even after masking for inaccessible regions and excluding windows with few SNPs. We found that soft sweeps were abundant in all four lineages, accounting for > 73% of all individual sweeps, whereas hard sweeps were relatively rare (Table 5, <https://github.com/brandcm/Dissertation: File S22>). This pattern held true even when more stringent posterior probabilities were applied to consider a region a sweep and at least 30% of hard sweep windows and 76% of soft sweep windows were called with 50% or greater posterior probability (<https://github.com/brandcm/Dissertation: File S21>). Genomic regions linked to sweeps were also quite pervasive in all four lineages (Table 5); particularly among eastern chimpanzees, where roughly 86% of the genome was classified as linked to selective sweeps.

Table 5. Selective sweep summary per population.

Lineage	Number / Percent of Windows per Class Type						Number and Percent of Sweep Type		
	Hard	Linked-hard	Neutral	Linked-soft	Soft	Total	Hard	Soft	Total
<i>P. paniscus</i>	85 (0.4%)	1,576 (6.5%)	7,488 (30.8%)	13,168 (54.1%)	2,002 (8.2%)	24,319	81 (4.9%)	1,585 (95.1%)	1,666
<i>P. t. ellioti</i>	573 (2.4%)	6,358 (26.1%)	1,389 (5.7%)	14,498 (59.6%)	1,505 (6.2%)	24,323	488 (26.9%)	1,323 (73.1%)	1,811
<i>P. t. schweinfurthii</i>	32 (0.1%)	696 (2.9%)	1,835 (7.5%)	20,179 (83.0%)	1,581 (6.5%)	24,323	32 (2.3%)	1,376 (97.7%)	1,408
<i>P. t. troglodytes</i>	224 (0.9%)	1,746 (7.2%)	5,483 (22.5%)	15,121 (62.2%)	1,749 (7.2%)	24,323	184 (10.6%)	1,557 (89.4%)	1,741

We examined overlap in windows classified as either a hard or soft sweep across lineages, which may reflect either ancestral or parallel adaptation. Most hard sweep windows were unique to each lineage; however, we did find some shared windows across lineages (Figure 2). Central and Nigeria-Cameroon chimpanzees shared the highest number of sweep windows ($N = 33$) but when weighted by the total possible number of windows, the highest overlap for hard sweeps was between eastern and Nigeria-Cameroon chimpanzees ($7/32$ or ~ 0.21). No hard sweeps windows were shared across all lineages. Like hard sweeps, most soft sweep windows were also unique to each lineage (Figure 3). Among pairs of lineages there was remarkable consistency in the number of shared soft sweep windows ($N = 111-147$), even when the total possible number of shared windows is considered. One exception is eastern and central chimpanzees who shared nearly twice the number of soft sweep windows ($N = 267$). The highest number of shared soft sweep windows between three lineages occurred in the three chimpanzee subspecies ($N = 80$). Only 19 windows were shared across all four lineages.

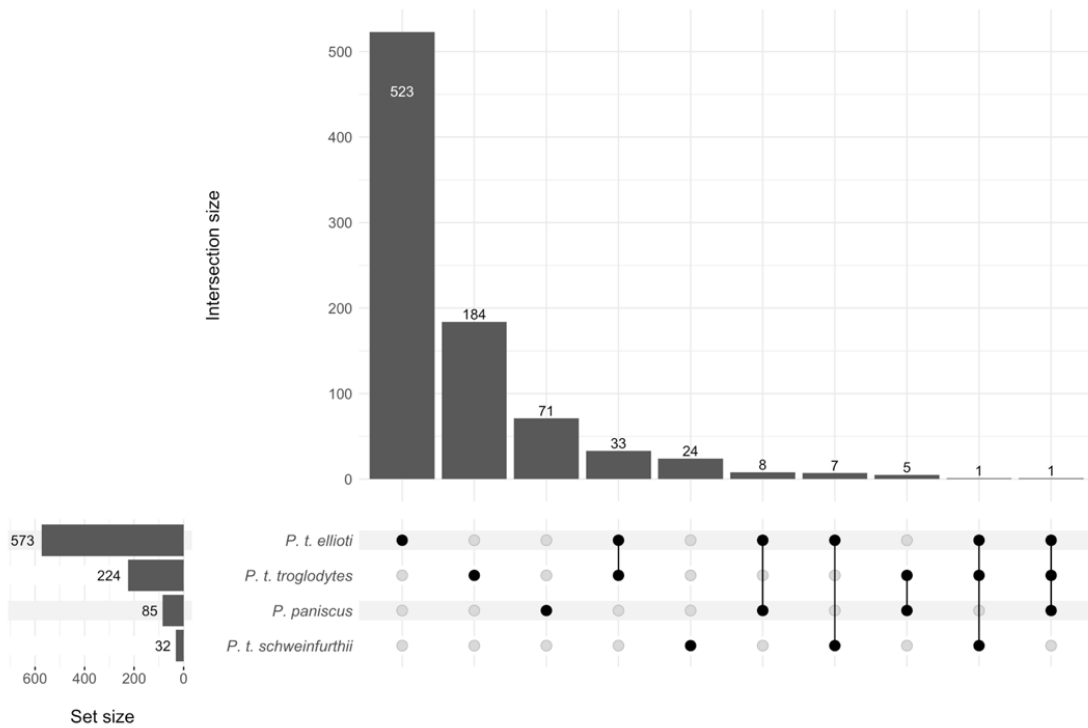


Figure 2. Unique and shared hard sweep windows. The frequency of windows shared by two or more lineages should be considered relative to the total possible number of shared windows (i.e., the set size of the lineage with the smallest set size).

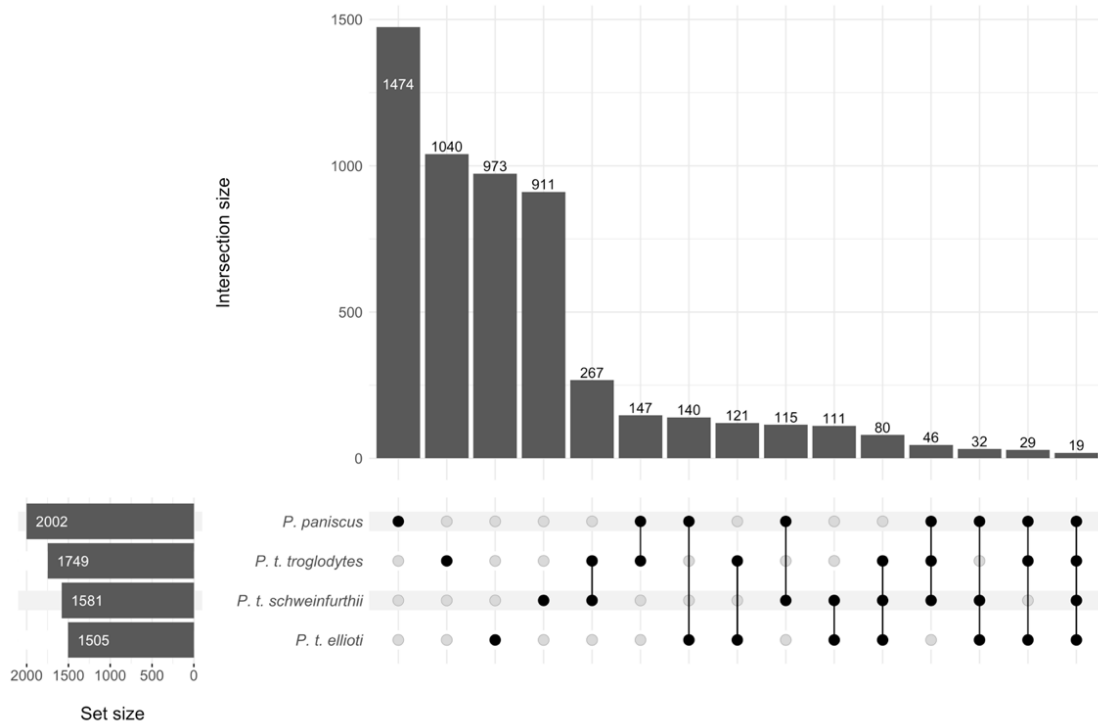


Figure 3. Unique and shared soft sweep windows. The frequency of windows shared by two or more lineages should be considered relative to the total possible number of shared windows (i.e., the set size of the lineage with the smallest set size).

After excluding genes that were > 50% masked, we identified 1,671 candidate genes in bonobo hard and soft sweeps, 1,761 genes in central chimpanzee sweeps, 1,372 genes in eastern chimpanzee sweeps, and 1,844 genes in Nigeria-Cameroon chimpanzee sweeps (<https://github.com/brandcm/Dissertation>: File S23). After correcting for multiple testing using the Benjamini-Hochberg method across all lineages, we identified only two significantly enriched pathways in central chimpanzees: nervous system development and central nervous system development (<https://github.com/brandcm/Dissertation>: File S24).

Discussion

Our study contributes to the emerging picture of recent evolution in *Pan* and adaptation more broadly. Contrary to the predictions of a mutation-limitation hypothesis, yet concordant with recent results for humans (e.g., Hernandez et al. 2011; Schrider and Kern 2017) and flies (Garud et al. 2015), we find soft sweeps to overwhelmingly predominate regions of the genome experiencing selective sweeps in both bonobos and the three chimpanzee subspecies we could analyze. These results confirm the prediction from Schmidt et al. (2019) who speculated that soft sweeps played a major role in the evolution of eastern and central chimpanzees. Those authors also posit that hard sweeps should be more frequent in western chimpanzees relative to other subspecies because of their low effective population size. While western chimpanzees are estimated to have the lowest effective population size, it is estimated to be only slightly lower than that of bonobos for which we found a high

number (95.1%) of soft sweeps (e.g., Prado-Martinez et al. 2013; de Manuel et al. 2016). It is curious that Nigeria-Cameroon chimpanzees exhibit the most hard sweeps in this analysis. While this could be the result of a multitude of factors, it is particularly interesting because this lineage has experienced a rather stable effective population size in recent evolutionary time as estimated by PSMC (Prado-Martinez et al. 2013; de Manuel et al. 2016), whereas a scenario with dramatic population decline would be expected to “harden” soft sweeps as haplotypes are stochastically lost, resulting in more hard sweeps (B.A. Wilson et al. 2014).

Our analysis of shared hard and soft sweeps found that most sweeps of both types were unique to each lineage. However, there was a high number of hard sweep windows shared between central and Nigeria-Cameroon chimpanzees as well as between eastern and Nigeria-Cameroon chimpanzees when the total possible number of shared sweeps was considered. Further, there were nearly twice the number of shared soft sweep windows shared between eastern and central chimpanzees. These results are similar to another recent study that found a large number of candidate sweep regions to be shared between those taxa (Nye et al. 2020). It is impossible to discern whether or not the overlap in hard sweeps between central and Nigeria-Cameroon chimpanzees and the overlap in soft sweeps for eastern and central chimpanzees is the result of shared ancestry and/or similar environmental conditions because both pairs of lineages share a geographic boundary: the Ubangi river for eastern and central chimpanzees and Sanaga river for central and Nigeria-Cameroon chimpanzees. The overlap in hard sweeps between eastern and Nigeria-Cameroon chimpanzees is more puzzling because they are not sister taxa and share a common ancestor ~ 600 ka (de Manuel et al. 2016). Therefore, parallel adaptation via similar physical and/or social environments may serve as a more likely hypothesis. While the

lowest in overall frequency, we also identified a number of soft sweep windows that were shared across three lineages as well as 19 windows that occurred in all four.

Future work should further investigate these shared sweep windows.

As mentioned above, soft sweeps are not exclusively the result of selection on standing genetic variation (Pennings and Hermisson 2006a; Pennings and Hermisson 2006b). However, given the estimated mutation rate for bonobos and chimpanzees, it appears unlikely that recurrent *de novo* mutations explain the majority of these soft sweeps. We did not explicitly model for different types of soft sweeps in our analysis. However, while soft sweeps from standing genetic variation and *de novo* mutations may exhibit similar genomic signatures, this must be tested before any additional conclusions are drawn. Hartfield and Bataillon (2020) recently suggested differences in diversity (as measured by π) at the selected locus may be used to differentiate soft sweep types, although this may be more difficult to accomplish in outcrossing species. Nonetheless, our results reveal a major role of standing genetic variation, and thus changes in the physical and social environment, in driving recent adaptations in *Pan*.

A few recent studies have considered the impact of effective population size on adaptive evolution in the great apes (Cagan et al. 2016; Nam et al. 2017). Theory predicts that the rate of adaptive evolution should be positively correlated with effective population size when $N_e s$ is $\gg 1$ (Gossmann et al. 2012). Both Cagan et al. (2016) and Nam et al. (2017) found a positive association between effective population size and the rate of adaptive evolution, measured by proportion of adaptive substitutions and the number of selective sweeps, respectively. However, we observed no clear linear relationship between the number of sweeps (hard, soft, or both) estimated from this analysis and the estimated effective population sizes for these four lineages (see <https://github.com/brandcm/Dissertation>: File S18 for population sizes).

This descriptive result should be considered cautiously because of the limited number of lineages analyzed here and the potential confounding effect of phylogeny. It is possible that this relationship may not be driven by the number of sweeps, but rather the strength of sweeps a population experiences (Nam et al. 2017). Estimates of selection strength are generally lacking for the great apes so this relationship remains a question for further study.

In addition to characterizing broad patterns in the genomic landscape for bonobos and chimpanzees, the results of this study also highlight thousands of candidate regions and genes for further analysis. We also find additional support for previous selection candidates. For example, disease has been long thought to shape evolution in primates (Nakajima et al. 2008; van der Lee et al. 2017). The potential for disease transmission between non-human primates and humans has also prompted much research, particularly focusing on the genomic underpinnings of host responses to lentiviruses, which include HIV and SIV (Gao et al. 1999; Van Heuverswyn et al. 2006; Compton et al. 2013; Nakano et al. 2020). Cagan and colleagues (2016) found evidence of recent positive selection within *IDO2*, a T-cell regulatory gene, among all four-chimpanzee subspecies and bonobos. We identified a candidate soft sweep region for eastern chimpanzees that overlaps this gene. However, this window had one of the lowest posterior probabilities in this lineage (49.7%) and there was a nearly equally high probability that this window was linked to a soft sweep (43.8%). Clearly, additional work is needed to understand the potential role of *IDO2* in *Pan* evolution. Schmidt et al. (2019) recently described three chemokine receptor genes—*CCR3*, *CCR9*, and *CXCR6*—had a significant number of highly differentiated SNPs in central chimpanzees. We could evaluate all three of these genes in our analysis but only one fell within a candidate sweep window: *CXCR6*. The window containing this gene was

confidently called as a soft sweep with a posterior probability of 85.5%. It is not known as to whether or not SIV_{cpz} uses *CXCR6* to enter chimpanzee host cells (Wetzel et al. 2018). However, multiple lines of evidence for selection either at this locus or within the window overlapping this gene prompt a closer examination of this genomic region. Finally, *TRIM5* fell within a hard sweep window in central chimpanzees. *TRIM5* is a well-known retrovirus restriction factor that appears subject to ancient, multi-episodic positive selection in primates (Sawyer et al. 2005).

Recent attention has focused on admixture between lineages in the genus *Pan* and the potential adaptiveness of introgressed genomic elements. de Manuel and colleagues (2016) identified 221 genes that fell within putatively introgressed elements in central chimpanzees from admixture with bonobos. Some of this admixture is estimated to occur < 200 ka, thus within the timeframe that the present analysis can detect selective sweeps. While we could not evaluate six of these 221 genes, five fell within candidate sweep regions in central chimpanzees from our study: *CDK8*, *EIF4E3*, *GRID2*, *PTPRM*, and *TRIM5*. As described above, *TRIM5* was unique to central chimpanzees. We found *CDK8* in sweep windows for bonobos, eastern chimpanzees, and Nigeria-Cameroon chimpanzees. In humans, *CDK8* mutations have been associated with multiple phenotypic effects including hypotonia, behavioral disorders, and facial dysmorphism (Calpena et al. 2019). We also identified *EIF4E3* in candidate sweeps for bonobos whereas *GRID2* and *PTPRM* were found in eastern chimpanzees. *EIF4E3* is a translation initiation factor (Osborne et al. 2013) while *PTPRM* is a member of the protein phosphatase family (PTP) and has multiple functions including cell proliferation and differentiation (Sun et al. 2012). *GRID2* generates ionotropic glutamate receptors and mutations have been associated with abnormalities of the cerebellum (Lalouette et al. 1998).

The gene ontology analysis produced only two statistically significant terms, nervous system development and central nervous system development, for a single *Pan* lineage: central chimpanzees. While cognitive and neurological differences are widely considered to differentiate bonobos and chimpanzees (e.g., Rilling et al. 2012; Stimpson et al. 2016; Staes et al. 2019), we are unaware of any studies that investigate variation among chimpanzee subspecies that may explain enrichment for nervous system and central nervous system development related genes specifically in central chimpanzees. We note that compared to other gene ontology analyses, our level of enrichment is quite low. While we excluded a large number of genes from our analysis due to poor coverage, our use of a custom background should increase, rather than decrease, statistical power.

The results from our analysis should be interpreted with some caution. First, while our classifiers achieved a high degree of accuracy, it is possible that some selective sweeps in each lineage were not detected or regions were incorrectly identified as such (Tables S2 - S5). We also note that we did not model small selection coefficients ($s < 0.02$) as we could not accurately classify sweeps under weak selection, which may be the result of the large window size (1.1 Mb) used here. One consequence may be that if weakly beneficial hard sweeps are present in the empirical data, they may have been sometimes classified as soft (Harris et al. 2018). Nonetheless, our classifiers were overall quite good at identifying moderately selected hard and linked-hard sweeps with both at approximately 95% accuracy across all lineages. Neutral and linked-soft regions were the most difficult to recognize with neutral regions typically being classed as soft-linked when they did not appear neutral. This suggests that the neutral portion of the genome for each lineage is slightly underestimated here. Finally, some moderately selected soft sweeps were

identified as hard sweeps in each of our classifiers, suggesting that some portion of identified hard sweeps in each lineage are, in fact, soft sweeps. The low false positive rates demonstrate the overall accuracy of the observed genomic patterns (i.e., the proportion of hard and soft sweeps) for these taxa. However, this point underscores the need to conduct subsequent analyses of the candidate regions and genes to confirm such the proposed mode of adaptation and investigate any functional consequences of that adaptation. In the ‘era of -omics’, the generation of candidate regions for any type of selection across populations and species appears to overwhelmingly outpace the confirmation of such patterns. Avenues of research that investigate these candidate genes in more detail are thus well poised to provide a deeper and more accurate understanding of lineage-specific adaptations.

Second, background selection, the loss of a linked neutral site from purifying selection on a deleterious allele, can potentially mimic patterns of selective sweeps and thus may impact the results of this study (Charlesworth et al. 1993). We did not explicitly model background selection in our analysis; however, evidence from simulations in various taxa demonstrate that this pattern of selection does not substantially increase the rate of false positives in selective sweep analyses (Schridder and Kern 2017; Schridder 2020). Further, Nam et al. (2017) considered the effect of background selection on genomic diversity in extant apes, including all five *Pan* lineages, and note that background selection alone does not produce the observed diversity reduction near genic regions in these lineages. While background selection may not largely affect certain selective sweep analyses, it may impact estimations of demography that are inferred using PSMC/MSMC approaches (Johri, Riall, et al. 2020; Johri, Charlesworth, et al. 2020). The demographic strings calculated from PSMC used in this analysis also broadly agree in population size shape with other

demographic estimates generated using other methods (e.g., Becquet and Przeworski 2007; Hey 2010), therefore, background selection unlikely affects the demographic models used in analysis. Yet, this issue should be strongly considered in future studies where demography is only inferred from PSMC/MSMC.

Further, sampling bias can reduce the accuracy of identifying selective sweeps. If multiple haplotypes are present in a population but only individuals sharing one haplotype are sampled, then the sweep would be classified as a hard sweep when it is a soft sweep. However, this scenario would only underestimate the degree of recent adaptation from soft sweeps. Therefore, if this sampling bias is present in this analysis, then soft sweeps may predominate recent *Pan* evolution to an even larger degree than described here. Population structure adds further complications to the classification of hard sweeps. Parallel adaptation produces multi-origin soft sweeps at the global population level that would appear to be hard in local populations, although even local samples may sometimes appear to be soft sweeps (Ralph and Coop 2010). Thus, if samples stemmed from one or few local populations then global soft sweeps may be misclassified as hard. A previous analysis estimated the geographic origin of individuals used in this analysis (de Manuel et al. 2016). These authors found that individuals from both eastern and central chimpanzee populations were sampled from multiple countries across the geographic range for both subspecies. Therefore, any hard sweeps detected in these populations are likely accurate at the subspecies level. The precise geographic origin could not be assessed for any of the bonobos or all of the Nigeria-Cameroon chimpanzees used in this analysis (de Manuel et al. 2016). As such, sampling or geographic bias may partially explain the high degree of hard sweeps observed in Nigeria-Cameroon chimpanzees, if they were sampled from a smaller geographic area than the other subspecies. We encourage future studies to

consider this potential bias when hard sweeps are encountered in existing data and during study design.

This analysis focuses on signatures of positive selection at single loci. However, there is theoretical and empirical evidence that a number of adaptive traits have a complex, multilocus architecture (Pritchard et al. 2010; Yang et al. 2017; Bergey et al. 2018). For these polygenic traits, shifts in the physical or social environment might result in allele frequency changes at many loci, of which, according to models, few to none of which would reach fixation (Pritchard et al. 2010). This may, in part, explain why hard sweeps appear to be rare in humans and other species if it represents a dominant mode of adaptation in these taxa. Unfortunately, at this point, we lack the data and methods to investigate the extent of polygenic selection across the genome in many non-model taxa such as *Pan*. Another factor to consider is dominance. Here, we assumed advantageous alleles were codominant; however, there is evidence that dominance may influence patterns of selective sweeps when variants occur via *de novo* mutation or recurrent mutation (Hartfield and Bataillon 2020). It is also worthwhile to address that this analysis explicitly focused on modelling very recent completed selective sweeps. Another future avenue of study in these lineages is the identification of incomplete or partial sweeps using existing approaches (Ferrer-Admetlla et al. 2014; Vy and Kim 2015) as well as explicitly modelling both incomplete and complete sweeps to address potential “temporal misclassification” (Zheng and Wiehe 2019).

Finally, while our approach to identifying hard and soft sweeps is a logical first step, future work should consider sweeps within subspecies to assess population-level (i.e., local), rather than lineage-specific (i.e., global) adaptations. This is underscored by the extensive phenotypic variation among chimpanzees, particularly

that of behavioral variation, which includes key characteristics that are often used to dichotomize bonobos and chimpanzees (Wilson et al. 2014). Further investigation is also clearly warranted in bonobos, whose overall phenotypic variation is likely underappreciated compared to chimpanzees (Hohmann and Fruth 2003b; Sakamaki et al. 2016; Beaune et al. 2017; Wakefield et al. 2019).

This study highlights the importance of changes in physical and/or social environment via soft selective sweeps in the recent evolution of our closest living relatives, chimpanzees and bonobos. Our results also yield further support for the ubiquity of soft, rather than hard, sweeps in adaptation. We contribute candidate regions and genes that may help identify unique phenotypes in each *Pan* lineage. Our findings also prompt many new questions including the estimation of selection strength coefficients and the degree of haplotypic diversity in candidate sweep regions. While our study focuses on these lineages broadly, this point also underscores the need for high-coverage genomic data collected using non-invasive methods at more local geographies.

CHAPTER IV

ESTIMATION OF *PAN* DEMOGRAPHY FROM SITE PATTERNS

Frances White, Alan Rogers, Timothy Webster, and I conceived of this analysis. The assembly and mapping of the genomic data in this analysis was conducted by Timothy Webster. Alan Rogers provided some code for this analysis. I performed the data analyses and wrote the initial draft of the manuscript. Frances White, Alan Rogers, Timothy Webster, and I edited the manuscript.

Introduction

The study of hybridization and admixture has a deep history, particularly for plants. This research not only contributes to our broader understanding of evolutionary processes but can shed light on past environmental conditions or population ranges that facilitated such admixture. Further, introgression can introduce novel advantageous alleles into a population on which positive selection can act (Hedrick 2013). This adaptive introgression is potentially faster than positive selection acting on *de novo* mutations, although it may be slower than adaptation from standing genetic variation (Hedrick 2013). Recent work using whole genome sequencing data points to the increasing ubiquity of introgression in the evolutionary history of large mammals, including hominins (Wall and Hammer 2006; Browning et al. 2018; Villanea and Schraiber 2019; Gokcumen 2020; Rogers et al. 2020), elephants (Palkopoulou et al. 2018), and bears (Cahill et al. 2015).

Our closest living relatives, bonobos (*Pan paniscus*) and chimpanzees (*P. troglodytes*), have been long examined for genomic signatures of admixture. These species can hybridize in captivity (Vervaecke and Van Elsacker 1992) but wild

populations are completely separated by the Congo River. Chimpanzees are also poor swimmers (Angus 1971) and are afraid of water (Kano 1992). Interestingly, bonobos do not share this fear of water (Kano 1992) and are known to forage in swamps (Uehara 1990; Hohmann et al. 2019). The geographic distribution of *Pan* prompted early speculation that the formation of this river, which was dated at the time to ~ 1.5 - 3.5 Ma, coincided with or prompted speciation in this genus (Horn 1979; Beadle 1981; Myers Thompson 2003). However, some early genetic studies of *Pan* speciation consistently dated their divergence to younger than 1.5 Ma (Won and Hey 2005; Becquet and Przeworski 2007; Caswell et al. 2008, but see Stone et al. 2002; Yu et al. 2003; Wegmann and Excoffier 2010). Further, it now appears that the Congo River is considerably older than previously thought, possibly up to 34 Ma (Leturmy et al. 2003; Lucazeau et al. 2003; Anka et al. 2010). Admixture between bonobos and chimpanzees would thus require a connection between the north and south banks of the Congo River unless a *Pan* population ranged south enough to travel around the headwaters of the Congo River, although the distance makes this scenario less likely assuming the historical ranges of *Pan* are identical to their current ranges (Takemoto et al. 2015). We note that the impermeability of this geographic barrier is partially a function of river discharge, which can vary widely in both space and time. There is some evidence that river discharge has varied in the recent past, which could create an opportunity for bonobos and chimpanzees to diverge and subsequent opportunities for gene flow (Takemoto et al. 2015). This scenario is the most plausible given the current evidence. Such riverine barriers also separate three of the four chimpanzee subspecies while western chimpanzees occur west of a large forest-savannah mosaic known as the Dahomey Gap (Lester et al. 2021). These rivers also likely experience

variation in discharge, which may facilitate introgression between geographically proximate subspecies.

Early analyses for gene flow in *Pan* yielded inconsistent results and did not include data from Nigeria-Cameroon chimpanzees. Won and Hey (2005) described evidence of gene flow from western chimpanzees into central chimpanzees which was reported in subsequent studies (Caswell et al. 2008; Hey 2010; Wegmann and Excoffier 2010). Western chimpanzees may have also admixed with other lineages earlier in *Pan* evolutionary history. For example, introgression may have occurred from western chimpanzees into the ancestor of eastern and central chimpanzees (Hey 2010). In contrast, Becquet and Przeworski (2007) did not find evidence of gene flow in *Pan* with the possible exception between bonobos and eastern chimpanzees.

Admixture between bonobos and both the ancestor of eastern and central chimpanzees and the ancestor of all chimpanzees has also been described (Wegmann and Excoffier 2010). Hey (2010) also noted eastern-central chimpanzee gene flow, suggesting it occurred from central chimpanzees into eastern chimpanzees. One model from this study also indicated gene flow between eastern and western chimpanzees (Hey 2010). This scenario is puzzling given the current *Pan* biogeography, but this assumes that the ranges of the four subspecies have remained relatively static since the chimpanzee common ancestor.

Admixture inference from whole genome sequences has replicated many of these conflicting earlier results. Prado-Martinez et al. (2013) was the first whole genome analysis to consider gene flow across all five lineages and noted admixture between eastern and Nigeria-Cameroon chimpanzees as well as eastern and western chimpanzees. Given the large number of demographic parameters needed to be estimated under a complex evolutionary history (i.e., many introgression events), de

Manuel et al. (2016) estimated parameters using two sets of populations: one that included all lineages except for western chimpanzees and one that included all lineages except for Nigeria-Cameroon chimpanzees. These authors found 1) additional evidence for bonobo introgression into chimpanzees, 2) evidence of earlier gene flow between bonobos and chimpanzees, and 3) evidence for admixture between chimpanzee lineages. The most robust evidence from introgression of bonobos into chimpanzees suggested two events: one that occurred between 200 and 550 ka into the ancestor of eastern and central chimpanzees and a second event < 180 ka, after the eastern and central chimpanzee split (de Manuel et al. 2016). Kuhlwilm et al. (2019) further detected introgression from an extinct *Pan* species into bonobos between 377 and 1,627 ka. This ghost lineage was estimated to diverge from the bonobo and chimpanzee common ancestor > 3 Ma.

Here, we apply a recently developed method to compare previously proposed models for *Pan* evolutionary history and estimate 1) divergence times, 2) effective population sizes, and the 3) timing and degree of introgression. This approach employs site pattern frequencies to infer deep population history by simultaneously estimating all parameters. There are a few advantages to this approach compared to other commonly used methods for demography. First, within-population variation is ignored and recent changes in population size therefore cannot affect analyses. This results in fewer parameters that must be estimated. Second, the uncertainty introduced by statistical identifiability (i.e., when more than one model fits the data well) that is commonly encountered when ascertaining complex demographies can be incorporated into confidence intervals via model averaging.

Methods

Genomic Data

We retrieved raw short read data on bonobos and all four chimpanzee subspecies from the Great Ape Genome Project (GAGP) (Prado-Martinez et al. 2013). This dataset contained high coverage genomes (<https://github.com/brandcm/Dissertation>: File S0: Figures S1, S2) from 13 bonobos (*P. paniscus*), 18 central chimpanzees (*P. troglodytes troglodytes*), 19 eastern chimpanzees (*P. t. schweinfurthii*), 10 Nigerian chimpanzees (*P. t. ellioti*), and 11 western chimpanzees (*P. t. verus*). We retrieved short read data on a high-coverage human female, HG00513, collected as part of the 1000 Genomes Project (Auton et al. 2015) to use as an outgroup sequence to determine ancestral alleles per locus (Biosample ID: SAME123526).

Read Mapping and Variant Calling

Initial quality assessments in fastqc (Andrews 2010) and multiqc (Ewels et al. 2016) indicated a number of quality issues, including failed runs, problematic tiles, and substantial variation in base quality. We removed adapters and trimmed all reads with BBduk (<https://sourceforge.net/projects/bbmap/>). For trimming, we used the parameters “ktrim=r k=21 mink=11 hdist=2 qtrim=rl trimq=15 minlen=50 maq=20” for all reads and added “tpe and tpo” for paired reads.

We used XYalign (Webster et al. 2019) to create versions of the chimpanzee reference genome, panTro6 (Kronenberg et al. 2018), for male- and female-specific mapping. Specifically, the version of the reference for female mapping has the Y chromosome completely masked, as its presence can lead to mismapping (Webster et al. 2019). We then mapped reads with BWA MEM (Li 2013) and used SAMtools (Li

et al. 2009) to fix mate pairs, sort BAM files, merge BAM files per individual, and index BAM files. We use Picard (Broad Institute 2018) to mark duplicates with default parameters, before calculating BAM statistics with SAMtools. We next measured depth of coverage with mosdepth (Pedersen and Quinlan 2018), removing duplicates and reads with a mapping quality less than 30 for calculations.

We used GATK4 (Poplin et al. 2018) for joint variant calling across all samples. We used default settings for all steps—HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs—with three exceptions. First, we turned off physical phasing for computational efficiency and downstream VCF compatibility with filtering tools. Second, because multiple samples in this dataset suffer from contamination from other samples both within and across taxa (Prado-Martinez et al. 2013), we employed a contamination filter to randomly remove 10% of reads during variant calling. This should have the effect of reducing confidence in contaminant alleles. Finally, we output non-variant sites to allow equivalent filtering of all sites in the genome and more accurate assessments of callability.

The above quality control, assembly, and variant calling steps are all contained in an automated Snakemake (Köster and Rahmann 2012) available on GitHub (https://github.com/thw17/Pan_reassembly). The repository also contains a Conda environment with all software versions and origins, most of which are available through Bioconda (Grüning et al. 2018).

Variant Filtration

We considered only autosomes for this analysis as the X and Y chromosome violate many of the assumptions for the following methods (Webster and Wilson Sayres 2016). We also excluded unlocalized scaffolds (N = 4), unplaced contigs (N =

4,316), and the mitochondrial genome from any downstream analyses. Additional filtration steps were completed using bcftools (Li 2011); command line inputs are provided in parentheses. We first normalized variants by joining biallelic sites and merging indels and SNPs into a single record (“norm -m +any”) using the panTro6 FASTA. We only included SNPs (“-v snps”) that were biallelic (“-m2 -M2”). On a per sample basis within each site, we marked genotypes where sample read depth was less than 10 and/or genotype quality was less than 30 as uncalled (“-S . -i FMT/DP \geq 10 && FMT/GT \geq 30”). To ensure that missing data did not bias our results, we further excluded any sites where less than ~ 80% of individuals (N = 56) were confidently genotyped (“AN \geq 112”). We also removed any positions that were monomorphic for either the reference or alternate allele (“AC > 0 && AC \neq AN”). While lack of or low coverage at a locus is problematic, loci with excessive coverage are also of concern. These sites may yield false heterozygotes that are usually the result of copy number variation or paralogous sequences (Li 2014). As our data exhibit a high degree of inter-individual and inter-chromosomal variation in mean coverage (Brand et al. 2021), we applied Li's (2014) recommendation for a maximum depth filter ($d + 4\sqrt{d}$) to the mean chromosomal coverage of the individual in our sample (*Pan* or *Homo*) with the highest coverage and excluded any loci that exceeded this value (“filter -e FMT/DP > $d + 4\sqrt{d}$ ”) (<https://github.com/brandcm/Dissertation: File S2>). These filtrations steps yielded between 2,413,791,600 and 2,493,198,004 SNVs for our downstream analyses (<https://github.com/brandcm/Dissertation: File S25>).

After filtration, we generated reference allele frequency (RAF) files for each population.

Null Model of Demography

We first constructed a null model with all five populations and no introgression events. As the topology of this model is well supported, we use it in each of the alternative models (see below). Demographic modelling was conducted using Legofit (Rogers 2019; Rogers et al. 2020; Rogers 2021). Legofit requires at least one “fixed” parameter to set the molecular clock, so we chose to set the divergence between bonobos and chimpanzees to the median value as estimated from de Manuel et al. (2016). This value (1.88 Ma) was input in generation units (75,200), based on a generation time of 25 years (Langergraber et al. 2012). While each of the remaining nodes were set with the median estimate from de Manuel et al. (2016), we designated these parameters to be “free” in order to generate parameter estimates. We also estimated population size by setting them to free and using rough estimates as initial values.

Alternative Demography Models

We then constructed a set of models using all subsets of previously described introgression events. We did not include events that would be uninformative for site pattern analysis (e.g., admixture between eastern and central chimpanzees, although such an event or state would simply broaden the confidence intervals for the divergence time parameter). We initially included four introgression events: α , β , γ , and δ . α denotes introgression from a ghost *Pan* lineage into bonobos (Kuhlwilm et al. 2019). β denotes introgression from bonobos into the ancestor of eastern and central chimpanzees (de Manuel et al. 2016). γ denotes introgression from the ancestor of eastern and central chimpanzees into Nigeria-Cameroon chimpanzees (de Manuel et al. 2016). δ denotes introgression from bonobos into central chimpanzees (de Manuel

et al. 2016). For models with multiple admixture events, we used the estimated order of events from oldest to youngest when naming the model (de Manuel et al. 2016; Kuhlwilm et al. 2019). We also reversed the direction of introgression for γ , γ_r , and considered whether this may yield a better fitting model. We did so because site patterns will reflect the net admixture such that bidirectional gene flow will yield a positive value if the direction of gene flow is correctly specified. As γ may have been bidirectional (de Manuel et al. 2016) we included models where the net introgression was larger in either direction (from Nigeria-Cameroon chimpanzees into the ancestor of eastern and central chimpanzees and vice-versa) to ensure that we considered the full range of possible scenarios. Finally, we also considered gene flow from western chimpanzees into the ancestor of eastern and central chimpanzees (defined as ϵ), as suggested by demographic analyses of STRs (Hey 2010; Wegmann and Excoffier 2010). As the timing of this event relative to γ is unclear, we considered scenarios where either γ or ϵ was first and the other followed. The order of these events is reflected in the model name. In total, we considered 42 demographic models including the null model (Figure 4).

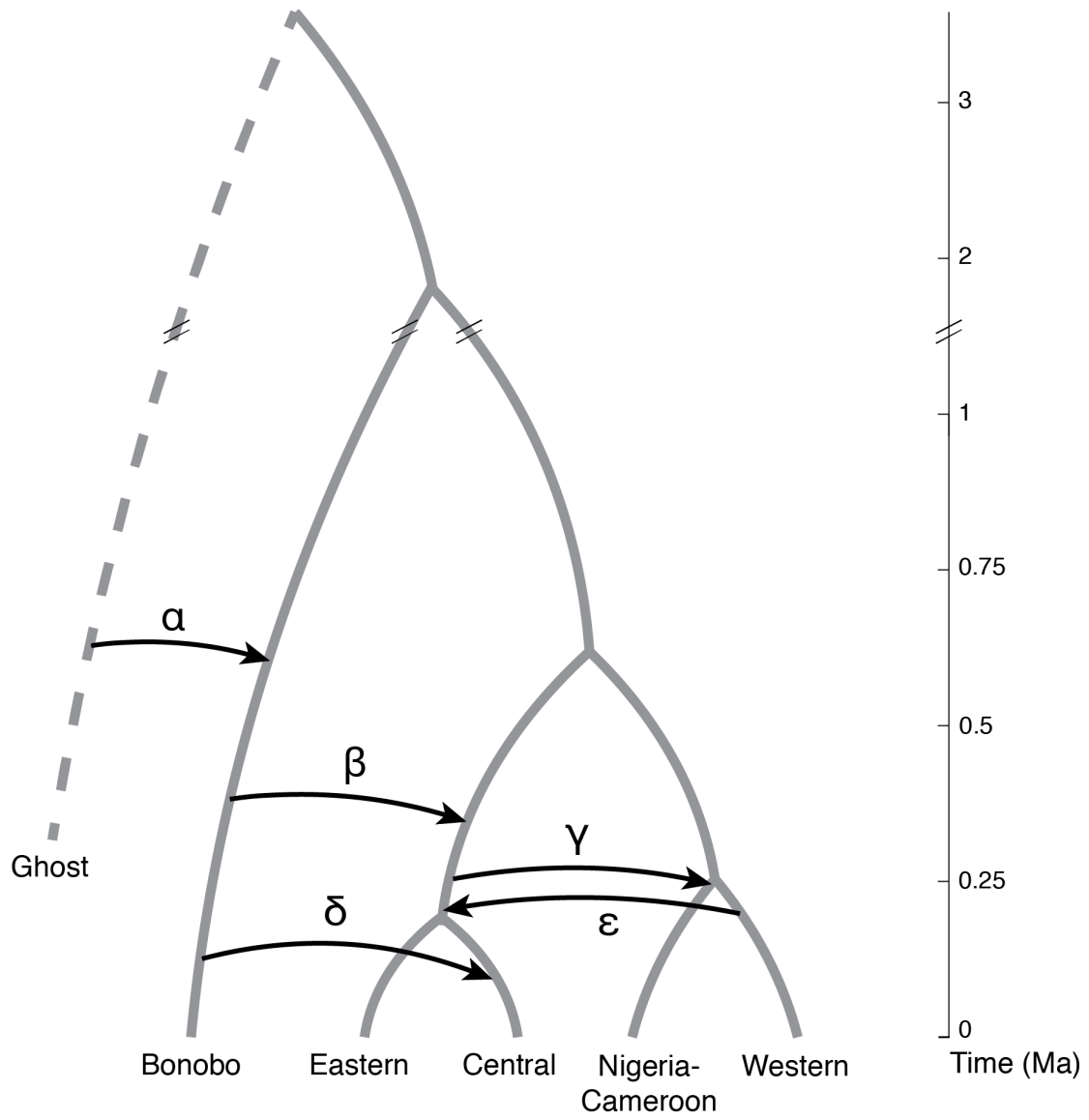


Figure 4. Demographic model and introgression events considered in this analysis. Ghost refers to the extinct *Pan* lineage proposed by (Kuhlwilm et al. 2019).

Analysis

We used Legofit (Rogers 2019; Rogers et al. 2020; Rogers 2021) to estimate demographic history in the five extant lineages of bonobos and chimpanzees. We first used the “sitepat” function to 1) call ancestral alleles, 2) tabulate site patterns from the RAF files including singletons and 3) generate 50 bootstrap replicates. Site patterns are calculated by sampling one haploid genome from each population and the contribution of a given site pattern is the probability that a subsample would exhibit this site pattern.

Once site patterns have been tabulated, Legofit uses these data and a population model (described above and stored as a .lgo file) to estimate parameters by maximizing the composite likelihood via the “legofit” function. Full likelihood is not maximized because information on linkage disequilibrium is not considered. Legofit employs differential evolution (DE) to maximize composite likelihood. Uncertainty is measured via moving-blocks bootstrap. Loci that are linked are not statistically independent, so Legofit resamples blocks of 500 SNVs.

Legofit can be run using one of two algorithms: deterministic and stochastic (Rogers 2021). For computational efficiency, we employed the deterministic algorithm in all but our two most complex models ($\alpha\beta\gamma\delta\epsilon$ and $\alpha\beta\epsilon\gamma\delta$) where we used the stochastic algorithm. References below to “modest precision” apply only to the stochastic algorithm.

We ran the “legofit” function per demographic model on our real data and each of the 50 bootstrap replicates. We conducted this in several stages following Rogers et al. (2020). In Stage 1, points in the DE swarm were scattered widely across parameter space and the objective function was evaluated with modest precision. As some legofit jobs may converge on different local maxima of the composite likelihood

surface, each of the legofit jobs wrote its own swarm of points to a state file. In Stage 2, each legofit job initialized its DE swarm by reading all of the state files produced in Stage 1, enabling legofit to choose among local optima discovered in Stage 1. The evaluation of the objective function was done to high precision in Stage 2. At this point, we used the “pclgo” function to re-express free variables as principal components. Some free parameters may be tightly correlated and this can result in broader confidence intervals because there are fewer dimensions than parameters. This issue can be addressed by reducing the dimension of the parameter space. Our early analyses used a value of 0.001 (“--tol 0.001”) such that principal components were only retained if they explained $> 0.001\%$ of the variance. However, as the exclusion of dimensions may introduce bias, we retained the full dimension. Re-expression of dimensions as principal components can also improve model fit because it allows legofit to operate on uncorrelated dimensions (Rogers 2021). This step produces a new model file (.lgo file). We then repeated Stages 1 and 2 as Stages 3 and 4 using the new .lgo file.

We tested for potential bias in the parameter estimates from our best fitting model by generating simulations using msprime (Kelleher et al. 2016) and fitted those simulated data to the δ model. We used parameter point estimates from our fitted model, the previous fixed time parameter (75,200 generations or 1.88 Ma for the *Pan* common ancestor), and used median effective population sizes from Prado-Martinez et al. (2013) for lineages where we did not have an estimate for N_e from our model (*P. t. ellioti*, *P. t. schweinfurthii*, and *P. t. verus*). We simulated 1×10^4 chromosomes, each 2×10^6 bp in length, and used a mutation rate of 1.4×10^{-8} and a recombination rate of 1×10^{-8} . This was repeated to generate a total of 50 simulated data sets to which we fit the δ model using all four stages of the deterministic approach described

above. We then visually compared the model's point estimates to these simulated bootstraps to assess parameter bias.

All models and scripts for this analysis are available on GitHub (https://github.com/brandcm/Pan_Demography). Many figures were made in R, version 3.6.3 (R Core Team 2020) using `ggplot2`, version 3.3.3 (Wickham 2016) and correlations between the estimated parameters for the best fit model were visualized using 'corrplot' (Wei and Simko 2021).

Data Availability

The raw data underlying this article are previously published (Prado-Martinez et al. 2013; de Manuel et al. 2016) and are available from the Sequence Read Archive (PRJNA189439 and SRP018689) and the European Nucleotide Archive (PRJEB15086).

Results

Legofit aligned 2,366,070,805 loci across all six lineages and determined the ancestral allele for 52,809,700 sites. These sites were used to determine site pattern frequencies in the data and 50 bootstrap replicates (Figure 5).

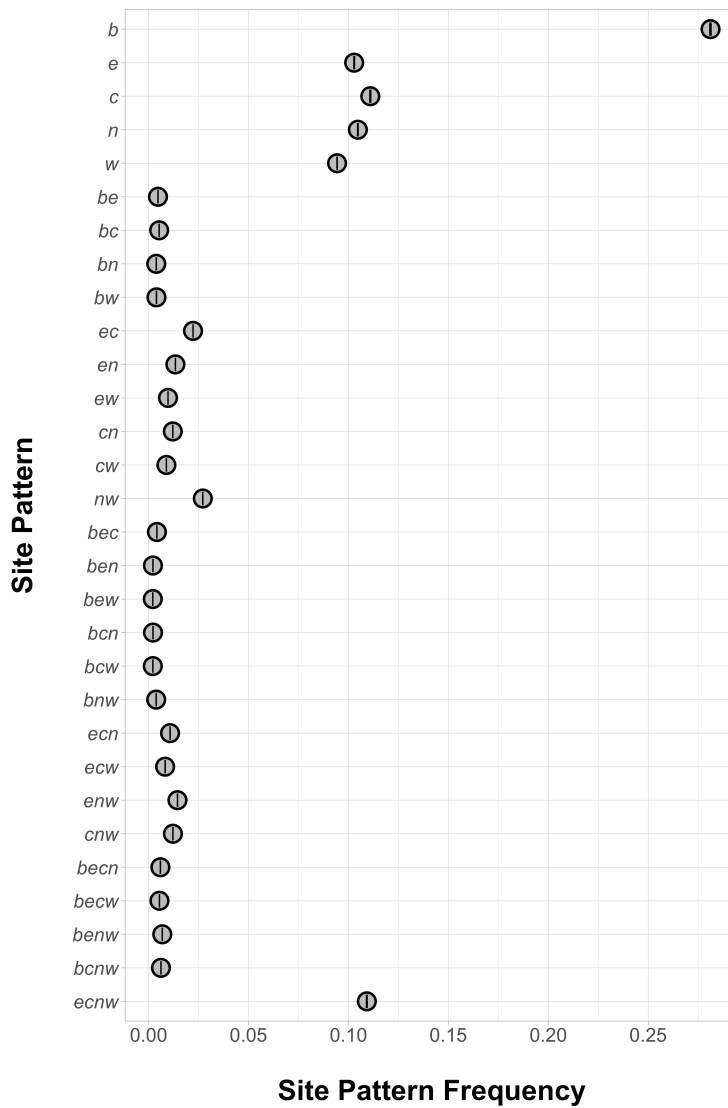


Figure 5. Observed site patterns. The width of vertical line for each point represents the 95% CI.

After comparing models, we found a single model that best fit the observed site patterns: model δ . This model includes a single episode of introgression from bonobos into central chimpanzees. It had small residuals and exhibited the smallest bepe value (1.3×10^{-5}) and a booma weight of 1 (<https://github.com/brandcm/Dissertation>: File S26), therefore model averaging was not invoked.

Point estimates and confidence intervals for the δ model parameters are provided in Table 1. This model estimated the age for the common ancestor of all chimpanzees to be 895 ka (95% CI: 892 - 898 ka), while the ancestor for western and Nigerian-Cameroon chimpanzees dates to 183 ka (95% CI: 178 - 195 ka) and the ancestor of eastern and central chimpanzees was dated to 142 ka (95% CI: 136 - 152 ka). The model also estimated effective population size to vary considerably over time with approximately 40,000 individuals at the time of *Pan* divergence (95% CI: 36,849 - 37,011) and ~17,100 chimpanzees (95% CI: 17,029 - 17,203) immediately prior to the divergence of the chimpanzee common ancestor. Both lineages subsequently increased in size. We found that ~ 2.3% (95% CI: 2.26 - 2.36%) of the central chimpanzee genome was introgressed from bonobos and dated this event to approximately 71 ka.

Table 6. Model parameter estimates. δ = introgression from *P. paniscus* into *P. t. troglodytes*, ec = ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*, nw = ancestor of *P. t. ellioti* and *P. t. verus*, ecnw = common ancestor of all *P. troglodytes* lineages, becnw = *Pan* common ancestor.

		Point estimate	Lower bound	Upper bound
Admixture	δ	0.023298	0.0226364	0.0236331
	b	605000	585000	630000
	c	1002.24	1.42472	3055.845
Population Size	ec	166113	162911	166727
	nw	75548	74368.5	76928.5
	ecnw	17124.5	17029.25	17203
	becnw	36937.2	36849.6	37010.55
Time	δ	71127.25	68233.5	76227.75
	ec	142254.5	136467	152455.75
	nw	183462.25	177824.75	195139.5
	ecnw	894540	892297.5	898012.5

After simulating data using the best fitting model, we found minimal bias in our parameter estimates for admixture and the effective population size of older events (Figure 6). Population sizes for bonobos and central chimpanzees were under- and over-estimated, respectively, whereas the population size for the ancestor of eastern and central chimpanzees was slightly under-estimated. Point estimates for divergence times exhibited some bias to older ages although the age for the common ancestor of all chimpanzees agreed with the simulated data.



Figure 6. Parameter estimate bias. The orange points represent point estimates for the parameters from the δ model. Open gray circles represent 50 values estimated by legofit using site patterns generated from data simulated with the δ model parameters using msprime. If the simulated data are $<$ the point estimate, the point estimate is underestimated, while if the simulated data are $>$ the point estimate, the point estimate is overestimated. δ = introgression from *P. paniscus* into *P. t. troglodytes*, *ec* = ancestor of *P. t. schweinfurthii* and *P. t. troglodytes*, *nw* = ancestor of *P. t. ellioti* and *P. t. verus*, *ecnw* = common ancestor of all *P. troglodytes* lineages, *becnw* = *Pan* common ancestor.

Discussion

We evaluated previously proposed models for the evolutionary history of the genus *Pan*. While mitochondrial, Y-chromosomal, and nuclear DNA have yielded some consistent parameter estimates, many others remain imprecise or may suffer from bias. Our results suggest a more simple evolutionary history: the best fitting model only includes one introgression event from bonobos into central chimpanzees. As we did not explicitly model introgression events that would not affect site patterns (e.g., subsequent gene flow between recently diverged lineages), we cannot speak to that aspect of *Pan* demographic history. However, if such events occurred shortly after divergence, we would expect the resulting confidence interval to be quite large. This may be the case for the ancestors of both eastern and central as well as Nigeria-Cameroon and western chimpanzees. Yet, this interval is quite small for the common ancestor of all chimpanzees.

We estimate that approximately 2.3% of central chimpanzee DNA is derived from bonobos and that this event dated to ~ 71 ka. To our knowledge, there is no data on the discharge of the Congo River for this time period. Presently, this part of the river is one of the deepest and widest sections, with little seasonal variation in discharge near Kinshasa (Takemoto et al. 2015). This suggests that direct contact between bonobos and central chimpanzees would be difficult; however, our result and those from others (de Manuel et al. 2016) strongly suggest that this contact occurred. Evidence of gene flow from bonobos into central chimpanzees is not only consistent with previous reports but is further evidenced by the possible adaptiveness of introgressed bonobo alleles in central chimpanzees, potentially related to reproduction in males (Nye et al. 2018). Identification of the sites with shared site patterns between

these lineages identified here could be informative to confirm and expand candidate regions for adaptive introgression.

Our estimate of the time of divergence for the ancestor of eastern and central and the ancestor of Nigeria-Cameroon and western chimpanzees is similar to other past results (Becquet and Przeworski 2007; Hey 2010; Prado-Martinez et al. 2013; de Manuel et al. 2016 but see Wegmann and Excoffier 2010). Assessment of parameter bias suggests that the point estimates may be slightly overestimated but not considerably so. Our analysis did not find bias in our estimate of the age of the common chimpanzee ancestor. This particular parameter is much older than others (e.g., 544 - 633 ka (de Manuel et al. 2016)). While the phenotypic differences between chimpanzees subspecies are likely still emerging, previously described differences, particularly between eastern and western chimpanzees, may support an older divergence date for the chimpanzee common ancestor. The absence of a shared positive selection signal in chimpanzees, as described in Chapter II, also tentatively supports a deep divergence for common chimpanzees.

Estimates for population size largely support previous findings (Prado-Martinez et al. 2013; de Manuel et al. 2016). Following divergence, the common ancestor of all chimpanzees experienced a period of decline. This was followed by substantial increases in both the ancestor of Nigeria-Cameroon and western chimpanzees, and particularly the ancestor of eastern and central chimpanzees, which we estimate to be approximately 76,000 and 166,000 individuals, respectively. The estimated N_e for each lineage suggests that each subspecies experienced a population decline after divergence with their common ancestor. However, two of our population size estimates are puzzling. We found a small population size for central chimpanzees at the time of introgression from bonobos and a large population size for bonobos at

the same time. This may represent an instance of statistical identifiability where parameters are correlated, resulting in a broader confidence interval (Rogers 2019). However, neither parameter is tightly correlated with any others (<https://github.com/brandcm/Dissertation>: File S0: Figure S10). Further, the genetic diversity of both subspecies does not support these numbers. Regardless of genetic diversity, it seems implausible the central chimpanzees would experience a bottleneck ~ 70 ka and generate a recent N_e estimate of $\sim 36,000$ individuals. In fact, our analysis of bias in parameter estimates suggests that the present estimate of $\sim 1,000$ is a generous overestimate. There appears to be a more plausible explanation for the bonobo population size at the time of admixture. Such a high parameter value could be explained by geographic population structure (Nei and Takahata 1993). The geographic origin of the bonobo genomes used in this analysis are unknown. However, some structure has been inferred from cranio-dental morphology (Pilbrow and Groves 2013) and mitochondrial haplotypes (Kawamoto et al. 2013, but see Eriksson et al. 2004). Another potential line of support for structure in bonobos comes from the curious geographic restriction of bonobo malarial infection to individuals east of the Lomami River (Liu et al. 2017). Therefore, it is possible that our unusually large parameter estimate is driven by bonobo population structure.

We note that the parameters estimated from this analysis were generated by setting one fixed parameter (the *Pan* divergence date or Tbecnw) to set the molecular clock. We chose this parameter because there is increasing consensus from *Pan* genomic data relative to other time parameters. However, the point estimate used in this analysis was the median of a range from de Manuel et al. (2016). Thus, if the true divergence date is different to that used here, our parameter estimates would change

as well. Additional genomic data from bonobos and chimpanzees may yield more accurate estimates of this critical parameter.

The site patterns of derived alleles in bonobos and chimpanzees confirm multiple aspects of their evolutionary history while offering new insights into other facets. We find support for a single introgression event from bonobos into central chimpanzees although the biogeography of this event remains difficult to explain. Collectively, the best fit demographic model is simpler than more recently proposed models. Finally, our results point to a deeper divergence time for common chimpanzees. Additional genomic and paleoenvironmental data would be immensely informative in deciphering the evolutionary history of our closest living history and may provide insight into the evolution of other taxa in this region during this time period, including humans.

CHAPTER V

CONCLUSION

Our closest living relatives are two species in the genus *Pan*: bonobos and chimpanzees. The phylogenetic proximity of these taxa to humans highlights their importance as models for human evolution. Studies of living bonobos and chimpanzees are essential to this goal. However, the virtual absence of a bonobo and chimpanzee fossil record means that genomic data provide the best window into their evolutionary past to better understand how bonobos and chimpanzees diverged and came to be the lineages we know today. This dissertation uses reassembled and remapped autosomal genomic data from all five *Pan* lineages to answer questions and test hypotheses about adaptation and demography in these apes.

The evolutionary history of chimpanzees and bonobos has provided many time points during which positive selection could drive the phenotypes observed today in all five lineages. In Chapter II, we leverage genomic data on synonymous and nonsynonymous within lineage polymorphic and divergent loci between *Pan* and humans to identify candidates under positive selection at deeper times scales. We also apply a modification of this test to increase statistical power for detecting selection candidates using genome wide averaged parameters and inferring other evolutionary parameters for identified genes. We found a range of candidate genes for adaptation in each lineage ranging from 7 in western chimpanzees to 54 in central chimpanzees. Many candidates were unique to each lineage. We found only one gene unique to all chimpanzees and another two genes were found to be shared across all five lineages. Together, these candidates may have phenotypic impacts on various traits including the brain, immunity, musculature, reproduction, and the skeletal system. Analysis of

individual exons also may provide insight into regions of the coding sequence that are under selection and also generated additional candidates for multiple lineages. We do not find evidence in support of the THV hypothesis, although a thyroid receptor gene emerged as a candidate in bonobos, which may contribute to their behavior and biology.

While older positive selection may have driven key differences between bonobos and chimpanzees, recent selection likely also contributes to their similarities and differences. In Chapter III, we used supervised machine learning to assess the extent to which this adaptation is the result of *de novo* mutation or standing genetic variation. There remains considerable debate on whether positive selection at single loci is predominantly shaped by hard or soft sweeps. Previous empirical tests stemmed from data in humans, *Drosophila*, and HIV. We demonstrate that similar to humans, *Drosophila*, and some cases of HIV, soft sweeps are much more common than hard sweeps in four of the five lineages we could examine. These results underscore the role of the physical and/or social environment in shaping *Pan* adaptations during the late Pleistocene. Most candidate sweeps were unique to each lineage although there was some overlap, particularly for soft sweeps. We found 19 candidate soft sweep windows shared across all four lineages. While plentiful, the genes in these windows may be important in bonobo and chimpanzee phenotypes.

Considerable attention has been paid to estimating demography in these species. However, many of the past studies rely on parameter heavy approaches and may suffer from parameter bias, particularly estimates of admixture. In Chapter IV, we consider previous models for the demographic history of *Pan* using site patterns. This approach allows for both parameter estimation and model selection. Among 42 different models, we fit the best fit model is rather simple and includes a single

introgression event from bonobos into central chimpanzees. We also estimate the common ancestor of chimpanzees is ~ 900 ka, much older than previously suggested.

The results of this dissertation expand our understanding of the evolutionary history of the genus *Pan*, particularly related to adaptation and demography. Such findings also prompt many new questions. These include analyses of the candidate genes described in Chapters II and III to identify variants and characterize the potential functional consequences. The GAGP dataset has and will continue to yield immense insight. Yet, additional genomic data would be massively beneficial for many reasons. First, no small number of newer genomic methods leverage massive sample sizes to detect different types of selection and demographic events. For example, large samples are particularly needed for confidently inferring haplotypes (Browning and Browning 2011). Second, the advent of sequencing genomes in wild primate (and other animal) populations using non-invasive methods is a critical development (Chiou and Bergey 2018; Ozga et al. 2021). These approaches will be able to shed light on many topics, including local adaptation. Information on individual life history and fine-scale environmental data can better tease apart complex gene-environment interactions. Although not necessarily restricted to wild primates, larger samples with information on phenotype may facilitate analysis of polygenic selection beyond humans and other model organisms. Results from Chapter IV may point to population structure in bonobos. Additional genomes from across their geographic range may help confirm or reject this possibility. Finally, future studies on gene expression and structural variation in *Pan* will likely fill in critical gaps regarding phenotypic differences in this genus. Beyond genomics, the findings from Chapters II, III, and IV also highlight the need for more ecological and paleoenvironmental data, particularly from the Congo Basin. In my future research, I

hope to build upon the findings of this dissertation with additional genomic data and methods to better understand our closest living relatives and their evolutionary history.

REFERENCES CITED

- Anding AL, Wang C, Chang T-K, Sliter DA, Powers CM, Hofmann K, Youle RJ, Baehrecke EH. 2018. Vps13D encodes a ubiquitin-binding protein that is required for the regulation of mitochondrial size and clearance. *Curr Biol* 28:287-295.e6.
- Andolfatto P. 2008. Controlling type-I error of the McDonald–Kreitman test in genomewide scans for selection on noncoding DNA. *Genetics* 180:1767.
- Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, White TJ, Green ED, Bustamante CD, et al. 2009. Targets of balancing selection in the human genome. *Mol Biol Evol* 26:2755–2764.
- Andrews S. 2010. FASTQC. A quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Anestis SF. 2004. Female genito-genital rubbing in a group of captive chimpanzees. *Int J Primatol* 25:477–488.
- Anestis SF, Webster TH, Kamilar JM, Fontenot MB, Watts DP, Bradley BJ. 2014. *AVPR1A* variation in chimpanzees (*Pan troglodytes*): Population differences and association with behavioral style. *Int J Primatol* 35:305–324.
- Angus S. 1971. Water-contact behavior of chimpanzees. *Fol Primatol* 14:51–58.
- Anka Z, Séranne M, Primio R di. 2010. Evidence of a large upper-Cretaceous depocentre across the Continent-Ocean boundary of the Congo-Angola basin. Implications for palaeo-drainage and potential ultra-deep source rocks. *Marine and Petroleum Geology* 27:601–611.
- Antón SC, Potts R, Aiello LC. 2014. Evolution of early *Homo*: An integrated biological perspective. *Science* 345:1236828.
- Arsic N, Rajic T, Stanojic S, Goodfellow PN, Stevanovic M. 1998. Characterisation and mapping of the human SOX14 gene. *Cytogenet Cell Genet* 83:139–146.
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74.
- Baptiste A. 2015. gridExtra: Miscellaneous Functions for “Grid” Graphics. Available from: <http://CRAN.R-project.org/package=gridExtra>
- Barratt CD, Lester JD, Gratton P, Onstein RE, Kalan AK, McCarthy MS, Bocksberger G, White LC, Vigilant L, Diegues P, et al. 2020. Late Quaternary habitat suitability models for chimpanzees (*Pan troglodytes*) since the Last Interglacial (120,000 BP). *bioRxiv* [Internet]. Available from: <http://biorxiv.org/content/early/2020/05/25/2020.05.15.066662>

- Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet* 12:767–780.
- Basabose AK. 2002. Diet composition of chimpanzees inhabiting the Montane forest of Kahuzi, Democratic Republic of Congo. *Am J Primatol* 58:1–21.
- Beadle L. 1981. The inland waters of tropical Africa: An introduction to tropical limnology. New York: Longman
- Beaune D, Hohmann G, Serckx A, Sakamaki T, Narat V, Fruth B. 2017. How bonobo communities deal with tannin rich fruits: Re-ingestion and other feeding processes. *Behav Process* 142:131–137.
- Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res* 17:1505–1519.
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLOS Biol* 5:e310.
- Behringer Verena, Deschner T, Murtagh R, Stevens JMG, Hohmann G. 2014. Age-related changes in thyroid hormone levels of bonobos and chimpanzees indicate heterochrony in development. *J Hum Evol* 66:83–88.
- Behringer Verena, Deschner Tobias, Deimel Caroline, Stevens JMG, Hohmann G. 2014. Age-related changes in urinary testosterone levels suggest differences in puberty onset and divergent life history strategies in bonobos and chimpanzees. *Horm Behav* 66:525–533.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300.
- Bergey CM, Lopez M, Harrison GF, Patin E, Cohen JA, Quintana-Murci L, Barreiro LB, Perry GH. 2018. Polygenic adaptation and convergent evolution on growth and cardiac genetic pathways in African and Asian rainforest hunter-gatherers. *Proc Natl Acad Sci USA* 115:E11256.
- Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH. 2019. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol* 3:286–292.
- Boesch C. 1996. Social grouping in Tai chimpanzees. In: McGrew WC, Marchant LF, Nishida T, editors. *Great Ape Societies*. Cambridge: Cambridge University Press. p. 101–113.
- Boesch C, Boesch-Achermann H. 2000. *The chimpanzees of the Tai Forest: Behavioral ecology and evolution*. Oxford: Oxford University Press

- Boesch C, Crockford C, Herbinger I, Wittig R, Moebius Y, Normand E. 2008. Intergroup conflicts among chimpanzees in Tai National Park: lethal violence and the female perspective. *Am J Primatol* 70:519–532.
- Boesch C, Head J, Tagg N, Arandjelovic M, Vigilant L, Robbins MM. 2007. Fatal Chimpanzee Attack in Loango National Park, Gabon. *Int J Primatol* 28:1025–1034.
- Boesch C, Hohmann G, Marchant LF eds. 2001. Behavioural Diversity in Chimpanzees and Bonobos. Cambridge: Cambridge University Press
- Boose KJ, White FJ, Meinelt A. 2013. Sex differences in tool use acquisition in bonobos (*Pan paniscus*). *American Journal of Primatology* 75:917–926.
- Bradley BJ. 2008. Reconstructing phylogenies and phenotypes: a molecular view of human evolution. *Journal of Anatomy* 212:337–353.
- Brand CM, Marchant LF. 2019. Social hair plucking is a grooming convention in a group of captive bonobos (*Pan paniscus*). *Primates* 60:487–491.
- Brand CM, White FJ, Ting N, Webster TH. 2021. Soft sweeps predominate recent positive selection in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*). *bioRxiv*:2020.12.14.422788.
- Broad Institute. 2018. Picard Tools. Available from: <http://broadinstitute.github.io/picard/>
- Browning SR, Browning BL. 2011. Haplotype phasing: existing methods and new developments. *Nat Rev Genet* 12:703–714.
- Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* 173:53–61.e9.
- Bullinger AF, Burkart JM, Melis AP, Tomasello M. 2013. Bonobos, *Pan paniscus*, chimpanzees, *Pan troglodytes*, and marmosets, *Callithrix jacchus*, prefer to feed alone. *Anim Behav* 85:51–60.
- Cagan A, Theunert C, Laayouni H, Santpere G, Pybus M, Casals F, Prüfer K, Navarro A, Marques-Bonet T, Bertranpetit J, et al. 2016. Natural selection in the great apes. *Mol Biol Evol* 33:3268–3283.
- Cahill JA, Stirling I, Kistler L, Salamzade R, Ersmark E, Fulton TL, Stiller M, Green RE, Shapiro B. 2015. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol Ecol* 24:1205–1217.

- Calpena E, Hervieu A, Kaserer T, Swagemakers SMA, Goos JAC, Popoola O, Ortiz-Ruiz MJ, Barbaro-Dieber T, Bownass L, Brilstra EH, et al. 2019. De novo missense substitutions in the gene encoding CDK8, a regulator of the mediator complex, cause a syndromic developmental disorder. *Am J Hum Genet* 104:709–720.
- Campbell CJ. 2006. Lethal intragroup aggression by adult male spider monkeys (*Ateles geoffroyi*). *Am J Primatol* 68:1197–1201.
- Castellano D, Macià MC, Tataru P, Bataillon T, Munch K. 2019. Comparison of the full distribution of fitness effects of new amino acid mutations across great apes. *Genetics* 213:953–966.
- Caswell JL, Mallick S, Richter DJ, Neubauer J, Schirmer C, Gnerre S, Reich D. 2008. Analysis of chimpanzee history based on genome sequence alignments. *PLOS Genet* 4:e1000057.
- Chang T-C, Yang Y, Yasue H, Bharti AK, Retzel EF, Liu W-S. 2011. The expansion of the *PRAME* gene family in Eutheria. *PLOS One* 6:e16867.
- Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303.
- Cheng X, DeGiorgio M. 2020. Flexible mixture model approaches that accommodate footprint size variability for robust detection of balancing selection. *Mol Biol Evol* 37:3267–3291.
- Chiou KL, Bergey CM. 2018. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Sci Rep* 8:1975.
- Choi H, Jin S, Kwon JT, Kim Jihye, Jeong J, Kim Jaehwan, Jeon S, Park ZY, Jung K-J, Park K, et al. 2016. Characterization of mammalian *ADAM2* and its absence from human sperm. *PLOS One* 11:e0158321.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6:80–92.
- Compton AA, Malik HS, Emerman M. 2013. Host gene evolution traces the evolutionary history of ancient primate lentiviruses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368:20120496.
- Coolidge HJ. 1933. *Pan paniscus*. Pigmy chimpanzee from south of the Congo river. *Am J Phys Anth* 18:1–59.
- Crockford SJ. 2003. Thyroid rhythm phenotypes and hominid evolution: a new paradigm implicates pulsatile hormone secretion in speciation and adaptation changes. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 135:105–129.

- Cronin KA, De Groot E, Stevens JMG. 2015. Bonobos show limited social tolerance in a group setting: A comparison with chimpanzees and a test of the relational model. *Fol Primatol* 86:164–177.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
- Darolti I, Wright AE, Pucholt P, Berlin S, Mank JE. 2018. Slow evolution of sex-biased genes in the reproductive tissue of the dioecious plant *Salix viminalis*. *Mol Ecol* 27:694–708.
- deMenocal PB. 2004. African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth Planet Let Sci* 220:3–24.
- Doran DM. 1993. Comparative locomotor behavior of chimpanzees and bonobos: The influence of morphology on locomotion. *Am J Phys Anth* 91:83–98.
- Drake AG. 2011. Dispelling dog dogma: an investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape. *Evol Dev* 13:204–213.
- Eilertson KE, Booth JG, Bustamante CD. 2012. SnIPRE: Selection inference using a Poisson random effects model. *PLOS Comput Biol* 8:e1002806.
- Eriksson J, Hohmann G, Boesch C, Vigilant L. 2004. Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Mol Ecol* 13:3425–3435.
- Etienne L, Nerrienet E, LeBreton M, Bibila GT, Foupouapouognigni Y, Rousset D, Nana A, Djoko CF, Tamoufe U, Aghokeng AF, et al. 2011. Characterization of a new simian immunodeficiency virus strain in a naturally infected Pan troglodytes troglodyteschimpanzee with AIDS related symptoms. *Retrovirology* 8:4.
- Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.
- Eyre-Walker A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics* 162:2017.
- Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610–618.
- Fawcett K, Muhumuza G. 2000. Death of a wild chimpanzee community member: Possible outcome of intense sexual competition. *Am J Primatol* 51:243–247.
- Fay JC, Wu C-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.

- Fedigan LM. 1983. Dominance and reproductive success in primates. *Am J Phys Anth* 26:91–129.
- Feldblum JT, Wroblewski EE, Rudicell RS, Hahn BH, Paiva T, Cetinkaya-Rundel M, Pusey AE, Gilby IC. 2014. Sexually coercive male chimpanzees sire more offspring. *Curr Biol* 24:2855–2860.
- Ferguson W, Dvora S, Fikes RW, Stone AC, Boissinot S. 2012. Long-term balancing selection at the antiviral gene *OAS1* in central African chimpanzees. *Mol Biol Evol* 29:1093–1103.
- Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol* 31:1275–1291.
- Fruth B, Hickey J, André C, Furuichi T, Hart JA, Hart TB, Kuehl H, Maisels F, Nackoney J, Reinartz GE, et al. 2016. *Pan paniscus* (errata version published in 2016). Available from: <https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T15932A17964305.en>
- Fruth B, Hohmann G. 2018. Food sharing across borders. *Hum Nat* 29:91–103.
- Furuichi T. 1989. Social interactions and the life history of female *Pan paniscus* in Wamba, Zaire. *Int J Primatol* 10:173–197.
- Furuichi T. 2009. Factors underlying party size differences between chimpanzees and bonobos: a review and hypotheses for future study. *Primates* 50:197–209.
- Furuichi T. 2011. Female contributions to the peaceful nature of bonobo society. *Ev Anth* 20:131–142.
- Furuichi T, Hashimoto C, Tashiro Y. 2001. Fruit availability and habitat use by chimpanzees in the Kalinzu Forest, Uganda: Examination of fallback foods. *Int J Primatol* 22:929–945.
- Furuichi T, Idani G, Ihobe H, Hashimoto C, Tashiro Y, Sakamaki T, Mulavwa MN, Yangozene K, Kuroda S. 2012. Long-term studies on wild bonobos at Wamba, Luo Scientific Reserve, D. R. Congo: Towards the understanding of female life history in a male-philopatric species. In: Kappeler PM, Watts DP, editors. *Long-Term Field Studies of Primates*. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 413–433. Available from: https://doi.org/10.1007/978-3-642-22514-7_18
- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, et al. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 397:436–441.
- Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLOS Genet* 11:e1005004.

- Gauthier J, Meijer IA, Lessel D, Mencacci NE, Krainc D, Hempel M, Tsiakas K, Prokisch H, Rossignol E, Helm MH, et al. 2018. Recessive mutations in *VPS13D* cause childhood onset movement disorders. *Ann Neurol* 83:1089–1095.
- Gerloff U, Hartung B, Fruth B, Hohmann G, Tautz D. 1999. Intracommunity relationships, dispersal pattern and paternity success in a wild living community of bonobos (*Pan paniscus*) determined from DNA analysis of faecal samples. *Proc R Soc Lond B Biol Sci* 266:1189–1195.
- Gokcumen O. 2020. Archaic hominin introgression into modern human genomes. *Am J Phys Anth* 171:60–73.
- Good JM, Wiebe V, Albert FW, Burbano HA, Kircher M, Green RE, Halbwax M, André C, Atencia R, Fischer A, et al. 2013. Comparative population genomics of the ejaculate in humans and the great apes. *Mol Biol Evol* 30:964–976.
- Goodall J. 1986. The chimpanzees of Gombe: Patterns of behavior. Cambridge, MA: Belknap Press
- Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The Effect of Variation in the Effective Population Size on the Rate of Adaptive Molecular Evolution in Eukaryotes. *Genome Biology and Evolution* 4:658–667.
- Gossmann TI, Waxman D, Eyre-Walker A. 2014. Fluctuating selection models and McDonald-Kreitman type analyses. *PLOS One* 9:e84540.
- de Groot NG, Heijmans CMC, Zoet YM, de Ru AH, Verreck FA, van Veelen PA, Drijfhout JW, Doxiadis GGM, Remarque EJ, Doxiadis IIN, et al. 2010. AIDS-protective HLA-B*27/B*57 and chimpanzee MHC class I molecules target analogous conserved areas of HIV-1/SIV_{cpz}. *Proc Natl Acad Sci USA* 107:15175.
- Groves CP. 2001. Primate Taxonomy. Washington DC: Smithsonian Institution Press
- Groves CP. 2005. Geographic variation within eastern chimpanzees (*Pan troglodytes cf schweinfurthii* Giglioli, 1872). *Australas Primatol* 17:19–46.
- Gruber T, Clay Z. 2016. A comparison between bonobos and chimpanzees: A review and update. *Ev Anth* 25:239–252.
- Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J, The Bioconda Team. 2018. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat Methods* 15:475–476.
- Hamilton WD. 1964. The genetical evolution of social behaviour I & II. *J Theor Biol* 7:1–52.

- Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ. 2007. Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm receptor gene *PKDREJ*. *Am J Hum Genet* 81:44–52.
- Han S, Andrés AM, Marques-Bonet T, Kuhlwilm M. 2019. Genetic variation in *Pan* species is shaped by demographic history and harbors lineage-specific functions. *Genome Biol Evol* 11:1178–1191.
- Hare B, Kwetuenda S. 2010. Bonobos voluntarily share their own food with others. *Curr Biol* 20:R230–R231.
- Hare B, Melis AP, Woods V, Hastings S, Wrangham R. 2007. Tolerance allows bonobos to outperform chimpanzees on a cooperative task. *Curr Biol* 17:619–623.
- Hare B, Wobber V, Wrangham R. 2012. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Anim Behav* 83:573–585.
- Hare B, Yamamoto S. 2015. Moving bonobos off the scientifically endangered list. *Behaviour* 152:247–258.
- Harris RB, Sackman A, Jensen JD. 2018. On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. *PLOS Genet* 14:e1007859.
- Hartfield M, Bataillon T. 2020. Selective sweeps under dominance and inbreeding. *G3 Genes Genom Genet* 10:1063.
- Hashimoto C. 1997. Context and development of sexual behavior of wild bonobos (*Pan paniscus*) at Wamba, Zaire. *Int J Primatol* 18:1–21.
- Hedrick PW. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol Ecol* 22:4606–4618.
- Henn BM, Cavalli-Sforza LL, Feldman MW. 2012. The great human expansion. *Proc Natl Acad Sci USA* 109:17758.
- Hermisson J, Pennings PS. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.
- Hermisson J, Pennings PS. 2017. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol Evol* 8:700–716.
- Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, Project 1000 Genomes, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331:920–924.

- Hey J. 2010. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol Biol Evol* 27:921–933.
- Hof J, Sommer V. 2010. Apes Like Us: Portraits of a Kinship. Edition Panorama. Germany: Mannheim
- Hohmann G. 2001. Association and social interactions between strangers and residents in bonobos (*Pan paniscus*). *Primates* 42:91–99.
- Hohmann G, Fruth B. 2000. Use and function of genital contacts among female bonobos. *Anim Behav* 60:107–120.
- Hohmann G, Fruth B. 2003a. Lui Kotal - A new site for field research on bonobos in the Salonga National Park. *Pan Africa News* 10:25–27.
- Hohmann G, Fruth B. 2003b. Culture in bonobos? Between-species and within-species variation in behavior. *Curr Anthropol* 44:563–571.
- Hohmann G, Fruth B. 2011. Is blood thicker than water? In: Robbins MM, Boesch C, editors. Among African Apes: Stories and Photos from the Field. Berkeley: University of California Press. p. 61–76.
- Hohmann G, Ortmann S, Remer T, Fruth B. 2019. Fishing for iodine: what aquatic foraging by bonobos tells us about human evolution. *BMC Zool* 4:5.
- Holloway AK, Lawniczak MKN, Mezey JG, Begun DJ, Jones CD. 2007. Adaptive gene expression divergence inferred from population genomics. *PLOS Genet* 3:e187.
- Horn AD. 1979. The taxonomic status of the bonobo chimpanzee. *Am J Phys Anth* 51:273–281.
- Huang DW, Sherman BT, Lempicki RA. 2008. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37:1–13.
- Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
- Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153.
- Huerta-Sanchez E, Durrett R, Bustamante CD. 2008. Population genetics of polymorphism and divergence under fluctuating selection. *Genetics* 178:325.
- Humle T, Maisels F, Oates JF, Plumptre AJ, Williamson EA. 2016. *Pan troglodytes* (errata version published in 2018). e.T15933A129038584. Available from: <https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T15933A17964454.en>.
- Idani G. 1991. Social relationships between immigrant and resident bonobo (*Pan paniscus*) females at Wamba. *Fol Primatol* 57:83–95.

- Inogwabini B-I. 2020. Wild Bonobos and Wild Chimpanzees and Human Diseases. In: Inogwabini B-I, editor. *Reconciling Human Needs and Conserving Biodiversity: Large Landscapes as a New Conservation Paradigm: The Lake Tumba, Democratic Republic of Congo*. Cham: Springer International Publishing. p. 109–121. Available from: https://doi.org/10.1007/978-3-030-38728-0_9
- Inogwabini B-I, Matungila B, Mbende L, Abokome M, Tshimanga T wa. 2007. Great apes in the Lake Tumba landscape, Democratic Republic of Congo: newly described populations. *Oryx* 41:532–538.
- Ishizuka S, Kawamoto Y, Sakamaki T, Tokuyama N, Toda K, Okamura H, Furuichi T. 2018. Paternity and kin structure among neighbouring groups in wild bonobos at Wamba. *R Soc Open Sci* 5:171006.
- Jaeggi AV, De Groot E, Stevens JMG, Van Schaik CP. 2013. Mechanisms of reciprocity in primates: testing for short-term contingency of grooming and food sharing in bonobos and chimpanzees. *Evol Hum Behav* 34:69–77.
- Jaeggi AV, Stevens JMG, Van Schaik CP. 2010. Tolerant food sharing and reciprocity is precluded by despotism among bonobos but not chimpanzees. *Am J Phys Anth* 143:41–51.
- Jensen JD. 2014. On the unfounded enthusiasm for soft selective sweeps. *Nat Commun* 5:5281.
- Johnson SC, Bonnefille R, Chivers DJ, Groves CP, Horn AD, Jungers WL, Kimura T, McHenry HM, Prasad KN, Schwartz JH, et al. 1981. Bonobos: Generalized hominid prototypes or specialized insular dwarfs? *Curr Anthropol* 22:363–375.
- Johri P, Charlesworth B, Jensen JD. 2020. Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. *Genetics* 215:173.
- Johri P, Riall K, Becher H, Charlesworth B, Jensen JD. 2020. The impact of purifying and background selection on the inference of population history: problems and prospects. *bioRxiv*:2020.04.28.066365.
- Jungers WL, Susman RL. 1984. Body size and skeletal allometry in African apes. In: Susman RL, editor. *The Pygmy Chimpanzee Evolutionary Biology and Behavior*. New York: Plenum Press. p. 131–178.
- Kaburu SSK, Inoue S, Newton-Fisher NE. 2013. Death of the alpha: Within-community lethal violence among chimpanzees of the Mahale Mountains National Park. *Am J Primatol* 75:789–797.
- Kahlenberg SM, Emery Thompson M, Wrangham RW. 2008. Female competition over core areas in *Pan troglodytes schweinfurthii*, Kibale National Park, Uganda. *Int J Primatol* 29:931.

- Kahlenberg SM, Thompson ME, Muller MN, Wrangham RW. 2008. Immigration costs for female chimpanzees and male protection as an immigrant counterstrategy to intrasexual aggression. *Anim Behav* 76:1497–1509.
- Kalan AK, Kulik L, Arandjelovic M, Boesch C, Haas F, Dieguez P, Barratt CD, Abwe EE, Agbor A, Angedakin S, et al. 2020. Environmental variability supports chimpanzee behavioural diversity. *Nat Commun* 11:4451.
- Kamada F, Aoki Y, Narisawa A, Abe Y, Komatsuzaki S, Kikuchi A, Kanno J, Niihori T, Ono M, Ishii N, et al. 2011. A genome-wide association study identifies *RNF213* as the first Moyamoya disease gene. *J Hum Genet* 56:34–40.
- Kano T. 1992. The last ape: Pygmy chimpanzee behavior and ecology. Stanford: Stanford University Press
- Kawamoto Y, Takemoto H, Higuchi S, Sakamaki T, Hart JA, Hart TB, Tokuyama N, Reinartz GE, Guislain P, Dupain J, et al. 2013. Genetic structure of wild bonobo populations: Diversity of mitochondrial DNA and geographical distribution. *PLOS One* 8:e59660.
- Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, Wilson ML, Li Y, Learn GH, Beasley TM, Schumacher-Stankey J, et al. 2009. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* 460:515–519.
- Kelleher J, Etheridge AM, McVean G. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLOS Comput Biol* 12:e1004842.
- Kenigsberg S, Lima PDA, Maghen L, Wyse BA, Lackan C, Cheung ANY, Tsang BK, Librach CL. 2017. The elusive MAESTRO gene: Its human reproductive tissue-specific expression pattern. *PLOS One* 12:e0174873.
- Kern AD, Schrider DR. 2016. Discoal: flexible coalescent simulations with selection. *Bioinformatics* 32:3839–3841.
- Kern AD, Schrider DR. 2018. diploS/HIC: An updated approach to classifying selective sweeps. *G3 Genes Genom Genet* 8:1959–1970.
- Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press
- Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28:2520–2522.
- Kovalaskas S, Rilling JK, Lindo J. 2020. Comparative analyses of the *Pan* lineage reveal selection on gene pathways associated with diet and sociality in bonobos. *Genes Brain Behav* n/a:e12715.
- Krassowski M. 2020. ComplexUpset. Available from: <https://doi.org/10.5281/zenodo.3700590>

- Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, Underwood JG, Nelson BJ, Chaisson MJP, Dougherty ML, et al. 2018. High-resolution comparative analysis of great ape genomes. *Science* [Internet] 360. Available from: <https://science.sciencemag.org/content/360/6393/eaar6343>
- Kuhlwilm M, Han S, Sousa VC, Excoffier L, Marques-Bonet T. 2019. Ancient admixture from an extinct ape lineage into bonobos. *Nat Ecol Evol* 3:957–965.
- Kuroda S, Nishihara T, Suzuki S, Oko RA. 1996. Sympatric chimpanzees and gorillas in the Ndoki Forest, Congo. In: McGrew WC, Marchant LF, Nishida T, editors. *Great Ape Societies*. Cambridge: Cambridge University Press. p. 71–81.
- Lalouette A, Guénet J-L, Vríz S. 1998. Hotfoot mouse mutations affect the $\delta 2$ glutamate receptor gene and are allelic to *lurcher*. *Genomics* 50:9–13.
- Langergraber KE, Mitani JC, Vigilant L. 2007. The limited impact of kinship on cooperation in wild chimpanzees. *Proc Natl Acad Sci USA* 104:7786.
- Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci USA* 109:15716.
- van Leeuwen KL, Hill RA, Korstjens AH. 2020. Classifying chimpanzee (*Pan troglodytes*) landscapes across large-scale environmental gradients in Africa. *Int J Primatol* [Internet]. Available from: <https://doi.org/10.1007/s10764-020-00164-5>
- Lefebvre V, Behringer RR, de Crombrughe B. 2001. L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway. *Osteoarthr Cartil* 9:S69–S75.
- Lester JD, Vigilant L, Gratton P, McCarthy MS, Barratt CD, Dieguez P, Agbor A, Álvarez-Varona P, Angedakin S, Ayimisin EA, et al. 2021. Recent genetic connectivity and clinal variation in chimpanzees. *Commun Biol* 4:283.
- Leturmy P, Lucazeau F, Brigaud F. 2003. Dynamic interactions between the gulf of Guinea passive margin and the Congo River drainage basin: 1. Morphology and mass balance. *J Geophys Res Solid Earth* [Internet] 108. Available from: <https://doi.org/10.1029/2002JB001927>
- Lex A, Gehlenborg N, Strobel H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics* 20:1983–1992.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.

- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*:1303.3997.
- Li H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30:2843–2851.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li YF, Costello JC, Holloway AK, Hahn MW. 2008. “Reverse ecology” and the power of population genomics. *Evolution* 62:2984–2994.
- Lieberman DE, Carlo J, Ponce de León M, Zollikofer CPE. 2007. A geometric morphometric analysis of heterochrony in the cranium of chimpanzees and bonobos. *J Hum Evol* 52:647–662.
- Liu W, Morito D, Takashima S, Mineharu Y, Kobayashi H, Hitomi T, Hashikata H, Matsuura N, Yamazaki S, Toyoda A, et al. 2011. Identification of *RNF213* as a susceptibility gene for Moyamoya disease and its possible role in vascular development. *PLOS One* 6:e22542.
- Liu W, Sherrill-Mix S, Learn GH, Scully EJ, Li Y, Avitto AN, Loy DE, Lauder AP, Sundararaman SA, Plenderleith LJ, et al. 2017. Wild bonobos host geographically restricted malaria parasites including a putative new *Laverania* species. *Nat Commun* 8:1–14.
- Lucazeau F, Brigaud F, Leturmy P. 2003. Dynamic interactions between the Gulf of Guinea passive margin and the Congo River drainage basin: 2. Isostasy and uplift. *J Geophys Res Solid Earth* [Internet] 108. Available from: <https://doi.org/10.1029/2002JB001928>
- Lucchesi S, Cheng L, Janmaat K, Mundry R, Pisor A, Surbeck M. 2020. Beyond the group: How food, mates, and group size influence intergroup encounters in wild bonobos. *Behav Ecol* 31:519–532.
- Malenky RK, Kuroda S, Vineberg EO, Wrangham RW. 1994. The significance of terrestrial herbaceous foods for bonobos, chimpanzees, and gorillas. In: Wrangham RW, McGrew WC, de Waal FBM, Heltne PG, editors. *Chimpanzee Cultures*. Cambridge: Harvard University Press. p. 59–75.
- Malenky RK, Wrangham RW. 1994. A quantitative comparison of terrestrial herbaceous food consumption by *Pan paniscus* in the Lomako Forest, Zaire, and *Pan troglodytes* in the Kibale Forest, Uganda. *Am J Primatol* 32:1–12.
- de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-Rodriguez J, Dupanloup I, Lao O, Hallast P, et al. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354:477–481.

- Mao Y, Catacchio CR, Hillier LW, Porubsky D, Li R, Sulovari A, Fernandes JD, Montinaro F, Gordon DS, Storer JM, et al. 2021. A high-quality bonobo genome refines the analysis of hominid evolution. *Nature* [Internet]. Available from: <https://doi.org/10.1038/s41586-021-03519-x>
- Marzec AM, Kunz JA, Falkner S, Atmoko SSU, Alavi SE, Moldawer AM, Vogel ER, Schuppli C, van Schaik CP, van Noordwijk MA. 2016. The dark side of the red ape: male-mediated lethal female competition in Bornean orangutans. *Behav Ecol Sociobiol* 70:459–466.
- Matsuzawa T, Humle T. 2011. Bossou: 33 Years. In: Matsuzawa T, Humle T, Sugiyama Y, editors. *The Chimpanzees of Bossou and Nimba*. Tokyo: Springer Japan. p. 3–10. Available from: https://doi.org/10.1007/978-4-431-53921-6_2
- Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* 23:23–35.
- McBrearty S, Jablonski NG. 2005. First fossil chimpanzee. *Nature* 437:105–108.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
- Medkour H, Castaneda S, Amona I, Fenollar F, André C, Belais R, Mungongo P, Muyembé-Tamfum J-J, Levasseur A, Raoult D, et al. 2021. Potential zoonotic pathogens hosted by endangered bonobos. *Sci Rep* 11:6331.
- Melin AD, Janiak MC, Marrone F, Arora PS, Higham JP. 2020. Comparative ACE2 variation and primate COVID-19 risk. *Commun Biol* 3:641.
- Melin AD, Orkin JD, Janiak MC, Valenzuela A, Kuderna L, Marrone III F, Ramangason H, Horvath JE, Roos C, Kitchener AC, et al. 2021. Variation in predicted COVID-19 risk among lemurs and lorises. *Am J Primatol* n/a:e23255.
- Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol* 28:659–669.
- Mitani JC. 2009. Male chimpanzees form enduring and equitable social bonds. *Anim Behav* 77:633–640.
- Mitani JC, Watts DP, Amsler SJ. 2010. Lethal intergroup aggression leads to territorial expansion in wild chimpanzees. *Curr Biol* 20:R507–R508.
- Mitteroecker P, Gunz P, Bookstein FL. 2005. Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. *Evol Dev* 7:244–258.

- Moniaux N, Junker WM, Singh AP, Jones AM, Batra SK. 2006. Characterization of Human Mucin MUC17: COMPLETE CODING SEQUENCE AND ORGANIZATION. *J Biol Chem* 281:23676–23685.
- Moscovice LR, Douglas PH, Martinez-Iñigo L, Surbeck M, Vigilant L, Hohmann G. 2017. Stable and fluctuating social preferences and implications for cooperation among female bonobos at LuiKotale, Salonga National Park, DRC. *Am J Phys Anthropol* 163:158–172.
- Mughal MR, DeGiorgio M. 2019. Localizing and classifying adaptive targets with trend filtered regression. *Mol Biol Evol* 36:252–270.
- Muller MN, Kahlenberg SM, Emery Thompson M, Wrangham RW. 2007. Male coercion and the costs of promiscuous mating for female chimpanzees. *Proc R Soc Lond B Biol Sci* 274:1009–1014.
- Muller MN, Thompson ME, Kahlenberg SM, Wrangham RW. 2011. Sexual coercion by male chimpanzees shows that female choice may be more apparent than real. *Behav Ecol Sociobiol* 65:921–933.
- Myers Thompson JA. 2003. A model of the biogeographical journey from Proto-pan to *Pan paniscus*. *Primates* 44:191–197.
- Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T, Kimura A. 2008. Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* 60:727–735.
- Nakano Y, Yamamoto K, Ueda MT, Soper A, Konno Y, Kimura I, Uriu K, Kumata R, Aso H, Misawa N, et al. 2020. A role for gorilla APOBEC3G in shaping lentivirus evolution including transmission to humans. *PLOS Pathog* 16:e1008812.
- Nam K, Munch K, Mailund T, Nater A, Greminger MP, Krützen M, Marquès-Bonet T, Schierup MH. 2017. Evidence that the rate of strong selective sweeps increases with population size in the great apes. *PNAS* 114:1613–1618.
- Narasimhan VM, Rahbari R, Scally A, Wuster A, Mason D, Xue Y, Wright J, Trembath RC, Maher ER, Heel DA van, et al. 2017. Estimating the human mutation rate from autozygous segments reveals population differences in human mutational processes. *Nat Commun* 8:1–7.
- Nei M, Takahata N. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. *J Mol Evol* 37:240–244.
- Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86:641–647.
- Nishida T. 1983. Alpha status and agonistic alliance in wild chimpanzees (*Pan troglodytes schweinfurthii*). *Primates* 24:318–336.

- Nishida T. 2011. Chimpanzees of the lakeshore: Natural history and culture at Mahale. Cambridge: Cambridge University Press
- Nye J, Laayouni H, Kuhlwilm M, Mondal M, Marques-Bonet T, Bertranpetit J. 2018. Selection in the introgressed regions of the chimpanzee genome. *Genome Biol Evol* 10:1132–1138.
- Nye J, Mondal M, Bertranpetit J, Laayouni H. 2020. A fully integrated machine learning scan of selection in the chimpanzee genome. *NAR Genom Bioinform* [Internet] 2. Available from: <https://doi.org/10.1093/nargab/lqaa061>
- Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* 23:263–286.
- Oleksyk TK, Smith MW, O’Brien SJ. 2010. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc B Biol Sci* 365:185–205.
- Osborne MJ, Volpon L, Kornblatt JA, Culjkovic-Kraljacic B, Baguet A, Borden KLB. 2013. eIF4E3 acts as a tumor suppressor by utilizing an atypical mode of methyl-7-guanosine cap recognition. *Proc Natl Acad Sci USA* 110:3877.
- Ozga AT, Webster TH, Gilby IC, Wilson MA, Nockerts RS, Wilson ML, Pusey AE, Li Y, Hahn BH, Stone AC. 2021. Urine as a high-quality source of host genomic DNA from wild populations. *Mol Ecol Resour* 21:170–182.
- Palagi E. 2006. Social play in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*): Implications for natural social systems and interindividual relationships. *Am J Phys Anth* 129:418–426.
- Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S, Karpinski E, Ivancevic AM, To T-H, Kortschak RD, et al. 2018. A comprehensive genomic history of extinct and living elephants. *Proc Natl Acad Sci USA* 115:E2566.
- Paoli T, Palagi E, Tarli SMB. 2006. Reevaluation of dominance hierarchy in bonobos (*Pan paniscus*). *Am J Phys Anth* 130:116–122.
- Parish AR. 1994. Sex and food control in the “uncommon chimpanzee”: How Bonobo females overcome a phylogenetic legacy of male dominance. *Ethol Sociobiol* 15:157–179.
- Parish AR. 1996. Female relationships in bonobos (*Pan paniscus*). *Hum Nat* 7:61–96.
- Parish AR, de Waal FBM, Haig D. 2000. The other “closest living relative”: How bonobos (*Pan paniscus*) challenge traditional assumptions about females, dominance, intra- and intersexual interactions, and hominid evolution. *Ann N Y Acad Sci* 907:97–113.
- Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34:867–868.

- Pennings PS, Hermisson J. 2006a. Soft sweeps II—Molecular population genetics of adaptation from recurrent mutation or immigration. *Mol Biol Evol* 23:1076–1084.
- Pennings PS, Hermisson J. 2006b. Soft sweeps III: The signature of positive selection from recurrent mutation. *PLOS Genet* 2:e186.
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39:1256–1260.
- Pilbrow V. 2006. Population systematics of chimpanzees using molar morphometrics. *J Hum Evol* 51:646–662.
- Pilbrow V, Groves C. 2013. Evidence for divergence in populations of bonobos (*Pan paniscus*) in the Lomami-Lualaba and Kasai-Sankuru regions based on preliminary analysis of craniodental variation. *Int J Primatol* 34:1244–1260.
- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*:201178.
- Porubsky D, Sanders AD, Höps W, Hsieh P, Sulovari A, Li R, Mercuri L, Sorensen M, Murali SC, Gordon D, et al. 2020. Recurrent inversion toggling and great ape genome evolution. *Nat Genet* 52:849–858.
- Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, et al. 2009. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer’s disease. *PLOS One* 4:e6501.
- Potts R. 1998. Variability selection in hominid evolution. *Ev Anth* 7:81–96.
- Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O’Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499:471–475.
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 20:R208–R215.
- Pruetz JD, Ontl KB, Cleaveland E, Lindshield S, Marshack J, Wessling EG. 2017. Intragroup lethal aggression in West African chimpanzees (*Pan troglodytes verus*): Inferred killing of a former alpha male at Fongoli, Senegal. *Int J Primatol* 38:31–57.
- Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature* 486:527–531.

- Przeworski M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–1189.
- Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* 59:2312–2323.
- Pusey A, Murray C, Wallauer W, Wilson M, Wroblewski E, Goodall J. 2008. Severe aggression among female *Pan troglodytes schweinfurthii* at Gombe National Park, Tanzania. *Int J Primatol* 29:949.
- Pusey A, Williams J, Goodall J. 1997. The influence of dominance rank on the reproductive success of female chimpanzees. *Science* 277:828.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing Available from: <https://www.R-project.org/>
- Ralph P, Coop G. 2010. Parallel adaptation: One or many waves of advance of an advantageous allele? *Genetics* 186:647–668.
- Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol* 13:735–748.
- Reichert KE, Heistermann M, Keith Hodges J, Boesch C, Hohmann G. 2002. What females tell males about their reproductive status: Are morphological and behavioural cues reliable signals of ovulation in bonobos (*Pan paniscus*)? *Ethology* 108:583–600.
- Rilling JK, Scholz J, Preuss TM, Glasser MF, Errangi BK, Behrens TE. 2012. Differences between chimpanzees and bonobos in neural systems supporting social cognition. *Soc Cogn Affect Neurosci* 7:369–379.
- Rogers AR. 2019. Legofit: Estimating population history from genetic data. *bioRxiv*:613067.
- Rogers AR. 2021. An efficient algorithm for estimating population history from genetic data. *bioRxiv*:2021.01.23.427922.
- Rogers AR, Harris NS, Achenbach AA. 2020. Neanderthal-Denisovan ancestors interbred with a distantly related hominin. *Sci Adv* 6:eaay5483.
- Sakai N, Hiromi Terami, Suzuki S, Megumi Haga, Ken Nomoto, Nobuko Tsuchida, Ken-ichirou Morohashi, Naoaki Saito, Maki Asada, Megumi Hashimoto, et al. 2008. Identification of *NR5A1* (*SF-1/AD4BP*) gene expression modulators by large-scale gain and loss of function studies. *J Endocrinol* 198:489–497.

- Sakamaki T, Kasalevo P, Bokamba MB, Bongoli L. 2012. Iyondji Community Bonobo Reserve: A recently established reserve in the Democratic Republic of Congo. *Pan Africa News* 19:16–19.
- Sakamaki T, Maloueki U, Bakaa B, Bongoli L, Kasalevo P, Terada S, Furuichi T. 2016. Mammals consumed by bonobos (*Pan paniscus*): new data from the Iyondji forest, Tshuapa, Democratic Republic of the Congo. *Primates* 57:295–301.
- Sakamaki T, Ryu H, Toda K, Tokuyama N, Furuichi T. 2018. Increased frequency of intergroup encounters in wild bonobos (*Pan paniscus*) around the yearly peak in fruit abundance at Wamba. *Int J Primatol* 39:685–704.
- Sánchez-Villagra MR, Geiger M, Schneider RA. 2016. The taming of the neural crest: a developmental perspective on the origins of morphological covariation in domesticated mammals. *R Soc Open Sci* 3:160107.
- Sandel AA, Reddy RB. 2021. Sociosexual behaviour in wild chimpanzees occurs in variable contexts and is frequent between same-sex partners. *Behaviour* 158:249–276.
- Sandel AA, Watts DP. 2021. Lethal coalitionary aggression associated with a community fission in chimpanzees (*Pan troglodytes*) at Ngogo, Kibale National Park, Uganda. *Int J Primatol* 42:26–48.
- Sarich VM, Wilson AC. 1967. Immunological time scale for hominid evolution. *Science* 158:1200.
- Sawyer SL, Wu LI, Emerman M, Malik HS. 2005. Positive selection of primate *TRIM5α* identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci USA* 102:2832.
- Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–175.
- Scarry CJ, Tujague MP. 2012. Consequences of lethal intragroup aggression and alpha male replacement on intergroup relations and home range use in tufted capuchin monkeys (*Cebus apella nigrinus*). *Am J Primatol* 74:804–810.
- Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 46:919–925.
- Schmidt JM, Manuel M de, Marques-Bonet T, Castellano S, Andrés AM. 2019. The impact of genetic adaptation on chimpanzee subspecies differentiation. *PLOS Genet* 15:e1008485.
- Schrider DR. 2020. Background selection does not mimic the patterns of genetic diversity produced by selective sweeps. *Genetics* 216:499–519.

- Schrider DR, Kern AD. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol Biol Evol* 34:1863–1877.
- Schrider DR, Mendes FK, Hahn MW, Kern AD. 2015. Soft shoulders ahead: Spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics* 200:267–284.
- Schwarz E. 1929. Das Vorkommen des Schimpansen auf den linked Kongo-Ufer. *Rev Zool Bot Afr* 16.
- Seong E, Insolera R, Dulovic M, Kamsteeg E-J, Trinh J, Brüggemann N, Sandford E, Li S, Ozel AB, Li JZ, et al. 2018. Mutations in *VPS13D* lead to a new recessive ataxia with spasticity and mitochondrial defects. *Ann Neurol* 83:1075–1088.
- Serckx A, Huynen M-C, Bastin J-F, Hambuckers A, Beudels-Jamar RC, Vimond M, Raynaud E, Kühl HS. 2014. Nest grouping patterns of bonobos (*Pan paniscus*) in relation to fruit availability in a forest-savannah mosaic. *PLOS One* 9:e93742.
- Sharp PM, Hahn BH. 2011. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb Perspect Med* [Internet] 1. Available from: <http://perspectivesinmedicine.cshlp.org/content/1/1/a006841.abstract>
- Shea B. 1983. Paedomorphosis and neoteny in the pygmy chimpanzee. *Science* 222:521.
- Simons EA, Frost SR. 2020. Ontogenetic allometry and scaling in catarrhine crania. *J Anat* [Internet] n/a. Available from: <https://doi.org/10.1111/joa.13331>
- Smith RJ, Jungers WL. 1997. Body mass in comparative primatology. *J Hum Evol* 32:523–559.
- Soto DC, Shew C, Mastoras M, Schmidt JM, Sahasrabudhe R, Kaya G, Andrés AM, Dennis MY. 2020. Identification of structural variation in chimpanzees using optical mapping and nanopore sequencing. *Genes* 11.
- Staes N, Koski SE, Helsen P, Franssen E, Eens M, Stevens JMG. 2015. Chimpanzee sociability is associated with vasopressin (*Avpr1a*) but not oxytocin receptor gene (*OXTR*) variation. *Horm Behav* 75:84–90.
- Staes N, Smaers JB, Kunkle AE, Hopkins WD, Bradley BJ, Sherwood CC. 2019. Evolutionary divergence of neuroanatomical organization and related genes in chimpanzees and bonobos. *Cortex* 118:154–164.
- Staes N, Stevens JMG, Helsen P, Hillyer M, Korody M, Eens M. 2014. Oxytocin and vasopressin receptor gene variation as a proximate base for inter- and intraspecific behavioral differences in bonobos and chimpanzees. *PLOS One* 9:e113364.

- Staes N, Weiss A, Helsen P, Korody M, Eens M, Stevens JMG. 2016. Bonobo personality traits are heritable and associated with vasopressin receptor gene 1a variation. *Sci Rep* 6:38193.
- Stanford CB. 1998. The social behavior of chimpanzees and bonobos: Empirical evidence and shifting assumptions. *Curr Anthropol* 39:399–420.
- Sterck EHM, Watts DP, van Schaik CP. 1997. The evolution of female social relationships in nonhuman primates. *Behav Ecol Sociobiol* 41:291–309.
- Stevens JMG, Vervaecke H, Van Elsacker L. 2010. The bonobo's adaptive potential: social relations under captive conditions. In: Furuichi T, Thompson J, editors. *The Bonobos: Behavior, Ecology, and Conservation*. New York: Springer. p. 19–38.
- Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, Great Ape Genome Project, Bustamante CD, Hammer MF, Wall JD. 2015. The time scale of recombination rate evolution in great apes. *Mol Biol Evol* 33:928–945.
- Stimpson CD, Barger N, Tagliabata JP, Gendron-Fitzpatrick A, Hof PR, Hopkins WD, Sherwood CC. 2016. Differential serotonergic innervation of the amygdala in bonobos and chimpanzees. *Soc Cogn Affect Neurosci* 11:413–422.
- Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Mol Biol Evol* 28:63–70.
- Stone AC, Griffiths RC, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc Natl Acad Sci USA* 99:43.
- Stumpf RM. 2011. Chimpanzees and bonobos: Inter- and intraspecies diversity. In: Campbell CJ, Fuentes A, MacKinnon KC, Bearder SK, Stumpf RM, editors. *Primates in perspective*. New York: Oxford University Press. p. 340–356.
- Stumpf RM, Boesch C. 2005. Does promiscuous mating preclude female choice? Female sexual strategies in chimpanzees (*Pan troglodytes verus*) of the Tai National Park, Côte d'Ivoire. *Behav Ecol Sociobiol* 57:511–524.
- Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, Mohajeri K, Kondova I, Bontrop RE, Persengiev S, et al. 2013. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res* 23:1373–1382.
- Sugiyama Y. 1999. Socioecological factors of male chimpanzee migration at Bossou, Guinea. *Primates* 40:61–68.
- Sugiyama Y, Fujita S. 2011. The demography and reproductive parameters of Bossou chimpanzees. In: Matsuzawa T, Humle T, Sugiyama Y, editors. *The Chimpanzees of Bossou and Nimba*. New York: Springer. p. 23–34.

- Sun P-H, Ye L, Mason MD, Jiang WG. 2012. Protein tyrosine phosphatase μ (PTP μ or PTPRM), a negative regulator of proliferation and invasion of breast cancer cells, is associated with disease prognosis. *PLoS One* 7:e50183.
- Surbeck M, Boesch C, Crockford C, Thompson ME, Furuichi T, Fruth B, Hohmann G, Ishizuka S, Machanda Z, Muller MN, et al. 2019. Males with a mother living in their group have higher paternity success in bonobos but not chimpanzees. *Curr Biol* 29:R354–R355.
- Surbeck M, Coxe S, Lokasola AL. 2017. Lonoa: The establishment of a permanent field site for behavioural research on bonobos in the Kokolopori Bonobo Reserve. *Pan Africa News* 24:13–15.
- Surbeck M, Langergraber KE, Fruth B, Vigilant L, Hohmann G. 2017. Male reproductive skew is higher in bonobos than chimpanzees. *Curr Biol* 27:R640–R641.
- Surbeck M, Mundry R, Hohmann G. 2011. Mothers matter! Maternal support, dominance status and mating success in male bonobos (*Pan paniscus*). *Proc R Soc Lond B Biol Sci* 278:590–598.
- Susman RL ed. 1984. The pygmy chimpanzee: Evolutionary biology and behavior. New York: Springer
- Tajsharghi H, Darin N, Rekabdar E, Kyllerman M, Wahlström J, Martinsson T, Oldfors A. 2005. Mutations and sequence variation in the human myosin heavy chain IIa gene (*MYH2*). *Eur J Hum Genet* 13:617–622.
- Takemoto H, Kawamoto Y, Furuichi T. 2015. How did bonobos come to range south of the Congo river? Reconsideration of the divergence of *Pan paniscus* from other *Pan* populations. *Ev Anth* 24:170–184.
- Takemoto H, Kawamoto Y, Higuchi S, Makinose E, Hart JA, Hart TB, Sakamaki T, Tokuyama N, Reinartz GE, Guislain P, et al. 2017. The mitochondrial ancestor of bonobos and the origin of their major haplogroups. *PLoS One* 12:e0174851.
- Talebi MG, Beltrão-Mendes R, Lee PC. 2009. Intra-community coalitionary lethal attack of an adult male southern muriqui (*Brachyteles arachnoides*). *Am J Primatol* 71:860–867.
- Tan J, Hare B. 2013. Bonobos share with strangers. *PLoS One* 8:e51922.
- Terio KA, Kinsel MJ, Raphael J, Mlengeya T, Lipende I, Kirchhoff CA, Gilagiza B, Wilson ML, Kamenya S, Estes JD, et al. 2011. Pathologic lesions in chimpanzees (*Pan troglodytes schweinfurthii*) from Gombe National Park, Tanzania, 2004–2010. *J Zoo Wildl Med* 42:597–607.

- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.
- Thompson JAM. 2001. On the nomenclature of *Pan paniscus*. *Primates* 42:101–111.
- Thompson-Handler N, Malenky RK, Badrian N. 1984. Sexual behavior of *Pan paniscus* under natural conditions in the Lomako Forest, Equateur, Zaire. In: Susman RL, editor. *The Pygmy Chimpanzee: Evolutionary Biology and Behavior*. Boston, MA: Springer US. p. 347–368. Available from: https://doi.org/10.1007/978-1-4757-0082-4_14
- Tian X, Pascal G, Monget P. 2009. Evolution and functional divergence of *NLRP* genes in mammalian reproductive systems. *BMC Evol Biol* 9:202.
- Tokuyama N, Furuichi T. 2016. Do friends help each other? Patterns of female coalition formation in wild bonobos at Wamba. *Anim Behav* 119:27–35.
- Tokuyama N, Sakamaki T, Furuichi T. 2019. Inter-group aggressive interaction patterns indicate male mate defense and female cooperation across bonobo groups at Wamba, Democratic Republic of the Congo. *Am J Phys Anth* 170:535–550.
- Townsend SW, Slocombe KE, Emery Thompson M, Zuberbühler K. 2007. Female-led infanticide in wild chimpanzees. *Curr Biol* 17:R355–R356.
- Toyoda S, Miyazaki T, Miyazaki S, Yoshimura T, Yamamoto M, Tashiro F, Yamato E, Miyazaki J. 2009. *Sohlh2* affects differentiation of KIT positive oocytes and spermatogonia. *Dev Biol* 325:238–248.
- Tratz EP, Heck H. 1954. Der afrikanische Anthropoide “Bonobo”: Eine neue Menschenaffengattung. *Säugetierkundliche Mitteilungen* 2:97–101.
- Turley K, Frost SR. 2014. The appositional articular morphology of the talo-crural joint: The influence of substrate use on joint shape. *Anat Rec* 297:618–629.
- Tutin CEG. 1979. Mating patterns and reproductive strategies in a community of wild chimpanzees (*Pan troglodytes schweinfurthii*). *Behav Ecol Sociobiol* 6:29–38.
- Tutin CEG, Fernandez M, Rogers ME, Williamson EA, McGrew WC, Altmann SA, Southgate DAT, Crowe I, Whiten A, Conklin NL, et al. 1991. Foraging profiles of sympatric lowland gorillas and chimpanzees in the Lopé Reserve, Gabon. *Philos Trans R Soc B Biol Sci* 334:179–186.
- Uehara S. 1990. Utilization patterns of a marsh grassland within the tropical rain forest by the bonobos (*Pan paniscus*) of Yalosidi, Republic of Zaire. *Primates* 31:311–322.

- Unger S, Gónna MW, Le Béché A, Do Vale-Pereira S, Bedeschi MF, Geiberger S, Grigelioniene G, Horemuzova E, Lalatta F, Lausch E, et al. 2013. *FAM111A* mutations result in hypoparathyroidism and impaired skeletal development. *Am J Hum Genet* 92:990–995.
- Valero A, Schaffner CM, Vick LG, Aureli F, Ramos-Fernandez G. 2006. Intragroup lethal aggression in wild spider monkeys. *Am J Primatol* 68:732–737.
- de Valles-Ibáñez G, Hernandez-Rodriguez J, Prado-Martinez J, Luisi P, Marqués-Bonet T, Casals F. 2016. Genetic load of loss-of-function polymorphic variants in great apes. *Genome Biol Evol* 8:871–877.
- Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, Loul S, Butel C, Liegeois F, Bienvenue Y, et al. 2006. SIV infection in wild gorillas. *Nature* 444:164–164.
- van der Lee R, Wiel L, van Dam TJP, Huynen MA. 2017. Genome-scale detection of positive selection in nine primates predicts human-virus evolutionary conflicts. *Nucleic Acids Res* 45:10634–10648.
- Vervaecke H, Van Elsacker L. 1992. Hybrids between common chimpanzees (*Pan troglodytes*) and pygmy chimpanzees (*Pan paniscus*) in captivity. *Mammalia* 56:667–669.
- Vigilant L, Hofreiter M, Siedel H, Boesch C. 2001. Paternity and relatedness in wild chimpanzee communities. *Proc Natl Acad Sci USA* 98:12890.
- Villanea FA, Schraiber JG. 2019. Multiple episodes of interbreeding between Neanderthal and modern humans. *Nat Ecol Evol* 3:39–44.
- Vy HMT, Kim Y. 2015. A composite-likelihood method for detecting incomplete selective sweep from population genomic data. *Genetics* 200:633.
- Vyklicka L, Lishko PV. 2020. Dissecting the signaling pathways involved in the function of sperm flagellum. *Curr Opin Cell Biol* 63:154–161.
- de Waal FBM. 1989. Peacemaking among primates. Cambridge, MA: Harvard University Press
- de Waal FBM, Lanting F. 1997. Bonobo: The forgotten ape. Berkeley: University of California Press
- Wakefield ML, Hickmott AJ, Brand CM, Takaoka IY, Meador LM, Waller MT, White FJ. 2019. New observations of meat eating and sharing in wild bonobos (*Pan paniscus*) at Iyema, Lomako Forest Reserve, Democratic Republic of the Congo. *Fol Primatol* 90:179–189.

- Walker K, Hare B. 2017. Bonobo baby dominance: Did female defense of offspring lead to reduced male aggression? In: Hare B, Yamamoto S, editors. *Bonobos: Unique in Mind, Brain and Behavior*. Oxford: University of Oxford Press. p. 49–64.
- Wall JD, Hammer MF. 2006. Archaic admixture in the human genome. *Curr Opin Genet Dev* 16:606–610.
- Waller MT, White FJ. 2016. The effects of war on bonobos and other nonhuman primates in the Democratic Republic of the Congo. In: Waller MT, editor. *Ethnoprimatology: Primate Conservation in the 21st Century*. New York: Springer International Publishing. p. 179–192. Available from: https://doi.org/10.1007/978-3-319-30469-4_10
- Watts DP. 2004. Intracommunity coalitionary killing of an adult male chimpanzee at Ngogo, Kibale National Park, Uganda. *Int J Primatol* 25:507–521.
- Webster TH, Couse M, Grande BM, Karlins E, Phung TN, Richmond PA, Whitford W, Wilson MA. 2019. Identifying, understanding, and correcting technical artifacts on the sex chromosomes in next-generation sequencing data. *Gigascience* [Internet] 8. Available from: <https://academic.oup.com/gigascience/article/8/7/giz074/5530326>
- Webster TH, Wilson Sayres MA. 2016. Genomic signatures of sex-biased demography: progress and prospects. *Curr Opin Genet* 41:62–71.
- Wegmann D, Excoffier L. 2010. Bayesian inference of the demographic history of chimpanzees. *Mol Biol Evol* 27:1425–1435.
- Wei T, Simko V. 2021. R package “corrplot”: Visualization of a Correlation Matrix. Available from: <https://github.com/taiyun/corrplot>
- Weigand H, Leese F. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zool J Linn Soc* 184:528–583.
- Wetzel KS, Yi Y, Yadav A, Bauer AM, Bello EA, Romero DC, Bibollet-Ruche F, Hahn BH, Paiardini M, Silvestri G, et al. 2018. Loss of CXCR6 coreceptor usage characterizes pathogenic lentiviruses. *PLOS Pathog* 14:e1007003.
- White FJ. 1986. Behavioral ecology of the pygmy chimpanzee.
- White FJ. 1988. Party composition and dynamics in *Pan paniscus*. *Int J Primatol* 9:179–193.
- White FJ. 1992. Activity Budgets, feeding behavior, and habitat use of pygmy chimpanzees at Lomako, Zaire. *Am J Primatol* 26:215–223.
- White FJ. 1996. *Pan paniscus* 1973 to 1996: Twenty-three years of field research. *Ev Anth* 5:11–17.

- White FJ. 1998. Seasonality and socioecology: The importance of variation in fruit abundance to bonobo sociality. *Int J Primatol* 19:1013–1027.
- White FJ. 2012. *The Quarterly Review of Biology* 87:171–172.
- White FJ, Burgman MA. 1990. Social organization of the pygmy chimpanzee (*Pan paniscus*): Multivariate analysis of intracommunity associations. *Am J Phys Anth* 83:193–201.
- White FJ, Wood KD. 2007. Female feeding priority in bonobos, *Pan paniscus*, and the question of female dominance. *Am J Primatol* 69:837–850.
- Wickham H. 2016. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag Available from: <https://ggplot2.tidyverse.org>
- Wickham H. 2019. stringr: Simple, Consistent Wrappers for Common String Operations.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. 2019. Welcome to the tidyverse. *J Open Source Softw* 4:1686.
- Wilson BA, Petrov DA, Messer PW. 2014. Soft selective sweeps in complex demographic scenarios. *Genetics* 198:669.
- Wilson ML, Boesch C, Fruth B, Furuichi T, Gilby IC, Hashimoto C, Hobaiter CL, Hohmann G, Itoh N, Koops K, et al. 2014. Lethal aggression in *Pan* is better explained by adaptive strategies than human impacts. *Nature* 513:414–417.
- Wilson ML, Wrangham RW. 2003. Intergroup relations in chimpanzees. *Annu Rev Anthropol* 32:363–392.
- Wobber V, Wrangham R, Hare B. 2010. Bonobos exhibit delayed development of social behavior and cognition relative to chimpanzees. *Curr Biol* 20:226–230.
- Won Y-J, Hey J. 2005. Divergence population genetics of chimpanzees. *Mol Biol Evol* 22:297–307.
- Wrangham RW. 1986. Ecology and social relationships in two species of chimpanzee. In: Rubenstein DI, Wrangham RW, editors. *Ecological aspects of social evolution: Birds and mammals*. Princeton, NJ: Princeton University Press. p. 354–378.
- Wrangham RW. 1999. Evolution of coalitionary killing. *Am J Phys Anth* 110:1–30.
- Wrangham RW, Chapman CA, Clark-Arcadi AP, Isabirye-Basuta G. 1996. Social ecology of Kanyawara chimpanzees: implications for understanding the costs of great ape groups. In: McGrew WC, Marchant LF, Nishida T, editors. *Great Ape Societies*. Cambridge: Cambridge University Press. p. 45–57.

- Wrangham RW, Clark AP, Isabirye-Basuta G. 1992. Female social relationships and social organization of Kibale Forest chimpanzees. In: Nishida T, McGrew WC, Marler P, Pickford M, de Waal FBM, editors. *Topics in Primatology*, Vol. 1 Human Origins. Tokyo: University of Tokyo Press. p. 81–98.
- Wrangham RW, Conklin NL, Chapman CA, Hunt KD, Milton K, Rogers E, Whiten A, Barton RA, Widdowson EM, Whiten A, et al. 1991. The significance of fibrous foods for Kibale Forest chimpanzees. *Philos Trans R Soc B* 334:171–178.
- Wrangham RW, Pilbeam D. 2001. African apes as time machines. In: Galdikas B, Briggs N, Sheeran L, Shapiro G, Goodall J, editors. *All Apes Great and Small*. New York: Plenum. p. 5–17.
- Wright SI, Andolfatto P. 2008. The impact of natural selection on the genome: Emerging patterns in *Drosophila* and *Arabidopsis*. *Annu Rev Ecol Evol Syst* 39:193–213.
- Wu Z, Jiang H, Zhang L, Xu X, Zhang X, Kang Z, Song D, Zhang J, Guan M, Gu Y. 2012. Molecular analysis of *RNF213* gene for Moyamoya disease in the Chinese Han population. *PLOS One* 7:e48179.
- Yamakoshi G. 1998. Dietary responses to fruit scarcity of wild chimpanzees at Bossou, Guinea: Possible implications for ecological importance of tool use. *Am J Phys Anth* 106:283–295.
- Yamakoshi G. 2004. Food seasonality and socioecology in *Pan*: Are West African chimpanzees another bonobo? *Afr Study Monogr* 25:45–60.
- Yang C-W, Chang CY-Y, Lai M-T, Chang H-W, Lu C-C, Chen Y, Chen C-M, Lee S-C, Tsai P-W, Yang S-H, et al. 2015. Genetic variations of *MUC17* are associated with endometriosis development and related infertility. *BMC Med Genet* 16:60.
- Yang J, Jin Z-B, Chen J, Huang X-F, Li X-M, Liang Y-B, Mao J-Y, Chen X, Zheng Z, Bakshi A, et al. 2017. Genetic signatures of high-altitude adaptation in Tibetans. *Proc Natl Acad Sci USA* 114:4189.
- Yerkes R. 1925. *Almost Human*. New York: Century
- Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li W-H. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–1518.
- Zheng Y, Wiehe T. 2019. Adaptation in structured populations and fuzzy boundaries between hard and soft sweeps. *PLOS Comput Biol* 15:e1007426.
- Zihlman AL, Bolter DR. 2015. Body composition in *Pan paniscus* compared with *Homo sapiens* has implications for changes during human evolution. *Proc Natl Acad Sci USA* 112:7466.

Zihlman AL, Cramer DL. 1978. Skeletal differences between pygmy (*Pan paniscus*) and common chimpanzees (*Pan troglodytes*). *Fol Primatol* 29:86–94.