# IMPACTS OF ENVIRONMENTAL CHANGE ON WILD NON-HUMAN PRIMATES:

## BEHAVIOUR AND THE HOLOBIONT, WITH IMPLICATIONS FOR DISEASE EMERGENCE

By

**Jan Frederik Gogarten**

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Biology Department
McGill University
Montreal, Canada

March 2017

## DEDICATION

This thesis, like my last, is dedicated to the next generation, especially the young Jolyon Max Gogarten, Oona Daniella Grabowska, Radha Bance Vidich, Rupert James Gogarten, Seren Alana Schönfeld, Thomasina Adele Gogarten, Simone Bance Vidich, Wilfred George Gogarten, and my own Alva Matilde Gogarten. I hope that as a species, we will learn to balance a world where we can reduce human suffering and maximize human joy, while creating a space for wildlife and wilderness, where animals, plants, and microorganisms can co-exist and partake in the joy of existence. It is my deepest hope that humanity leaves you a vibrant and interesting world to enjoy and explore.

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

**A BRIEF ABSTRACT IN ENGLISH**

Environments are not homogeneous or stable, even on small spatial and temporal scales. Rather, abiotic and biotic components of the environment can change and impact one another in major ways and lead to long-term shifts in the stable states of ecosystems. In this thesis I examine how environmental changes impact non-human primate populations. To this end I use historic long-term data and my own field-based studies in Uganda and the Côte d'Ivoire to provide insights into the cascading impacts environmental changes have on non-human primate behaviour, ecology and ultimately microorganism transmission. I find that changing climates and food availability have changed grouping behaviour in some species, and that these changes have cascading impacts on other behaviours including activity budgets, diets, and competitive regimes. I examine gut microbiomes of nine sympatric wild non-human primate species in a community where a strong hunter-prey relationship exists between chimpanzees and colobines. Despite sharing an environment, regular interactions, and this hunter-prey relationship, I find that individuals harbour unique and persistent microbiomes that are influenced by the host's species, social group, and their mother, but not grooming, aggression or spatial proximity. I also find a branching order of host-species phylogenies built using the composition of their microbial communities as characters, which suggests incongruence with known non-human primate phylogenetic relationships, with chimpanzees sister to their colobine prey, possibly due to broader bacterial exposure through hunting. Results suggest that changing environments, influencing changing sociality and in turn predation rates, might have major impacts on wild non-human primate microbiomes. To understand how these changes might impact humans living in proximity to these wildlife populations through the disease emergence process, I present a new method to generate full genomes from wildlife samples containing *Treponema pallidum*, a pathogen responsible for syphilis and yaws disease in humans. I find that this pathogen is widely distributed across primate species in Africa and that this pathogen is closely related to human yaws disease, likely representing a major reservoir. Ultimately, this suggests that environmental changes and corresponding changes in non-human primate behaviour and sociality might be changing disease risk for humans living in proximity to these wildlife populations.

## A BRIEF ABSTRACT IN FRENCH

Les environnements ne sont guère homogènes ou stables, même à petites échelles spatiales ou temporelles. Au contraire, les éléments abiotiques et biotiques de l'environnement peuvent changer et avoir un impact réciproque majeur provoquant des changements à long terme sur l'état initialement stable des écosystèmes. Dans cette thèse, j'examine comment les changements environnementaux influencent des populations de primates (en excluant les humains). À cette fin, j'utilise des données historiques de longue date, ainsi que mes propres études de terrain menées en Ouganda et en Côte d'Ivoire pour fournir un aperçu des impacts des changements environnementaux sur le comportement des primates, l'écologie et finalement la transmission des microorganismes. Nous avons constaté que le changement du climat et la disponibilité des ressources ont modifié le comportement de regroupement chez certaines espèces, et que ces changements ont eu des effets en cascade sur d'autres comportements tels que les le temps alloué aux activités, la diète et les intéractions compétitives. Nous avons étudié les microbiomes intestinaux de neuf espèces de primates sauvages sympatriques vivant dans une communauté où existe une forte relation de chasseurs-proies entre chimpanzés et colobes. Malgré le partage d'un environnement commun où les interactions sont régulières à cause de cette relation de chasseurs-proies, nous avons constaté que les individus possèdent des microbiomes uniques et tenaces influencés par, entre autres, l'espèce de l'hôte, le groupe social, et non le toilettage, l fréquence 'agression ou la proximité spatiale. On trouve également un ordre dans l'enbranchement des phylogénies des espèces auxquelles appartiennent les hôtes construit en utilisant la composition de leurs communautés microbiennes en tant que caractère. Ceci suggère une incongruence avec les relations phylogénétiques déjà connues entre les primates, avec des chimpanzés proches des colobes qui sont leur proie, probablement due à une exposition bactérienne plus large lors de leur chasse. Les résultats suggèrent que les changements de l'environnement influencent un changement de socialité, qui à son tour influence le taux de prédation, et donc, pourraient avoir des impacts majeurs sur les microbiomes des primates vivant dans la forêt. Pour comprendre comment ces changements pourraient avoir un impact sur les humains vivant à proximité de ces populations de primates à travers le processus de maladie émergente, je présente une nouvelle méthode pour générer des génomes complets à partir de ces échantillons de primates contenant *Treponema pallidum*, un agent pathogène responsable de la syphilis et de la maladie du pian chez les humains. Nous constatons que cet agent pathogène est largement répendu chez les espèces de

primates en Afrique et qu'il est phylogénétiquement proche de la maladie du pian humain, représentant probablement un important réservoir. Enfin, cela suggère que les répercutions des changements environnementauxsur le comportement et la socialité des primates pourraient modifier le risque de transmission de cette maladie vis-à-vis des humains vivant à proximité de ces populations de primates.

## PREFACE

### Thesis style

This thesis is written in a manuscript-based format. It consists of five data driven investigative chapters (2, 3, 4, 5, 6), each represented by a manuscript. To help bridge the link between chapters 4 and 5, I have included a review article that was published in *Virology* as an appendix. The style of each chapter follows the scientific journal *Ecology Letters*. All data chapters use data collected from my own fieldwork and long-term studies in Kibale National Park, Uganda and Taï National Park, Côte d'Ivoire. Chapter 6 is augmented with additional samples collected by colleagues in Bijilo Forest Park, Gambia, Niokolo-Koba National Park, Senegal and Lake Manyara National Park, Tanzania.

**Chapter 2:** Gogarten, J.F.*, Jacob, A.L.*, Ghai, R.R., Rothman, J.M., Twinomugisha, D., Wasserman, M.D. & Chapman, C.A. (2015). Group size dynamics over 15+ years in an African forest primate community. *Biotropica*, 47, 101-112.
*These authors contributed equally to the manuscript

**Chapter 3:** Gogarten, J.F., Bonnell, T.R., Brown, L.M., Campenni, M., Wasserman, M.D. & Chapman, C.A. (2014). Increasing group size alters behaviour of a folivorous primate. *International Journal of Primatology*, 35, 590-608.

**Chapter 4:** Gogarten, J.F., Davies, J., Gogarten, J.P., Graf, J., Mielke, A., Mundry, R., Nelson, M.C., Wittig, R., Leendertz, F.H. & Calvignac-Spencer, S. (*In prep.*). Hunting captures more than prey: evidence for microbiome assimilation by chimpanzees.

**Chapters 5:** Gogarten, J.F., Düx, A., Schuenemann, V.J., Nowak, K., Boesch, C., Wittig R.M., Krause, J., Calvignac-Spencer, S.*, & Leendertz F.H.* (2016) Tools for opening new chapters in the book of *Treponema pallidum* evolutionary history. *Clinical Microbiology and Infection*, 22, 916-921.
*These authors contributed equally to the manuscript

**Chapter 6:** Knauf S*, Gogarten JF*, Schuenemann VJ*, De Nys HM*, Düx A, Strouhal M, Mikalová L, Bos KI, Armstrong R, Batamuzi EK, Chuma IS, Davoust B, Diatta G, Fyumagwa RD, Kazwala RR, Keyyu JD, Lejora IAV, Levasseur A, Liu H, Mayhew MA, Mediannikov O, Raoult D, Sinclair B, Wittig RM, Roos C, Leendertz FH, Šmajs D, Nieselt K, Krause J*, Calvignac-Spencer S*. (*In prep.*). Non-human primates as reservoirs for yaws
*These authors contributed equally to the manuscript

**Appendix Chapter:** Gogarten, J.F.*, Akoua-Koffi, C.*, Calvignac-Spencer, S.*, Leendertz, S.A.J.*, Weiss, S.*, Couacy-Hymann, E., Koné, I., Peeters, M., Wittig, R.M., Boesch, C., Hahn,

B.H. & Leendertz, F.H. (2014). The ecology of primate retroviruses–An assessment of 12 years of retroviral studies in the Taï National Park area, Côte d′ Ivoire. *Virology*, 460, 147-153.
*These authors contributed equally to the manuscript

**Contribution of Co-authors and acknowledgements**

This thesis is a compilation of my original contributions. These projects were interdisciplinary and ambitious in both temporal scope and spatial scale. As such, during each of these projects, I frequently relied on the technical experience, long-term data, samples, equipment, funding, and insights of a number of key collaborators. However, I was a major driving force for generating the research questions, project design, analysis of data, and preparation of each of the associated manuscripts.

**Chapter 2**: In this chapter, in collaboration with Aerin Jacob, Ria Ghai, Jessica Rothman, Dennis Twinomugisha, Michael Wasserman, and Colin Chapman, I examined whether there was evidence for dynamism in the group sizes of non-human primates (hereafter NHPs) in Kibale National Park, Uganda over more than 15 years and what environmental and ecological factors might be responsible for observed changes. This work has been published in *Biotropica* (Gogarten *et al.* 2015) and I was largely responsible for integrating the long-term data, designing, and executing the data analysis, and writing the bulk of this manuscript. Aerin Jacob was instrumental in sparking the idea for this project together with Colin Chapman, and we worked closely together on compiling and cleaning the long-term forest data; she is acknowledged as an equally contributing author for these valuable contributions. For the collection of the recent census and forest data, I worked closely with Colin Chapman to organise and assist the trained staff of the Kibale Monkey Project (KMP) headed by Dennis Twinomugisha and Patrick Omeja. Historical data were collected by the KMP and Colin Chapman. All co-authors were integral sources of ideas and for the interpretation of results. All were involved in improving and writing parts of the manuscript, though I integrated all comments and suggestions and I am responsible for the majority of this text. Fred Babweteera, David Marcogliese, Emilio Bruna, Lauren Chapman, Jonathan Davies, Katie Milton, Louis Lefebvre, Marty Lechowicz, Charles Nunn, Tom Struhsaker and two anonymous reviewers provided helpful insights, while David Mills shared his observations of golden cats and Johanna Bleecker helped in creating the map. Richard Wrangham was instrumental in establishing the

tree plots in 1989 and we thank him for his long-term collaboration. I am extremely indebted to the long-term KMP field assistants, particularly Tusiime Lawrence who established the tree plots and has helped monitor them ever since.

**Chapter 3**: In this chapter, I examined the impact of changing group size on the behaviour and ecology of a group of red colobus monkeys (*Procolobus rufomitratus*) in Kibale National Park, Uganda, in collaboration with Tyler Bonnell, Leone Brown, Marco Campenni, Michael Wasserman, and Colin Chapman. This manuscript has been published in the *International Journal of Primatology* (Gogarten *et al.* 2014b). I was responsible for generating the idea for this project, the study design, data analysis, and majority of the writing for this manuscript. This work was conducted using the long-term data from a habituated red colobus group, which were generously provided by Colin Chapman and collected in large part by the personnel of the KMP, though I contributed to the collection of a portion of this data set during my two field seasons in Uganda. My co-authors were sources of valuable discussion that were helpful in guiding these analyses and interpretation of the results. Magnus Bein, Carl Boodman, Carola Borries, Johanna Bleecker, Lauren Chapman, Jonathan Davies, Caitlin Friesen, Ria Ghai, Aerin Jacob, Laura Johnson, Charles Nunn, David Marcogliese, Joanna Setchell, and two anonymous reviewers made suggestions for analyses and provided helpful comments on this manuscript. Catherine Potvin, Jon Sakata, and Christopher Solomon gave particularly valuable advice and recommendations regarding statistical analyses and detailed comments on this manuscript.

**Chapter 4**: In this chapter, I examine factors influencing the bacterial gut microbiome of NHP species living in Taï National Park, Côte d'Ivoire in collaboration with Jonathan Davies, Peter Gogarten, Joerg Graf, Alexander Mielke, Roger Mundry, Michael C. Nelson, Roman M. Wittig, Fabian H. Leendertz, and Sébastien Calvignac-Spencer. This manuscript has not yet been submitted for publication. I designed the sample collection scheme, collected behavioural data, developed and carried out much of the laboratory analyses, and was responsible for the majority of the data analysis and interpretation, and writing of this manuscript. This work is based primarily on samples and data that were collected during my two field seasons in Taï National Park, where I worked closely with the Taï Chimpanzee Project's various Ivorian assistants and international students, some of whom assisted with long-term data collection of fecal samples and the behavioural data used in this analysis. I was responsible for the bulk of the study design

and developing the data analysis and methods, with useful and important input from Jonathan Davies, Peter Gogarten, Joerg Graf, Roger Mundry, Michael C. Nelson, Roman M. Wittig, Fabian H. Leendertz, and Sébastien Calvignac-Spencer. Alexander Mielke was instrumental in helping collect the behavioural data from sooty mangabeys used in this analysis. In addition, the chimpanzee fecal samples used in this chapter were collected by various Ivorian assistants and international students that were working as part of this organization's long-term research program. The Microbial Analysis, Resources and Services (MARS) facility at the University of Connecticut was instrumental in sequencing the 16S amplicon data. Roger Mundry and Will Pearse were supportive in discussing and developing statistical approaches for the analysis of the bacterial community data, while Linda Vigalant and Jack Lester were responsible for generating the mangabey and chimpanzee relatedness data used in this analysis. Alex Mielke, a PhD student at the Max Planck Institute for Evolutionary Anthropology, was extremely helpful in developing the behavioural data collection protocols and assisted in the field component of this project.

**Chapter 5**: In this chapter I reviewed our current knowledge on *Treponema pallidum* evolution and natural history and find that research efforts have been hampered by the difficulty of culturing and propagating *Treponema pallidum*. I worked in collaboration with Ariane Düx, Verena Schuenemann, Kathrin Nowak, Christophe Boesch, Roman M. Wittig, Johannes Krause, Sébastien Calvignac-Spencer, and Fabian H. Leendertz to develop a novel culture-free method for studying *Treponema pallidum* in wildlife by coupling hybridisation capture and next generation sequencing technologies. I was largely responsible for conducting the literature review, designing the study and method, conducting the laboratory and bioinformatics analysis and writing the manuscript. This work has been published in *Clinical Microbiology and Infection* (Gogarten *et al.* 2016). I present data generated with such an approach suggesting that asymptomatic bones from non-human primates occasionally contain enough treponemal DNA to recover large fractions of their genomes. These methods can be applied to other metagenomic samples such as modern biopsies and ancient human bones that can be used to resolve long-standing controversies, while our bone results show that the host range of *Treponema pallidum* is much larger than previously appreciated.

**Chapter 6**: This project was part of a major interdisciplinary and multi-institute effort for which I am an equally contributing first author. I investigated *Treponema pallidum* infections of

wildlife using a novel in-solution hybridization capture approach developed in Chapter 6, which enabled full genome characterization of this pathogen in a multitude of NHP species. I have been involved in collecting samples, designing methods, conducting laboratory and bioinformatics analyses, and writing this manuscript. As part of my dissertation work in Taï National Park, I noticed sooty mangabeys with lesions and symptoms reminiscent of *Treponema pallidum* infections in humans and I worked with the veterinarians Helene De Nys, Fabian Leendertz, and Arianne Düx to collect samples and symptom data. I worked together with Helene De Nys to conduct PCR testing of tissue and fecal samples to confirm the infectious agent. Unfortunately, the regions amplified by these PCRs did not allow us to confirm the subspecies of *Treponema pallidum* or understand how it was related to other confirmed cases in humans or wildlife. This prompted me to develop a novel in-solution hybridization capture approach with Sebastien Calvignac-Spencer that allowed us to generate high-coverage genomes of the pathogen. During this phase, long-term primate researchers working in Senegal and the Gambia described similar symptoms and provided the Leendertz laboratory with samples for testing; I again conducted the laboratory analyses to confirm the bacterial infection and used our new approach to generate whole genomes from *Treponema pallidum* strains infecting these animals. Sascha Knauf has been spearheading an effort to examine the *Treponema pallidum* infections of baboons and had developed a collaboration with David Šmajs to sequence its genome with long-range PCRs. They reached out to the Krause and Schuenemann laboratories to replicate their attempts to generate full genomes from these bacteria using a microarray approach, with great success. To unify our approaches, we confirmed our results using the Krause and Schuenemann' labs microarray approach and decided to combine the data from all of these approaches and different study species into one high impact manuscript, incorporating data from the Krause and Schuenemann lab on the mutation rate of these bacteria generated from human samples. The multitude of equally contributing first authors is a testament to the scale and interdisciplinarity of this research, but I have been heavily involved in this project from the beginning, collecting samples in the field, analyzing them at the bench, working in silica to clean and analyze the next generation sequence data, and ultimately interpreting our results and writing up our findings.

**Appendix Chapter**: In this chapter I conducted a literature review of the available data on retrovirus ecology and evolution of the primates of Taï National Park in collaboration with Chantal Akoua-Koffi, Sebastien Calvignac-Spencer, Siv Aina Leendertz, Sabrina Weiss,

Emmanuel Couacy-Hymann, Inza Koné, Martine Peeters, Christophe Boesch, Beatrice Hahn, and Fabian Leendertz. I was largely responsible for writing up these ideas and bringing together the comments from the many contributors to this manuscript. This work has been published in *Virology* (Gogarten *et al.* 2014a) with Chantal Akoua-Koffi, Sebastien Calvignac-Spencer, Siv Aina Leendertz, and Sabrina Weiss recognized as equally contributing authors in recognition of their interdisciplinary expertise, which was needed to interpret results on this scale and their generous support with ideas, comments, and literature for me to review. This review and integration of long-term data provided important insights into the prevalence, within- and cross-species transmission of primate retroviruses (including to local human populations), and the importance of virus-host interactions in determining cross-species transmission risk. This manuscript highlights how the ecology and evolution of retroviruses may change in a shifting environment and identifies key avenues for future research.

**Statement of originality**: This thesis drew on historic and original datasets to show, for the first time in non-human primates, that group sizes can shift at a large spatial and temporal scale, even while population numbers as a whole remain stable. This has the important implication that for some species we cannot use historic group sizes to estimate population sizes from group density estimates, and also suggests that group sizes may be an important aspect of a species' biology that can respond to changing environments such as the recovering forest in Kibale National Park. I show that changing group sizes also impact the activity budgets and diets of animals, and importantly that these behavioural changes seem to allow individuals to adapt to the shifting demands imposed by living in different group sizes, as fitness remained stable. For the first time, I present evidence that hunting might be transferring gut bacterial between primate species, and also that species and social groups maintain distinct microbiomes, even when they share a common environment and regularly interact. The sharing of gut microorganisms between group members represents an important aspect of the biology of organisms that was likely influenced by group size increases in Kibale National Park. Lastly, I show that hybridization capture methods are able to generate genomic information on *Treponema pallidum* from a diversity of sample types, including bones, which may be extremely useful to answer questions about the origins of these pathogens in ancient human populations. I apply this approach to samples collected from non-human primates across the range of yaws endemicity in humans and show that a number of species across Africa have *Treponema pallidum* infections that are closely

related to human yaws, and that this pathogen likely repeatedly jumped between primate hosts. My findings suggest that non-human primates represent a high-diversity reservoir for this pathogen, and that this reservoir will likely impact global eradication efforts.

**Acknowledgments**

I would like to thank my parents, Maria and Peter Gogarten, for their enthusiasm and excitement about the sciences and unyielding encouragement for my pursuit of strange ideas and dreams. Their input into my career as a primatologist and biologist has proven invaluable, both as scientists and collaborators, travelers, and as citizens of the world, but also as revolutionaries. Their joy at a new idea and in questioning the establishment and doctrine, serves as a ceaseless reminder for me to stay true to my ideals and keep an open mind. They have reminded me that kindness and curiosity should drive scientific pursuits, rather than the displays of ego and power that have reared their head as possible role models along the way. Peter's ability to raise a family with such love and dedication, while pursuing a career in academia at the forefront of so many interesting ideas is inspiring. His own non-linear random walk through ideas and methods has given me the courage to take joy in the deterministic chaos driving my own scientific trajectory and life. I am so grateful for his time, friendship, love, and patience. Maria has always been a steadfast supporter and confidant, trading ideas and trying to figure out the best strategies to bring them to fruition. I apologize for her worry as her child learned to do necropsies on carcasses that died of unknown causes in the depth of remote jungles, and for those late night malaria phone calls followed by days of silence as the cell-phone network went down. They have both been instrumental in introducing me to interesting ideas like the holobiont, horizontal gene transfer, and the microbiome. Maria's valiant and exemplary transition to a life with blindness during the years I worked on this thesis has been inspirational and I apologize for having been absent for so many of these hard life stages. So much happens in the time it takes to write a thesis.

My brother, Jacob Gogarten, has been an inspiration and great friend; joining me in the move to Berlin and on the quest of re-learning the German language, he has been a source of seemingly endless lively discussions and insights. He demonstrated the daring to switch his focus after he completed his degree in music and philosophy, now studying to be a doctor at the Charité. He is the embodiment of the Renaissance man and his knowledge spans so many disciplines. Together we have struggled with combining a love for ideas and the esoteric with a deep-seated want for practice and action. We have tried to confront and be conscious of our privilege as white males in academia, and he has challenged me to maintain my optimism and a dedication for positive

change in the face of the multitude of deep-seated problems in the world. A strong proponent of logic and discussion, he has been instrumental in creating a community of rigor and dialogue. These have been balanced with the levity of making music together and many strudel and falafel filled afternoons. I am extremely grateful that we have had these years living together in the same city as adults.

My sister, Jennifer Morris Gogarten, her wife Stephanie Morris Gogarten, and their four wonderful children, Jolyon, Rupert, Thomasina and Wilfred have been a source of great joy and support. I am so thankful for their strong convictions and for sharing the enthusiasm and energy of their four children. Though it is a reminder that a great deal of time has passed, it has been such a pleasure to watch them be born and grow up over the course of this dissertation. You are a beautiful family of optimists that always believed that these projects and ideas would be fruitful. Your enthusiasm is contagious and you are role models for excelling at both academic and family life. Your bravery and stoicism as a same-sex family, which inherently meant standing up to the bigotry and prejudice that is still rampant in the world, is astounding and a source of inspiration. Your love and support for one another has made me excited to begin a similar adventure.

The years I worked on this dissertation were exciting times; I have fallen in love, got married, and moved in with a wonderful friend and partner. I am extremely thankful for the support and love from my wife, Linn Gogarten. Your patience with long days of work, long absences for fieldwork, and teaching in the tropics, have allowed me to enjoy this process all the more. Having a friend to discuss these ideas and experiences with and get insights into very different ways of seeing the world has made my life so much richer. I look forward to sharing these journeys in the years to come, and to times when we are both done with our formal educations and are freer to roam and develop new projects together. You are a critical thinker, though highly sensitive to the feelings and emotions of others, and have shown me new ways to see and interact with the world. Linn gave birth to our wonderful, curious and intrepid daughter, Alva Matilde Gogarten, and I am excited to share my future explorations with them both. I am so thankful and grateful for the hours of discussions about the ideas in this thesis, but also about the nature of people and their feelings of wonder and awe for the world. Thank you for stoking that feeling in me.

This thesis has taken me to a great many places, from Montreal, Quebec to Berlin, Germany, to Kibale National Park, Uganda, and Taï National Park in the Côte d'Ivoire. I had the opportunity to learn methods at the Center for International Collaboration and Advanced Studies in Primatology at Kyoto University, Japan and attend meetings in Vietnam, Mexico, France, England, the United States of America, and the opportunity to teach American and Malagasy students at Ranomafana National Park and around the country in Madagascar. All of these have put me into contact with many wonderful people, who have been instrumental in not only the science, but my enjoyment of life as well. Beyond McGill University, the Robert Koch Institute and The Max Planck Institute for Evolutionary Anthropology have been instrumental to this research.

I would like to thank my advisor, Jonathan Davies for his help and support throughout my graduate school experience. He has given me the freedom to explore new ideas and work independently and I will be eternally grateful for his kindness and generosity in taking me in when I was an advisor-less and lost student. His enthusiasm and willingness to help certainly enabled me to stay in academia. I am extremely thankful for those late nights coding and generating ideas with Max Farrell at the Quebec Centre for Biodiversity Science's Biodiversity Science Intensive School at the Gault Nature Reserve at Mont-Saint-Hilaire. He is a man of great morals and outstanding character. Thank you.

I would also like to thank the advisor with whom I started this thesis, Colin Chapman. His perseverance through disease and troubled times has proven to me the resilience of humans; I certainly bear no grudge for troubled times, and am grateful for his inspiration, support and willingness to field my bizarre ideas. Some have certainly been fruitful, all have been interesting, and the process a pleasure. I remember meeting Colin at a dinner party held for the McGill fish community, with whom I had been working on weakly electric fish at the time, where he got me started down this long path of primatology, conservation biology, and disease ecology. He encouraged me to do my first fieldwork with primates at Lomas Barbudal, where he had conducted the first primate censuses. Since then, he has always had an open door and encouraged me ever step of the way. I would certainly not have had these wonderful opportunities without him. The scope of his contributions to the field of primatology and forest ecology are enormous

and I am grateful to count myself among those whose lives he has touched, and as his friend, protégé, and colleague.

Given the non-traditional path of my thesis, I am extremely thankful to have had such a caring and supportive committee. Charlie Nunn has been looking out for me since my first scared visit to his office at Harvard in 2008. I am grateful for his willingness to challenge me and my ideas, while facilitating my work with primates and helping me build a research community. His book, *Infectious diseases in primates: behaviour, ecology and evolution* was an early source of inspiration for these ideas and he supported my first forays into phylogenetic comparative methods and learning how to do statistics in R. His writing since has been constant source of provocative ideas. Louis Lefebvre was extremely important in helping me make the transition between advisors and his support has also facilitated my travels and pursuits of other interests. His work on animal innovations and the transmission of ideas through social groups was an inspiration to me as I began thinking about the importance of behaviour for coping with environmental change; he was always there with a novel insight or perspective to a fundamental problems. David Marcogliese has challenged me to think outside the box; his ideas about the importance of parasites in food webs fundamentally changed the way I think about microorganisms and he has driven me to consider ecosystems and organisms in entirely new ways. All of their support and considerateness through trying times has played a pivotal part of my happiness with these projects and ideas.

I am extremely thankful for Fabian Leendertz's willingness to take on and financially and intellectually support a lost graduate student from Montreal. He has played a decisive role in the development of my skillsets, ideas, connections, and joy with science. His ceaseless support for his students is inspiring and I hope to pass on the favour to the next generation of students. His involvement in cutting edge science and big ideas is matched by a dedication to genuine collaboration with and support for research in poorer countries. At the Robert Koch Institute, Sébastien Calvignac-Spencer has been a major source of rigor for the scientific method but also a constant voice of support to try new methods and ideas. He has championed the idea that molecular biology is robust and that I need not be fearful of making a mistake. Thank you for always finding time to answer my questions and discuss ideas.

Roman Wittig and the Max Planck Institute for Evolutionary Anthropology also played a major role in helping me begin fieldwork at an incredible field site and exploring behavioural data in new and creative ways. Thank you for challenging me and fostering a community of discussions and rigorous science and opening doors to amazing datasets and study animals. A special thanks to Roger Mundry, who provided me with shelter in Berlin, training and support in statistical methods and all things related to R, and most importantly friendship and lively discussions. Your comments and ideas following a long walk echoed in my head long after we parted ways. Your commitment to your ideas and ideals is an inspiration and I hope to be able to emulate these in my own life down the line.

The McGill Biology Department has been incredible, and a special thanks goes out to the staff that make these sorts of projects feasible. An equally important expression of thanks to the Dean of Science, Martin Grant, who financially legitimized my planned transition to new projects and labs; he has played a central role in making these experiences possible. To this end, the Chair of the Biology Department, Graham Bell, negotiated on my behalf and went above the call of duty to ensure my studies were not interrupted or disturbed by the medical leave of my then advisor, Colin Chapman.

The broader community of primatologists and ecologists have been so generous in supporting me through difficult transitions and nurturing and exploring new ideas. I am especially grateful to Andreas Koenig, Andrew Macintosh, Jessica Rothman, Patricia Wright, and Thomas Gillespie. I would also like to thank Carola Borries who always lent a helping hand and offered the sound advice ‘focus and finish’, which I admired, but fully admit to not having been able to put into practice. I thank my old Stony Brook colleagues and friends, Andrea Baden, Clara Scarry, David Fernandez, James Herrera, Jessica Lodwick, Leone Brown, Matthew Banks, Rachel Jacobs, Roberta Salmi, and Wendy Erb, whom I regularly ask for advice and guidance and am always excited to see at meetings or the occasional visit.

My colleagues in the Chapman Lab were extremely important in developing these ideas and surviving the many transitions involved in this dissertation. Particular thanks to Michael Wasserman, Ria Ghai, Aerin Jacob, Tyler Bonnell, Sarah Turner, Johanna Bleecker, and Chesley Walsch for their support, camaraderie and friendship. The broader Montreal community whom I was lucky enough to call friends and colleagues were instrumental to surviving hardships and

enjoying life to its fullest. This includes but is certainly not limited to: Adam Herman, Amy Vaillancourt, Andrea Reid, Beth Nyboer, Caitlin Friesen, Carl Boodman, Dalal Hanna, Darryl Singh, David Rempel, Edward Wong, Felipe Pérez Jvostov, Victor Frankel Vilches, Gabrielle Abdelahad-Acosta, Georgia McNeil, Grace Bell Herman, Ian Hatton, Jann Tomaro, Jessica Morgan Fahey, Kristian Jebsen, Kyle Bobiwash, Magnus Bein, Matthew Osmond, Marie-Hélène Greffard, Malie Lessard-Therrien, Mark Heffernan, Max Farrell, Morgan Boenke, Nick Schofield, Patrick Sabourin, Corey Chivers, Shahin Muttalib, Josie Iacarella, Carly Ziter, Benjamin Pyenson, Suzanne Gray, Chelsea Chisholm, Ted Daigle, Tammy Elliott, Siobhan McPherson, Sophia Kehler, Tibor Timar, Vanessa Finnie, Vincent Fugère. Many of you have been instrumental to my life and patient with me while I tried to learn French and to navigate the Canadian system; you are a wonderful community and great friends and I remember my time with you all with great warmth in my heart.

Similarly in Berlin, my colleagues in the Leendertz Lab were a source of laughter and great discussions. Particular thanks to: Alexander Lang, Andreas Sachse, Constance Hoffman, Doris Wu, Fee Zimmerman, Grit Schubert, Helene De Nys, Kim Grützmacher, Maja Kovacev-Wegener, Markus Ulrich, Maude Pauly, Sabrina Weiß, Siv Aina Leendertz, Susanne Köhler, Therese Löhrich, and Verena Keil. A special thanks to Ulla Thiesen and Kevin Merkel; who have always showed such patience with me as I learned to navigate a molecular lab, who have encouraged me to make mistakes and have helped me clean up the messes, who have created a encouraging and safe space to learn, and dealt with me as a nervous and worried non-molecular biologist who was tasked with setting up and running a laboratory in the field. Ulla has driven me to laugh and enjoy the often-overwhelming ride and shared the joy of learning French along the way. I am grateful to Freda Madinda and Malie Lessard-Therrien for their help and assistance translating the abstract to French and always providing interesting thoughts and ideas on these projects. I am extremely thankful for Colleen Gault, who started me along this path of primatology by hiring me to work in Lomas Barbudal and live in a small house at the edge of tropical dry forest in Costa Rica. I am grateful that our life paths have taken us both to Germany and that we have continued to be able to spend so much time together; your way of seeing the world is a beautiful one and it has been fun having you as a mentor and friend along this journey. For giving me the opportunity to move and keeping me from turning into a vegetable during the lab work, analysis, and writing phases of this dissertation, I thank my ultimate Frisbee teammates

in *Parkscheibe* and *UltimateConnects*. The German community was incredibly welcoming and I have made a great many friends in the time I got to spend there. For their patience as I learned the German language and making my transition to life in Germany so smooth and pleasurable, I thank my dear friends: Katja Galkina, Klaus Schmidt, Marcus Cieslak, Rosa Gerhards, Sabine Graudenz and Thomas John. Also a great thanks to the occupied house project on the Nehringstraße, where I was able to live with and experience an alternative community and transition into life in Berlin; the times were wonderful and I am grateful for your welcoming generosity.

I certainly would not have enjoyed my time in Germany, or likely met my wife and maintained such a good work life balance without my good friend Mathias Oster; our regular sauna evenings helped me survive the winters in Berlin, the discussions and dinners drove me to rapidly improve my German but also expand my horizons and to be more critical about the role of science in our society. His love for nature and dedication to movement and exploration have kept me healthy and of sound mind, and his friendliness to strangers seems to know no bounds. We have had so many beautiful adventures and I hope that wherever my travels take me, he knows I would like to return the favour and share my food, social network, time, and housing for as long as he would like to stay. He has set the gold standard for how I will try to treat travelers and those seeking to build their homes in new places. Thank you for improving my life so thoroughly.

Having worked in so many parks and projects has meant that there are a great deal of people to thank for their support. Individuals who assisted with specific projects are acknowledged in the following chapters, but special thanks for field support and comradery in Taï National Park to Alexander Mielke, Bekah Myers, Jonathan Müller-Tiburtius, the camp managers Vera Leinert and Sylvain Lemoine, and the mangabey field assistants, Simone T. Kannieu, Gnimion, and Daniel. In Uganda, Patrick Omeja and Dennis Twinomugisha were good friends and instrumental to making the work there smooth and possible at all. The Chapman's field teams in Kibale National Park and Nabugabo were generous with their time and skills and a joy to work and live with.

A special thanks to Ariane Düx, whom I have had the joy to spend a great deal of time with over the past several years; she has been instrumental to this thesis and also become a great friend. This experiences spanned the globe; late nights over a nutella jar in camp at Taï Forest, teaching

me how to do a necropsy in tyvek suit in the middle of the jungle with midges biting our faces around our eyes, to long discussions about food and any topic under the sun during our days following chimpanzees and managbeys during our two shared field seasons there, to her 'assistance' in the laboratory where she quickly surpassed my own knowledge and skills and drove the lab component of these studies into interesting directions. The discussions have been interesting, your scepticism refreshing, and I hope we can collaborate on laboratory and field science for the rest of our careers; it has been an honour working with you.

And of course, I will not forget my roots and my old childhood friends who remain some of my closest confidants: Zbigniew Grabowski, David Vidich, Roston Kirk, and Nate Ash-Morgan and their expanded network of Sara Swetzoff and Jessie Bance. Their children Oona Daniella Grabowska and Radha Bance Vidich are beautiful. The discussions are always lively and fruitful and I hope our future adventures bring us to similar places and with more time to share. Our times at Tupper lake, my childhood and adventures in the woods of New England, around bonfires and climbing the hills of Mansfield, have echoed in my mind as a sort of Eden, something I try to emulate or remind myself of when times are difficult or friends harder to find. Thank you for your courage and inspiration to travel widely, to throw yourselves onto opportunities, to spur me on to test my limits, how you foster your artistic sides, which has inspired me to not give up on my own, for coming to Japan and climbing mountains and sharing hot tubs and snow adventures with Hichan, for your wildness and laughter. I hear echoes of your voices long after we have parted ways. Thank you for your continued support and friendship and for always taking me in when I show up on your doorsteps.

While conducting these studies, I was supported by a Graduate Research Fellowship from the National Science Foundation, Vanier Canada Graduate Fellowship awarded by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research Systems Biology Training Program, the German Academic Exchange Service's (Deutscher Akademischer Austausch Dienst) long-term research grant to Germany, a McGill University Biology Graduate Excellence Fellowship, and the McGill Biology Department's Writing-Award. I received generous research support from the Explorer Club's Eddie Bauer Young Explorer Grant, two McGill University Research Travel Awards, three Systems Biology Travel Awards, five Quebec Centre for Biodiversity Science Excellence Awards, a Graduate

Research Mobility Award from McGill University, the Cooperative Research Program of the Center for International Collaboration and Advanced Studies in Primatology at Kyoto University, McGill University's Dean of Science's Research Fund, the Systems Biology Rotation Support Award, and the Göttingen Social Structure Summer School Travel Grant.

I have had the pleasure of spending a great deal of time with many non-human primates; I thank them for their patience and tolerance. At times I wish I could have broken down the objectivity and reached out to console you in times of trouble, treat an injury, or get involved in the grooming market. Alas, I have tried my best to stay out of your social lives and remained a passive observer, but certainly do not underestimate the impact I must have had on your daily lives. You have certainly greatly impacted mine. Lastly, I would like to acknowledge the forests and ecosystems where I have had the pleasure to spend so much time and learn such a great deal. I am eternally grateful to the giant trees and shady corners in these forests I have been lucky enough to call my temporary home. I sincerely hope that some of these ideas will ultimately lead to practice and change and ensure your persistence into the future.

For those people and organizations I have forgotten; it likely lies not in the magnitude of your influence on these ideas, projects and my life, but on the frailty of my memory. I apologize. For all my friends and family, I am grateful for your willingness to let me go into the world and pursue these adventures, yet aware that these travels and pursuits have inherently cost me time with those I love. My sincerest gratitude and thanks.

# CHAPTER 1

## INTRODUCTION

## 1. 1 OVERVIEW

In this thesis I examined how environmental changes impact non-human primate (NHP) populations, the transmission of microorganisms within and among them, and ultimately their conservation status. To this end I used both historic long-term data and my own field-based studies in Uganda and the Côte d'Ivoire to provide insights into the cascading impacts environmental changes have on microorganism transmission within these diverse NHP communities and ultimately on the humans living in proximity to these wildlife populations through disease emergence processes. I examined the impact of environmental change on NHP behaviour, health, and fitness and the mechanisms by which these may influence and interact with NHP pathogens and symbionts.

### 1.1.1 Environmental change

Increasing evidence suggests that environments are not homogeneous or stable, even on relatively small spatial or temporal scales (Turner *et al.* 1993; Turkington 2009; Chapman *et al.* 2010a; Mori 2011). These changes go beyond seasonality and are taking place in both the abiotic and biotic environment. The abiotic and biotic components of an environment impact one another in major ways and can lead to long-term shifts in the stable states of ecosystems (Scheffer *et al.* 2001). The stability of ecosystems to changes has been a major area of interest for ecologists, though our predictive capacity remains extremely limited (Holling 1973; Ives & Carpenter 2007). Unfortunately, environmental change is only expected to increase in the coming decades as a result of accelerating rates of anthropogenic disturbance and global climate change (Opdam & Wascher 2004). Darwinian evolution by natural selection may be too slow to allow animals, particularly for those with a slow life histories and small population sizes, to cope and adapt to these rapidly changing environments. Other adaptive processes that may allow some species' to cope with rapidly changing environments are therefore of increasing interest (Stockwell *et al.* 2003).

#### *1.1.1.1 Abiotic factors*

Abiotic change has been well documented over the last few decades. As result of increasing greenhouse gases in the atmosphere, due in large part to human use of fossil fuels, global

temperatures have already risen dramatically since the industrial revolution. Projections for future change vary depending on the scale of the human response, though even the most dramatic interventions representing a 'best case' scenario, suggest we will only be able to limit the rise in global temperatures to 2°C. Climate warming is predicted to have major impacts on many life forms by changing chemical and enzymatic processes, with recent changes having been linked to shifting species' ranges and phenologies (Parmesan 2006). Anthropogenic disturbance is not limited to the emissions of greenhouse gases and particulates; habitat transformation has had equal or even larger impacts historically. Deforestation has major impacts not only on the biotic components of an ecosystem through habitat lost, but also changes many abiotic factors in the adjacent remaining forest. For example, wind, air moisture, soil moisture, and sunlight all are different between forest edges and the interior (Murcia 1995). Deforestation and changing land cover can lead to erosion and soil loss that is difficult to recover.

#### *1.1.1.2 Biotic factors*

Interactions between species additionally shape their abundance and distribution. Organisms play major roles in modifying and creating habitats, with some species serving as ecosystem engineers (Jones *et al.* 1994, 1997; Wright & Jones 2006). Increasing rates of anthropogenic biodiversity loss are thought to be the result of myriad factors including the abiotic changes listed above, but also including bushmeat hunting, overharvesting, and other forms of biotic resource extraction such as logging (Campbell *et al.* 2011; N'Goran *et al.* 2012). Shifts in the abiotic environment will likely drive change important ecosystem processes including the phenology of these systems (Sparks & Carey 1995; Parmesan 2006). Beyond phenological changes in fruit and leaf availability through time, there may also be changes in nutrient and energy availability across landscapes (Gogarten *et al. 2012b;* Rothman *et al.* 2015). Pathogen and symbiont dynamics represent one such important ecosystem process that is likely to change in response to changing community composition and abiotic environment (Lafferty 2009).

#### *1.1.1.3 Pathogens and symbionts*

Theory suggests that parasites may play major roles in stabilizing ecosystems (Marcogliese 2005), but many parasites are also threatened with extinction. Removal of some parasites may result in unexpected consequences for fitness of hosts at a population level: for example removing helminths from individual wild African buffaloes improved the health and well-being

of the treated individuals, but increased the spread of tuberculosis among the population as a whole, ultimately decreasing herd fitness. Those worm-free buffalo appeared to live longer but stayed infected with tuberculosis and spread the infection more in the herd (Ezenwa & Jolles 2015). Decreasing population sizes and fragmentation of landscapes may also reduce the genetic resistance available in populations to respond to disease perturbations. Similarly pathogen ranges and animal densities and behaviour are known to vary as a result of food availability (Chapman *et al.* 2010b; Chapman *et al.* 2015a; Chapman *et al.* 2015b), hunting pressure (Rosenbaum *et al.* 1998; Wright 2003), and disease (Milton 1996) and many animal populations have been declining globally. There are conflicting predictions about whether parasites and symbionts will become less or more virulent with decreasing host animal densities (Bolzoni & De Leo 2013; Chapman *et al.* 2015a; Chapman *et al.* 2015b) or whether these microorganisms will simply go extinct, as suggested by comparative studies showing that threatened species have less parasites (Altizer *et al.* 2007). Interestingly, in some locations, animal populations are being compressed into smaller areas due to emigration from disturbed areas into remaining suitable habitat (Cowlishaw 1999); factors increasing host proximity and contact rate are generally thought to increase parasite transmission among hosts and may relax counter-selection on parasite virulence (Anderson 1978; Altizer *et al.* 2003; Bonnell *et al.* 2010). To date, little research is available to inform our understanding of how microorganism transmission is changing in the face of changing host distributions and contact rates with humans and wildlife. Biotic responses to changing ecosystems can be complex and interactions between life forms can both stabilize or destabilize a system (Holling 1973).

### 1.1.2 Compensatory behavioural change

In light of the major environmental changes facing organisms, behavioural change represents one avenue for animals to rapidly adapt to changing environments (Sih *et al.* 2011). Behavioural plasticity may be particularly important for long-lived organisms, which must cope with a variety of environments and different selection pressures in their lifetime (Miner *et al.* 2005; Krützen *et al.* 2011; van Woerden *et al.* 2011). There are many difference types of behaviours that animals can modify, but social behaviours represent an important facet that influences many other aspects of an animals' ecology and evolution (Elgar 1989; Altizer *et al.* 2003; Ezenwa 2004; Borries *et*

*al.* 2008; Kuehl *et al.* 2008). Primates represent a particularly social order; compared to non-Primate mammals; ~33% of non-Primate mammal genera form year round associations, while 75% of Primate genera live in year round associations (van Schaik & Kappeler 1997). Inter- and intra-specific variation in social group size and their degree of sociality is a major feature of NHP societies that impacts many aspects of their ecology and evolution (Altizer *et al.* 2003; Nunn *et al.* 2009; Griffin & Nunn 2012; Rifkin *et al.* 2012). Grouping confers many costs and benefits (Alexander 1974; van Schaik 1983; Wrangham *et al.* 1993; Janson & Goldsmith 1995). Theory suggests that there may be some optimum group size for any given environment, at which groups are expected to stay at (or slightly above: Sibley 1983). Changing their group sizes thus might be one rapid way that animals cope with changing biotic and abiotic environments. As a result of such changes, many aspects of animals' social organization could be changing on a faster time scale than previously appreciated. Such changes are expected to have cascading impacts on health and ultimately survival of individuals living in different social contexts. The link between sociality and health has received considerable research interest (reviewed in: Kappeler *et al.* 2015; Nunn *et al.* 2015), though much less is known about how such systems will respond to the environmental changes discussed above (Chapman *et al.* 2015b).

### 1.1.3 Microbiomes and the holobiont

Changes in the composition of the holobiont (i.e., the host with all of its symbiotic associated organism, along the entire mutualism–parasitism continuum) might be another way that animals are able to adapt to rapidly changing environments (Dheilly 2014). The microbiome has been estimated to make up 90% of cells in an organism and to represent 99% of the genes present in an organism (though much debate remains surrounding the consistency and validity of these numbers: Sender *et al.* 2016). This microbiome impacts a broad array of processes including a host's ability to access nutrients (Tremaroli & Bäckhed 2012), development and tissue maturation (Collins *et al.* 2012), and health by processes such as pathogen exclusion and immune system priming (Hooper *et al.* 2012). This microbial diversity seems to be versatile, which may allow organisms to adapt more rapidly to changing conditions and function optimally in a particular habitat (Zilber-Rosenberg & Rosenberg 2008; Amato *et al.* 2014). For example, the gut microbiome of mice kept at low temperatures diverged from those kept at high temperatures,

and when the gut microbiome of cold-exposed mice were transplanted to germ-free mice, the recipients produced more beige fat cells burning higher amounts of energy and generating more heat than white fat cells (Chevalier *et al.* 2015). Yet we know surprisingly little about what influences microorganism community assembly and how differences in community structure may impact a host's fitness. The rather intriguing hologenome theory of evolution considers the holobiont as the unit of selection in evolution (Zilber-Rosenberg & Rosenberg 2008); while considerations of the individual as a holobiont are not new, technological advances now allow us to consider these ideas empirically. Interestingly, first insights into the broader holobiont of NHPs suggest non-vertical transmission of components of the holobiont and hologenomes (i.e., the sum genetic information of the NHP host and all of its microbiota) does occur (Tung *et al.* 2015; Moeller *et al.* 2016b). Changing sociality and group sizes could therefore have major impacts on the distribution and composition of both mutualistic and pathogenic microorganism within and between social groups.

The comparative work spearheaded by Nunn and Altizer *et al.* examining the influence of behaviour, physiology, and ecology on NHP parasite communities has made exciting forays into understanding factors influencing holobiont assembly in NHPs (Altizer *et al.* 2003; Nunn & Heymann 2005; Altizer *et al.* 2007; Gillespie *et al.* 2008; Griffin & Nunn 2012; Rifkin *et al.* 2012). These studies though, are limited in that the databases used in these analyses have focused largely on a narrow aspect of the holobiont (i.e., largely non-pathogenic helminths) and data collection has been opportunistic, with the parasite identification methods used to date possibly missing large amounts of cryptic diversity (Ghai *et al.* 2014). While representing the best information currently available, much of the data used in these studies comes from different ecosystems and represent population level aggregations, leaving many of the mechanisms for observed patterns and how they play out in an ecosystem, poorly resolved. Helminth infections are known to ameliorate inflammatory diseases and intriguing studies on model organisms suggest that this effect is mediated by changes in the gut microbiome community; transferring the gut microbiota of helminth infected mice to animals without helminth infections through co-habitation, while ensuring no helminth transfer was possible as adult worms could not replicate within their hosts nor infect new hosts, was sufficient to provide protection against inflammatory diseases (Zaiss *et al.* 2015).

Recent research suggests that viral communities in a NHP species are non-randomly assembled across a landscape (Anthony *et al.* 2015); this analysis was unable to determine the factors driving this pattern, but suggests well-designed studies may be able to identify the factors shaping variance in microorganism communities between hosts across a landscape (Ostfeld *et al.* 2005). Few community microbiome studies using standardized methods are available from NHPs to date, though these few have provided exciting first insights into how this vastly understudied diversity is assembling. For example, aspects of the microbiome may be readily transmitted between individuals and that social interactions, environments, and relatedness may influence these communities (Tung *et al.* 2015). Phylogenetic relatedness plays a major role in predicting the similarity between species (Moeller *et al.* 2014), though not many examinations of multiple species within a single ecosystem are available to partition out the effect of environment from phylogeny (Moeller *et al.* 2013a; McCord *et al.* 2014; Fogel 2015). Captivity seems to have a major effect on the microbiome, as does habitat type and seasonality (Amato *et al.* 2015). Habitat disturbance seems to mirror many of these effects (Amato *et al.* 2013; Barelli *et al.* 2015), though this has not been observed in all ecosystems (McCord *et al.* 2014), possibly reflecting the resilience of more diverse NHP communities to perturbation. Disease also seems to impact the gut bacterial community for some species and pathogens (Moeller *et al.* 2013b), but not for others (Moeller *et al.* 2015). Critically, the finding of antibiotic resistance genes in wild NHP micobiomes, likely resulting from antibiotic use in human communities living in proximity to these NHP populations, suggests perturbations may be subtle and widespread (Rwego *et al.* 2008). To date, the scale and importance of intra- and inter-species microorganism transmission for human and NHP health remains largely unknown.

### 1.1.6 The importance of non-human primates for conservation and human health

The majority of human infectious diseases have their origins in wildlife populations (Jones *et al.* 2008); for example the current West Africa Ebola virus disease outbreak is thought to have resulted from a spillover from a bat reservoir (Marí Saéz *et al. 2015*) and the human immunodeficiency viruses HIV-1 and 2 seem to have simian origins (Hahn *et al.* 2000). Understanding wildlife disease ecology and evolution represents a major hurdle for predicting and mitigating human disease risk and ensuring global biosecurity. The close evolutionary

relationship and similar physiology of NHPs and humans make NHPs a likely source for zoonotic transmission of pathogens (Calvignac-Spencer *et al.* 2012b). Research suggests that rates of zoonotic disease transmission may be increasing (Jones *et al.* 2008), but understanding the complicated process of disease emergence, particularly in light of a changing environments, is still in its infancy. Understanding what microrganisms represent pathogens in NHP populations is a vital first step for understanding and mitigating risk.

Studies over the past decade have demonstrated that infectious diseases have joined bushmeat hunting and habitat loss as major drivers of population declines in wildlife populations, and are of particular concern for a number of NHP species (Daszak *et al.* 2000; Leendertz *et al.* 2006; Ryan & Walsh 2011). These findings suggest that disease is a strong force of selection driving evolutionary and ecological processes and is a major conservation concern (Freeland 1976; Williams *et al.* 2008). The finding of antibiotic resistance genes (Rwego *et al.* 2008), gastrointestinal parasites from humans and livestock (Salyer *et al.* 2012), and human respiratory viruses (Köndgen *et al.* 2008) in wild NHP populations, suggests anthropogenic disturbance is increasing disease risk and human disease spillover into wildlife populations. Further, an understanding of the evolution and between species transmission of these pathogens between NHPs and humans living at the human-wildlife interface will pave the way for developing mitigation efforts to improve the health of both humans and wildlife and providing baseline levels that could be useful for detecting change.

### 1.2 THESIS OUTLINE

The overarching objective of this thesis is to examine how environmental changes might be impacting wild NHP populations. The studies presented represent a combination of historical long-term data and my own field-based studies, while the interdisciplinary nature of the questions addressed has forced me to integrate ecological methods and theory with novel molecular approaches and spatially and temporally broad datasets. I had the opportunity to conduct field studies at two of the longest running field sites targeting diverse NHP communities in tropical rainforests. I used long-term data to examine the stability of NHP food availability, their forest's composition and structure, their diseases, and their predators in Kibale National

Park, Uganda; Kibale represents some of the highest primate biomass and diversity in the world and has been the focus of long-term study for nearly a century, greatly facilitating such longitudinal insights. I sought to understand how NHPs are changing their group sizes in response to changes in important ecological factors (**Chapter 2**) and how these changes in social group size in turn impact competition, activity budgets, and ultimately fitness of these animals (**Chapter 3**). I looked to understand how these changes in social group size and behaviour might be impacting the holobiont, specifically the bacterial gut microbiomes of NHP populations; to accomplish this, I conducted field work in Taï National Park, Côte d'Ivoire, and examined how social networks and between species interactions, particularly the hunting colobus monkeys by chimpanzees, impact the bacterial gut microbiome of the NHP community (**Chapter 4**). I then looked to see how the changes observed in Kibale National Park might be impacting microorganisms spilling over to the neighbouring human population in this region; I used field studies on the retroviruses of NHP populations in Taï National Park and conducted a literature review to access how environmental changes may impact retroviruses circulating in this NHP community (**Appendix Chapter**). Lastly, to solidify the importance of understanding the ecology and evolution of NHP holobionts for both NHPs and human health and conservation, I identified a bacterial species associated with an extremely severe etiology, and made the important discovery that a widely distributed human pathogen infecting hundreds of thousands of people globally despite persistent large scale eradication efforts, *Treponema palladium pertenue,* is likely the product of recent zoonotic transmission. To this end, I developed a novel method for the sequencing of full genomes from metagenomic samples and demonstrate its utility on NHPs from Taï National Park (**Chapter 6**). Furthermore, using these novel methods, I document that this microorganism has recently radiated across the Primate order, causing severe symptoms in many NHP species; continuous spillover is a major risk and could be responsible for the failure of global eradication efforts to date (**Chapter 7**).

### 1.2.1 Chapter 2

In my second chapter, I examined the causes and consequences of changing group sizes in a NHP community, with an emphasis on changing food availability, disease, and forest structure (Gogarten *et al.* 2015). Group size can impact many aspects of the ecology and social

organization of animals, and is thus an important ecological feature of NHP societies. I used historical long-term data to investigate group size stability for five NHP species in Kibale National Park, Uganda from 1996-2011. Surprisingly, I found that while group sizes did not change for most species, red colobus monkeys' (*Procolobus rufomitratus*) group sizes increased at all spatial scales for which I was able to collect data, though small sample sizes for some species examined suggest that the power to detect changes for these species was small. To augment this broad-scale survey data, I analyzed several years of demographic data from three habituated groups of redtail monkeys (*Cercopithecus ascanius*), eight groups of black-and-white colobus (*Colobus guereza*), and one red colobus group. In this dataset, the red colobus group size increased from 59 to 104 individuals, while redtail monkey and black-and-white colobus group sizes were stable, mirroring the broader survey results. To understand mechanisms behind the observed group size changes in red colobus versus the stability observed in other NHP species, I examined forest dynamics at two spatial scales between 1990 and 2013, considered changes in predator population sizes, and explored evidence of disease dynamics. Using these diverse sources of data, I found that the cumulative size of all trees and red colobus food trees increased over 24 years, suggesting that changing food availability might be one of the factors related to the group size changes observed for red colobus, while data on predation and disease suggest these factors played less of a role. Overall, the NHP and forest results from this chapter suggest that Kibale is in a non-equilibrium state and underwent major changes over the last two decades. I discuss ways that future conservation and management can take these important insights into consideration.

### 1.2.2 Chapter 3

In my third chapter, I examined how changes in group size influence NHP behaviour and ecology, with an emphasis on competition, activity budgets, and ultimately fitness (Gogarten *et al.* 2014b). Much of what is known about the effects of group size on the behavioural ecology of animals has come from comparisons across multiple groups of different sizes. These findings may be biased, because behavioural differences across groups may be more indicative of how environmental variation influences animal behaviour, rather than the variance in group size itself. To circumvent this limitation, I used longitudinal behavioural and survey data to examine how

changes in group size across time affect the behaviour of the folivorous red colobus monkey in Kibale National Park, Uganda. Based on six years of data on a group that increased from 57 to 98 members and controlling for food availability, I found that increasing group size resulted in changing activity budgets. Specifically, as group size increased, individuals spent less time feeding and socializing, more time traveling, and increased the diversity of their diet. These changes appear to allow animals to compensate for the greater scramble competition apparent at larger group sizes, as increasing group size did not show the predicted relationship with lower female fecundity. These results support recent findings documenting feeding competition in folivorous NHPs, which had previously been predicted to not have to compete over resources due to the abundance of leaves in rainforest ecosystems. My findings also document the behavioural flexibility of these animals, an important trait that allows social mammals to maximize the benefits of sociality (e.g., increased vigilance), while minimizing costs (e.g., increased feeding competition) in changing environments. Concurrently I was involved as a co-author in a series of studies not included in this dissertation, which examined how these changes impact gastrointestinal parasites (Chapman *et al.* 2012), group genetic structure and dispersal (Miyamoto *et al.* 2013), the cascading impact of these changes for population sizes (Chapman *et al.* 2015a; Chapman *et al.* 2015b), polyspecific associations and between-species interactions (Gogarten and Chapman, *In prep*), and the broader ecosystem because these NHP species seem to serve as ecosystem engineers (Chapman *et al.* 2013).

### 1.2.3 Chapter 4

In my fourth chapter, I explored how observed changes in behaviour and sociality might impact the holobiont. I examined how social group membership, social behaviour, and hunting behaviour influence the gut bacterial microbiome community of wild NHPs in Taï National Park, Côte d'Ivoire. Predation rates are known to vary with group sizes, in addition, relatedness, social behaviour, and hunting may also be influenced by group size, and were thus of particular interest (Mitani & Watts 1999; Miyamoto *et al.* 2013; Gogarten *et al.* 2014b; Gogarten *et al.* 2015). I collected and analyzed fecal samples from nine wild NHP species regularly forming mixed-species associations or interacting in Taï National Park. I generated amplicons covering the 16S V4 hypervariable region and sequenced them using an Illumina MiSeq to examine the bacterial

taxa present in this ecosystem. To understand factors influencing within-species variation, I concentrated sampling on sooty mangabeys (*Cercocebus atys*, *N*=276) and chimpanzees (*Pan troglodytes verus*, *N*=98), for which much more metadata were available. This analysis provided important insights into factors influencing microbiome composition in a wild NHP community; despite sharing an environment, regular interactions, and a hunter-prey relationship between chimpanzees and red colobus, we find that individual NHPs harbor unique and persistent microbiomes, that are influenced by the host's species, social group, and relatedness. Surprisingly, we found no evidence that the social network of sooty mangabeys influenced their gut microbiome similarity. We demonstrated a branching order of host-species phylogenies built using the composition of their microbial communities as characters, which suggests incongruence with known NHP phylogenetic relationships; rather chimpanzees were the sister taxa to their colobine prey. In contrast to strong signals suggesting phylogenetic clustering in the microbiomes of all monkeys, chimpanzee microbiomes show evidence of deep phylogenetic evenness or overdispersion. This suggests unique ecological processes driving community assembly, possibly due to broader bacterial exposure through hunting.

### 1.2.2 Appendix Chapter

For a fifth chapter, I had initially sought to examine how environment and behavioural changes might be impacting the transmission of microorganisms known to have a major impact on human health, retroviruses. I accessed 12 years of research on NHP retroviruses in the Taï National Park area, Côte d'Ivoire and collected my own data on the prevalence and evolution of these retroviruses in sooty mangabeys over two field seasons (Gogarten *et al.* 2014a). I compiled data on the prevalence, within- and cross-species transmission of primate retroviruses (including towards local human populations), and discuss the evidence of important virus-host interactions that determine cross-species transmission risk.

The simian immunodeficiency virus of sooty mangabeys (SIVsmm) is thought to be the origin of the human immunodeficiency virus type 2 (HIV-2), but little is known about the epidemiology and standing genetic diversity of SIVsmm in wild populations. With the exception of the zoonotic infection of humans with SIVsmm, the SIVs in this NHP community exhibit strict host specificity. Chimpanzees, despite high exposure, are SIV free in this ecosystem. For other

viruses this is not always the case: for example, the Simian T-lymphotropic virus type 1 (STLV-1) strains circulating in these NHPs are not strictly species-specific, with one strain of STLV-1 infecting sooty mangabeys, and another red colobus and chimpanzees, forming two relatively homogeneous clades. Similarly simian foamy viruses also appear to be able to cross the species boundary, with chimpanzee being infected with SFV from red colobus. This suggests that there is ample opportunity for transmission, but not all retroviruses are able to cross the species boundary. Despite huge exposure through hunting, chimpanzees appear to be resistant to SIVwrc infecting red colobus; I discuss host resistance mechanisms and viral integration factors that might be responsible for this observation, which may explain why humans also appear resistant to the SIVwrc infecting red colobus. The lack of chimpanzee SIVsmm infections, despite this virus' clear ability to infect humans (SIVsmm infecting humans is referred to as HIV-2), suggests they may have a yet undescribed restriction factor that confers resistance to SIVsmm or it could reflect the rarity with which chimpanzees hunt sooty mangabeys in this ecosystem. I discuss how major changes in monkey populations in this park might impact within and between-species transmission.

For a fifth chapter I initially sought to extend the findings of Santiago et al. (2005), who predicted that SIVsmm super infections must be common among sooty mangabeys, given observed recombination rates in their study, but were unable to determine the prevalence of these superinfections or identify the individuals that harboured these infections. Further they were unable to understand whether horizontal or vertical transmission is the norm for this virus in this host. Thus many questions remained regarding within group transmission of SIVsmm and how this is influenced by behaviour. I had intended to examine these questions using new methodological advances and infrastructure available in Taï National Park, Côte d'Ivoire. These could have allowed for the construction of retroviral relatedness networks and the detection of superinfections from non-invasively collected fecal samples, as was demonstrated for chimpanzees (Goffe et al. 2012). I collected several thousand fecal samples and set up and contributed to the habituation of a group of sooty mangabeys to human presence (i.e., the term habituation is used here to indicate repeated applications of a stimulus, in this case human presence, that results in a decreased response to the stimulus, in this case a fear response or reaction to human presence) and collection of behavioural data from this habituated group of sooty mangabeys. Unfortunately, as seems to be the case for many non-invasive fecal detection

assays, particularly molecular and serological approaches (pers. comm. Ahidjo Ayouba, February 9th, 2016), detection of our target pathogen, SIVsmm, was not as sensitive in our study group as had been described in previous studies. There are myriad reasons why we may have failed to detect SIVsmm, including lack of co-infections or dietary stress that decreased shed viruses in feces, decreased virulence of a specific strain reducing copy numbers, inhibition by compounds in the diet, or specific host genetic factors that decreased shedding in feces. As a result of these failed attempts, I shifted my efforts to the sooty mangabey simian foamy virus, SFVsmm, where the detection methods were much more efficient. Unfortunately a selective sweep in the study group meant that there was not enough diversity to examine transmission processes as a function of primate behaviour or construct viral relatedness networks from these data as had been done for chimpanzees in this study system (Goffe et al. 2012). Due to the shortcomings of these efforts, this chapter, included as an appendix, focuses on this published literature review, which incorporates some of my own data from sooty mangabeys, to address how environment and behavioural changes might be impacting the transmission of retroviruses in this NHP community.

### 1.2.4 Chapter 5

In my fifth chapter, I sought to understand how members of the holobiont community, specifically bacterial taxa, might influence the health and fitness of a NHP host. I developed a method to generate whole genome sequences from metagenomic samples by coupling of hybridisation capture of the *Treponema pallidum* genome and next generation sequencing. I show the feasibility of this tool and the widespread nature of infections by screening DNA extracts from bones from the Leendertz laboratory's collection with three independent PCR systems and document *Treponema pallidum* in chimpanzees (*Pan troglodytes verus*) and red colobus (*Procolobus badius*) from Tai National Park, Côte d'Ivoire. Bones were targeted because a previous study was able to sequence DNA from a 200 year old skeleton (Kolman *et al.* 1999) and *Treponema pallidum* causes bone lesions and may sequester in the skeleton, though much debate remains about the feasibility of using bones to detect syphilis infections from skeletal remains (von Hunnius *et al.* 2007). That this pathogen is shared between a predator and

prey in this ecosystem, suggests that changing behaviour and species interactions might ultimately shape pathogen dynamics.

### 1.2.4 Chapter 6

Using the molecular hybridization approach presented in my fifth chapter, I documented *Treponema pallidum* infections associated with severe skin lesions, in some cases including dramatic soft tissue destruction and cartilage damage of the nose, in wild sooty mangabeys in Taï National Park, Côte d'Ivoire, and green monkeys (*Chlorocebus sabaeus*) from Bijilo Forest Park, Gambia and Niokolo-Koba National Park, Senegal. To examine the evolutionary relatedness of these treponemes with those previously documented in baboons (*Papio cynocephalus anubis*) from Lake Manyara National Park, Tanzania and those infecting humans and causing yaws (*Treponema pallidum pertenue*), endemic syphilis (*T. p. endemicum*), and venereal syphilis (*T. p. pallidum*), I conducted a hybridization capture experiment to enrich *Treponema pallidum* DNA in prepared DNA libraries. This approach allowed us to sequence the full genomes from the tissue samples of the *Treponema pallidum* strains infecting *Cercocebus atys*, *Chlorocebus sabaeus* and *Papio cynocephalus anubis.*

Phylogenomic analyses revealed that all *Treponema pallidum* strains infecting NHPs are most closely related to the sub-species *T. p. pertenue*. Strains infecting humans and NHPs do not appear to be reciprocally monophyletic. This discordance in the phylogeny of the *Treponema pallidum* strains infecting these primates and the primate phylogeny, suggest cross-species transmission between primates, a recent zoonotic transmission event, and the recent spread of *T.p. pertenue* in human populations. This greatly broadens the known host range of *T.p. pertenue* and suggests this pathogen can cross over between distantly related species. The large number of infected individuals found in two groups of sooty mangabeys in this park and the associated severe etiology suggest this bacterial taxon can have a major impact on individual and group fitness. Further studies are needed to understand the transmission mode for this pathogen, though the large number of infected young animals suggests play or grooming might be important routes. Results from this sixth chapter highlight the need for monitoring primates in countries where yaws control efforts are underway to ensure continued spillover does not hamper

eradication efforts and broadly highlights the importance of understanding the distribution, ecology, and evolution of NHP symbionts and parasites for both human and primate health.

### 1.2.4 Chapter 7

In this thesis, I sought to examine the impact of environmental change on the behaviour, health, and fitness of NHP communities. In chapter two, I found that some primates are changing their group sizes in response to environmental changes. In chapter three I examined how these changes in social group size in turn impact competition, activity budgets, and ultimately fitness of these animals. In chapter four, I then examined how these same changes in social group size and behaviour might be impacting the holobiont, specifically the bacterial gut microbiome. In an appendix chapter, I sought to understand how these environmental changes might be impacting the retroviruses circulating in a NHP community. In chapter five, I wanted to understand how members of the holobiont community might influence NHP health, so I develop a novel approach for understanding pathogen evolution in wild NHP populations. In chapter six I use this tool to help understand the evolution of a NHP parasite that appears to be geographically widespread and may serve as a reservoir for humans. I identified a bacterial species associated with an extremely severe etiology in the Taï National Park NHP community and showed that it has recently radiated across the Primate order possibly corresponding to the large scale environmental changes associated with the human species' global footprint. As this pathogen is either directly transmitted through social contact or by fly vectors that are at higher densities in social groups, this pathogen may be impacted by changes in sociality as a response to changing environments. I conclude this thesis with final thoughts and synthesis of themes that arise from this work and discuss key avenues for future research.

# CHAPTER 2

## GROUP SIZE DYNAMICS OVER 15+ YEARS IN AN AFRICAN FOREST PRIMATE COMMUNITY

Jan F. Gogarten[1,2*], Aerin L. Jacob[1,3*], Ria R. Ghai[1], Jessica M. Rothman[4], Dennis Twinomugisha[5], Michael D. Wasserman[6], and Colin A. Chapman[5,7]

A version of this chapter is published in *Biotropica*, Volume 47, Issue 1, pp. 101-112

*Authors contributed equally to this work.

[1]Dept. of Biology, McGill University, 1205 Docteur Penfield, Montreal, QC, H3A 1B1, Canada;

[2]Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher
[1]Dept. of Biology, McGill University, 1205 Docteur Penfield, Montreal, QC, H3A 1B1, Canada;
[2]Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103, Germany and Research group Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institut, Nordufer 20, 13353, Berlin, Germany.

[3]Department of Geography, University of Victoria, PO Box 3060 STN CSC, Victoria, BC, Canada, V8W 3R4

[4]Department of Anthropology, Hunter College of the City University of New York, and New York Consortium in Evolutionary Primatology, 695 Park Avenue, New York, NY 10065, USA;

[5]Makerere University Biological Field Station, P.O. Box 967, Fort Portal, Uganda;

[6]School of Environmental Science & Policy, St. Edward's University, 3001 South Congress Ave., Austin, Texas 78704-6489, USA.

[7]McGill School of Environment and Dept. of Anthropology, McGill University, Montreal, Quebec, H3A 2T7, Canada and the Wildlife Conservation Society, 2300 Southern Blvd, Bronx, New York, 10640, USA.

## 2.1 ABSTRACT

Group size affects many aspects of the ecology and social organization of animals. We investigated longitudinal group size stability for five primate species in Kibale National Park, Uganda from 1996-2011 at three nested spatial scales. Survey data indicated that group sizes did not change for most species, with the exception of red colobus monkeys (*Procolobus rufomitratus*), in which group size increased at all spatial scales. Mangabey (*Lophocebus albigena*) group size increased in old-growth forest, but the sample size and increase were small. To augment this survey data, we collected several years of demographic data on three habituated groups of redtail monkeys (*Cercopithecus ascanius*), eight groups of black-and-white colobus (*Colobus guereza*), and one red colobus group. The red colobus group increased from 59 to 104 individuals, while redtail monkey and black-and-white colobus group sizes were stable, mirroring our survey results. To understand mechanisms behind group size changes in red colobus versus stability in other primates, we monitored forest dynamics at two spatial scales between 1990 and 2013, considered changes in predator population, and explored evidence of disease dynamics. The cumulative size of all trees and red colobus food trees increased over 24 years, suggesting that changing food availability was driving group size changes for red colobus, while predation and disease played lesser roles. Overall, our results and evidence of changing primate densities suggest that the Kibale primate community is in a non-equilibrium state. We suggest future conservation and management efforts take this into consideration.

## 2.2 INTRODUCTION

The size of animal groups can alter stress levels (Pride 2005), susceptibility to diseases (Freeland 1976; Snaith *et al.* 2008), reproductive and developmental rates (Borries *et al.* 2008), individual and group behaviour (Koenig 2002; Nunn *et al.* 2009), and group survival (Heg *et al.* 2005). To effectively conserve threatened species requires understanding how threats may affect group size through time. For example, the long-term viability of populations with the same number of individuals may differ based on how these individuals are distributed in groups, with smaller groups conferring different benefits and consequences than larger groups (*e.g.*, rates of reproduction and development vary with group size; Borries *et al.* 2008). Typically, however, the average group size of non-migratory species in a region is considered relatively stable (Wrangham *et al.* 1993; Janson & Goldsmith 1995), thus it has not generally been considered in conservation planning.

Grouping confers predictable benefits (Alexander 1974; van Schaik 1983), thus differences in size can be attributed to variation in the costs of grouping (Wrangham *et al.* 1993). One such cost is foraging efficiency, which decreases with increasing group size (Janson & Goldsmith 1995). These ideas have been formalized in the Ecological Constraints Model (Wrangham *et al.* 1993; Chapman & Chapman 2000) that predicts average group size should be stable in regions with stable environments. Yet there is accumulating evidence that forest environments are not stable (Turner *et al.* 1993; Turkington 2009; Chapman *et al.* 2010a; Mori 2011). While non-equilibrium dynamics are a central concept in modern ecological theory (Mori 2011), their implications for group size are infrequently considered with the exception of sudden catastrophic change (e.g., disease: Gulland 1992; hurricanes: Pavelka *et al.* 2003).

Kibale National Park, in Western Uganda, represents one of the few well-studied tropical forest ecosystems for which long-term data on plant and animal communities are available. Many areas of Kibale have been well protected since the 1930s and its forest and wildlife have been intensively studied since the 1970s. Integrating this existing long-term data suggest either ecosystem stability or instability depending on the components examined and time scale considered (Mitani *et al.* 2000; Chapman *et al.* 2005; Chapman *et al.* 2010a; Struhsaker 2010; Lwanga *et al.* 2011). (Chapman *et al.* 2010a) analyzed changes in tree recruitment and growth over 18 years and concluded that the old-growth forest is in a non-equilibrium state and likely

recovering from a large disturbance within the last several hundred years. Similar changes in tree species composition have been recorded in other old-growth tropical forests in other areas of Africa, including Cameroon (Hawthorne 1996), Uganda (Sheil *et al.* 2000), Gabon (Tutin & Oslisly 1995), and the Republic of Congo (Brncic *et al.* 2007), pointing to the influences of large- and small-scale human disturbance, fluctuating elephant populations, and climate change (Bongers *et al.* 2009). This dynamism is evident in animal populations as well. For example, the size of blue monkey groups (*Cercopithecus mitis*) from central Kibale (Ngogo) suggested that group sizes were not in equilibrium (Angedakin & Lwanga 2011), yet another Ngogo study suggested long-term stability in group sizes of redtail monkeys (*Cercopithecus ascanius*), black-and-white colobus (*Colobus guereza*), and mangabeys (*Lophocebus albigena*; Teelen 2007) .

The objective of our research was to examine group size dynamics in five species of diurnal primates over 15 years (1996-2011) at different spatial scales. We consider: red colobus (*Procolobus rufomitratus*), black-and-white colobus, redtail monkeys, mangabeys, and blue monkeys. We also gathered detailed demographic data from one well-habituated group of red colobus, three groups of redtail monkeys, and eight groups of black-and-white-colobus observed for six, four, and four years, respectively. We explore potential explanations for changes in group size using long-term data on forest dynamics.

## 2.3 METHODS

### 2.3.1 Study site

Kibale National Park (795 $km^2$; $0^o$ 13' - $0^o$ 41' N and $30^o$ 19' - $30^o$ 32' E) is a moist-evergreen forest in western Uganda (Fig. 2.1). In 1932, Kibale was designated a Crown Forest Reserve; in 1993 it became a National Park. Anthropogenic disturbances created a mosaic of old-growth and regenerating forest habitats throughout the park (Struhsaker 1997; Chapman & Lambert 2000). In the late 1960s, much of northern Kibale was logged (Struhsaker 1997) including two study areas used in this research: the 405-ha forestry compartment K14 was logged at 14 $m^3$ /ha (approximately 5.1 stems/ha) and the 347-ha compartment K15 was logged at 21 $m^3$/ha (approximately 7.4 stems/ha). Although extraction rates for Sebatoli, a northern region of the park, are not available, stand structure indicates it was logged at similar levels to K15 (Chapman

unpubl. data). Compartment K30, immediately south of K14, is a 282-ha area that was not commercially harvested; although a few large stems (0.03–0.04 trees/ha) were cut by pitsawyers, this seems to have had little impact on the forest (Struhsaker 1997). Other areas included in this study are believed to have been similarly impacted in a minor way by pitsawyers, but have been less extensively studied.

### 2.3.2 Study subjects

Our study was conducted on five co-occurring primate species, including two colobines (red colobus and black and white colobus). Colobines are considered to be predominantly folivorous, and overlap considerably in diet, with young leaves making up the majority of food eaten. However, red colobus groups are often substantially larger in size than black and white colobus groups (Chapman & Pavelka 2005) and exhibit substantial differences in tree species and parts consumed (Oates 1977; Chapman & Chapman 2002; Harris & Chapman 2007; Struhsaker 2010). We also examined three species of cercopithecine monkeys (blue monkeys, redtail monkeys, and mangabeys) that are predominantly frugivorous, although insects and young leaves also compose parts of their diet (Struhsaker 1978). At the broadest scale, we included data from two additional primate species: L'hoest monkeys *(Cercopithecus lhoesti)* and olive baboons *(Papio anubis)*. Both species relying primarily on fruit, but are more terrestrial than other five monkey species.

### 2.3.3 Quantifying group size

We counted primate group sizes in two periods (July 1996 - May 1998 and July 2010 – May 2011; $N$ = 268 group counts across all scales) at three nested spatial scales: (1) unhabituated groups throughout the park (*broad scale*), (2) unhabituated and habituated groups in adjacent logged (K14 and K15) and old-growth (K30) forest (*intermediate scale*), and (3) unhabituated and habituated groups occurring only in old-growth forest (K30) (*fine scale*) (Fig. 2.1; Table 2.1-2.3). The broad scale spanned the entire park, but centered around four locations each approximately 12-15 km apart along a north-south gradient (Fig. 2.1). The Kanyawara study area provided access to K14, K15, and K30 (Chapman & Chapman 1997; Struhsaker 1997). At the fine and intermediate scales, we used long-term (~24 years) data on tree species composition and structure to explore relationships between food abundance, nutritional quality, and group size.

To obtain accurate primate group count, three observers selected a study area for eight days per month. When a primate group was found in the designated area, we recorded the location and

attempted to count all individuals. The time spent with each group was variable, but we monitored a group as long as was necessary to ensure that we were confident that the group count was accurate; the maximum time spent with a single group was ten hours. To ensure count accuracy, observers waited until the group made a single-file movement across a canopy opening, such as a treefall gap or road, where it is possible to easily count individuals. Repeat counts were made of the same group. We found this method to be effective for all species, regardless of level of habituation and species-specific behaviours (*e.g.*, canopy height selection). Differences between species, such as their density, home range size, and habituation, influenced the ease with which we could accurately count groups. To ensure this did not affect our data, we took a conservative approach and only included counts for which we were totally confident in accuracy and precision. Due to logistic constraints, fewer groups were counted in the 2010-2011 census than in the 1996-1998 census period; we accounted for these differences in sample sizes statistically (see: **Analysis of group size data**).

As a means of verifying changes in group counts at the three nested spatial scales, we examined changes in group size using detailed demographic data from habituated study groups of red colobus, black-and-white colobus, and redtail monkeys that ranged in logged and old-growth forest around Kanyawara. These groups have each been studied for at least four years, group size and composition were known, and all adults are individually recognizable. We repeatedly counted the number of individuals in one group of red colobus between July 2006 and September 2011 ($N$ = 28 counts), in eight groups of black-and-white colobus between February 2008 and January 2012 ($N$ = 83 counts), and in three groups of redtail monkeys between August 2008 and January 2012 ($N$ = 6 counts).

### 2.3.4 Analysis of group size data

We tested group size for normality using the Shapiro-Wilks test and normalized it with a square root transformation (Sokal & Rohlf 1995). We tested for changes in group sizes on the broad scale for each primate species between the two time periods (1996-98 and 2010-11) using Welch's two sample t-tests. We present back-transformed means and 95% confidence limits following Sokal and Rohlf (1995). Statistical comparisons between sample periods were not possible for three primate species for the following reasons: (1) we did not count any L'Hoest monkey groups in 2010-11, (2) we did not count any baboon or L'Hoest monkey groups in

compartments K14, K15, or K30 in either time period, and (3) we counted only one group of mangabeys in K30 in each period. We counted few blue monkey groups in each period as these animals are widely dispersed, typically at very low density, and secretive (Butynski 1990), so these results are interpreted with caution.

To assess the impact of sample size differences between sampling periods, we used delete-d jackknifing without replacement to down-sample the 1996-1998 data to the sample size of the 2010-2011 survey ($N$=10,000 replicates). We present the mean of this down-sampled data along with the percentage of these replicates for which the mean is greater than the mean from the 2010-2011 survey (Table 2.1). For each replicate we conducted a Welch's two sample t-test and present the percentage of these t-tests that were significant at the $p<0.05$ level. These results are statistically conservative (*i.e.*, high probability of not finding a statistical effect when there is one) as they repeatedly discard a large proportion of the 1996-1998 data, but are presented to allow the reader to access the importance of sample size differences between study periods.

For the habituated group of red colobus, we used a linear regression to determine whether group size increased through time. To test for changes in group sizes for the habituated groups of black-and-white colobus and redtail monkeys, we divided the study (2008-2012) in two equal periods, calculated the mean group size for each period, and compared them using a paired t-test.

### 2.3.5 Quantifying forest change

To identify relationships between primate food abundance, nutrition, and group size at the fine and intermediate spatial scale, we analyzed data from permanent tree plots in the Kanyawara area (200 x 10 m; total area = 5.2 ha). These plots were established and surveyed in December 1989 and located at random places along the existing trail system. They were re-surveyed in May 2000, September-November 2006, and January-May 2013. In each plot, trees diameter-at-breast-height (DBH) ≥ 10 cm were identified to species-level, individually marked with a uniquely numbered aluminum tag, and measured for DBH. Voucher specimens for all trees were given to Makerere University Biological Field. During each re-survey, we relocated and measured all tagged trees, recorded tree deaths, and included new trees recruiting into the ≥ 10 cm DBH size class. We measured the tree's DBH 1.2 m above the ground using parameters established previously in the study area (Chapman *et al.* 2010a).

### 2.3.6 Analysis of forest data

The DBH of a tree varies reliably with both fruit and leaf biomass, is practical and easy to measure, and has low inter-observer error (Catchpole & Wheeler 1992; Chapman *et al.* 1994; FAO 1997; Enquist & Niklas 2001, 2002). We calculated the $\log_{10}$(DBH) of all trees in each plot and summed it (*i.e.*, the cumulative $\log_{10}$(DBH)) to assess whether forest structure changed over time. We summed $\log_{10}$(DBH) because of the allometric relationship between DBH and plant productivity; we used cumulative $\log_{10}$(DBH) as an index of food availability (Snaith & Chapman 2008).

Primate populations are likely more influenced by changes in the abundance of food trees than the abundance of all trees in an area. We followed (Chapman *et al.* 2010a) and used dietary data to determine food trees for each primate species and conducted a separate analysis on cumulative $\log_{10}$(DBH) of major food tree species for each primate species, in each plot, in each time period. We defined major food tree species as those that accounted for ≥ 4 percent of feeding time, as reported by Rudran (1978) and Butynski (1990) for blue monkeys*,* Waser (1975) and Olupot (1994) for mangabeys*,* Harris and Chapman (2007) and Oates (1977) for black-and-white colobus, Rode *et al.* (2006: unpublished data) and Stickler (2004: unpublished data) for redtail monkeys, and Chapman and Chapman (2002: unpublished data) and Struhsaker (1975, 2010) for red colobus.

The preceding analyses test for changes in *quantity* of food available to primates; however, analyses of the ecological determinants of red colobus abundance clearly indicate that the *quality* of available foods is also important (Chapman & Chapman 2002; Wasserman & Chapman 2003; Chapman *et al.* 2004). As a measure of food quality for red colobus, we used the protein-to-fiber ratio, which is a good predictor of folivore leaf choice (Milton 1979) that has been shown to predict colobine biomass at local and regional scales (Milton 1979; Waterman *et al.* 1988; Oates *et al.* 1990; Chapman & Chapman 2002; Ganzhorn 2002; Chapman *et al.* 2004; Gogarten *et al. 2012b*). The relationship between the protein-to-fiber ratio and colobus biomass has been demonstrated with the overall protein-to-fiber ratio of mature leaves in an area. Since young leaves constitute a larger portion of the red colobus diet than mature leaves (Struhsaker 1975; Ryan *et al.* 2013), we ran the analysis to measure the effect of the protein-to-fiber ratio of mature versus young leaves. For further discussion of the application of the protein-to -fiber ratio see (Oates *et al.* 1990; Chapman *et al.* 2004), and for details of sample collection, processing, and

the determination of protein and fiber see Chapman and Chapman (2002), Rothman *et al.* (2012) and (Gogarten *et al. 2012b*).

To test for temporal variation in food abundance parameters, we compared repeat samples of the 11 permanent tree plots (K30 - fine scale) and 26 plots (K30, K14, K15 - intermediate scale) between the four surveys using a linear mixed effect model, with sampling periods included as fixed effects and vegetation plot included as a random effect. These models were implemented in the R package 'nlme' (Pinheiro *et al.* 2012; R Development Core Team 2012). Additionally, for each of the five primate species with detailed dietary data, we calculated the percent change in cumulative $\log_{10}$(DBH) of food species in each plot during each of the four surveys. To incorporate the protein-to-fiber ratio of mature and young leaves of species eaten by red colobus into the measure of food availability, we re-ran the analysis with cumulative $\log_{10}$(DBH) weighted by the protein-to-fiber ratio of each major food tree species.

### 2.4 RESULTS

At the broad scale we found a significant increase in red colobus group size between 1996-98 and 2010-11; we did not detect a significant increase for any other species (Table 2.1). When we examined groups at the intermediate scale (K30, K14, K15), we found that average group size for red colobus increased from 35.3 to 47.5 individuals (Table 2.2) and for mangabeys from 12.0 to 17.3. However, mangabey results should be interpreted with caution due to the small number of groups sampled in 2010-2011 ($N$ = 3; Table 2.2). We found similar trends for these two species at the fine scale in the old-growth forest (K30), but the smaller sample size resulted in marginal significance for the red colobus and did not allow us to statistically test the change in mangabey group size (2010-2011 $N$ = 1; Table 2.3). Chapman *et al.* (2010b) found that density of red colobus groups in K30 decreased between 1996 and 2006 (1996 = 5.5 groups/km$^2$; 2006 = 4.2 groups/km$^2$); however, since we document an increase in average group size, these results suggest that individual density remained relatively constant (1996 = 204 individuals/km$^2$; 2006 = 219 individuals/km$^2$).

For the long-term study groups with detailed demographic data, we did not detect a significant change in group size of black-and-white colobus (2008-9 mean= 7.1, 2010-12 mean = 8.7, $t$=1.766, $df$ =7, $p$=0.121; Fig. 2.2a) or redtail monkeys (2008-9 mean = 24.4, 2010-12 mean =

28.5, $t$ = 1.452, $df$ = 2, $p$=0.284; Fig. 2.2b). In contrast, the long-term red colobus study group increased from 59 to 104 individuals from 2006-11 ($R^2$=0.863, $F_{[1,26]}$=171.7, $p$<0.001), with an estimated increase of 7.6 individuals/year (SE = 0.580, $t$=13.104, $p$<0.001; Fig. 2.2c); this corroborates our survey data.

A linear mixed effects model detected a significant increase in cumulative $\log_{10}$(DBH) of all trees during the 2013 tree survey at the fine scale (Table 2.4) and intermediate scale as well as an increase in 2006 at the intermediate scale (Table 2.5). No significant changes in the availability of mangabey or blue monkey foods were detected across tree surveys at either scale. Compared to previous surveys, there was less black-and-white colobus food available in 2006 and 2013 at the fine scale in the old-growth forest, but no change in food availability at the intermediate scale, which included two areas that had been logged (K14 and K15) (Tables 2.4, 2.5). We documented more redtail monkey and red colobus foods available in 2013 than in previous times at both the small and intermediate scales. We documented an increase in red colobus and red tail food available for both species in 2006 at the intermediate scale. When red colobus food availability was weighted by the protein-to-fiber ratio of mature leaves, there was an increase in availability of quality foods in 2013 at both scales. In contrast, food availability weighted by the protein-to-fiber ratio of young leaves remained similar across the four surveys in the old-growth forest (Table 2.4), but increased at the intermediate scale that included regenerating areas (Table 2.5).

## 2.5 DISCUSSION

Variation in primate group sizes have been documented across species (Janson & Goldsmith 1995), space (Stanford 1995), and time (Angedakin & Lwanga 2011; Strier & Mendes 2012). However, to our knowledge our study represents the first systematic analysis of stability in primate group sizes on large temporal and spatial scales. For most primate species we examined, average group sizes remained stable across time at the park-wide scale. The only species for which we detected a change in group sizes at this broad spatial scale was red colobus. This increase was also observed at the intermediate scale in the logged and old-growth forest compartments and at the fine scale in the old-growth forest. Detailed demographic data from the long-term red colobus group support this trend, with average group size increasing by 7.6

individuals/year. In contrast, other detailed data from redtail monkey and black-and-white colobus groups suggest stability in group sizes (Mitani *et al.* 2000; Teelen 2007; Chapman *et al.* 2010b). Despite small sample size in 2010-2011, we also detected an increase in mangabey group sizes between the two sampling periods at the intermediate spatial scale.

Socioecological theory suggests that grouping strategies change when food resources change. Specifically, group size is expected to increase with increasing food availability (Milton 1984; Chapman & Chapman 2000). With respect to predation, from an evolutionary perspective an increase in predation pressure is expected to increase group size to increase group protection through vigilance or dilution effects; however, predators can decrease group size through overexploitation (Alexander 1974; van Schaik 1983; Delm 1990; Teelen 2008). Average group sizes might be expected to change if populations are recovering from a large disturbance, such as disease or natural disaster (Gulland 1992; Pavelka *et al.* 2003). Isolating particular factors responsible for the observed changes in group sizes through time is difficult because of a paucity of long-term data on all potential factors.

### 2.5.2 Potential drivers of observed changes: changing food availability and quality

Both stability and dynamism were evident in food availability and food quality depending on the spatial scale and primate species being considered (Tables 2.4 and 2.5). Overall primate food availability in K30 appears to have increased between 1996 and 2013. Socio-ecological theory suggests that resource distributions can have major impacts on primate sociality (Clutton-Brock & Harvey 1977; Wrangham 1980). Whether folivores, like red colobus, defend resources and exhibit competition over resources remains a point of considerable debate (Fashing *et al.* 2007; Snaith & Chapman 2008; Isbell 2012), but observed changes in red colobus food availability and quality (high protein-to-fiber ratio) may have changed within- and between-group competition for resources. An increase in food abundance and quality might favour larger groups if resources are defensible, there is increased competition over resources, and large groups have a competitive advantage over smaller groups that outweighs increases in within-group competition that can occur with increasing group size. Fashing (2001) found evidence that male black and white colobus defended resources as part of a mate defense strategy, demonstrating the importance of resource distribution on grouping behaviour; the finding that black and white colobus food availability remained stable across all periods at the larger spatial scale may

explain the observed stability in group sizes. Given that food for both redtail monkeys and red colobus appears to have increased at both spatial scales, it is puzzling that redtail monkey group size did not increase, while red colobus group sizes did. Other factors that influence food quality such as minerals, toxins, and phytoestrogens (Wasserman & Chapman 2003; Rode *et al.* 2006; Rothman *et al.* 2012; Wasserman *et al.* 2012) might explain the stability in red tail group sizes. Overall, however, it appears that increases in the availability and quality of red colobus food resources is a likely mechanism driving the observed increases in red colobus group size across all three spatial scales.

The changes in food availability we documented highlight the dynamism of forest composition, even in a relatively well-protected old-growth forest, which may reflect forest succession (Eggeling 1947; Chapman *et al.* 2010a). Changes in forest composition or structure have been recorded in other forests including Budongo National Park, Uganda (Sheil *et al.* 2000), La Selva, Costa Riliebca (Lieberman & Lieberman 1987; Norden *et al.* 2009), and Lambir Hills National Park (Russon *et al.* 2005) and Sungei Menyala Forest Reserve (Manokaran & Kochummen 1987), Malaysia. Other long-term studies highlight the importance of considering unpredictable factors in forest succession, including tree species-specific reproduction events and dispersal limitation on Barro Colorado Island, Panama (Dent *et al.* 2013) and ancient, as well as recent, natural and anthropogenic disturbances (van Gemerden *et al.* 2003; Mori 2011). What roles equilibrium and non-equilibrium factors play in forest succession in Kibale is not yet clear; further study is needed to determine the rates of change and drivers of forest composition, and their interaction with animal populations, including the roles of land use history (Synnott 1971) and an expanding elephant population (Laws 1970; Omeja *et al.* 2014).

### 2.5.3 Potential drivers: predation and disease

Predation is hypothesized to be an important driver of ecological and evolutionary processes, particularly with regard to sociality (van Schaik 1989; Isbell 1994), since even low rates of predation can have major impacts on primates with slow life histories (Cheney & Wrangham 1987; Isbell 1994). Although data are scarce, it is possible that group size could vary as a function of changing predation pressure (van Schaik & van Hooff 1983; Isbell 1994). In Kibale, known primate predators include leopards (*Panthera pardus*), golden cats (*Profelis aurata),* crowned hawk-eagles (*Stephanoaetus coronatus*: (Struhsaker & Leakey 1990; Mitani *et al.*

2001)), and chimpanzees (*Pan troglodytes* (Mitani & Watts 2001; Teelen 2008)). Bushmeat hunting of primates by humans is rare or absent altogether in the region (Struhsaker 1975). While research on felids, crowned hawk eagles, and chimpanzees does show that primates are primary prey resources, the overall predation pressure in the Kanyawara region is very low with respect to all predators. It does not appear that predation pressure has changed significantly over our study period (Skorupa 1989; Struhsaker & Leakey 1990; Mitani & Watts 1999; Teelen 2008; Lwanga *et al.* 2011 Chapman unpubl. data; Nakazawa *et al.* 2013).

Similarly, disease can cause rapid reductions in population size and group sizes (Collias & Southwick 1952; Milton 1996). The red colobus in Kanyawara have been observed extensively since 1970 (Struhsaker 1975; Chapman *et al.* 2010b; Struhsaker 2010) and are known to harbour a number of parasites and viruses (Gillespie *et al.* 2005; Goldberg *et al.* 2008; Goldberg *et al.* 2009; Lauck *et al.* 2011; Bailey *et al.* 2014). These pathogens may impact fitness, but there has not been an observed disease outbreak in the last forty years that would directly implicate recovery from an epidemic in red colobus group size increases. These observations suggest that neither predation nor disease adequately explain the stability in group sizes of most primate species, or the increases in red colobus group sizes.

### 2.5.4 Consequences of changing group sizes for primate ecology and conservation

The observed increase in red colobus group sizes will affect various aspects of their ecology and conservation (Gogarten *et al.* 2014b). When Borries *et al.* (2008) studied how development and reproductive rates varied with group size in the folivorous Phayre's leaf monkey (*Trachypithecus phayrei*), they found that infants in large groups weaned later and females had longer inter-birth intervals than in smaller groups. This suggests that large groups of arboreal folivorous monkeys have slower reproduction and ultimately lower female fitness than smaller groups (if survival rates were similar). This in turn suggests that a general increase in group size, may result in a slower increase in population size for a folivorous primate, although there is conflicting evidence from this population of red colobus (Snaith & Chapman 2008; Gogarten *et al.* 2014b). The observed changes in food availability and group sizes are likely changing primate ranging patterns (Chapman & Chapman 2000), stress levels (Pride 2005), diets (Snaith & Chapman 2008; Gogarten *et al.* 2014b), activity budgets (Gogarten *et al.* 2014b), population genetic structure (Miyamoto *et al.* 2013), and disease dynamics within- and between-species (Freeland

1976; Kuehl *et al.* 2008; Snaith & Chapman 2008; Caillaud *et al.* 2013; Gogarten *et al.* 2014a). These changes, in turn may have major cascading impacts on the entire ecosystem as both folivorous and frugivorous primates have been argued to play major roles as ecosystem engineers (Chapman *et al.* 2013).

The documented changes in red colobus group size compared to the relative stability of group size in other primate species - despite apparent increases in food - suggest that Kibale primate populations and some forest habitats may be in a non-equilibrium state. If indeed primate populations in Kibale are not at equilibrium, conserving their populations and habitats requires integrating unpredictability and instability into management plans to maximize ecosystem resilience and withstand unforeseen change (Hamilton *et al.* 1986; Mori 2011). Currently, habitat management in Kibale largely focuses on returning "natural forest" to areas degraded by logging, fire, or human encroachment with the goal of increasing populations of forest-dependent species of conservation concern (Uganda Wildlife Authority 2003). However, it is unclear what "natural forest" means, as forest in Kibale has almost certainly been changing from anthropogenic forces for the last several thousand years (Hamilton *et al.* 1986), as have other African tropical rainforests (van Gemerden *et al.* 2003; Brncic *et al.* 2007). Managing to reduce habitat heterogeneity ignores the dynamic nature of disturbance in animal and plant population dynamics; more homogenous landscapes may be less resilient to large-scale disturbances. Instead, it may be better to manage ecosystems by incorporating small- and large-scale disturbances (Mori 2011), as well as using non-equilibrium theory in conservation planning.

## 2.6 CONCLUSION

Our data suggest that red colobus group sizes are increasing in Kibale across all measured scales. In measuring both food abundance and quality, we find that an increase in overall food resources may be driving this increase, with larger groups conferring benefits that are not being offset by increased competition over food. Group size has remained stable in all other primates studied, regardless of changing food resources for some species. Despite stability in group sizes, changes in group density has been recorded for some species (Chapman *et al.* 2010b). Our data suggests that the Kibale primate community is in a non-equilibrium state.

## 2.7 ACKNOWLEDGEMENTS

We thank the Office of the President, Uganda, the Uganda National Council for Science and Technology, and the Uganda Wildlife Authority for permission to conduct this research. Long-term funding came from the Wildlife Conservation Society, Natural Science and Engineering Research Council (NSERC), Fonds Québécois de la Recherche sur la Nature et les Technologies, National Institutes of Health (NIH) grant TW009237 as part of the NIH-NSF Ecology of Infectious Disease program, and the UK Economic and Social Research Council. JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336), the Canadian Institutes of Health Research's Strategic Training Initiative in Health Research's Systems Biology Training Program, an NSERC Vanier Canada Graduate Scholarship (CGS), and a long-term Research Grant from the German Academic Exchange Service (DAAD-91525837-57048249). ALJ was supported by an NSERC-CGS and the Canadian Federation of University Women. MDW was supported by a Tomlinson Postdoctoral Fellowship. We thank Fred Babweteera and Emilio Bruna, Lauren Chapman, Katie Milton, Tom Struhsaker and two anonymous reviewers for helpful insights, David Mills for sharing his observations of golden cats, Johanna Bleecker for making the map, and Tim Davenport for help with the KiSwahili translation. Comments from Jonathan Davies, Marty Lechowicz, Louis Lefebvre, David Marcogliese, and Charles Nunn greatly improved the manuscript. Richard Wrangham was instrumental in establishing the tree plots in 1989 and we thank him for his long-term collaboration. We are particularly grateful to CAC's field assistants, particularly Tusiime Lawrence who established the tree plots and has helped monitor them ever since.

**Table 2.1**: **Primate group sizes between two sampling periods across all of Kibale National Park, Uganda (broad scale).**

| | 1996-1998 Group Counts | | | 2010-2011 Group Counts | | | Delete-d jackknifed 1996-1998 Group Counts Welch | | Welch Two Sample T-test[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Species | N groups | Mean group-size[b] ($\bar{x}$) | 95% confidence limit[b] | N groups | Mean group-size[b] ($\bar{x}$) | 95% confidence limit[b] | Mean group-size* | *Percent of t-tests significant* | *t* | *df* | *p* |
| Baboon | 6 | 28.49 | 13.75-48.55 | 3 | 32.66 | 29.87-35.58 | 28.73 (34.0%) | 0.00 % | 0.59 | 5.08 | 0.58 |
| Blue monkey | 11 | 9.31 | 6.08-13.23 | 3 | 10.60 | 1.76-26.89 | 9.50 (32.3%) | 0.00 % | 0.39 | 3.53 | 0.72 |
| Black and white colobus | 61 | 8.22 | 7.41-9.08 | 27 | 7.84 | 7.18-8.53 | 8.23 (79.5%) | 1.54 % | 0.73 | 82.75 | 0.47 |
| L'Hoest monkey | 4 | 19.63 | 8.90-34.55 | - | - | - | - | - | - | - | - |
| Grey-cheeked mangabey | 17 | 13.75 | 10.86-16.97 | 8 | 16.52 | 12.47-21.14 | 13.77 (4.1%) | 13.22 % | 1.20 | 17.13 | 0.25 |
| Red colobus | 55 | 28.44 | 24.25-32.97 | 27 | 46.63 | 39.11-54.81 | 28.50 (0%) | 99.97 % | 4.32 | 53.71 | <0.001 |
| Red-tailed monkey | 34 | 19.29 | 16.14-22.73 | 14 | 19.18 | 12.75-26.92 | 19.38 (52.9%) | 0.00 % | 0.03 | 19.61 | 0.98 |

[a] Comparing group sizes between the sampling periods, 1996-1998 and 2010-2011. Group size was square root transformed to improve normality. [b] Values were back-transformed following Sokal and Rohlf (1995) for square root transformed data. *To account for different sample sizes in the two surveys, we used delete-d jackknifing to down-sample the 1996-1998 data to the number of samples in the 2010-2011 survey (10,000 replicates). We present the mean of this down-sampled data along the percentage of these replicates that are greater than the mean from the 2010-2011 survey.

**Table 2.2: Primate group sizes during two sampling periods in K30, K15, and K14 in Kibale National Park, Uganda (intermediate scale).**

| | 1996-1998 Group Counts | | | 2010-2011 Group Counts | | | Delete-d jackknifed 1996-1998 Group Counts Welch | | Welch Two Sample T-test[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Species | N groups | Mean group-size[b] ($\bar{x}$) | 95% confidence limit[b] | N groups | Mean group-size[b] ($\bar{x}$) | 95% confidence limit[b] | Mean group-size* | *Percent of t-tests significant* | *t* | *df* | *p* |
| Blue monkey | 9 | 9.74 | 5.69-14.88 | 3 | 10.60 | 1.76-26.89 | 9.98 (40.1%) | 0.00% | 0.24 | 4.27 | 0.82 |
| Black and white colobus | 45 | 8.89 | 7.90-9.93 | 13 | 7.89 | 7.26-8.55 | 8.90 (90.2%) | 10.51% | 1.74 | 54.62 | 0.09 |
| Grey-cheeked mangabey | 11 | 12.05 | 9.87-14.45 | 3 | 17.31 | 13.75-21.29 | 12.08 (0.0%) | 17.89% | 3.79 | 9.95 | 0.0036 |
| Red colobus | 33 | 35.26 | 30.51-40.34 | 16 | 47.47 | 38.56-57.31 | 35.31 (0.0%) | 61.28% | 2.52 | 27.49 | 0.018 |
| Redtail monkey | 20 | 20.54 | 17.14-24.24 | 7 | 13.99 | 6.18-24.96 | 20.61 (100.0%) | 1.06% | 1.45 | 7.66 | 0.19 |

[a] Comparing group sizes between the sampling periods, 1996-1998 and 2010-2011. Group size was square root transformed to improve normality. [b] Values were back-transformed following Sokal and Rohlf (1995) for square root transformed data. * To account for different sample sizes in the two surveys, we used delete-d jackknifing to down-sample the 1996-1998 data to the number of samples in the 2010-2011 survey (10,000 replicates). We present the mean of this down-sampled data along the percentage of these replicates that are greater than the mean from the 2010-2011 survey.

**Table 2.3**: **Primate group sizes during two sampling periods in K30 in Kibale National Park, Uganda (fine scale).**

| | 1996-1998 Group Counts | | | 2010-2011 Group Counts | | | Delete-d jackknifed 1996-1998 Group Counts Welch | | Welch Two Sample T-test[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Species | N groups | Mean group-size[b] ($\bar{x}$) | 95% confidence limit[b] | N groups | Mean group-size[b] ($\bar{x}$) | 95% confidence limit[b] | Mean group-size* | *Percent of t-tests significant* | *t* | *df* | *p* |
| Blue monkey | 3 | 12.16 | 1.77-31.88 | 2 | 10.40 | - | 12.30 (67.10%) | 67.10% | 0.28 | 1.85 | 0.80 |
| Black and white colobus | 17 | 7.82 | 5.92-9.99 | 3 | 7.66 | 6.28-9.18 | 7.95(48.3%) | 48.34% | 0.16 | 18 | 0.87 |
| Mangabey | 1 | 16 | - | 1 | 19 | - | - | - | - | - | - |
| Red colobus | 14 | 37.17 | 27.47-48.34 | 11 | 52.07 | 39.50-66.38 | 37.25 (0.0%) | 0.00% | 1.95 | 22.24 | 0.065 |
| Redtail monkey | 6 | 21.05 | 11.40-33.65 | 5 | 17.05 | 7.72-30.04 | 21.13 (100%) | 100.00% | 0.68 | 8.82 | 0.52 |

[a] Comparing group sizes between the sampling periods, 1996-1998 and 2010-2011. Group size was square root transformed to improve normality. [b] Values were back-transformed following Sokal and Rohlf (1995) for square root transformed data. * To account for different sample sizes in the two surveys, we used delete-d jackknifing to down-sample the 1996-1998 data to the number of samples in the 2010-2011 survey (10,000 replicates). We present the mean of this down-sampled data along the percentage of these replicates that are greater than the mean from the 2010-2011 survey. This would need to be less than 2.5% to be significant at the $p<0.05$ level.

**Table 2.4 – Results of the linear mixed effects models to test for changes in food availability (cumulative $log_{10}$(DBH)) in 11 plots between the four survey periods. Means, 95% confidence intervals ($CI_{95}$, 1.96 times SE) and t-values of fixed effects (sampling periods) are given. Significant values are in bold. Vegetation plots included as random effect.**

| Food abundance measure (cumulative $log_{10}$(DBH) | Intercept (mean +- $CI_{95}$) | *t* | 1999 (mean +- $CI_{95}$) | *t* | 2006 (mean +- $CI_{95}$) | *t* | 2013 (mean +- $CI_{95}$) | *t* |
|---|---|---|---|---|---|---|---|---|
| All trees | **125.1 +- 16.6** | **14.74*****| -4.3 +- 7.8 | -1.09 | 3.1 +- 7.8 | 0.79 | **9.3 +- 7.8** | **2.35*** |
| BWC food trees | **26.8 +- 5.9** | **8.92***** | -2.5 +- 2.7 | -1.84 | **-3.1 +- 2.7** | **-2.28*** | **-3.3 +- 2.7** | **-2.45*** |
| MG food species | **28.8 +- 7.4** | **7.66***** | -0.2 +- 3.0 | -0.11 | -0.3 +- 3.0 | -0.19 | -1.2 +- 3.0 | -0.82 |
| RT food species | **38.7 +- 10.9** | **6.99***** | 0.6 +- 5.6 | 0.21 | 3.1 +- 5.6 | 0.28 | **5.95 +- 5.6** | **2.06*** |
| BM food species | **61.8 +- 19.4** | **6.23***** | -3.3 +- 3.6 | -1.79 | -2.4 +- 3.6 | -1.32 | -2.8 +- 3.6 | 1.55 |
| RC food species | **49.7 +- 12.0** | **8.14***** | -0.5 +- 5.8 | -0.18 | 3.8 +- 5.8 | 1.29 | **9.5 +- 5.8** | **3.21**** |
| RC food species weighted by protein:fiber of ML | **31.9 +- 6.4** | **9.74***** | 0.4 +- 3.0 | 0.24 | 3.1 +- 3.0 | 1.99 | **6.7 +- 3.0** | **4.35***** |
| RC food species weighted by protein:fiber of YL | **48.6 +- 7.6** | **12.48***** | -2.0+- 4.6 | -0.87 | -0.5+- 4.6 | -0.20 | 1.4+- 4.6 | 0.58 |

BWC = black-and-white colobus, MG = mangabey, RT = redtail monkey, BM = blue monkey, RC = red colobus

YL = young leaves, ML = mature leaves. *** ($P < 0.001$), ** ($P < 0.01$) and * ($P < 0.05$).

**Table 2.5**: **Results of the linear mixed effects models to test for changes in food availability (cumulative $log_{10}$(DBH)) in 22 plots between the four survey periods in K30, K15 and K14. Means, 95% confidence intervals ($CI_{95}$, 1.96 times SE) and t-values of fixed effects (sampling periods) are given. Significant values are in bold. Vegetation plots included as random effect.**

| Food abundance measure (cumulative $log_{10}$(DBH) | Intercept (mean +- $CI_{95}$) | *t* | 1999 (mean +- $CI_{95}$) | *t* | 2006 (mean +- $CI_{95}$) | *t* | 2013 (mean +- $CI_{95}$) | *t* |
|---|---|---|---|---|---|---|---|---|
| All trees | **111.7 +- 13.7** | **16.6*****| -1.9 +- 5.1 | -0.72 | **7.3 +- 5.1** | **2.83**** | **12.4 +- 5.1** | **4.78***** |
| BWC food trees | **28.0 +- 4.8** | **11.4***** | -0.9 +- 1.9 | -0.93 | -0.9 +- 1.9 | -0.99 | -1.2 +- 1.9 | -1.28 |
| MG food species | **28.2 +- 4.6** | **12.10***** | -0.5 +- 1.8 | -0.51 | -0.1 +- 1.8 | -0.09 | -0.9 +- 1.8 | -0.91 |
| RT food species | **35.6 +- 6.9** | **10.07***** | 0.0 +- 3.0 | 0.02 | **3.2 +- 3.0** | **2.05*** | **4.7 +- 3.0** | **3.03**** |
| BM food species | **57.4 +- 11.3** | **9.99***** | -2.0 +- 2.8 | -1.40 | -0.8 +- 2.8 | -0.55 | -0.8 +- 2.8 | -0.53 |
| RC food species | **46.6 +- 9.7** | **9.44***** | 1.1 +- 3.3 | 0.65 | **5.4 +- 3.3** | **3.20**** | **8.6 +- 3.3** | **5.09***** |
| RC food species weighted by protein:fiber of ML | **30.8 +- 5.9** | **10.27***** | 1.6+- 1.9 | 1.70 | **4.5+- 1.9** | **4.68***** | **6.9+- 1.9** | **7.12***** |
| RC food species weighted by protein:fiber of YL | **46.4 +- 7.6** | **12.03***** | 0.0 +- 2.8 | 0.01 | 2.0 +- 2.8 | 1.39 | **3.2 +- 2.8** | **2.21*** |

BWC = black-and-white colobus, MG = grey-cheeked mangabey, RT = red-tailed monkey, BM = blue monkey, RC = red colobus

YL = young leaves, ML = mature leaves. *** ($P < 0.001$), ** ($P < 0.01$) and * ($P < 0.05$).

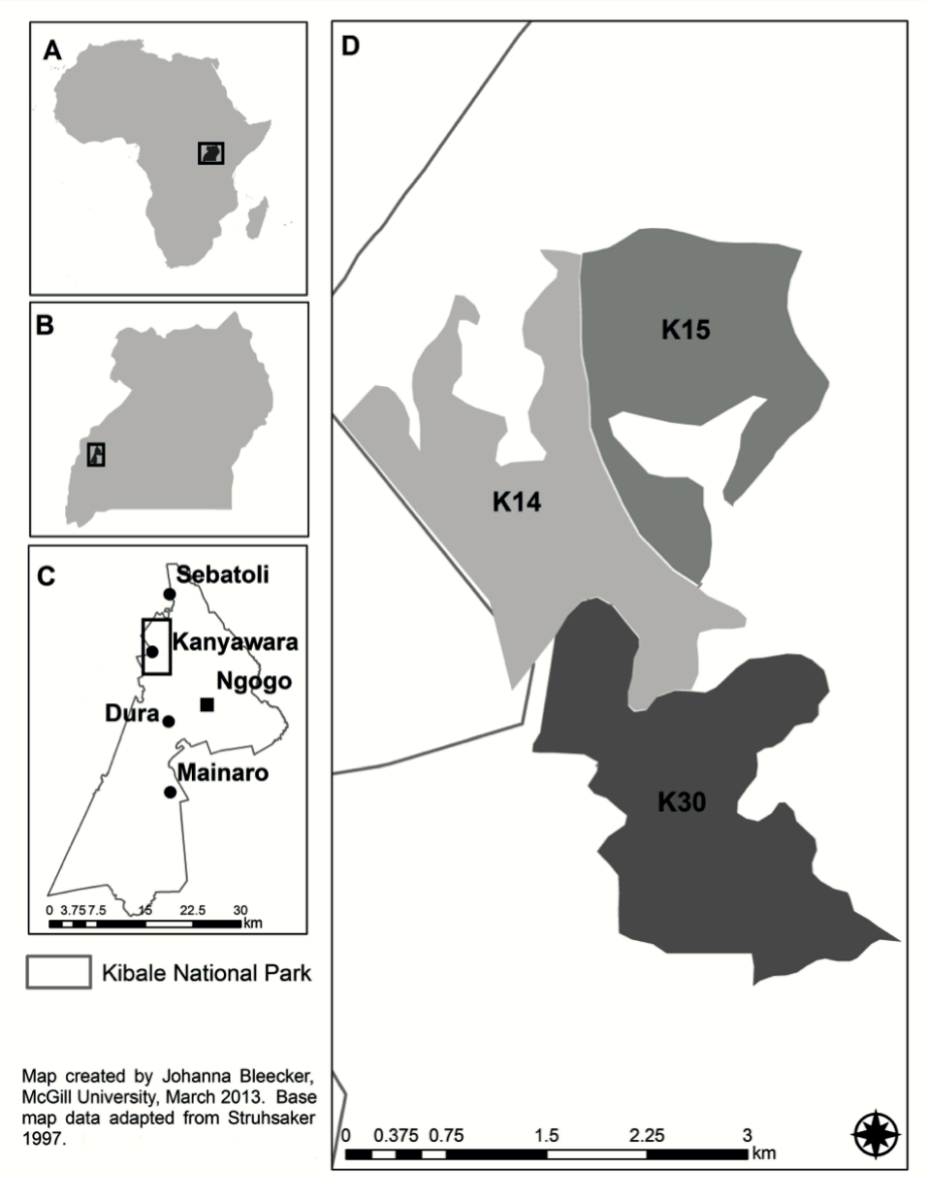



**Figure 2.1: Map indicating study locations.** The locations of a) Uganda, b) Kibale National Park, c) the study areas Sebatoli, Kanyawara, Dura, and Mainaro, and d) forestry compartments K30 (old-growth forest), K14 and K15 (logged forest).

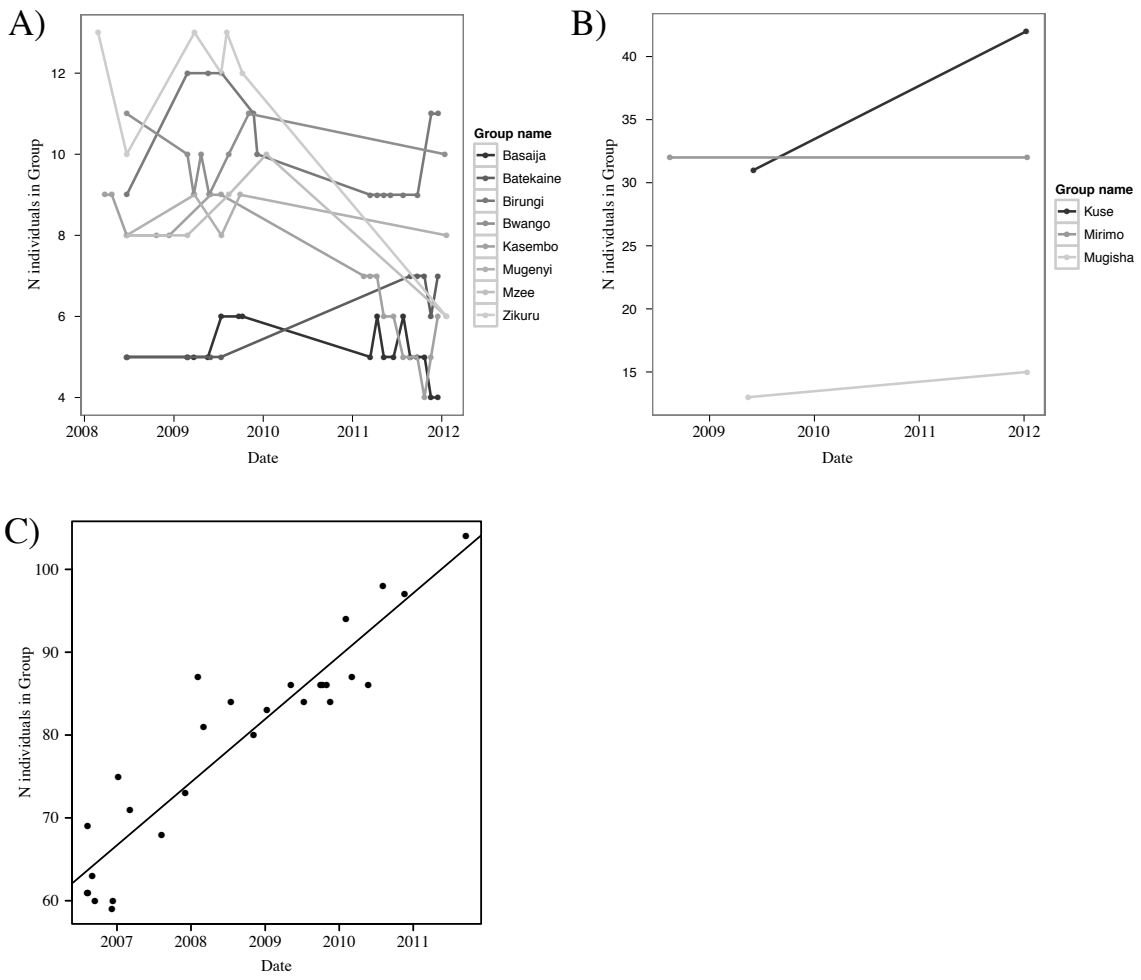



**Figure 2.2: Group sizes of primate group sizes of primate species in Kibale National Park.** A) Group size of eight habituated groups of black and white colobus through time. B) Group sizes of three habituated groups of redtail monkeys through time. C) Group size of one habituated group of red colobus through time; a solid line represents the linear regression of group size on time.

# LINKING STATEMENT 1

In Chapter 2 I documented major changes in red colobus group sizes on a large spatial and temporal scale. This was in stark contrast to the relative stability in average group sizes observed for all other primate species in the park for which we were able to collect data. While the exact causes of these changes were difficult to ascertain, we did find an increase in overall food resources that may have been related to the observed increase in group size in red colobus. These findings suggested that individuals living in these changing groups might be subjected to rapidly changing pressures and challenges. Group sizes have been shown to influence many aspects of an animal's life, including stress levels, disease transmission, reproductive rates, behaviour and ultimately fitness.

These observations on group size motivated us to explore how differences in group size might impact individual behaviour and fitness. To date, much of what is known about the effects of group size on the behavioural ecology of individuals living in these groups, has come from comparisons across multiple groups of different sizes. In chapter 3, I use an alternatively approach that draws on longitudinal data collected from a single group of wild red colobus over six years, during which group size increased dramatically. I examined how changing group sizes affect the behaviour of folivorous red colobus, specifically feeding, socializing, traveling, and dietary diversity. We predicted that larger group sizes are associated with increased time spent traveling (Koenig 2002; Snaith & Chapman 2008), decreased time spent feeding and socializing (van Schaik *et al.* 1983), higher dietary diversity (Snaith & Chapman 2007, 2008), and lower female fecundity (Beehner *et al.* 2006; Borries *et al.* 2008; Snaith & Chapman 2008). We expected that red colobus spend less time traveling, more time feeding and socializing, and have lower dietary diversity with increased food availability, so we control for variation in this factor in our analysis.

# CHAPTER 3

## INCREASING GROUP SIZE ALTERS BEHAVIOUR OF A FOLIVOROUS PRIMATE

Jan F. Gogarten[1,2], Tyler R. Bonnell[3], Leone M. Brown[4], Marco Campenni[2], Michael D. Wasserman[5], and Colin A. Chapman[6]

A version of this chapter is published in the *International Journal of Primatology*, Volume 35, Issue 2, pp. 590–608

[1]Dept. of Biology, McGill University, 1205 Docteur Penfield, Montreal, Quebec, H3A 1B1, Canada and Research Group Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Nordufer 20, 13353, Berlin, Germany.

[2]Dept. of Primatology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103, Germany.

[3]Dept. of Geography, McGill University, 805 Sherbrooke Street West, Montreal, Quebec, H3A 0B9, Canada.

[4]Dept. of Biology, Tufts University, 163 Packard Ave., Medford, Massachusetts, 02155, USA.

[5]Environmental Science and Policy, School of Behavioural and Social Sciences, St. Edward's University, 3001 South Congress Ave., Austin, Texas 78704, USA and Dept. of Anthropology, McGill University, 855 Sherbrooke Street West, Montreal, Quebec, H3A 2T7, Canada.

[6]McGill School of Environment and Dept. of Anthropology, McGill University, Montreal, Quebec, H3A 2T7, Canada and Wildlife Conservation Society, Bronx, New York, New York, 10460, USA.

## 3.1 ABSTRACT

Group size influences many aspects of mammalian social life, including stress levels, disease transmission, reproductive rates, and behaviour. However, much of what is known about the effects of group size on behavioural ecology has come from comparisons across multiple groups of different sizes. These findings may be biased because behavioural differences across groups may be more indicative of how environmental variation influences animal behaviour, rather than group size itself. To partially circumvent this limitation, we used longitudinal data to examine how changes in group size across time affect the behaviour of the folivorous red colobus monkey (*Procolobus rufomitratus*) of Kibale National Park, Uganda. Controlling for food availability, we demonstrated that increasing group size resulted in altered activity budgets, based on six years of data on a group that increased from 57 to 98 members. Specifically, as group size increased, individuals spent less time feeding and socializing, more time traveling, and increased the diversity of their diet. These changes appear to allow the animals to compensate for greater scramble competition apparent at larger group sizes, as increasing group size did not show the predicted relationship with lower female fecundity. Our results support recent findings documenting feeding competition in folivorous primates. Our results also document behavioural flexibility, an important trait that allows many social mammals to maximize the benefits of sociality (e.g., increased vigilance), while minimizing costs (e.g., increased feeding competition).

## 3.2 INTRODUCTION

Group size affects many aspects of social organization and behaviour (Pulliam & Caraco 1984; Janson & Goldsmith 1995). The size of social groups has been implicated in the stress levels of individuals (Takeda *et al.* 2003; Pride 2005; Snaith *et al.* 2008), disease susceptibility (Davies *et al.* 1991; Cote & Poulin 1995; Loehle 1995; Nunn & Heymann 2005; Rifkin *et al.* 2012), reproductive and developmental rates (Creel & MacDonald 1995; Mann *et al.* 2000; Borries *et al.* 2008), and individual and group behaviour (Brown & Brown 1996; Griffiths & Magurran 1997; Koenig 2002; Carpenter 2007; Nunn *et al.* 2009). Inter- and intra-species variation in group sizes has been documented across broad spatial and temporal scales (Holmes & Price 1986; Poulin 1995; Weldon *et al.* 2004), and the causes and impacts of this variation have been the focus of intense study (Elgar 1989; Brown & Brown 1996; Ezenwa 2004; Croft *et al.* 2008; Griffin & Nunn 2012).

Among the numerous explanations proposed to explain the large variation in social systems observed across vertebrates (Wrangham 1980; Elgar 1989; Isbell 1991; Sterck *et al.* 1997; Snaith & Chapman 2007), the influence of group size on feeding competition between and within groups has been proposed as a major player in the evolutionary ecology of sociality (Eubank *et al.* 2004). A key relationship used in the development of theoretical models of group size determinants was the dependence of the rate of energy intake on the number of individuals in a group (Pulliam & Caraco 1984; Wrangham *et al.* 1993; Croft *et al.* 2008; Perkins *et al.* 2009). The increased nutritional requirements of larger groups are thought to cause larger groups to travel farther to visit more food patches that are depleted more rapidly (Janson & van Schaik 1988; Chapman *et al.* 1995; Janson & Goldsmith 1995; Chapman & Chapman 2000). Thus, individuals in larger groups are predicted to have to increase energy expenditure and travel farther to maintain energy intake (i.e., scramble competition: van Schaik & van Hooff 1983; van Schaik 1989). The increased energetic costs of living in a large group may be a sensible investment, however, if larger groups experience decreased predation or increased success in inter-group encounters (Croft *et al.* 2008; Perkins *et al.* 2009). Given the pros and cons of increases in group size, there may be an ideal group size that is most appropriate for a given environment where the benefit-to-cost ratio of group-living is maximized (van Schaik & van Hooff 1983). While some groups of animals may have more flexibility to adapt to changes in group size, herbivorous mammals are of particular interest because their fibrous diet, subsequent

slow digestion, and large body sizes (Demment & Van Soest 1985) make it difficult to adjust time budgets to changing competitive regimes (Borries *et al.* 2008).

The apparent abundance of resources (i.e., leaves) for folivorous primates in tropical forests has led to the suggestion that scramble competition should be limited or absent, and thus not a constraint on group size in folivorous primates (Isbell 1991). This hypothesis has been supported by numerous between-group comparisons of folivorous primates, which have found little or no evidence of scramble competition (Clutton-Brock & Harvey 1977; Struhsaker & Leland 1987; Janson & Goldsmith 1995; Yeager & Kirkpatrick 1998; Yeager & Kool 2000). In a comparative analysis of primates, however, Janson and Goldsmith (1995) found that folivores exhibited smaller group sizes than similarly sized frugivores in similar areas. This apparent contradiction between theory and empirical data has been coined the 'folivore paradox' (Janson & Goldsmith 1995; Steenbeek & van Schaik 2001; Snaith & Chapman 2005; Snaith & Chapman 2007); however, there is mounting evidence that folivores may exhibit significant scramble competition for quality leaf resources. Folivores are resource limited in larger groups (Borries *et al.* 2008), deplete patches as a function of group and patch size (Snaith & Chapman 2005), their total biomass is predicted by the quality of resources (Oates *et al.* 1990; Chapman *et al.* 2002), and the size of their daily range increases with group size (Snaith & Chapman 2008). These findings suggest that resources are limiting to folivores, and that scramble competition limits group sizes (Snaith & Chapman 2007; Borries *et al.* 2008; Snaith & Chapman 2008), as it does for frugivores.

In an attempt to resolve these contradictory findings, we conducted a longitudinal analysis of group size, food availability, and behaviour in red colobus monkeys (*Procolobus rufomitratus*) of Kibale National Park, Uganda. The mean group size of red colobus across Kibale has increased over the past 15 years, suggesting high levels of plasticity in this social variable (Gogarten *et al., In review*). Most studies examining the effects of group size on the behaviour of social mammals have compared multiple groups with different group sizes, but this method may be problematic if primates adjust group sizes in response to ecological conditions to avoid the costs of scramble competition, or if large groups are in high quality areas, negating the need to increase day range or modify activity budgets (Snaith & Chapman 2005). Our analysis of a single group through time reduces the number of confounding variables and provides an alternate means of examining the effect of group size on behaviour (Symington 1988). We examined how

changes in group size affect the behaviour of folivorous red colobus, specifically feeding, socializing, traveling, and dietary diversity. We predicted that larger group sizes are associated with increased time spent traveling (Koenig 2002; Snaith & Chapman 2008), decreased time spent feeding and socializing (van Schaik *et al.* 1983), higher dietary diversity (Snaith & Chapman 2007, 2008), and lower female fecundity (Beehner *et al.* 2006; Borries *et al.* 2008; Snaith & Chapman 2008). We expected that red colobus spend less time traveling, more time feeding and socializing, and have lower dietary diversity with increased food availability, so we control for variation in this factor in our analysis.

### 3.3 METHODS

We collected data from a well-habituated group of red colobus in Kibale National Park, Uganda (Struhsaker 1997; Chapman *et al.* 2000; Chapman *et al.* 2010b; Gogarten *et al.* 2012a). Kibale is a 795 $km^2$ park consisting of moist, mid-altitude forest located in the western part of Uganda in the foothills of the Rwenzori mountains (0°13'-0–0°41'N and 30°19'–30°32'E). Rainfall is seasonal with two rainy seasons (1,691 mm; data from 1990–2013, Chapman & Chapman, *unpubl. data*). Monthly rainfall data were collected immediately adjacent to the group's home range throughout the study.

We collected activity data over 56 months (July 2006 to February 2011). We collected group scan data by randomly selecting 5 adults every 30 minutes (N = 35,100 individual scans, monthly mean = 638 scans, range = 190-969 scans). During these scans, we recorded the individual's identity if known, sex, behaviour (e.g., feeding, traveling, being social [i.e., grooming, playing], resting; behavioural categories following: Struhsaker, 1975), the species and plant part being consumed if feeding, and the group spread. Individuals were classified as adults based on body size (Struhsaker 1975). Data were collected by C.A.C, J.F.G., and 5 well-trained field assistants; we conducted an intensive training period to standardize all data collection techniques and minimize inter-observer error prior to data collection.

Counts of the number of individuals in the group were conducted opportunistically when animals crossed open areas, facilitating a complete group count ($N_{counts}$ = 28). We estimated group sizes for each month from the available group counts by building a generalized additive model (GAM) with a smoothing spline over the studies' duration, using the R package 'mgcv' (Fig. 3.1; $R^2$ =

0.878, $p < 0.001$; Miyamoto *et al.,* 2013, Wood, 2011; Zuur *et al.,* 2009). We used a GAM model because we were interested in best fitting the available data and did not want to assume a linear relationship between time and group size (Zuur *et al.* 2009; Wood & Wood 2011; Wood 2011). For each group count, we calculated the infant to female ratio as an estimate of group-wide female fertility.

To estimate the availability of foods, we monitored the phenology of 309 trees representing 33 species each month (for details on tree species and methods: Chapman *et al.* 2005). We focused our estimate of food availability on young leaves, as young leaves constituted over 77.0% of the plant parts observed being consumed across the 56 months of observation. The remainder of the observed diet consisted of bark (representing 5.9% of time spent feeding), leaf petioles (5.5%), mature leaves (3.5%), unripe fruit (1.3%), leaf buds (1.2%), and flowers (1.2%; for a detailed analysis of the red colobus diet, see: Chapman *et al.*, 2002; Chapman & Chapman 2002). During each month, we estimated the availability of young leaves on each tree using a 0 to 4 scale (details in Chapman *et al.,* 2005). If a monitored tree died, it was replaced with another tree of the same species and approximate size to ensure a continuous sample size. As an index of food availability, we used the sum of the 0-4 scores divided by the number of trees monitored that month. Monthly phenology data were available for all months of the study except June 2008, June 2009, June 2010, and October 2010; for these months we used the average of the food availability index of the month before and after the missing month. In addition, while all of the species that we monitored were eaten by red colobus, we constructed a second food availability index consisting of nine of the ten most commonly consumed species, to capture aspects of phenology potentially most relevant to the red colobus (Table 3.1). These ten most commonly consumed species made up almost 60% of this group`s diet, but we excluded *Newtonia buchannani* because it is a rare, but preferred, species for which phenology data were not available.

When individuals were observed eating during a scan ($N$ = 15,774 scans), we recorded the species, age, and plant part being consumed (e.g., *Celtis durandii* young leaves, *Markhamia lutea* leaf petioles of a specific length). Due primarily to poor visibility, we were unable to identify the plant species being consumed in ~ 2% of feeding scans and these scans were excluded from the analysis of dietary diversity. For each month, we calculated the inverse of Simpson's index of diversity (Simpson 1949; Levins 1968) using the 'vegan' package for R

(Oksanen *et al.* 2012). We rarefied the data to the minimum number of scans during which individuals were observed eating in a given month ($N$ = 104) to eliminate any sample size effect (Soetaert & Heip 1990). This index ranges from 1 to a maximum equal to the number of species in the sample, with higher values indicating higher diversity.

While our behavioural data would ideally be analyzed using GLMMs with a multinomial response variable and controlling for potential temporal autocorrelation, we are unaware of any current implementation. Thus, we examined the influence of food availability and group size on behaviour using three separate generalized linear mixed effects models (GLMMs). Specifically, behaviour was treated as a binomial variable (feeding/not feeding, traveling/not traveling and socializing/not socializing) and we fitted models with a binomial error distribution, included monthly rainfall as a factor to control for aspects of environmental variability beyond food availability, and included month as a random factor. To account for temporal autocorrelation of months, as evidenced by moderate patterning in the residuals of these models, we used an AR1 correlation structure in our models (Zuur *et al.* 2009). Models were implemented in the 'nmle' (Pinheiro *et al.* 2012) and 'MASS' R packages (Venables & Ripley 2002) using the glmmPQL function, which implements generalized linear mixed-model with a penalized quasi-likelihood. To test whether the estimation of group size in months, rather than actual counts, impacted our analysis, we created three reduced GLMMs including only data from those months where group counts were available.

Because the GLMMs did not allow us to simultaneously analyze behavioural differences across categories, we built separate models to the explain variance in each of feeding, traveling, and social behaviours. Unfortunately these models do not allow the inclusion of random effects or temporal autocorrelation, so results should be interpreted with caution. Given that individuals had a choice between any number of behaviours at any given time, we analyzed behaviour as a nominal response to group size in a multinomial baseline-category logit model (Agresti 2002), implemented in the 'nnet' R package (Venables & Ripley 2002; Ripley 2011). We predicted that the behaviour of an adult would be a function of group size, food availability, and rainfall. We included resting behaviour in this multinomial baseline-category logit model to have a base-line behaviour against which to compare our three behaviours of interest (Agresti 2002). These four behaviours represented 97.8% of the scans.

We tested the influence of food availability, rainfall, and group size on dietary diversity using multiple regression. In addition, to specifically test for the effect of group size on dietary diversity, we constructed a simple linear model to explain variance in dietary diversity as a function of group size. To test whether the infant to female ratio, an indicator of female fecundity (van Schaik 1983; Fedigan 1986), was influenced by group size, we constructed a linear model explaining variance in the infant to female ratio as a function of group size. We examined the influence of group size and food availability on group spread using a GLMM with a poisson error distribution and included monthly rainfall as a factor to control for aspects of environmental variability beyond food availability, and included month as a random factor. As for the analysis of the behavioural data, we used an AR1 correlation structure to control for temporal autocorrelation in the data (Zuur *et al.* 2009). All statistical analyses were conducted in R version 2.14.2 (R Development Core Team 2012).

### 3.4 ETHICAL CONSIDERATIONS

This work conforms to the legal requirements of Canada and Uganda and was carried out under appropriate ethics and legal clearances. Specifically, permission to conduct this research was given by the National Council for Science and Technology, the Uganda Wildlife Authority, and the McGill Animal Care Committee.

### 3.5 RESULTS

Group size increased from 57 to 98 individuals over 56 months (Fig. 3.1). Animals spent most of their time feeding (45.8%), followed by resting (37.7%), traveling (7.9%), and socializing (6.4%). The monthly food availability index ranged from 0.852 to 2.041 (mean = 1.673, $\sigma$ = 0.247). The food availability index for the ten top red colobus foods ranged from 1.298 to 2.430 (mean = 1.898, $\sigma$ = 0.229), while monthly rainfall ranged from 34.7 to 376.3 mm (mean = 142.71, $\sigma$ = 84.85).

The percentage of time spent feeding ranged from 32.9 to 59.8% across months, the percentage of time spent traveling ranged from 0.7 to 23.3%, and the percentage of time spent socializing ranged from 1.5 to 13.2%. Monthly variation in behaviour was best explained by variation in

group size. Increasing group size was associated with a decreased percentage of time spent feeding and socializing, and an increased percentage of time spent traveling (Table 3.2). Controlling for changes in food availability and rainfall, as the observed group size increased (by *N*=41individuals), red colobus were 77% less likely to be observed feeding during a scan, 82% less likely to be observed socializing, and 239% more likely to be observed traveling. Increasing food availability was associated with an increase in the proportion of time spent feeding, while the relationships to time spent socializing (positive) and traveling (negative) were not significant (Table 3.2). Comparing the maximum to minimum food availability observed, while controlling for changes in food availability and rainfall, individuals were 111% less likely to be observed traveling, 54% more likely to be observed feeding, and 59% more likely to be socializing. Rainfall did not predict any behaviour. Similar results were found between food availability based on the ten most commonly consumed foods and group size: feeding was predicted by group size and food availability, but not rainfall; socializing was predicted by group size, but not food availability or rainfall; and traveling was predicted by group size and food availability, but not rainfall (Table 3.3). Results of the reduced models, including only months where group counts were available, are qualitatively extremely similar (Table 3.4), although in models explaining variance in traveling and socializing behaviour, group size was no a longer significant predictor, likely due to small sample size. The effects of previously significant predictors remained in the same direction as in the full models (i.e., the effect of previously positive predictors remained positive and previously negative predictors remained negative).

From results of the multinomial logit model, behaviour was most affected by group size (likelihood ratio $\chi^2 = 873.04$, $df = 3$, $p < 0.001$), followed by food availability (likelihood ratio $\chi^2 = 204.23$, $df = 3$, $p < 0.001$), and rainfall (likelihood ratio $\chi^2 = 11.59$, $df = 3$, $p = 0.009$). Increasing group size was associated with an increased probability that individuals were traveling and a decreased probability that they were socializing or feeding (Table 3.5; Figs. 3.2A and 3.2B), while increasing food availability was associated with an increased probability of individuals feeding and socializing and a decreased probability of traveling (Table 3.5; Fig. 3.2C). Similar results were observed using the availability of the most commonly consumed foods: behaviour was best predicted by group size ($\chi^2 = 840.58$, $df = 3$, $p < 0.001$), followed by food availability ($\chi^2 = 132.61$, $df = 3$, $p < 0.001$) and rainfall ($\chi^2 = 15.14$, $df = 3$, $p = 0.002$).

Increasing group size was associated with higher dietary diversity ($R^2$= 0.23, $F_{[1,53]}$ = 16.6, $p$ < 0.001; Fig. 3.3). This relationship was also the observed in the multiple regression analysis, explaining variance in dietary diversity as a function of group size, food availability, and rainfall ($R^2$ = 0.24, $F_{[3,51]}$ = 5.42, $p$ < 0.005), where food availability ($t$ = 0.470, $p$ = 0.640) and rainfall ($t$ = 0.074, $p$ = 0.941) were not significant factors in the model, but group size was ($t$ = 2.279, $p$ < 0.01). Increasing group size had no effect on the female to infant ratio ($R^2$= 0.045, $F_{[1,26]}$ = 1.22, $p$ = 0.279). While neither food availability ($t$ = -0.730, $df$ = 51, $p$ = 0.470) nor rainfall ($t$ = 1.012, $df$ = 51, $p$ = 0.316) were significant predictors of group spread, increasing group size was associated with increasing group spread (AR1 Phi =0.765, $t$ = 2.380, $df$ = 51, $p$ < 0.001).

### 3.6 DISCUSSION

Our results demonstrate that changes in group size impact the behaviour of a folivorous primate above and beyond the variance explained by environmental factors, such as rainfall or food availability. With increasing red colobus group size, group spread increased and individuals spent less time feeding and socializing and more time traveling. This is likely a result of increases in patch depletion and scramble competition (caused by higher within-group competition), necessitating an increased proportion of time spent traveling to meet the group's nutritional demands. Similarly, dietary diversity increased with group size, suggesting that within-group competition forces individuals to eat less-preferred foods. Taken together, these results suggest that there are increasing costs of within-group competition associated with increasing group size, including the energetics of travel, quality of the diet, and time available for sociality. This finding is in accord with a growing body of evidence that scramble competition in folivores limits group sizes (Snaith & Chapman 2005; Snaith & Chapman 2007; Borries *et al.* 2008; Harris *et al.* 2010). In fact, in the current study the impact of this scramble competition apparent at larger group sizes causes greater shifts in the activity budget than seasonal fluctuations in food availability or rainfall.

Our results suggest that scramble competition is occurring in folivores (Borries *et al.,* 2008; Snaith & Chapman, 2007, 2008). A remaining question is, given an abundant supply of leaves in tropical forests, why is there still competition among animals over this resource? Possible reasons include variance in resource quality driven by factors such as plant secondary

compounds (Schofield *et al.* 2001; Cardiff *et al.* 2007; Rothman *et al.* 2009), available energy (Danish *et al.* 2006; Rothman *et al.* 2011), protein-to-fiber ratio (Chapman *et al.* 2003; Felton *et al.* 2009), lack of nutrients such as sodium (Rode *et al.* 2003; Rothman *et al.* 2006; Reynolds *et al.* 2009; Irwin *et al.* 2010), or overall fiber content (Milton 1979). While the red colobus appear to be able to deal with increasing scramble competition by increasing their dietary diversity and changing their activity budget, individuals in the group may be obtaining lower quality foods as group size increases. A recent study using genetic data from two unequally sized neighbouring groups of red colobus demonstrated increased female relatedness within the smaller group, which may suggest that females are less likely to disperse when there is less within-group competition (Miyamoto *et al.* 2013). Future studies should examine variation in the nutritional content and chemical defenses of less preferred plants, and if individuals in larger group sizes are forced to consume lower quality foods. It is possible that the increasing dietary diversity observed at larger group sizes is simply a result of individuals in larger groups encountering a greater variety of foods due to increased group spread and potentially larger home ranges (Snaith & Chapman 2008). We found no relationship between the infant-to-female ratio and group size, which suggests that shifts in red colobus diet are not yet great enough to impact female fecundity and that this behavioural flexibility is able to compensate for increased scramble competition. Alternatively, there may be a lag time between the increased levels of competition and resulting changes in energy expenditure and diet, and their effects on female fecundity, which the current study, despite its long-term nature, was unable to capture.

Studying a single group longitudinally, rather than multiple groups in a cross-sectional study, allowed us to control for environmental variation and reduce the number of potentially confounding ecological and group-specific variables. However, it is important to acknowledge the limitations associated with studying a single group. It remains possible that the patterns observed in this group are an anomaly, thus further studies are needed to corroborate these findings in other groups. The costs and time necessary for habituating primates pose significant hurdles to replicating this study, and an alternative powerful approach is combining findings from both longitudinal and cross-sectional studies. The results of the current study closely mirror results from a cross-sectional study of red colobus in Kibale, where individuals in larger groups exhibited increased group spread and spent more time feeding and less time engaged in social behaviour (Snaith & Chapman 2008). Further, the changes in group size observed in this group

of red colobus are mirrored by changes in group sizes observed across Kibale National Park over the last 15 years (Gogarten *et al., In review*). The factors responsible for these large scale changes in group size remain unknown; changes in food availability or quality may be playing a role, but additional studies are needed. Alternatively, red colobus groups may be recovering from a major disturbance (e.g., disease), which led to smaller than optimal group sizes for the environment, with group sizes only now slowly recovering. Regardless of the causes of the observed changes in group sizes, these changes are modifying the behaviour of red colobus, which in turn may impact disease susceptibility, stress, reproductive rates, and ultimately population viability (Pride 2005; Borries *et al.* 2008; Snaith *et al.* 2008; Griffin & Nunn 2012; Rifkin *et al.* 2012).

Socioecological theory suggests that individuals will remain in a group so long as the benefits outweigh the costs, although average group size could rise above the optimum as the benefit for a solitary animal to be a member of a group is greater than the benefit of excluding an individual attempting to join a group (Giraldeau 1988; Purvis *et al.* 2000). While we found support for the possibility that ecological constraints will limit group size, a number of alternative hypotheses have been proposed, including: infanticide (Steenbeek & van Schaik 2001; Cardillo *et al.* 2005; Chapman & Pavelka 2005), disease (Chapman *et al.* 2006; Nunn & Altizer 2006; Godfrey & Irwin 2007; Chapman *et al.* 2009), social memory (Shumway & Stoffer 2000), and predation, the latter of which may limit group size via mortality, especially if larger groups are easier for predators to detect (Hairston *et al.* 1960; Isbell 1994). We have no evidence that infanticide or predation represent significant pressures for this red colobus group, although it has been observed in this species (Struhsaker & Leland 1985; Struhsaker & Leland 1987), but the importance of these and other factors in limiting group size is a continuing focus of our monitoring efforts of this group. The changes in behaviour documented here suggest that continued monitoring of this population is critical, as such changes have been shown to impact fitness (Borries *et al.* 2008) and disease susceptibility (Griffin & Nunn 2012) in primate populations.

### 3.7 ACKNOWLEDGEMENTS

We thank Magnus Bein, Carl Boodman, Carola Borries, Johanna Bleecker, Lauren Chapman, Caitlin Friesen, Ria Ghai, Aerin Jacob, Laura Johnson, Joanna Setchell, and two anonymous reviewers for helpful comments on this project. We are grateful to Catherine Potvin, Jon Sakata, and Christopher Solomon for advice and recommendations regarding statistical analyses and for their comments on this manuscript. JFG would like to particularly thank Charles Nunn, David Marcogliese and Jonathan Davies for their continued support, helpful comments, and suggestions throughout the development of this project and his dissertation. Funding for the research in Kibale was provided by the Canada Research Chairs Program, Natural Science and Engineering Research Council of Canada, National Science Foundation, National Geographic, and by NIH grant TW009237 as part of the joint NIH-NSF Ecology of Infectious Disease program and the UK Economic and Social Research Council. JFG was supported by a Graduate Research Fellowship from the National Science Foundation, the Canadian Institutes of Health Research's Systems Biology Training Program, a Vanier Canada Graduate Scholarship, and the Explorers Club – Eddie Bauer Youth Grant. JFG gives particular thanks to Linn Taubert and Maria, Peter, Jacob, Jennifer, Stephanie, Jolyon, Rupert, and Thomasina Gogarten for their love, support and generosity that ultimately made this project feasible.

**Table 3.1: Food species most commonly consumed by the red colobus study group in Kibale National Park, Uganda between July 2006 and February 2011.**

| Species name | Percent of diet |
|---|---|
| *Albizia grandibracteata* | 10.5 |
| *Trilepsium madagascariense*# | 9.4 |
| *Newtonia buchananii** | 8.7 |
| *Celtis africana* | 5.6 |
| *Celtis durandii* | 4.4 |
| *Millettia dura* | 3.8 |
| *Strombosia scheffleri* | 3.6 |
| *Parinari excelsa* | 3.5 |
| *Dombeya mukou* | 3.3 |
| *Macaranga schweinfurthii* | 3.2 |
| *Prunus africana* | 2.9 |
| Total | 58.9 |

# Formerly *Bosqueia phoberos*

*Phenology data not available for this species as it was very rare but preferred, so it was not included in the food availability index

**Table 3.2: Generalized linear mixed models explaining variance in probability that individual red colobus in the study group in Kibale National Park, Uganda were socializing, feeding, or traveling at a given scan between July 2006 and February 2011.**

| | | Independent variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Group size | | | Rainfall | | | Food availability | | |
| Dependent variable | AR1 Phi | Effect (std. err.) | *T* | *p*-value | Effect (std. err.) | *T* | *p*-value | Effect (std. err.) | *T* | *p*-value |
| Proportion of time spent socializing | 0.087 | -0.0200 (0.00702) | -3.912 | **<0.0001** | -0.000589 (0.000708) | -0.832 | 0.409 | 0.498 (0.309) | 1.609 | 0.114 |
| Proportion of time spent feeding | 0.409 | -0.0188 (0.00389) | -4.819 | **<0.0001** | 0.000183 (0.000398) | 0.460 | 0.648 | 0.537 (0.175) | 3.067 | **0.0035** |
| Proportion of time spent traveling | 0.153 | 0.0583 (0.0108) | 5.392 | **<0.0001** | 0.000898 (0.00110) | 0.818 | 0.417 | -0.932 (0.492) | -1.894 | 0.0639 |

*P*-values in bold indicate significant results (α = 0.05). AR1 Phi represents the estimated autocorrelation between consecutive months for the model. The df for all presented *T*-values = 51.

**Table 3.3: Generalized linear mixed models explaining variance in probability that individual red colobus in the study group in Kibale National Park, Uganda were socializing, feeding, or traveling at a given scan between July 2006 and February 2011, using a food availability index based only on the top ten red colobus foods.**

| | | Independent variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Group size | | | Rainfall | | | Food availability* | | |
| Dependent variable | AR1 Phi | Effect (std. err.) | *T* | *p*-value | Effect (std. err.) | *T* | *p*-value | Effect (std. err.) | *T* | *p*-value |
| Proportion of time spent socializing | 0.087 | -0.0164 (0.00616) | -2.661 | **0.010** | -0.000613 (0.000735) | -0.834 | 0.482 | 0.380 (0.295) | 1.285 | 0.205 |
| Proportion of time spent feeding | 0.408 | -0.0150 (0.00343) | -4.353 | **<0.0001** | 0.000146 (0.000417) | 0.350 | 0.728 | 0.430 (0.168) | 2.563 | **0.0134** |
| Proportion of time spent traveling | 0.153 | 0.0530 (0.00936) | 5.679 | **<0.0001** | 0.00110 (0.00113) | 0.975 | 0.334 | -0.890 (0.462) | -1.927 | 0.0595 |

*P*-values in bold indicate significant results (α = 0.05). AR1 Phi represents the estimated autocorrelation between consecutive months for the model. The df for all presented *T*-values = 51.

***** Models built using food availability index based on the ten most commonly consumed red colobus foods.

**Table 3.4: Reduced generalized linear mixed models explaining variance in probability that individual red colobus in the study group in Kibale National Park, Uganda were socializing, feeding, or traveling at a given scan between July 2006 and February 2011. Included only data from those months where group counts were available.**

| | | Independent variables | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Group size | | | Rainfall | | | Food availability | | |
| Dependent variable | AR1 Phi | Effect (std. err.) | *T* | *p*-value | Effect (std. err.) | *T* | *p*-value | Effect (std. err.) | *T* | *p*-value |
| Proportion of time spent socializing | 0.083 | -0.0133 (0.0116) | -1.139 | 0.271 | -0.00201 (0.00123) | -1.638 | 0.120 | 0.243 (0.457) | 0.532 | 0.602 |
| Proportion of time spent feeding | 0.412 | -0.0211 (0.00605) | -3.484 | **0.0028** | 0.00131 (0.000642) | 2.035 | 0.0578 | 0.588 (0.243) | 2.419 | **0.0271** |
| Proportion of time spent traveling | 0.134 | 0.0208 (0.0228) | 0.915 | 0.3728 | -0.000154 (0.00239) | -0.0643 | 0.950 | 0.0321 (0.928) | 0.0346 | 0.973 |

*P*-values in bold indicate significant results ($\alpha = 0.05$). AR1 Phi represents the estimated autocorrelation between consecutive months for the model. The df for all presented *T*-values =19.

**Table 3.5: Estimated parameters in the multinomial logit model for red colobus behaviour in Kibale National Park, Uganda between July 2006 and February 2011, with resting as the baseline category.**

| Logit | Intercept | Effect of group size | Effect of food availability | Effect of rainfall |
|---|---|---|---|---|
| $\log(\pi_f/\pi_r)$ | 0.442 | -0.0169 | 0.649 | 0.000251 |
| | (0.106) | (0.00145) | (0.0650) | (0.000149) |
| $\log(\pi_s/\pi_r)$ | -0.939 | -0.0290 | 0.934 | -0.000384 |
| | (0.204) | (0.00277) | (0.126) | (0.000291) |
| $\log(\pi_t/\pi_r)$ | -5.445 | 0.0589 | -0.694 | 0.000688 |
| | (0.224) | (0.00294) | (0.123) | (0.000271) |

SE values in parentheses. These parameters are combined to describe the log odds of performing a given behaviour instead of resting as a function of group size, food availability, and rainfall (e.g., the odds of feeding over resting is $\log(\pi_f/\pi_r) = 0.442 - 0.0169*$ Group size + 0.649 * Food availability + 0.000251 * rainfall). *f,* feeding; *r*, resting; *t*, traveling; *s,* socializing.

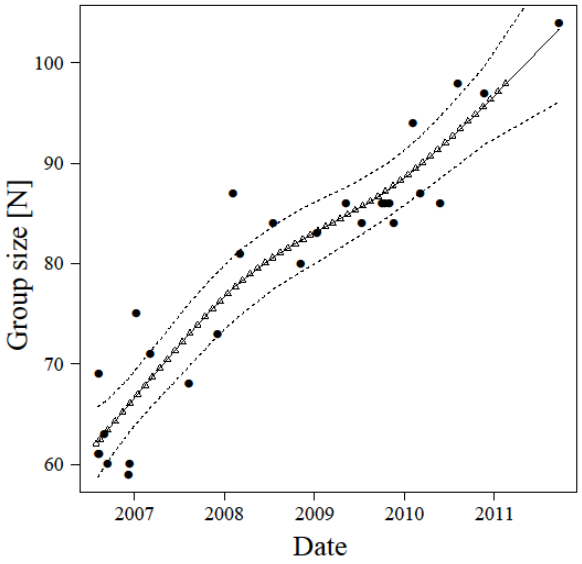



**Figure 3.1: Size of the study group of red colobus in Kibale National Park, Uganda between July 2006 and February 2011**. Black circles represent group counts, triangles represent the predicted group size for each month, the solid line represents the generalized additive model predicting group size through time, and the dashed line represents the 95% confidence bands.

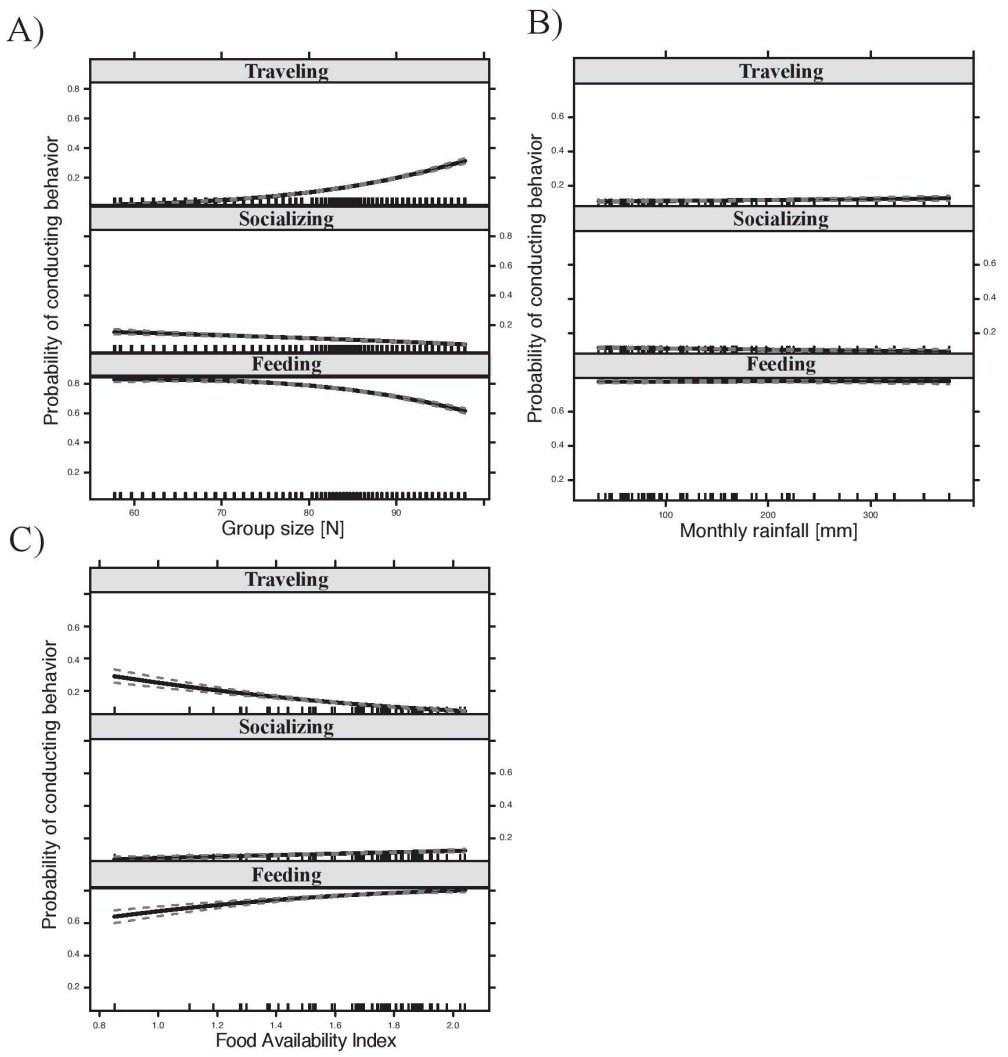



**Figure 3.2: Fitted probabilities from the multinomial logit model** showing the effect of A) group size, B) rainfall, and C) food availability on the probability of conducting different behaviours by the red colobus group studied in Kibale National Park, Uganda between July 2006 and February 2011. Dashed lines indicate 95 % point-wise confidence envelope around the fitted probabilities of the model. Tick marks along the x-axis represent data used in the construction of the model.

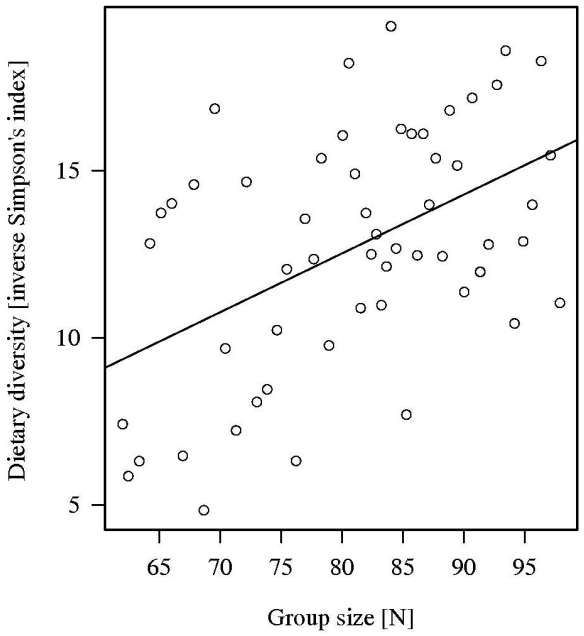



**Figure 3.3**: **Dietary diversity as a function of the size of the red colobus study group** in Kibale National Park, Uganda between July 2006 and February 2011. Solid line represents the regression explaining variance in dietary diversity as a function of group size.

# LINKING STATEMENT 2

Non-human primates living in protected areas are faced with rapidly changing social and ecological pressures. During my PhD, I was involved in a series of studies examining how these changes impact gastrointestinal parasite communities (Chapman *et al.* 2012), group genetic structure and dispersal (Miyamoto *et al.* 2013), the cascading impact of these changes on population sizes (Chapman *et al.* 2015a; Chapman *et al.* 2015b) and polyspecific associations (Gogarten and Chapman, unpubl. data). In the previous two chapters I documented major environmental and behavioural changes happening on a large spatial and temporal scale in Kibale National Park. The impact of changes in-group size documented in chapter 2, were examined in detail in chapter 3. In the following chapter I consider in the gut microbiome.

Non-primate life is estimated to make up 90% of the cells in a primate and to contribute as much as 99% of the unique genes present in an organism. This microbial diversity seems to be highly versatile, which may allow organisms to adapt more rapidly to changing conditions (Zilber-Rosenberg & Rosenberg 2008). On the other hand, major pathogens illustrate the potential problems that members of this microbiome can cause. If this microbiome is partially heritable, then it may also represent the actual unit of selection in evolution (Zilber-Rosenberg & Rosenberg 2008). Microbiome assembly might also be influenced by short-term ecological processes such as dispersal limitation (Costello *et al.* 2012). Dispersal limitation occurs by selective exposure to other hosts or environmental items, which in effect limits the potential source pool on which other ecological processes can act. Primates represent a particularly social order, with 75% of primate genera forming year round associations in contrast with ~33% of non-primate mammal genera (van Schaik & Kappeler 1997). These close social communities create opportunities for transmission of the gut microbiome and might create a 'pan-microbiome' of organisms that are shared by a social group, which could maintain differences observed between host species (Moeller *et al.* 2016b) while enhancing the community stability of microbiomes within groups and individuals via metacommunity dispersal dynamics (Leibold *et al.* 2004). To date we know little about how sociality impacts microorganism community composition in primates, and how stable these communities are through time. In the next chapter, I seek to address this knowledge gap by examining the bacterial gut microbiome of wild non-human primates in Taï National Park, Côte d'Ivoire.

# CHAPTER 4

## HUNTING CAPTURES MORE THAN PREY: EVIDENCE FOR MICROBIOME ASSIMILATION BY CHIMPANZEES

Jan F. Gogarten[1,2,3], Jonathan Davies[1], J. Peter Gogarten[4], Joerg Graf[4], Alexander Mielke[2], Roger Mundry[5], Michael C. Nelson[4], Roman M. Wittig[2,6], Fabian H. Leendertz[3], Sébastien Calvignac-Spencer[3]

[1]Department of Biology, McGill University, 855 Sherbrooke Street West, Montreal, QC, H3A2T7, Canada

[2]Primatology Department, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany

[3]Project Group Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Seestraße 10, 13353, Berlin, Germany

[4]Department of Molecular and Cell Biology, University of Connecticut, 91 North Eagleville Road, Unit 3125, Storrs, Connecticut, 06269-3125, USA

[5]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany

[6]Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, B.P.1303, Abidjan 01, Côte d'Ivoire

## 4.1 ABSTRACT

Microbiomes make up as much as 90% of cells and 99% of unique genes in their hosts. The microbiome impacts a variety of processes including a host's ability to access nutrients and maintain health. While species differences in microbiomes have been described across ecosystems, little is known about how microbiomes of individual hosts assemble, particularly in the ecological and social contexts in which they evolved. We examined gut microbiome assembly within nine sympatric wild non-human primate (NHP) species in a community where a strong hunter-prey relationship exists between chimpanzees (*Pan troglodytes verus*) and colobines. Despite sharing a common environment, and regular interspecific interactions, we found that individuals harbored unique and persistent microbiomes that were influenced by host species, social group, and parentage, but not social relationships among members of the same social group. We found a branching order of host-species networks constructed using the composition of their microbial communities as characters, which was incongruent with known NHP phylogenetic relationships, with chimpanzees sister to their colobine prey. In contrast to strong evidence of phylogenetic clustering in the microbiomes of all monkeys, chimpanzee microbiomes showed strong evidence of phylogenetic overdispersion. We suggest this reflects unique ecological processes driving microbiome assembly in chimpanzees, possibly due to broader exposure through hunting and microbiome assimilation of their primate prey. This comprehensive study of contemporaneous microbiomes of all sympatric diurnal NHPs in an ecosystem highlights the diverse transmission routes that shape these complex communities.

## 4.2 INTRODUCTION

Gut microbiomes of mammals represent complex communities. These microbiomes influence a broad array of processes including a host's ability to access nutrients (Tremaroli & Bäckhed 2012), development and tissue maturation (Collins *et al.* 2012), health via influencing processes such as pathogen exclusion and immune system priming (Hooper *et al.* 2012), and even behavior and scent (Ezenwa & Williams 2014). The importance of this and other microbiome communities (e.g., skin, vaginal, oral) has led to suggestion that the holobiont, i.e., the host plus the entirety of its associated microbiomes, is one of the units of biological organization on which natural selection acts (Zilber-Rosenberg & Rosenberg 2008; Bordenstein & Theis 2015).

Differences in the composition of the gut microbiome of hominine species appear to recapitulate their evolutionary relationships, possibly indicating co-divergence of the microbiome with their hosts (Ochman *et al.* 2010) and studies on the scale of individual bacterial taxa suggest that a number of bacterial lineages have co-diversified with hominines over the last 6 to 12 million years (Moeller *et al.* 2016a). However, such patterns can also emerge if cross-species transmission occurs and is facilitated by host phylogenetic proximity (i.e., preferential host-switching); an alternative hypothesis that is often difficult to disentangle from co-divergence (Charleston & Robertson 2002). Regardless of the underlying process, present-day humans seem to have depauperate microbiomes when compared to great apes and other non-human primates (NHPs) (Moeller *et al.* 2014). Studies of wild NHPs might thus provide important insights into the ancestral human gut microbiome, which was most certainly heavily modified by changes occurring recently in our history, such as the development of agriculture and the use of antibiotics (Gillings *et al.* 2015).

Microbiome assembly is influenced by ecological processes such as dispersal limitation and environmental selection (Costello *et al.* 2012). Dispersal limitation occurs by selective exposure to other hosts or the environment, which in effect limits the potential source pool of potential colonizers. For example, in humans vaginal birth exposes new-borns to unique bacteria which have a long-term effect on their microbiome structure (Dominguez-Bello *et al.* 2010). The order in which colonization occurs might thus condition persistence or exclusion of particular taxa; e.g., some bacteria may serve as ecosystem engineers and facilitate colonization by others. Primates represent a particularly social order, with 75% of primate genera forming year round

associations in contrast with ~33% of non-primate mammal genera (van Schaik & Kappeler 1997). These close social communities present opportunities for transmission of the gut microbiome, and might present a 'pan-microbiome' of microorganisms that are shared by a social group, while maintaining differences observed between host species (Moeller *et al.* 2016b) and enhancing the community stability of microbiomes within groups and individuals via metacommunity dispersal dynamics (Leibold *et al.* 2004). Most NHPs also live in highly diverse tropical ecosystems, where leaves from an individual tree can be host to more than 400 bacterial taxa (Kembel *et al.* 2014). Disturbances to this complex ecological habitat can additionally influence gut microbiomes of NHPs (Amato *et al.* 2013; Barelli *et al.* 2015), while removal to captivity causes even more extreme shifts in microbiome composition (Clayton *et al.* 2016). Environmental selection may also play a role through the availability of resources a bacterium can efficiently exploit (e.g., diet drives shifts in the human microbiome (David *et al.* 2014)) and the fit to specific host conditions (e.g., human genotypes (Goodrich *et al.* 2014) or immune system states (Hooper *et al.* 2012)). The balance of these ecological processes involved in microbiome assembly in humans and NHPs remains largely unknown.

Host-microbiome ecosystems are short-lived (with an absolute upper bound being a host's lifetime); however, bacteria are fast-evolving organisms capable of horizontal gene transfer which provides a means for rapid evolutionary change (Gogarten *et al.* 2002), and allows for strong eco-evolutionary feedback loops between the microbiome, which provide essential ecosystem services, and host (Theis *et al.* 2016). For example, the shift to carnivory may have significantly changed the composition of the human microbiome (Moeller *et al.* 2014), while the seaweed rich diet of Japanese populations has been accompanied by a shift in the genes of their microbiomes that allow these populations to digest recalcitrant polysaccharides in their food (Hehemann *et al.* 2012; Thomas *et al.* 2012). We know much less, however, about the ecological processes that drove the evolution of the human microbiome during most of our species' lifetime, that is when humans were exclusively tropical hunter-gatherers living in sympatry with diverse NHP communities. Two approaches allow us to explore this question. First, the study of present-day hunter-gatherer societies, which suggests that most modern human populations may have lost much of their microbial diversity (Schnorr *et al.* 2014; Clemente *et al.* 2015), though even these hunter-gatherer societies have also undergone major societal and environmental

changes. Second, comparative analyses of our closest relatives—NHPs—in their natural environment.

Here we seek to understand the factors influencing microbiome assembly in a diverse wild NHP community sharing a common environment and including a known predator-prey relationship. We examined the gut microbiomes of all nine sympatric diurnal NHP species present in Taï National Park, Côte d'Ivoire, and described the microbiome community structure among individuals within conspecific groups, among social groups within species, and among species in the context of their phylogenetic relationships. First, we tested whether social groups form biological islands with distinct gut microbiomes (Moeller *et al.* 2016b). Second, we contrasted the importance of evolutionary history versus diet in shaping species differences. If the microbiome community is primarily structured through co-evolution or preferential host-switching, we would expect the microbiome of colobines and cercopithecines to be most similar as these clades are sister to each other. In contrast, if diet is the primary driver of microbiome community structure, we would expect more omnivorous chimpanzees and sooty mangabeys (*Cercocebus atys atys*) to have more similar microbiomes, with folivorous colobines distinct from both. Third, we examined how behavior mediates both relationships, for example, by modifying dispersal probabilities via influencing contact rates between individuals. Uniquely, our study system also includes a complex hunter-prey relationship, where chimpanzees regularly predate on colobines (Boesch 1994; Boesch & Achermann 2000). Such trophic interactions have previously been shown to influence the transmission of retroviruses between species (Gogarten *et al.* 2014a), here we additionally examined whether hunting provides a novel dispersal route between prey and hunter for gut microbes. In addition, for sooty mangabeys, we examined neonatal colonization and whether similarity in mother-offspring gut microbiomes persists into adulthood, the importance of behaviors, such as grooming that mediate conspecific interactions, and spatial proximity.

## 4.3 METHODS

### 4.3.1 Site and study groups

The study was conducted on wild primates in the Taï National Park, Ivory Coast (6°20'N to 5°10'N and 4°20'W to 6°50° W). Two habituated species of primates inhabit the study area; a

group of sooty mangabeys (named the Audrenissrou group) and three neighboring groups of chimpanzees (named the North, South, and East group). The mangabey group was habituated starting in November 2012, while the chimpanzee groups have been under observation since 1979 (Boesch & Achermann 2000). Monkey species in Taï regularly interact, forming mixed species associations (McGraw & Bshary 2002), while chimpanzee hunting of colobines, particularly red colobus, is common (Boesch 1994; Boesch & Achermann 2000).

Samples were collected from both habituated groups and unhabituated groups, with efforts made to avoid resampling individuals by collecting a small number of samples relative to the group size under a tree occupied by the group. The microbiomes of samples from the habituated mangabeys and chimpanzees were analysed to examine whether: 1) individuals had distinct microbiomes, 2) whether there was turnover in the microbiome over time, 3), whether after a year, microbiomes from the same individual were more similar to their own samples than to other individuals, 4) whether individuals had microbiomes similar to their mothers, and 5) whether microbiomes from the same social group are more similar to one another than to those from different social groups. Samples from all species were used to examine the effect of host species on the relative abundances and phylogenetic structure of fecal bacteria. Samples from chimpanzees were also examined to evaluate the phylogenetic depth at which bacterial communities were structured.

### 4.3.2 Sample collection

To examine the stability of the microbiome we collected repeated samples from known individual chimpanzees and sooty mangabeys. To evaluate the importance of mother-offspring relations in shaping the gut microbiome, we examined 22 known mother-offspring sooty mangabey pairs. To evaluate the impact of social group membership on the gut microbiome, we collected samples from chimpanzees living in three neighboring groups and sooty mangabeys living in the Audrenissrou group and a neighboring unhabituated group. To explore cross species variation in gut microbiomes and phylogenetic structure of these microbiomes, we collected samples from all nine sympatric diurnal NHP species. From the Audrenissrou group, we collected repeated samples from 53 and 26 individuals in April, May and June of 2014 and 2015 respectively, with a single additional sample available from September 2013. An additional individual was sampled only once in the 2014 and 2015 sampling periods. For the South and

North group we collected multiple samples from 18 and 11 individuals respectively, with six and one individual sampled only once in these groups. We collected a single sample from 28 individuals in the East group. Samples from chimpanzees and unhabituated groups were collected over the same months as the Audrenissrou group in 2014, with the exception of 11 of 28 samples collected from the unhabituated neighboring sooty mangabey group, which were available from August and November in 2013.

Fecal samples (*N*=380; Fig. S4.1) were stored either by immediately mixing ~1ml feces with an equal volume of RNAlater, or ~2ml of feces were kept cool in a thermos in the field and put into liquid nitrogen at the end of the day. For habituated animals, collection occurred immediately after defecation, while for unhabituated animals, once a group was detected, we waited at a distance before searching the area where the group had been, collecting only a limited number of samples to avoid repeated sampling of the same individual. Samples stored in RNAlater were homogenized by mixing vigorously and stored for 5 days at ambient temperature (25-30°C) following manufacturer's instructions and then stored in liquid nitrogen until transport on dry ice to maintain a < -80°C temperature chain.

### 4.3.3 Generating bacterial gut community data

To characterize the microbiome, DNA was extracted from samples using the Matrix Stool DNA purification kit (Roboklon). DNA concentrations were quantified using a Syngery HT (Biotek, Winooski, VT) with the Quant-iT PicoGreen kit (Invitrogen, ThermoFisher Scientific). Following Nelson et al. (2014), at the University of Connecticut's Microbial Analysis, Resources, and Services facility, the 16S V4 hypervariable region was amplified using the 515F and 806R primers developed by Caporaso et al. (2012). Briefly, amplicons were generated from 5 to 20ng of extracted DNA in triplicate PCR reactions. These reactions were pooled, quantified and diluted to 4nM prior to pooling amplicons for all samples. Pooled libraries were sequenced on an Illumina MiSeq using a 2x250bp sequencing protocol. Returned reads were pre-processed according to protocols described by Nelson et al. (2014). Read pairs were merged to form single amplicons, and screened for quality and length. Resulting reads were processed using QIIME v.1.8 and the 13-08 release of the Greengenes reference sequences. Operational taxonomic units (OTUs) were determined by clustering reads to the Greengenes reference 16S rRNA gene reference dataset (DeSantis *et al.* 2006) at 97% identity, and then *de novo* OTU clustering

performed on reads that failed to cluster to a reference (McDonald *et al.* 2011; Nelson *et al.* 2014). Chimeras were removed and the dataset was filtered to remove reads of low quality (less than Q30), singleton and doubleton OTUs and then OTUs present at less than 0.0005% (Bokulich *et al.* 2012). From these sequences, we generated a De novo phylogeny using FastTree, which incorporates the GreenGenes reference OTU sequences.

### 4.3.4 Species confirmation for samples collected from unhabituated animals

To confirm the species identity of samples collected from unhabituated primates, we used a PCR targeting the mitochondrial ribosomal 16S RNAgene (Forward primer: CGGTTGGGGTGACCTCGGA; Reverse primer: GATGTCCTGATCCAACATCGAG) and the following conditions: 5 min at 95°C, 42 cycles [30 s at 95°C, 30 s at 64°C, 60 s at 72°C], 10 min at 72°C (Calvignac‐Spencer *et al.* 2013). Amplified products were sequenced using Sanger's sequencing and sequences compared to publicly available sequences in GenBank through BLAST (Altschul *et al.* 1990). This resulted in the reclassification of 14 of the 53 samples collected from unhabituated groups.

In addition, chimpanzees hunt red colobus in this ecosystem, and bacterial sequences in feces could simply represent passaging of the gut microbiome of their prey (De Nys *et al.* 2015). To test whether this could be a major contributor of 16S sequences in our chimpanzee samples, we tested for colobine DNA using two colobine specific PCR systems (Colobinae mt 12S rRNA and Colobinae mt CR: described in Minhós *et al.* 2013; Schubert *et al.* 2015). Six of the 98 chimpanzee samples, all from the East study group, were positive for colobine DNA. However, these samples did not cluster together in the non-metric dimensional scaling nor did they group together in the cluster or maximum parsimony approaches discussed below. Similarly, stored in RNAlater were not found to cluster separately. We therefore included all samples in our main analyses presented here.

### 4.3.5 Behavioral data

From the sooty mangabey study group, we collected behavioral data using one-hour focal follows of all adults and subadults from January 1$^{st}$, 2014 to June 30$^{th}$, 2015. We continuously collected all occurrences of aggressive (1,715 acts) and grooming (2,771 acts) behavior, using an ethogram modified from Range and Noe (2002). For focal individuals, we recorded all approaches to within 1m given or received (4,365 acts). We hypothesized that the behavior

preceding and overlapping with sampling would most strongly impact microbial composition, so for the fecal samples collected between April and June 2014 we analyzed the behavioral data from January to June 2014 and for the fecal samples collected between April and June 2015 we used data from January to June 2015. To quantify dyadic grooming, we used a dyadic grooming index: Grooming(A+B) / (GroomingA+GroomingB - Grooming(A+B)), where Grooming(A+B) is the total time A and B spent grooming one another, GroomingA is the total time spent grooming by individual A and GroomingB is the total time spent grooming by individual B (Wittig & Boesch 2003). Dyadic proximity and aggression indices were recorded as for grooming but we used the number of observations instead of time (Wittig & Boesch 2003). We dropped individuals from a particular analysis if they were not observed doing a behavior at least 20 times (or in the case of grooming, 20 minutes).

### 4.3.6 Statistical analyses

Unless otherwise indicated, statistical analyses were conducted in R version 3.2.3 (R Core Team 2015).

#### *4.3.6.1 Intraspecific examination of beta diversity*

We used the Bray-Curtis (Bray & Curtis 1957) and weighted UniFrac dissimilarity indices (Lozupone *et al.* 2011) to examine pairwise dissimilarity between the bacterial communities of chimpanzee and sooty mangabey microbiomes. Results from the Bray-Curtis dissimilarity index were nearly identical to the UniFrac approach and are presented in the Supplementary Material. The UniFrac dissimilarity index incorporates the phylogeny of the bacterial taxa by calculating the fraction of shared branch length on the phylogenetic tree between samples (Lozupone *et al.* 2011). A weighted implementation of the UniFrac index incorporates the abundance of specific taxa. Because the weighted UniFrac index can be influenced by sampling effort (Lozupone *et al.* 2011), we first rarefied the data to the minimum sampling effort in a given set of samples being compared. UniFrac dissimilarities range from 0 (i.e., all branch length is shared between communities) to 1 (no branch length is shared between communities) and were calculated using the R package *phyloseq* (McMurdie & Holmes 2013). Statistical significance was assessed using Mantel tests and Mantel-like permutations (Sokal & Rohlf 1995). We used data from sooty mangabeys to evaluate the following predictions pertaining to intraspecific variation in gut microbiomes:

1. Microbiomes from the same individual are more similar than microbiomes from different individuals for both sampling periods combined, and for the subsets of 2014 and 2015 data,
2. Microbiomes from the same individual are more similar to one another when collected within the same year compared to when collected in different years,
3. Microbiomes from different sampling periods from the same individual are more similar to one another than when collected from different individuals in different sampling periods,
4. Microbiomes from young individuals (≤3 years of age at time of sampling) are more similar to that of their mothers than to mothers of other offspring,
5. Microbiomes from individuals of any age are more similar to that of their mothers than to other mothers,
6. Microbiomes from the same social group are more similar to one another than to those from different social groups.

For chimpanzees we used a similar approach to test hypotheses 1 and 6. For each of these Mantel tests we employed 1,000 permutations, including the original data as one permutation. To account for non-independence of samples from the same individual, we permuted subject assignments when comparing between, for instance, groups or years. As a test statistic we used the absolute difference between the mean dissimilarities within and between groups, and determined the *P*-value as the proportion of permutations that resulted in a test statistic larger than or equal to that of the original data. To compare the similarity of microbiomes of individuals within and between years, we used a Wilcoxen test (Gehan 1965). We examined the relationship between behavior (grooming, aggression, and proximity) and sooty mangabey microbiomes using Mantel tests using the community dissimilarity matrix and the behavioral matrices (grooming, aggression, and proximity). To estimate significance, we determined the proportion of permutations that resulted in an absolute Spearman correlation greater than or equal to that of the original data. Because of observed differences detected between sampling years, we ran these correlations between community dissimilarity matrices and behavioral matrices, separately for samples collected in 2014 and 2015 and focused on samples from adults and subadults as these individuals were targeted by our behavioral sampling strategy.

*4.3.6.2 Interspecific differences in microorganism abundance*

To investigate whether the presence and abundance of particular bacteria differed between species, we fitted a Generalized Linear Mixed Model with a negative binomial error structure and log link function (McCullagh & Nelder 1996; Baayen 2008), implemented using the function glmer.nb of the R package *lme4* (Bates *et al.* 2015). To maximize model stability, we excluded four NHP species for which there were less than 5 samples and 144 bacterial taxa present in < 50 samples (Figure S4.1 provides sample sizes and read distributions). To control for seasonal and temporal variation, we focused on samples collected from March to June 2014 ($N$ = 289). For each of the remaining OTUs, we built a model with the number of reads assigned to a particular OTU as the response variable. Into these models we included species as a fixed effect and individual and social group as random effects. Variation in sampling effort in the form of sequences generated per sample was included as an offset term representing the total number of reads per sample (log-transformed). Models with a dispersion parameter ≥1.3 were excluded from subsequent analyses. To test for the effect of species, we compared the full model with a null model that lacked the fixed effect of species but included the same random effects structure as the full model (Forstmeier & Schielzeth 2011). We compared the null and full model using a likelihood ratio test (Dobson & Barnett 2008). To assess model stability, we compared estimates from our full dataset excluding levels of random effects one at a time. For 918 of our bacterial OTUs we were able to fit a Generalized Linear Mixed Model with negative binomial error structure. Of these 17 had a dispersion parameter >1.3 and were not examined further. An additional 20 models were excluded due to model instability. We quantified phylogenetic structure in the placement of bacterial OTUs with a significant species effect on the bacterial phylogeny using the D statistic calculated in the R package 'caper' (Orme *et al.*) and compared observed D to that expected under no phylogenetic signal (expected D = 0) and under Brownian Motion (expected D =1) using 100,000 permutations.

*4.3.6.3 Microbiome community structure*

We followed Ochman et al. (2010) to evaluate whether similarity in primate fecal microbiome mirrors the phylogenetic relatedness of the hosts. Working with the rarified dataset so that sampling effort in terms of sequencing effort was equal for all samples, we coded each bacterial OTU as an ordered multistate character based on orders of magnitude of the number of reads

assigned to that taxa (hereafter referred to as the threshold based approach). This character matrix was analyzed using PAUP v4.0b10 and a heuristic maximum parsimony based search with subtree pruning and regrafting (SPR), and 250 bootstrap replicates (Swofford 2002). Tree topology was assessed by examining the mean branch length between clades for each bootstrap replicate. We explored OTU clustering by estimating the best-fitting root with the heuristic residual mean squared function in the program TempEst (Rambaut *et al.* 2016), which minimizes the variance of root-to-tip distances.

As a formal test of the association between host species and position of samples on the maximum parsimony dendrograms, we used the program BaTS (Bayesian Tip-associated Significance testing) v2.0 (Parker *et al.* 2008). BaTS tests for a correlation between a trait state, in this case host species, and topological position, allowing for a comparison to a null distribution generated under the assumption that trait values are independent from topology. We report the monophyletic clade statistic (MC) as a measure of the strength of clustering of each species on the dendrograms (Parker *et al.* 2008). In addition, to test whether there was an association between social group and topological position, we reran the BaTS analysis including only mangabey and chimpanzee samples, and reporting the MC for each social group. We used BaTS to test for an association between individual and topological position, using a subset of the mangabey and chimpanzee data for which individuals were known. To further explore the clustering of NHP microbiomes, we constructed a tree from the pairwise weighted UniFrac distance matrix using hierarchical clustering and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA)(Felsenstein & Felenstein 2004) and SplitsTree4 (Huson & Bryant 2006) and examined the pairwise UniFrac distances between species.

To examine the phylogenetic structure of primate fecal microbiome communities we calculated the mean pairwise phylogenetic distance (MPD) among OTUs and compared these to a null model based on random assembly from a regional species pool (Webb *et al.* 2002). MPD can range from 0 to infinity, with small values indicating phylogenetic clustering of closely related species and large values indicating phylogenetic overdispersion, i.e. assembly of more distantly related species (Webb *et al.* 2002). To contrast community structure among samples, we estimated standardized effect sizes (SES) by subtracting the mean MPD of 1000 communities assembled randomly from the observed MPD, and then dividing by the standard deviation of the 1000 randomizations (Webb *et al.* 2002). Positive SES-MPD values indicate over dispersed

communities wherein communities are composed of species that are less related than expected under the null model, while negative values represent under-dispersion (Webb *et al.* 2002). Traditionally, overdispersion has been interpreted as indicative of competition (competitive displacement of close relatives), whereas underdispersion has been interpreted as evidence for environmental filtering (Webb *et al.* 2002). We examined variation in SES-MPD across samples using a linear mixed-effects model with a Gaussian error structure and host species as a fixed effect and individual and social group as random effects. We again excluded 4 NHP species for which there were less than 5 samples. Models were fitted based on Maximum likelihood and using the function lmer of the R package *lme4* (Bates *et al.* 2015), employing the same model diagnostics and stability tests used for the model testing for bacterial differences between NHP species discussed above.

To examine hierarchical structuring in microbiome communities, i.e., social groups nested within species, mother-offspring pairs nested within groups, and individuals nested within these mother-offspring pairs we evaluated SES-MPD metrics assuming nested species pools. If nesting is strong, phylogenetic clustering should get stronger as the source pool becomes more inclusive and more species are added to the bacterial phylogeny. In contrast, if there is no hierarchical nesting, then increasing the source pool should not impact the strength of phylogenetic clustering. We ran these analyses for both chimpanzees and mangabeys separately. Because chimpanzees uniquely exhibited evidence for phylogenetic overdispersion, we additionally evaluated the mean nearest phylogenetic taxon distance (MNTD) in chimpanzees to explore the phylogenetic depth of structuring. In general, SES-MNTD is more sensitive to differences in closely related taxa whereas SES-MPD is more sensitive to patterns deeper in the phylogeny (Erickson *et al.* 2014). SES-MPD and significance testing were calculated using an implementation in PEZ (Pearse *et al.* 2015), which draws on the Picante R package (Kembel *et al.* 2010).

### 4.4 RESULTS

We found 1107 OTUs present in the gut microbiomes of primates in TNP. Of these, 1076 could be assigned to a phylum, with most belonging to Firmicutes (636) and Bacteroidetes (188), and less to Tenericutes (70), Proteobacteria (55), Actinobacteria (35), Cyanobacteria (25),

Verrucomicrobia (22), and Spirochaetes (17). A few OTUs belonged to Euryarchaeota (9), Lengisphaerae (6) and Elusimicrobia (5) and Fibrobacteres (4), while only a single taxa was assigned to each of WPS-2, Synergistetes, Planctomycetes and Chloroflexi. The core microbiome, here considered as OTUs found in more than 80% of individuals of a species, varied between species, though all had significant proportions of Firmicutes and Bacteroidetes (Fig. 4.1). Of the 1107 OTUs 396 could be assigned to a previously described genus, with *Prevotella* (94), *Ruminococcus* (38), *Oscillospira* (27), *Faecalibacterium* (22), Coprococcus (20), *Blautia* (17), *Bifidobacterium* (14), *Treponema* (13), and RFN20 (13) represented by more than 10 OTUs.

### 4.4.1 Intraspecific beta diversity

Within sooty mangabeys, we found that samples from the same individual were more similar than samples from different individuals ($n_{samples}$ = 229, $n_{individuals}$ = 59, $\bar{x}_{different\ individuals}$ = 0.37, $\bar{x}_{same\ individual}$ = 0.31, Mantel test: permutation $P < 0.001$; Fig. 4.2A). This effect was more pronounced when years were analyzed separately ($n_{samples}$ = 154, $n_{individuals}$ = 58, $\bar{x}_{different\ individuals}$ = 0.376, $\bar{x}_{same\ individual}$ =0.273, Mantel test: permutation $P < 0.001$; $n_{samples}$ = 74, $n_{individuals}$ = 27, $\bar{x}_{different\ individuals}$ = 0.33, $\bar{x}_{same\ individual}$ = 0.29, Mantel test: permutation $P < 0.001$; for the 2014 and 2015 samples respectively). For the 26 individuals sampled in both years, samples from the same individual were more similar within years than from the same individual between years, suggesting turnover in gut bacterial community composition ($n_{samples}$ =146, $\bar{x}_{between\ sampling\ year}$ = 0.35, $\bar{x}_{within\ sampling\ years}$ = 0.28; Wilcoxon test, $T^{+}$ = 348, $N$=26, $P < 0.001$; Fig. 4.2C). However, the microbiomes of samples collected between years were more similar when they came from the same individual than different individuals, suggesting individual differences in microbiome persisted across years ($n_{samples}$ = 229, $n_{individuals}$ = 59, $\bar{x}_{different\ individuals\ different\ years}$ = 0.38, $\bar{x}_{same\ individual\ different\ years}$ = 0.35, Mantel test: permutation $P < 0.001$; Fig. 4.2B).

Within sooty mangabeys, we found that both familial relationships and group membership impacted the gut microbiome. Samples from the same study group were more similar to one another than samples from different groups ($n_{samples}$ = 258, $n_{individuals}$ = 87, $\bar{x}_{different\ group}$ = 0.400, $\bar{x}_{same\ group}$ = 0.37, Mantel test: permutation $P$ = 0.022). Bacterial communities from mother-offspring pairs were more similar to one another than to those from offspring and non-mothers ($n_{samples\ =}$ 117, $n_{mother\text{-}offspring\ pairs}$ = 18, $\bar{x}_{non\text{-}mother\text{-}offspring\ pair}$ = 0.37, $\bar{x}_{mother\text{-}offspring\ pair}$ = 0.33, Mantel

test: permutation $P = 0.046$; Fig. 4.2D), though we did not detect a significant difference when we included offspring older than 3 years of age at the time of sampling ($n_{samples} = 136$, $n_{mother\text{-}offspring\ pairs} = 22$, $\bar{x}_{mother\text{-}offspring\ pair} = 0.36$, $\bar{x}_{non\text{-}mother\text{-}offspring\ pair} = 0.38$, Mantel test: permutation $P = 0.22$). However, there was no correlation between dyadic frequencies of grooming, aggression, proximity or co-occurrence in a subgroup and dissimilarity in bacterial community composition (Table 4.1).

Chimpanzees showed similar trends to mangabeys: samples from the same individual were more similar to one another than samples from different individuals ($n_{samples} = 47$, $n_{individuals} = 24$, $\bar{x}_{different\ individuals} = 0.41$, $\bar{x}_{same\ individual} = 0.30$, Mantel test: permutation $P < 0.001$, Fig. 4.2E; $n_{samples} = 23$, $n_{individuals} = 12$, $\bar{x}_{different\ individuals} = 0.48$, $\bar{x}_{same\ individual} = 0.36$, Mantel test: permutation $P < 0.001$ Fig. 4.2F; South and North group respectively), and samples from the three study groups were more similar within groups than between groups ($n_{samples} = 98$, $n_{individuals} = 64$, $\bar{x}_{UniFrac\ different\ group} = 0.\ 44$, $\bar{x}_{UniFrac\ same\ group} = 0.41$, Mantel test: permutation $P < 0.001$).

The BaTS analysis of the trees built using bacterial abundance, including only chimpanzees or sooty mangabeys, found that social group membership and individual identity were significantly structured (Table 4.3), supporting the findings of the UniFrac based analyses described above.

### 4.4.2 Interspecific differences in microbiomes

There was overwhelming support for the importance of host species on the abundance of bacterial taxa ($n_{samples} = 375$, $n_{individuals} = 160$, $n_{primate\ species} = 5$, likelihood ratio test comparing full and null model, for 853 out of 881 models, $P < 0.05$). For 49 bacterial taxa, both chimpanzees and red colobus had effect size estimates greater than zero, when compared to the sooty mangabey reference level, indicating those OTUs were more abundant in both chimpanzees and their red colobus prey. There appeared to be no phylogenetic pattern in the placement of these significant bacterial taxa on the bacterial phylogeny when compared to randomizations ($D = 0.85$, $P_{no\ phylogenetic\ structure} = 0.16$, $P_{Brownian} < 0.001$; Figure S4.3A+B).

We found strong evidence for distinct chimpanzee, colobine, and cercopithecine clades (Fig. 4.3, Table 4.2), with the chimpanzee clade clustering more closely with the colobine clade, a relationship also supported by the shorter average branch length distance between samples from chimpanzees and colobines versus the distance from either to the cercopithecine clade (Fig. 4.4). The BaTS analysis of the threshold-based maximum parsimony phylogeny revealed strong

phylogenetic structure of samples on host species (Table 4.2), but only a single of 100 bootstrap replicates supported a chimpanzee outgroup, a relationship that would be predicted by the primate phylogeny. The remaining 99 bootstrap replicates were consistent with a chimpanzee-colobine clade, though for 6 of these, one colobine sample was placed in the cercopithecine clade and for four others, one sooty mangabey sample jumped into the colobine clade. Nearly identical results were observed for the presence-absence based maximum parsimony analysis (see supplementary results for details).

UPGMA clustering supported the grouping of bacterial OTUs by primate host, though samples from several of the cercopithecine species jumped around within the sooty mangabey clade (Fig. 4.5) and one chimpanzee sample was found closer to the colobines than to other chimpanzee samples. A similar pattern was found with the Splitstree network, with several cercopithecine samples jumping into the sooty mangabey clade and long chimpanzee and colobine branches (Fig. S4.9). While there was a great deal of variation within species, UniFrac dissimilarity scores indicate that sooty mangabeys have similar microbiomes to other cercopithecines, reflecting functionally similar communities that mirror their close phylogenetic relationships, but are distinct from both chimpanzees and colobines (Fig. S4.6).

The microbiomes of the majority of samples exhibited strong phylogenetic structure, with phylogenetic clustering dominant, with the exception of chimpanzees, for which phylogenetic overdispersion was evident (Table 4.4 and S4.9, Fig. 4.6; likelihood ratio test comparing full and null model: $\chi^2$= 34.34, df=4, $P$<0.001). Sooty mangabeys additionally showed a signal indicative of nesting, where strength of clustering increased with the scale of the source pool (Fig. 4.7A), such that sooty mangabeys had a subset of the bacteria present within all primates, sooty mangabey groups had a subset of the bacteria within the species, mother-offspring pairs had a subset of those within the group, and individuals a subset of the bacteria within their mother-offspring pair. In contrast, for chimpanzees, increasing the source pool does not expand the phylogenetic breadth of the gut bacterial community, thus individuals and groups tend to sample from overdispersed bacterial lineages (Fig. 4.7B). If anything, phylogenetic overdispersion appeared to increase as we expanded the source pool, suggesting these broader scales intersperse OTUs evenly across the bacterial phylogeny rather than introduce phylogenetically distinct bacterial clades. By contrasting metrics that capture structuring deeper in the tree (SES-MPD) with those that capture structure towards the tips of the phylogeny (SES-MNTD) we additionally

showed that phylogenetic clustering at the tips is much weaker (Fig. S4.8). The chimpanzee microbiome thus appears to consist of overdispersed clusters of bacterial OTUs. One explanation for this pattern would be if the chimpanzee microbiome represents a composite of bacterial clades, a subset of which with extra-chimpanzee origin that were subsequently integrated into the chimpanzee microbiome and passed between individuals.

## 4.5 DISCUSSION

We present a comparative study of the gut microbiome among a diverse community of sympatric, wild, NHPs that includes a hunter-prey relationship. Our results support previous work suggesting that sympatric wild NHPs have individually distinct gut microbiomes (McCord *et al.* 2014; Fogel 2015). In addition, we found evidence that both sooty mangabey and chimpanzee social groups maintain a 'pan-microbiome' of shared microorganisms and that mother-offspring transmission of the microbiome in sooty mangabeys is important during early years in life before individuals develop their own distinct microbiomes. However, in contrast to findings from studies of baboons (Tung *et al.* 2015), within mangabeys we found no evidence for close social partners having more similar gut microbiomes. Further, we observed incongruence between microbial communities and NHP evolutionary relationships, which contrasts with Ochman et al. (2010) who reported that the evolutionary relationships of the wild great apes was recapitulated by their fecal microbial communities. Contrary to phylogenetic expectations, we found that the gut microbiome of chimpanzees clusters more closely with colobines. Uniquely, chimpanzee microbiomes were also characterized by strong overdispersion, indicative of exposure to a broader diversity of primate gut bacteria. We suggest these patterns might be explained by hunting and consuming of colobine prey by chimpanzees. Going beyond transmission of specific pathogens (Eppinger *et al.* 2006), our study suggests assimilation of the microbiome from one species into the microbiome of another through hunting.

Healthy humans exhibit high intra-individual stability in their gut microbiomes (Faith *et al.* 2013); in wildlife settings, NHP intra-individual stability in the gut microbiome varies by species and habitat. For example, samples from yellow baboons (*Papio cynocephalus*) collected a few days apart were as different from each other as samples collected over 10 years apart (Ren *et al.* 2015). Similarly, repeated sampling of rufous mouse lemurs (*Microcebus rufus*) found high

intra-individual variation between years (Aivelo *et al.* 2016), whereas western lowland gorillas (*Gorilla gorilla gorilla*)(Moeller *et al.* 2015), eastern chimpanzees (*Pan troglodytes schweinfurthii*)(Degnan *et al.* 2012), and black howler monkeys (*Alouatta pigra*)(Amato *et al.* 2013) appear to have more stable microbiome communities, exhibiting much lower temporal turnover. We show that sooty mangabeys and chimpanzees in TNP exhibit individually distinct microbiomes that persisted through time, but which differed by social group. While habitat and diet are both likely to influence the stability of the gut microbiome, especially when bacteria are dispersed via the environment, social interactions might also be important. The sociality of chimpanzees and sooty mangabeys appeared to create opportunities for transmission of the gut microbiome between individuals, and ultimately a 'pan-microbiome' shared by a social group, likely enhancing the stability of microbiomes of individuals within groups (Leibold *et al.* 2004).

Collectively our results suggest that dispersal limitation has a major role in driving microbiome community in this ecosystem. In humans, transmission of the microbiome from mothers to their infants during vaginal birth shapes an individual's microbiome (Dominguez-Bello *et al.* 2010; Funkhouser & Bordenstein 2013). Our findings suggest a similar signal in wild primates, with young sooty mangabeys having more similar microbiomes to their mothers, though this similarity was no longer detected in older individuals. In baboons and chimpanzees, the microbiome appears to be transmitted in part by an individual's sociality later in life. More social chimpanzees exhibit more diverse microbiomes (Moeller *et al.* 2016b) and in baboons, close social partners exhibited more similar microbiomes (Tung *et al.* 2015). To our surprise, we found no evidence that close social partners were more similar in their gut microbiomes within mangabeys, though social group mates did share more similar microbiomes. One potential factor driving this difference with baboons is that baboon microbiomes exhibit extremely high turnover, on the order of days (Ren *et al.* 2015), and these labile microbiomes might be more susceptible to colonization by novel bacteria from close social partners. In contrast, we found that sooty mangabey microbiomes were stable over longer periods, at least a year, and perhaps this innate stability renders social interactions less important for bacterial colonization.

The different community phylogenetic structures observed in chimpanzees and monkeys in the ecosystem at TNP suggest that very different fundamental processes are shaping the assembly of their microbial communities. We suggest this is likely due to the ecological process of dispersal limitation, particularly the increased exposure to a broader aspect of the NHP bacterial

community by hunting. The majority of the NHP bacterial microbiomes we examined exhibited significant phylogenetic structure, with phylogenetic clustering dominant, wherein closely related bacteria were more likely to co-occur than expected by chance. This is a pattern that has been observed at broad scales across diverse communities, from plants (Kembel & Hubbell 2006) to freshwater bacteria (Horner-Devine & Bohannan 2006), and has frequently been interpreted as evidence for environmental filtering (Webb *et al.* 2002), but might also reflect local evolutionary radiations, in this case bacterial radiation within host guts (Horner-Devine & Bohannan 2006). The phylogenetic overdispersion observed in chimpanzee gut microbiomes was thus not expected. Phylogenetic overdispersion has sometimes been suggested to be indicative of competition – the competitive displacement of closely related species – and might therefore indicate stronger interspecific competition between bacterial taxa in chimpanzees than monkeys. However, we found that patterns of overdispersion were strongest deeper in the bacterial phylogeny, and weaker towards the tips of the tree, where we might predict competition would be most strong. The broad taxonomic dispersion of the chimpanzee gut microbiome is consistent with high exposure to diverse bacterial clades. We suggest that the hunting of other NHPs by chimpanzees may provide one route by which chimpanzees have been exposed to such high bacterial diversity.

Humans are hunters of NHPs in many ecosystems, which is thought to have facilitated the cross-species transmission of some pathogens from NHPs to humans (e.g., HIV-1 and 2 (Erickson *et al.* 2014)); whether this close ecological relationship also impacts the gut microbiomes of human hunters is an important area of future research. Hunting and scavenging likely played a major role in hominin evolution (Domínguez-Rodrigo 2002), though to our knowledge no studies have explicitly addressed how this transition in diet impacted our gut microbiomes. Studies of contemporary Hadza hunter-gathers have suggested these human communities have more diverse microbiomes than Europeans or North Americans (Schnorr *et al.* 2014), which could be due to various environmental factors, including contact with NHPs during hunting. For example, sex differences in the microbiome observed in Hadza populations could represent differential exposure through hunting (Schnorr *et al.* 2014). While we are largely ignorant about determinants of gut microbiome communities in humans, particularly in hunter-gather societies, there is increasing awareness that they are critically important for health (Hooper *et al.* 2012). Our results suggest dispersal limitation likely played a major ecological role in the function and

evolution our microbiomes and highlights the need for in depth studies of humans and NHPs in complex ecological ecosystems.

## 4.6 Conclusion

We present a comprehensive analysis of contemporaneous microbiomes of all sympatric diurnal NHP species in an ecosystem including a known predator-prey relationship. Despite sharing an environment and being exposed to largely the same bacterial source pool, we show that host species maintain distinct microbiomes. These findings lend support to the idea that genetic factors and differential exposure from conspecifics might be important in influencing the gut microbiome of NHPs. We find that individuals have persistent microbiomes that are influenced by social group and maternity, but surprisingly not grooming and proximity. Contrary to phylogenetic expectations, gut microbiomes of chimpanzees cluster closer with their colobine prey, a finding even more the remarkable given the highly specialized diet of the latter which has been accompanied by physiological and microbial adaptations that enable foregut fermentation. In addition, and in contrast to strong signals suggesting phylogenetic clustering in the microbiomes of all monkeys, chimpanzee microbiomes exhibit phylogenetic overdispersion, suggesting a unique ecological process drives their community assembly. Our analyses indicate that chimpanzees have assimilated bacteria from the gut microbiomes of their NHP prey into their own microbiomes.

## 4.7 Acknowledgements

This article represents a chapter in the doctoral dissertation of JFG, and has thus benefited greatly from the input of his supervisor committee, David Marcogliese, Charles Nunn, and Louis Lefebvre. We thank the Ivorian Ministry of Environment and Forests, the Ministry of Research, the directorship of the Taï National Park, the Office Ivoirien des Parcs et Réserves, the Centre Suisse de Recherche Scientifique, the Taï Chimpanzee Project and their teams of field assistants for their support. JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336), the Canadian Institutes of Health Research's Strategic Training Initiative in Health Research's Systems Biology Training Program, an NSERC Vanier Canada Graduate Scholarship (CGS),

and a long-term Research Grant from the German Academic Exchange Service (DAAD-91525837-57048249). Core-funding for the Taï Chimpanzee Project is provided by the Max Planck Society. For assistance in the field we thank Ariane Düx, Vera Leinert, Bekah Myers, Jonathan Müller-Tiburtius, Sylvain Lemoine, and especially the long-term mangabey field assistants Simon T. Kannieu, Daniel, and Gnimion for the dedication to the project. For assistance and support with sequencing at the University of Connecticut we are grateful to Kendra Maas. For their support in the laboratory, JFG thanks Ulla Thiesen, Kevin Merkel, Andreas Sachse, and the rest of the P3 group at the Robert Koch Institute. JFG thanks James Herrera and Will Pearse for discussions and generously sharing R code for phylogenetic analyses.

**Table 4.1: Spearman rank correlations between behavioral association and UniFrac bacterial community dissimilarity for sooty mangabey fecal microbiomes**

| Behavior | 2014 | | 2015 | |
|---|---|---|---|---|
| | $r_{sb}$ | *P* | $r_{sb}$ | *P* |
| Grooming | -0.056 | 0.814 | -0.015 | 0.609 |
| Aggression | -0.014 | 0.599 | -0.003 | 0.491 |
| Proximity | -0.045 | 0.781 | 0.012 | 0.407 |

**Table 4.2: Results of BaTS testing of host species relationship with phylogenetic position, using threshold based maximum parsimony phylogeny**

| Test statistic | Host species | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| AI | | 0.732 | 21.066 | 19.024 | 22.927 | **<0.001** |
| PS | | 9.000 | 109.780 | 104.000 | 115.000 | **<0.001** |
| MC | *Cercocebus atys* | 84.000 | 7.880 | 5.000 | 13.000 | **0.010** |
| MC | *Procolobus verus* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | *Colobus polykomos* | 6.000 | 1.010 | 1.000 | 1.000 | **0.010** |
| MC | *Procolobus.badius* | 1.000 | 1.030 | 1.000 | 1.000 | 1.000 |
| MC | *Pan troglodytes verus* | 98.000 | 2.910 | 2.000 | 4.000 | **0.010** |
| MC | *Cercopithecus diana* | 2.000 | 1.050 | 1.000 | 1.000 | 0.060 |
| MC | *Cercopithecus campbelli* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | *Cercopithecus petaurista* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | *Cercopithecus nictitans* | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Table 4.3: Results of BaTS testing for a correlation between group membership, individual and phylogenetic position, using threshold based maximum parsimony phylogeny**

| Test statistic | Social group | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| *Cercocebus atys* | | | | | | |
| Group: AI | | 0.758 | 5.966 | 4.536 | 7.208 | **<0.001** |
| Group: PS | | 11.000 | 26.520 | 25.000 | 28.000 | **<0.001** |
| MC | Audrenissrou | 39.000 | 17.890 | 11.000 | 33.000 | **0.020** |
| MC | Neighbor | 6.000 | 1.870 | 1.000 | 3.000 | **0.010** |
| Individual: AI | | 12.708 | 29.210 | 28.640 | 29.723 | **<0.001** |
| Individual: PS | | 132.000 | 205.040 | 201.000 | 209.000 | **<0.001** |
| *Pan troglodytes verus* | | | | | | |
| Group: AI | | 2.318 | 7.172 | 5.816 | 8.492 | **<0.001** |
| Group: PS | | 27.000 | 42.980 | 38.000 | 47.000 | **<0.001** |
| MC | East | 4.000 | 2.200 | 2.000 | 4.000 | 0.080 |
| MC | North | 6.000 | 2.080 | 1.000 | 4.000 | **0.010** |
| MC | South | 5.000 | 3.410 | 2.000 | 6.000 | 0.160 |
| Individual: AI | | 8.987 | 11.512 | 11.215 | 11.618 | **<0.001** |
| Individual: PS | | 81.000 | 90.110 | 88.000 | 92.000 | **<0.001** |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Table 4.4: Number of samples by species showing phylogenetic clumping or overdispersion based on SES-MPD**

| | *Cercocebus atys* | *Cercopithecus nictitans* | *Cercopithecus petaurista* | *Cercopithecus campbelli* | *Cercopithecus diana* | *Colobus polykomos* | *Pan troglodytes verus* | *Procolobus badius* | *Procolobus verus* |
|---|---|---|---|---|---|---|---|---|---|
| Phylogenetic clumping ($P < 0.025$) | 232 | 2 | 1 | 2 | 8 | 6 | 1 | 5 | 1 |
| Phylogenetic overdispersion ($P > 0.975$) | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 0 | 0 |
| Non-significant ($0.025 < P < 0.975$) | 25 | 0 | 0 | 0 | 0 | 0 | 61 | 0 | 0 |

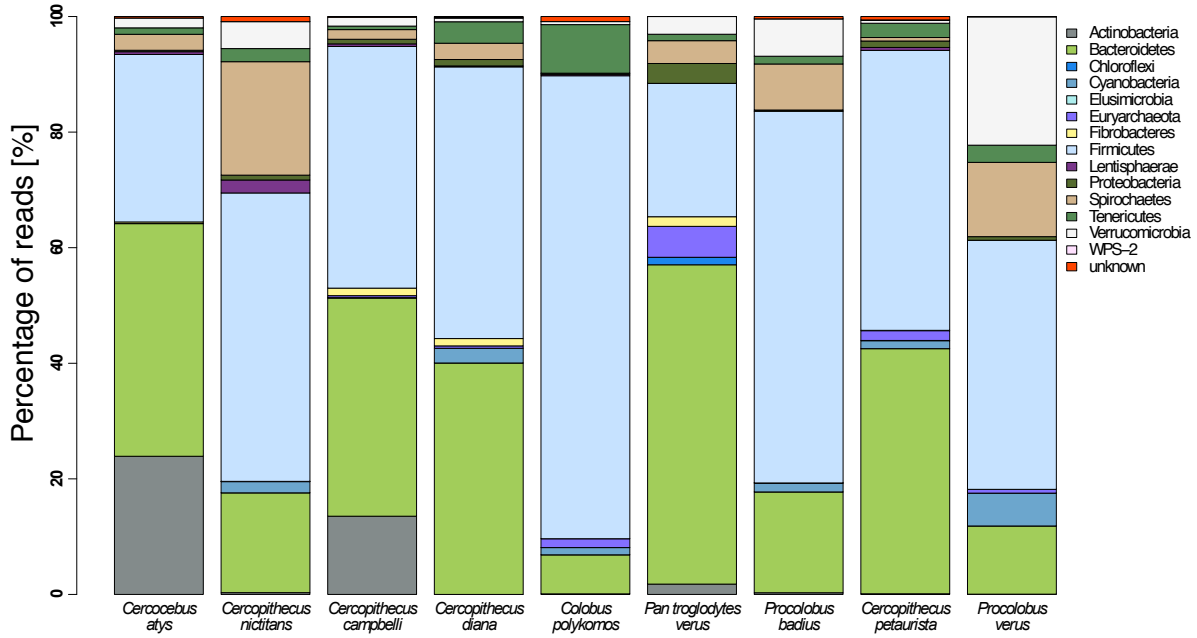
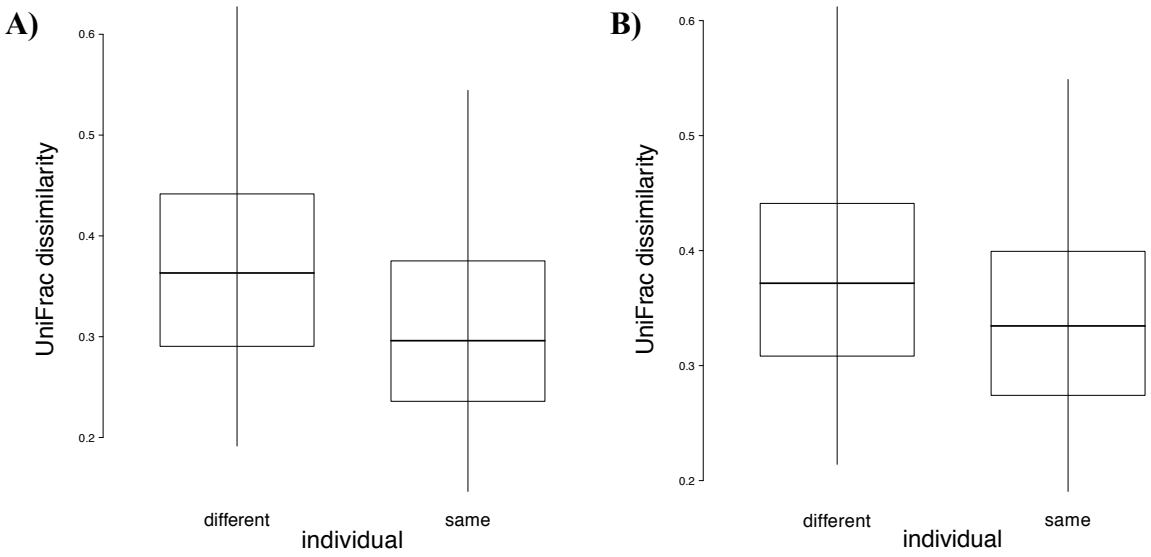
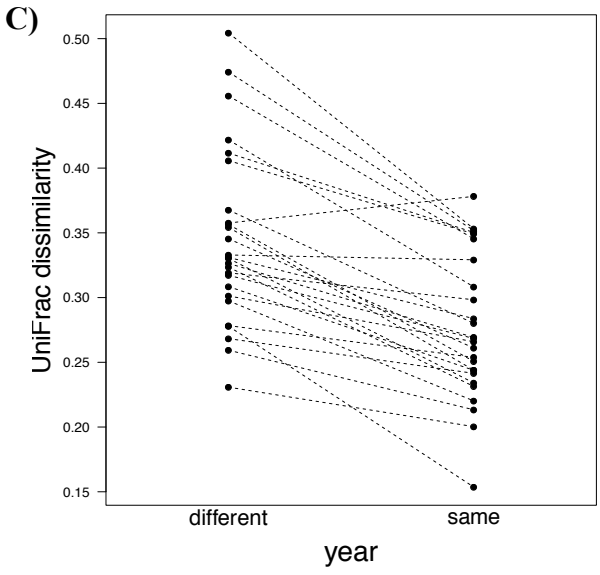
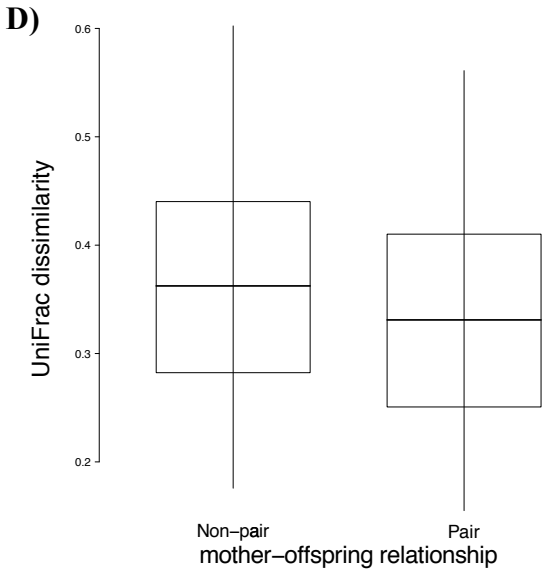


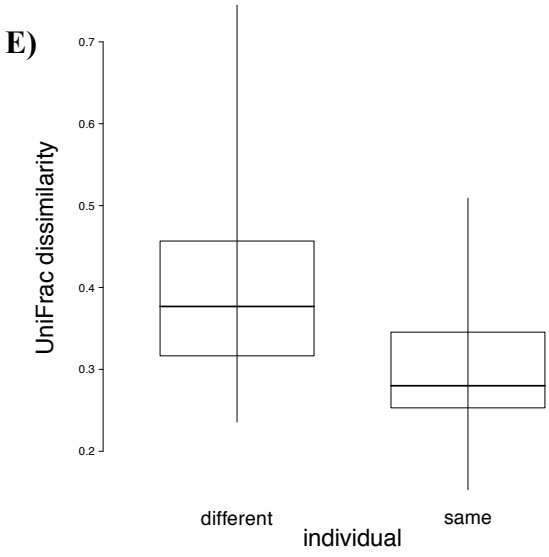

**Figure 4.1: Abundance of core taxa (i.e., those shared by at least 80% of individuals in a species) for each species, separated by phylum.**


A)
UniFrac dissimilarity
0.6
0.5
0.4
0.3
0.2
different
same
individual


B)
UniFrac dissimilarity
0.6
0.5
0.4
0.3
0.2
different
same
individual


C)
UniFrac dissimilarity
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
different
same
year


D)
UniFrac dissimilarity
0.6
0.5
0.4
0.3
0.2
Non−pair
Pair
mother−offspring relationship


E)
UniFrac dissimilarity
0.7
0.6
0.5
0.4
0.3
0.2
different
same
individual

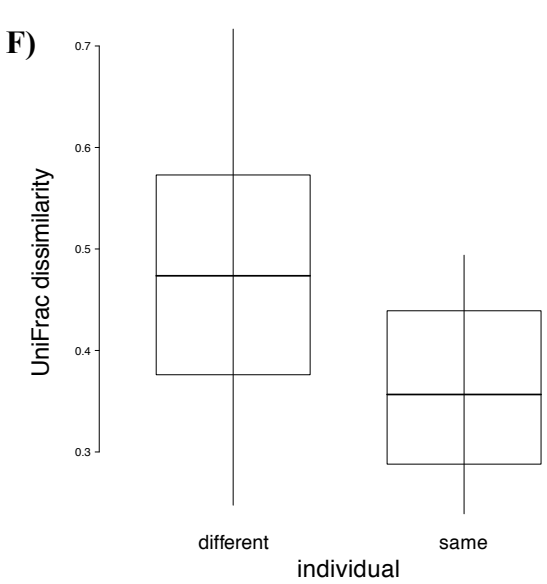

F)
UniFrac dissimilarity
0.7
0.6
0.5
0.4
0.3
different
same
individual

**Figure 4.2: Comparison of the UniFrac dissimilarities between samples** from sooty mangebeys in the Audrenissrou group, when stemming from the same or different individuals (**A**), between samples from different sampling years when stemming from the same or different individuals (**B**), for the same individuals sampled in the same versus different sampling years (**C**), when stemming from a mother offspring-pair or between a non-mother-offspring pair (**D**), between samples from chimpanzees in the South group (**E**) and North group (**F**) when stemming from the same or different individuals. The middle horizontal line represent the median while the rectangle shows the quartiles and the vertical line represents the 2.5 and 97.5% percentiles. Dashed lines in (**C**) indicate the paired nature of the dataset, connecting the dissimilarity for samples from each individual from the same or different sampling years.

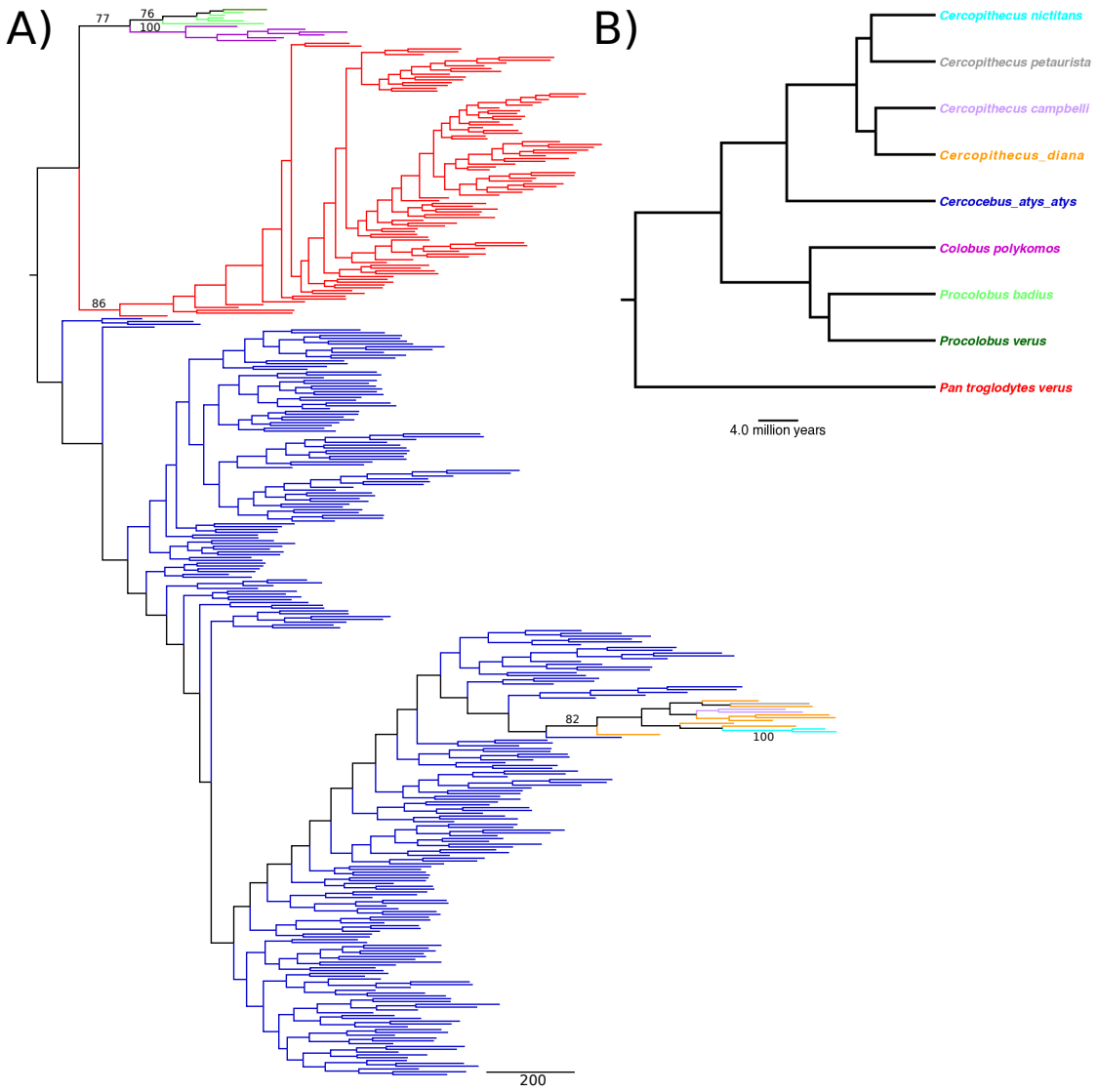



**Figure 4.3: Maximum parsimony phylogeny of samples colored by host species: A)** Phylogeny of samples estimated with a heuristic maximum parsimony using the threshold based scores of bacterial abundance as characters. Tips indicate a particular sample, while colors indicate the host's species, with black branches indicate branches not specific to a particular host. Bootstrap support is shown for the species clades and the root position was supported by our TempEst analysis. Some host species are represented by only a few samples; however, we do not find any evidence that sample size influenced our results. Host species largely group together and hosts with few samples do not appear to be unusual or driving the patterns observed. Metadata on host individual identity and social group are in the supplementary material. **B)** Phylogeny of the primate hosts based on eleven mitochondrial and six autosomal genes made available through the 10kTrees project (Arnold *et al.* 2010).

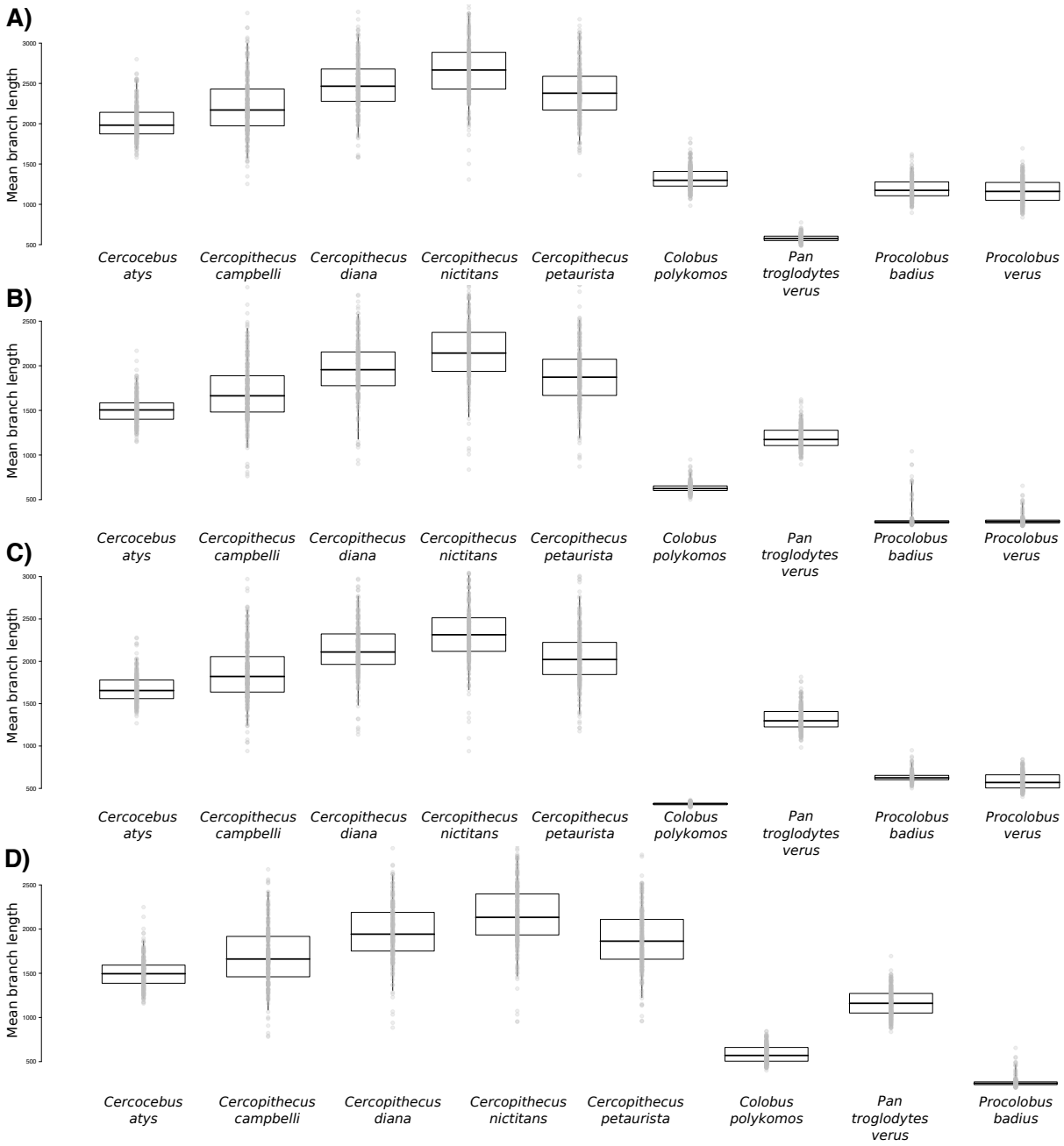



**Figure 4.4: Mean branch lengths between species for threshold based maximum parsimony analysis.** Based on comparisons of mean branch lengths of all samples to samples of **A)** chimpanzees, **B)** red colobus, **C)** king colobus, and **D)** olive colobus.

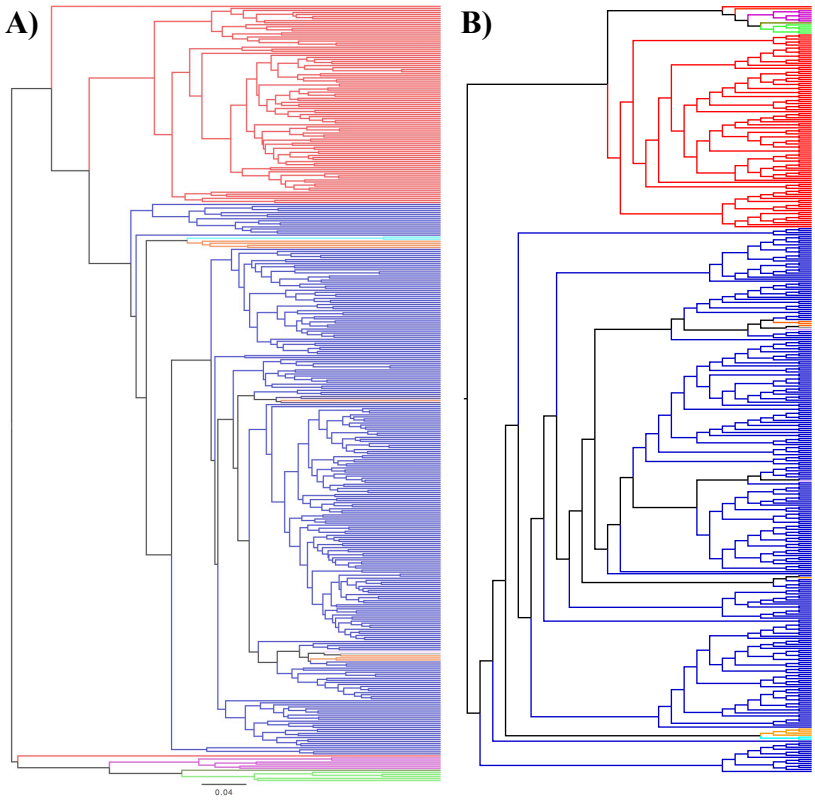



**Figure 4.5: A) UPGMA hierarchical clustering of samples based on UniFrac dissimilarity matrix**. Colors indicate the host's species as in Fig. 4.3B. **B)** A cladogram of the UPGMA clustering, using the root placement estimated from the TempEst analysis shown in Fig. 4.3, highlighting the congruence of these two analyses.

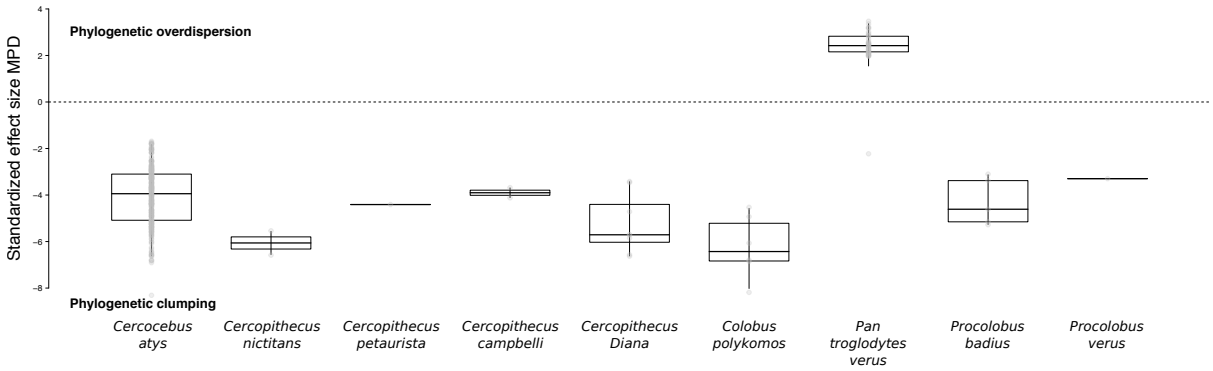



**Figure 4.6: Standard effect size of mean phylogenetic distance based on null model simulations of the bacterial community in each fecal sample, separated by host species.** The solid middle horizontal line of the rectangles represents the median, the rectangle shows the quartiles and the vertical line represents the 2.5 and 97.5% percentiles, while the values for each sample are indicated by overlapping gray circles. Values above the dashed line are those exhibiting phylogenetic overdispersion while those below the line exhibit phylogenetic clumping.

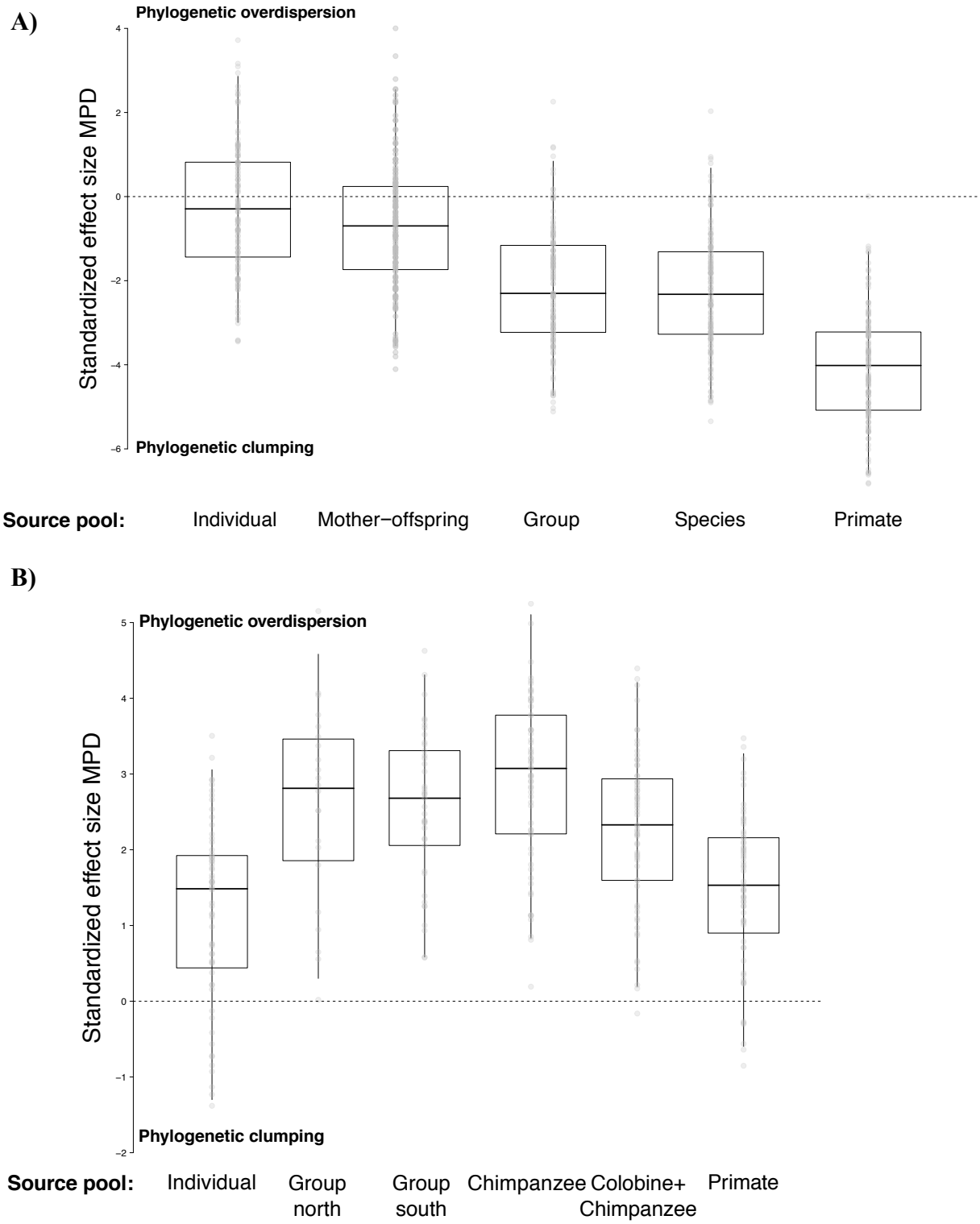



**Figure 4.7: Standard effect size of mean phylogenetic distance based on null model simulations of the bacterial community in each fecal sample using different bacterial source pools. A)** For sooty mangabeys for each individual that was included in a mother offspring pair, we used a source pool including only the bacteria found in any sample of a repeatedly sampled individual. We also considered a mother-offspring pool, using a source pool of only the bacteria found in any sample of the respective mother-offspring pair. For the

group pool we used a source pool of only the bacteria found in the social group, for the species, using a source pool of only the bacteria found in any sooty mangabey sample, and for the primate pool a source pool of all bacteria found in this study. **B)** For chimpanzees we ran a similar analysis; we did not include mother-offspring pairs and rather ran the group analysis for each of the two groups for which we had repeated sampling of individuals. In addition we included a source pool level that consisted of bacteria found in any colobine or chimpanzee sample. The solid middle horizontal line of the rectangles represents the median, the rectangle shows the quartiles and the vertical line represents the 2.5 and 97.5% percentiles, while the values for each sample are indicated by overlapping gray circles. Values above the dashed line are those exhibiting phylogenetic overdispersion while those below the line exhibit phylogenetic clumping.

# LINKING STATEMENT 3

In the last chapter, I discussed insights gained into the microbiomes of wild non-human primates (NHPs), showing that social groups form units with similar communities, and that the hunting behaviour of chimpanzees might be shaping the composition of these complex microbial communities. In my appendix chapter, I examined within- and cross-species transmission of primate retroviruses (including towards local human populations) and the importance of virus-host interactions in determining cross-species transmission risk. I discuss how retroviruses ecology and evolution might be changing in a shifting environment. These chapters serve as a stark reminder that, because of their evolutionary proximity to humans, NHPs can serve as reservoirs for human pathogens and symbionts. These results also highlight the need for methods to examine the diversity and evolution of other non-retrovirus NHP parasites in face of changing environments. In Chapter five I develop a new molecular tool for studying such wildlife pathogens. Many spirochete bacterial taxa are difficult to study, as they can be extremely difficult to culture. *Treponema pallidum* infections causing yaws disease and venereal syphilis are globally widespread in human populations, infecting hundreds of thousands and millions annually respectively. Though such treponemes have been shown to infect baboons, evidence in other NHP species is lacking, and despite millions of people being infected annually, only ten genomes of *Treponema pallidum* infecting humans have been published to date. I develop a hybridization capture enrichment approach and demonstrate its strength by generating genomic data from asymptomatic bones from non-human primates. Results suggest that similar to the retroviruses and bacterial transmitted between hunter and prey discussed in Chapter 5 and the Appendix Chapter, *Treponema pallidum* infections occur in both chimpanzees and red colobus. I discuss potential transmission modes for these pathogens and future avenues of research opened by this novel approach.

# CHAPTER 5

## TOOLS FOR OPENING NEW CHAPTERS IN THE BOOK OF *TREPONEMA PALLIDUM* EVOLUTIONARY HISTORY

Jan F. Gogarten[1,2,3], Ariane Düx[1, 4], Verena J. Schuenemann[5], Kathrin Nowak[1], Christophe Boesch[2], Roman M. Wittig[2,6], Johannes Krause[5,7], Sébastien Calvignac-Spencer[1,4*], and Fabian H. Leendertz[1*]

A version of this chapter is published in *Clinical Microbiology and Infection*, Volume 22, pp. 916-921.

*Authors contributed equally to this work.

[1]Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany.

[2]Primatology Department, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

[3]Department of Biology, McGill University, Montreal, Quebec, Canada.

[4]Viral Evolution, Robert Koch Institute, Berlin, Germany.

[5]Institute for Archeological Sciences, University of Tübingen, Tübingen, Germany.

[6]Taï Chimpanzee Project, CSRS, Abidjan, Ivory Coast.

[7]Max Planck Institute for the Science of Human History, Jena, Germany.

## 5.1 ABSTRACT

*Treponema pallidum* infections causing yaws disease and venereal syphilis are globally widespread in human populations, infecting hundreds of thousands and millions annually respectively; endemic syphilis is much less common and pinta has not been observed in decades. We discuss controversy surrounding the origin, evolution, and history of these pathogens in light of available molecular and anthropological evidence. These bacteria (or close relatives) seem to affect many wild African non-human primate species, though to date only a single non-human primate *Treponema pallidum* genome has been published, hindering detection of spillover events and our understanding of potential wildlife reservoirs. Similarly, only ten genomes of *Treponema pallidum* infecting humans have been published, impeding a full understanding of their diversity and evolutionary history. Research efforts have been hampered by the difficulty of culturing and propagating *Treponema pallidum.* Here we highlight avenues of research recently opened by the coupling of hybridisation capture and next generation sequencing. We present data generated with such an approach suggesting that asymptomatic bones from non-human primates occasionally contain enough treponemal DNA to recover large fractions of their genomes. We expect that these methods, which naturally can be applied to modern biopsies and ancient human bones, will soon considerably improve our understanding of these enigmatic pathogens and lay rest to old, yet unresolved controversies.

## 5.2 THE GENUS *TREPONEMA*

Organisms in the genus *Treponema* (phylum Spirochaetes, order Spirochetales, family Spirochaetaceae) are obligate parasites distributed across a broad range of animal hosts, though a few basal species may represent free living organisms (Norris *et al.* 2006). Treponemes have been detected across many of their host's body compartments: from the oral cavity (Griffen *et al.* 2012) to the hindgut (Bittar *et al.* 2014), and from the skin to the cartilage and bone (Norris *et al.* 2006). *Treponema* species can be pathogenic or non-pathogenic, with suggestions they may play a symbiotic role in some hosts. For example *Treponema* species found in termite guts perform $H_2$–$CO_2$ acetogenesis and nitrogen fixation, processes releasing carbon and energy or providing nitrogen for their host (Graber *et al.* 2004). Many *Treponema* species are abundant in gut microbiomes of humans and non-human primates, though their functional role in these communities remains unknown (Bittar *et al.* 2014; Schnorr *et al.* 2014). The *Treponema* genus is however best known for its pathogenic members, which are responsible for a large current and historic human global disease burden.

## 5.3 PATHOGENIC *TREPONEMA PALLIDUM*

Subspecies of the spirochete bacterium *Treponema pallidum* (*T. pallidum*) are responsible for yaws (*T. p. pertenue*), Bejel or endemic syphilis (*T. p. endemicum*), and venereal syphilis (*T. p. pallidum*). To date, the causative agent of pinta (currently classified as *T. carateum)* has not been cultured or isolated; this has precluded a determination of whether it represents a further member of the *T. pallidum* species or is divergent enough to warrant designation as its own species (Giacani & Lukehart 2014). Pinta disease is not discussed here in detail.

The clinical presentations associated with treponemal diseases share similarities but are still distinctive. Yaws is characterized by a primary cutaneous lesion, most often presenting on lower extremities, with secondary lesions developing across the body (Giacani & Lukehart 2014). In contrast, primary lesions of endemic syphilis infections are rarely observed, and when observed, are typically found in oral mucosa (Giacani & Lukehart 2014). Similarly, primary lesions of venereal syphilis are typically present on the genital, anal, or oral mucosae. All *T. pallidum* infections can become latent, hampering treatment efforts. Left untreated, endemic syphilis and yaws can lead to destructive osteitis of the nose, palate and nasal septum, with yaws often causing lesions on the feet (Giacani & Lukehart 2014). In contrast, late-stage venereal syphilis is more systemic, causing major neurological and

cardiovascular problems and growth of granuloma on many organs (Norris *et al.* 2006). Venereal syphilis can cause adverse outcomes for pregnant women including stillbirth, early fetal death, low birth weight, preterm delivery, neonatal death, and infection of the newborn (World Health Organization 2012). It was long thought that only venereal syphilis is involved with the central nervous system, though some evidence suggests this also occurs rarely with yaws and that yaws may also be transmitted vertically in some rare cases (Román & Román 1986; Giacani & Lukehart 2014). All *T. pallidum* are predominantly transmitted by direct contact with infectious lesions; yaws mostly by skin-to-skin contact, while endemic syphilis is also transmitted by contact with mucous membranes (Giacani & Lukehart 2014). In contrast, venereal syphilis is predominantly transmitted by sexual contact, though non-sexual transmission has been described (Krivatkin & Krivatkina 1997) and vertical transmission is a major concern (Newman *et al.* 2013). Consistent with their different transmission modes, yaws and endemic syphilis predominantly infect children 2 to 15 years of age, while venereal syphilis infects mainly adults and infants (Giacani & Lukehart 2014).

In 2009 there were 10.6 million new cases of venereal syphilis in adults, with more than 36.4 million adults thought to be infected (World Health Organization 2012); estimates for 2008 suggested more than half a million births were affected despite massive antenatal care efforts (Newman *et al.* 2013). While venereal syphilis is globally distributed, its prevalence varies by region, peaking in sub-Saharan Africa (World Health Organization 2012). Endemic syphilis is less well documented, though several cases have been reported across several countries in Africa and the Middle East (Giacani & Lukehart 2014), most recently in Iran in 2010 (Abdolrasouli *et al.* 2013). Yaws has been the target of major past and present eradication campaigns, which has reduced its confirmed distribution to 12 countries in Africa, Asia, and the South Pacific, where it still caused more than 300,000 new cases between 2008 and 2012 (Kazadi *et al.* 2014). Poor documentation and detection of latent infections combined with underreporting and misdiagnosis suggest that the magnitude of the disease burden for endemic syphilis and yaws is vastly underestimated. Though symptoms and progression of these infections varies, debate remains as to whether these are due to the pathogens themselves, or rather caused by differences between hosts, environmental variables, or mode of infection (Thornburg & Baseman 1983; Mulligan *et al.* 2008).

### 5.4 NON-HUMAN PRIMATE INFECTIONS

Non-human primates (NHPs) are also susceptible to *T. pallidum* infections and descriptions of symptoms (Baylet *et al.* 1970; Levréro *et al.* 2007; Knauf *et al.* 2012) combined with serological and morphological evidence suggest both syphilis and yaws like infections occur in a number of wild NHP populations across Africa (Baylet *et al.* 1970; Lovell *et al.* 2000; Knauf *et al.* 2013). The presence of *T. pallidum* in inflamed tissues has however only been demonstrated in baboons (Fribourg-Blanc *et al.* 1966; Baylet *et al.* 1970; Harper *et al.* 2012) (Knauf *et al.* 2012); conclusive evidence of infections in other species is still lacking. The recognition of *T. pallidum* infections in NHPs led to the hypothesis that human treponematoses are zoonotic in origin (Knauf *et al.* 2013), but whether human *T. pallidum* is the result of a single transmission event, continuous transmission, or whether these treponemes co-diverged with their primate hosts, is a source of debate (Hackett 1963; Harper *et al.* 2008b). It has been suggested that eradication initiatives for yaws might be hampered by continued spill-over from a NHP reservoir (Knauf *et al.* 2013), though evidence confirming transmission events, or even cross-NHP species transmission events in the wild, are lacking.

### 5.5 MOLECULAR DIAGNOSIS AND PHYLOGENETIC RELATIONSHIPS BETWEEN SUBSPECIES

Such conclusive evidence of zoonotic transmission is lacking in large part because, despite major advances in culture techniques, *T. pallidum* remains one of the last as-yet uncultured human pathogens (Lagier *et al.* 2015). The three recognized subspecies are morphologically indistinguishable and antigenically cross-reactive (Giacani & Lukehart 2014). Subspecies delineations in humans are therefore nearly always based on clinical and epidemiological data, rather than distinguishing molecular evidence (Mulligan *et al.* 2008). Most molecular diagnosis tools are PCR based but as *T. pallidum* is slowly evolving, the short regions amplified with these approaches fail to capture sufficient information to determine evolutionary history. A number of single nucleotide polymorphisms (SNPs) distinguishing subspecies have been proposed for diagnostics and untangling the evolutionary history of this species (Harper *et al.* 2008b), though the utility and validity of this approach has been questioned (Mulligan *et al.* 2008). The accumulation of sequences has cast doubt on a number of formerly recognized diagnostic SNPs which may be explained by mere sampling artefacts and/or recombination events (Mulligan *et al.* 2008).

Full genomes allow a resolution of evolutionary relationships that for the most part distinguishes subspecies in much the same manner as delineations based on symptoms, though misclassification has occurred when symptoms were used alone (Cameron *et al.* 2000) (Fig. 5.1). From extant human strains, molecular differences across genomes were described for only a few cases; 10 full genomes are published and these are >99.6% per cent identical (Šmajs *et al.* 2012). The whole genome of a treponeme infecting West African baboons (*Papio cynocephalus*) isolated in 1966 was recently sequenced; phylogenetic analysis suggested it is extremely similar to human *T. p. pertenue*, though more genomes of treponemes infecting NHPs are required to test the hypothesis that humans acquired this infection from NHPs or vice versa (Zobaníková *et al.* 2013). The diversity of non-pathogenic treponemes in humans and wildlife is even less well described, though preliminary insights suggest it is high (Graber *et al.* 2004; Bittar *et al.* 2014; Schnorr *et al.* 2014), complicating efforts to design pathogen specific primers and presenting challenges for the hybridisation capture approaches described below. This is particularly true for non-invasive sampling of NHP populations using faeces, where treponeme diversity is high (Bittar *et al.* 2014).

### 5.6 ORIGINS AND HISTORY OF *T. PALLIDUM* IN HUMAN POPULATIONS

Molecular tools and the availability of full genome data promises to shed light on the origins and evolutionary history of *T. pallidum,* though many questions remain unresolved with the currently available data. The origins and spread of venereal syphilis particularly has been the focus of much of the debate and controversy. The origins and spread of yaws and endemic syphilis are equally enigmatic but have mostly been discussed in relation to the question of venereal syphilis's origin. More specifically, controversy has surrounded the origin of venereal syphilis in European populations; the diseases seemingly appeared at the end of the 15$^{th}$ century, and quickly turned into a major epidemic that swept across the continent. Contemporaries began questioning the origins of the disease and the issue remains largely unresolved; three main hypotheses have been proposed.

The Columbian hypothesis posits that when Columbus and his crew returned from the New World in 1493, they brought with them not only tobacco and corn, but also a new infectious agent. Proponents of this theory recognize widespread syphilitic infections across a large temporal and spatial scale in the New World and argue that nothing comparable can be found in the European archaeological record prior to Colombus's return. They also consider the

very rapid rise of syphilis as an indication that a novel infectious disease was spreading in a naïve population (Hudson 1965).

Others have hypothesized that syphilis was present in the Old World long before Columbus's return and that it became more virulent around the time of Columbus or increased in prevalence and spread due to other social and geopolitical factors (Luger 1993). Proponents of the pre-Columbian hypothesis suggest that the progression of the disease makes it unlikely the crew would have been exhibiting infective lesions following their return to Europe, that the time interval between the sailors return from the New World and the start of a wide-spread outbreak across Europe is unrealistic, and draw on historic documents suggesting syphilis was present in Europe before Columbus's return (Luger 1993). Archaeological remains have been used to argue that pre-Columbian Old World skeletons show evidence of syphilitic infections (von Hunnius *et al.* 2006) and that New World pre-Columbian skeletons show evidence of lesions present in young individuals, which might be more consistent with yaws like infections than venereal syphilis (Baker & Armelagos 1988). Much of the debate surrounds whether bone lesions in different archaeological records in the Old and New World represent *T. pallidum,* and if so, which subspecies.

Another hypothesis presented is that yaws, venereal syphilis, and endemic syphilis actually represent the same pathogen, and that environmental and social conditions determine the outcome of the infection (Hudson 1965); proponents of this hypothesis draw on the low diversity and difficulties in identifying genetic differences between these subspecies. The full genome evidence discussed above suggests genetic differences do exist between these different pathogens, though small sample sizes might have missed a continuum of diversity in this species and species concepts are notoriously difficult to apply to bacterial lineages.

### 5.7 ANCIENT PATHOGEN DNA

Nucleic acids from archaeological and paleontological remains have proven a powerful tool for examining the phylogenetic relationships between historic and modern organisms and have the potential to resolve the aforementioned controversies (Poinar & Cooper 2000; Hofreiter *et al.* 2001; Schuenemann *et al.* 2013). While the treasure trove of information in ancient remains is appealing, the minute amounts of degraded genetic material calls for caution; contaminations have literally plagued the ancient DNA (aDNA) field. Great care is therefore required to ensure the replicability and authenticity of any findings; guidelines have

been proposed that include the need for replication by independent labs and the use of clean rooms (Poinar & Cooper 2000). The fragmented nature of aDNA means that only short reads can be generated with each PCR, precluding in-depth phylogenetic analyses for slowly evolving pathogens such as *T. pallidum*. The rise of next-generation sequencing (NGS) technologies provided a first means to generate large amounts of data from small amounts of starting material, circumventing many limitations of PCR based approaches. Interestingly, data from NGS can be leveraged to validate their own authenticity, e.g. by evidencing damage motifs characteristic of ancient samples (Briggs *et al.* 2007; Stoneking & Krause 2011; Jónsson *et al.* 2013). Ancient samples are metagenomic, containing a mix of host and environmental DNA and often only a low percentage of endogenous pathogen DNA. In some rare cases, shotgun NGS approaches have been powerful enough to generate complete or close-to-complete ancient bacterial genomes (e.g. *Tannerella forsythia* (Warinner *et al.* 2014) and *Mycobacterium leprae* (Schuenemann *et al.* 2013)), though the combination of NGS with hybridisation capture was the fundamental technical leap that revolutionized this field of microbiology. Enrichment experiments have thus far succeeded in shedding light on the history and evolution of a number of human bacterial pathogens including *Mycobacterium tuberculosis* (Bouwman *et al.* 2012), *Yersinia pestis* (Bos *et al.* 2011), *Vibrio cholerae* (Devault *et al.* 2014), and *Helicobacter pylori* (Maixner *et al.* 2016). These enrichment approaches also allowed researchers to generate sequence information for the pathogen causing leprosy (*Mycobacterium leprae*) from both ancient samples and modern samples obtained from patient biopsies (Schuenemann *et al.* 2013). As is the case for *T. pallidum,* propagation by culture is difficult, which has prohibited large scale sequencing of this pathogen; NGS approaches coupled with hybridisation capture have greatly expanded our understanding of modern *M. leprae* diversity and how this relates to ancient infections (Schuenemann *et al.* 2013).

Debate surrounding interpretation of archaeological evidence for *T. pallidum* infections would be clarified considerably if genome sequence information became available, particularly from pre-Columbian samples from the Old and New World. A short treponeme PCR fragment from a 200 year old mummy (Kolman *et al.* 1999) and from post-Columbian fetuses in Europe (Montiel *et al.* 2012) suggested aDNA approaches might prove effective; though in addition to being too modern to help inform the debate, the short fragments amplified precluded informative phylogenetic analyses. Many researchers have subsequently tried to use PCR based approaches to study ancient samples with evidence of *T. pallidum*

lesions, with very limited success, which has lead many to question the feasibility of using bones to study ancient *T. pallidum* infections (von Hunnius *et al.* 2007). This may partially be a product of the progression of venereal syphilis; in modern cases, the highest pathogen load is usually found in stage 1 when no bone lesions have developed; in adults bone deformations usually occur in stage 3 when it is difficult to detect the pathogen itself (Montiel *et al.* 2012). The availability of hybridisation capture coupled with NGS clearly represents a promising way forward; this approach should outperform PCR based approaches for screening samples with low concentrations and highly fragmented pathogen DNA, ultimately paving the way for the generation of genome-wide information from both ancient and modern samples.

## 5.8 Demonstrating feasibility of hybridisation capture enrichment for bone samples

We assessed the feasibility of such an approach using a set of samples that intuitively seemed unpromising; asymptomatic bones from non-human primates. The ability to recover treponemal DNA from bones, and more particularly from diagnosed bones, has been a matter of controversy (von Hunnius *et al.* 2007). On the other hand, as mentioned above, many NHP populations present clinical manifestations suggestive of treponemal infections, in some groups even at very high prevalence (Knauf *et al.* 2013).

For this study, we generated DNA extracts from contemporary non-lesioned NHP bones (*N*=51, 6 species from Taï National Park, Côte d'Ivoire) and first screened them with three independent PCR systems specific to *T. pallidum* (see *Chapter 5 - Supplementary material* for details). No bone was positive for all three PCR systems (12 tested positive for the shortest *PolA* sequence, and three were also positive for the longer *GDP* or *cfpA* fragments) suggesting *T. pallidum* DNA is highly fragmented or at low concentrations. Based on this screening, we selected three candidate extracts (i.e., those also positive for a longer fragment: Table 5.1) for which we conducted library preparation, enrichment, and sequencing in two separate laboratories using standard procedures to avoid contamination (Supplementary Fig. 5.1; see *Chapter 5 - Supplementary materials* for details). At the University of Tübingen we used a DNA microarray-based approach, with probes spanning the *T. p. pallidum* genome and post-capture pathogen enriched DNA was sequenced on an Illumina HiSeq 2500 (Arora *et al.* 2016). At the Robert Koch Institute, we used an in-solution capture approach, with baits spanning the *T. p. pertenue* Fribourg-Blanc genome and post-capture pathogen enriched

DNA was sequenced on an Illumina MiSeq. Hybridisation capture approaches have been used to enrich DNA up to 42% dissimilar to baits (Wylie *et al.* 2015); *T. pallidum* is characterized by extremely low diversity, suggesting these different bait sets would capture DNA from any subspecies of *T. pallidum*. Reads generated at both institutions were trimmed using Trimmomatic, mapped to *T. p. pertenue* Fribourg-Blanc using BWA-MEM (Li 2013), and deduplicated using Picard's *MarkDuplicates*. To ensure reads were not contributed by non-pathogenic treponemes we filtered reads; each mapped read was BLASTed against a local database of treponeme genomes and only reads which were a hit to every published *T. pallidum* genome and where the lowest bit score for a *T. pallidum* genomes was greater than the highest bit score for published non-*T. pallidum* treponeme genomes were kept. For comparing bit scores we used R version 3.2.3 (R Core Team 2015) with the package 'data.table' (Dowle *et al.* 2015). Surviving reads were mapped to a closely related out-group of *T. pallidum* infecting rabbits, *T. paraluiscuniculi,* (Šmajs *et al.* 2011), along with *T. p. pallidum* and *T. p. endemicum* that infect humans.

Both the array-based and in-solution-based approaches generated sequences spanning the *T. p. pertenue* Fribourg-Blanc genome (Table 5.1 and Fig. 5.2), though coverage was low for the two chimpanzee samples (range=0.036-1.8%) and moderate for the red colobus bone (7.2% and 13.9%; Table 6.1). Both laboratories, using unique indices and distinct library preparation methods and capture protocols, converged on similar findings for each of the bones. Combining data from the approaches resulted in 19.8% genome coverage from the red colobus bone (99.5% identical sites) and for the best chimpanzee sample, 3.4% coverage (98.9% identical sites). Sequence similarity to the most closely related outgroup, *T. paraluiscuniculi* was lower than for all three *T. pallidum* subspecies, suggesting chimpanzees and red colobus are infected with *T. pallidum* in the wild. That this approach worked for non-lesioned bones, suggests it should work efficiently for tissue samples or swabs from humans or wildlife, providing a cost-effective, culture-free means of generating whole-genome data. It also suggests aDNA studies might benefit from screening non-lesioned remains. The amount of sequence information generated represents a substantial increase compared to what has been feasible with PCR-based analyses, particularly from bone samples (Kolman *et al.* 1999; Montiel *et al.* 2012). Higher coverage could still be achieved by intensifying the sampling of DNA fragments, either through deeper sequencing of the same libraries and/or the generation and sequencing of further libraries. Capture experiments need not target entire genomes of *T. pallidum* and other bait designs, e.g. targeting unique but variable regions,

might increase the power of such an approach for screening samples; positive samples could subsequently be enriched with a genome wide approach to enable rigorous phylogenetic analysis. In any case, these experiments demonstrate the feasibility of capture enrichment in combination with NGS for studying the enigmatic pathogen, *T. pallidum*.

While field reports have suggested *T. pallidum* infections might be occurring in NHP species beyond baboons (Knauf *et al.* 2013), our results molecularly confirm these observations and expand the NHP reservoir to include our closest NHP relatives, chimpanzees. The strongest evidence for *T. pallidum* infection came from the preferred prey species of these chimpanzees, red colobus; transmission of microorganisms in this hunter-prey relationship has been documented (Gogarten *et al.* 2014a) though further studies are needed to understand the ecology and between-species transmission of *T. pallidum* in wildlife communities and to test the hypothesis of a NHP reservoir that is continuously spilling over into human populations.

## 5.9 Conclusion

Despite the availability of effective antibiotic treatments, *T. pallidum* infections causing venereal syphilis and yaws are still globally widespread, while endemic syphilis continues to reappear despite widespread eradication efforts. The paucity of genome data from *T. pallidum* infecting humans or wildlife has hampered a deep understanding of the ecology and evolution of this pathogen. This lack of genomic data is notably explained by the difficulty of culturing *T. pallidum*. We show the feasibility of *T. pallidum* DNA enrichment through hybridisation capture, and demonstrate that in combination with NGS technologies these approaches sometimes allow for the recovery of substantial parts of treponemal genomes. These approaches should be applicable to biopsy samples from symptomatic NHPs and humans and from ancient bone specimens of humans and NHPs. We expect they will provide a major contribution to our understanding of these enigmatic pathogens and help resolve long-standing controversy surrounding their ecology and evolution.

## 5.9 Acknowledgments

For assistance and support with sequencing at the RKI we are grateful to the Central Sequencing Laboratory, particularly Julia Hinzmann, Andreas Nitsche and Julia Tesch. For assistance with the processing of MiSeq reads and help accessing computing power at the

RKI we would like to thank Wojciech Dabrowski. For assistance with processing and accessing the Tübingen HiSeq reads we thank Kay Nieselt. For the Microarray-design of the *T. pallidum pallidum* array we thank Kirsten Bos, Kay Nieselt and Günter Jäger. For assistance in implementing the bioinformatics approach for filtering the reads generated in this project and generating the phylogeny from published genomes, we thank Peter Gogarten. For administrative assistance we are grateful to the help of Maja Kovacev-Wegener and for assistance in the laboratory at the RKI we thank Ulla Thiesen, Kevin Merkel, and Andreas Sachse, along with the entire P3 Project Group. JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336); the Canadian Institutes of Health Research's Strategic Training Initiative in Health Research's Systems Biology Training Program; an NSERC Vanier Canada Graduate Scholarship (CGS); the Quebec Centre for Biodiversity Science's (Centre de la Science de la Biodiversité du Québec) Excellence Awards Program; McGill University, particularly through Research Travel Awards, the Graduate Research Mobility Award Program, the Dean of Science's Research Fund, and the Excellence Awards Program; and a long-term Research Grant from the German Academic Exchange Service (DAAD-91525837-57048249). The bones tested in this study stem from Taï National Park, Côte d'Ivoire and were collected through the long-term research conducted with permission from the Ministère de la Recherche Scientifique, the Ministère de l'Environnement et des Eaux et Forets, the Direction Générale de la Recherche Scientifique et de l'Innovation Technique, and the Office Ivoirien des Parcs et Reserves; support from the Max Planck Society was integral for the long-term field presence of the Taï Chimpanzee Project, which enabled such an extensive collection and benefited from the assistance of all students and field assistants of the project. For comments on an earlier version of this manuscript, we thank Annemarrie Goedmakers.

**Table 5.1. Screening results for bone samples from Taï National Park, Côte d'Ivoire selected for *T. pallidum* enrichment**

| | | | | | **Array capture approach: Genome coverage (%) [Per cent identical sites (%)]** | | | | **In-solution capture: Genome coverage (%) [Per cent identical sites (%)]** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Species** | **Death year** | ***Pol A* seq.** | ***GDP* seq.** | ***cfp A* seq.** | ***T. paralui-scuniculi*** | ***T. p. endemic-cum*** | ***T. p. pallidum*** | ***T. p. perte-nue FB*** | ***T. paralui-scuniculi*** | ***T. p. endemic-cum*** | ***T. p. pallidum*** | ***T. p. perte-nue FB*** |
| *Proc. badius* | 2002 | + | + | - | 13.4 [97.8] | 13.8 [99.5] | 13.8 [99.4] | 13.9 [99.5] | 7.2 [98.6] | 7.2 [99.6] | 7.2 [99.5] | 7.2 [99.6] |
| *P. t. verus* | 1992 | + | - | + | 0.036 [97.3] | 0.036 [97.5] | 0.036 [97.5] | 0.036 [97.6] | 0.2 [95.6] | 0.2 [96.3] | 0.2 [96.3] | 0.2 [96.3] |
| *P. t. verus* | 1998 | + | + | - | 1.7 [97.5] | 1.7 [98.9] | 1.7 [98.9] | 1.8 [99.0] | 1.6 [98.1] | 1.7 [98.7] | 1.6 [98.6] | 1.7 [98.6] |

*P.t. versus* = *Pan troglodytes verus*; *Proc. badius* = *Procolobus badius*; + Indicates sanger sequence generated; - indicates no sequence generated; *FB* indicates *Treponema pallidum pertenue* Fribourg-Blanc (NC_021179) isolated from baboon. European Nucleotide Archive (ENA) study accession number PRJEB13855 (http://www.ebi.ac.uk/ena/data/view/PRJEB13855). Sample accession numbers for raw data generated at Tübingen University using a microarray-based approach: *Procolobus badius* - 2002 = ERS1138345, *Pan troglodytes verus* - 1992 = ERS1138347, and *Pan troglodytes verus* - 1998 = ERS1138349. Sample accession numbers for raw data generated at the Robert Koch Institute using the in-solution based capture approach: *Procolobus badius* - 2002 = ERS1138344, *Pan troglodytes verus* - 1992 = ERS1138346, and *Pan troglodytes verus* - 1998 = ERS1138348 and ERS1138350 (large number of reads split into two sets of paired reads). The following reference genomes were used for mapping and comparison of per cent identical sites: *T. paraluiscuniculi* (NC_015714), *T. p. endemiccum* (NZ_CP007548), *T. p. pallidum* (NC_021490), and *T. p. pertenue FB* (NC_021179).

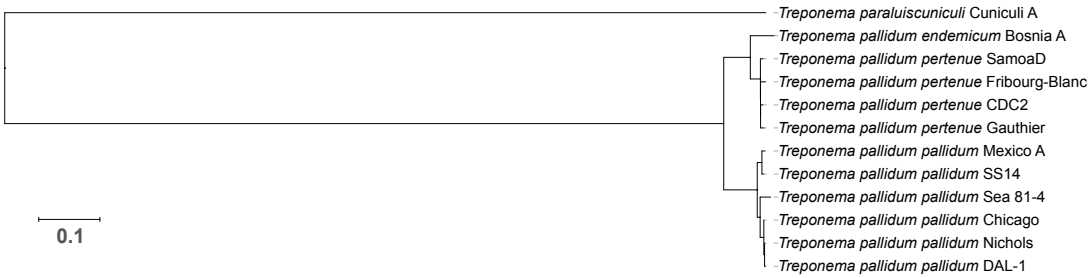



**Figure 5.1: Maximum likelihood phylogeny of *T. pallidum* subspecies and closely related Treponema paraluiscuniculi generated from full genome sequences.** Highly variable TPR genes were removed prior to phylogenetic analysis, as these are under strong positive selection and often recombine, suggesting these genes may be inappropriate for inferring phylogenetic history. Scale is in substitutions per variable site. Support values were calculated using Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-like aLRT) and all branches received values above 0.9, with the exception of the short branches within the *Treponema pallidum* pertenue clade between the CDC2 and Gauthier genomes and between these two genomes and *T. p. pertenue* Fribourg-Blanc, which received value of.0.422 and 0.098 respectively. Details on how phylogeny was generated in *Chapter 6 Supplementary materials*.

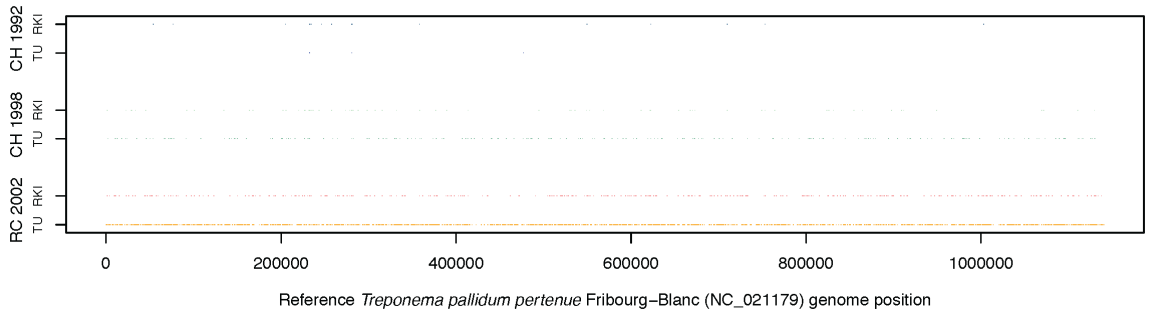



**Figure 5.2: Coverage of *T. p. pertenue* Fribourg-Blanc genome.** Each point represents coverage at that location on the *T. p. pertenue* Fribourg-Blanc genome for the red colobus bone (RC) and two chimpanzee bones (CH) using the two different enrichment approaches at the Robert Koch Institute (RKI) and Tübingen University (TU). Coverage appears randomly distributed across the genome from both approaches and from the three samples. Points are illegible in a printed version though readers are invited to zoom in on the electronic version to see these details.

Click on the image to be taken to a high resolution version available at the end of this document.

# LINKING STATEMENT 4

In Chapter 5 I developed a new method for studying the wildlife pathogen, *Treponema pallidum* and show that two non-human primate (NHP) species in Taï National Park, Côte d'Ivoire are infected. In Chapter 6, I use this approach to study this pathogen in multiple NHP species across Africa. Using these data, we show that a number of African NHPs exhibiting yaws- and syphilis-like symptoms in the wild, are infected with treponemes. Further, we conduct phylogenomic analyzes to determine whether a yaws or syphilis like strain is infecting each of these species. Using this genomic data allowed us to further analyze the evolution of this pathogen; results suggest that these pathogens radiated quickly across humans and NHPs and likely jumped repeatedly between their primate hosts. NHPs thus seem to represent a high-diversity reservoir for this pathogen, which may impact ongoing yaws eradication efforts. The changing behaviour observed in Chapters 2 and 3 might have major impacts on the prevalence of this pathogen, which in turn could greatly influence disease risk for humans living in proximity to these NHP populations.

# CHAPTER 6

## NON-HUMAN PRIMATES AS RESERVOIRS FOR YAWS

Sascha Knauf[1†], Jan F. Gogarten[2,3,4 †], Verena J. Schuenemann[5†], Hélène M. De Nys[2,^†], Ariane Düx[2,6], Michal Strouhal[7], Lenka Mikalová[7], Kirsten I. Bos[8], Roy Armstrong[9], Emmanuel K. Batamuzi[10], Idrissa S. Chuma[11], Bernard Davoust[12], Georges Diatta[13], Robert D. Fyumagwa[14], Reuben R. Kazwala[15], Julius D. Keyyu[14], Inyasi A. V. Lejora[16], Anthony Levasseur[12], Hsi Liu[16], Michael A. Mayhew[17], Oleg Mediannikov[12], Didier Raoult[12,13], Billy Sinclair[17], Roman M. Wittig[3,18], Christian Roos[19], Fabian H. Leendertz[2], David Šmajs[7*†], Kay Nieselt[20†], Johannes Krause[5,8*†], Sébastien Calvignac-Spencer[2,6*†]

[†]These authors contributed equally to this work.

[1]Work Group Neglected Tropical Diseases, Pathology Unit, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany.

[2]Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Seestraße 10, 13353 Berlin, Germany.

[3]Primatology Department, Max Planck Institute for Evolutionary Anthropology, Deutscher Pl. 6, 04103 Leipzig, Germany.

[4]Department of Biology, McGill University, 1205 Avenue du Docteur-Penfield, Montréal, QC H3A 1B1, Quebec, Canada.

[5]Institute for Archaeological Sciences, University of Tübingen, Ruemelinstr. 23, 72070 Tübingen, Germany.

[6]Viral Evolution, Robert Koch Institute, Seestraße 10, 13353 Berlin, Germany.

[7]Department of Biology, Faculty of Medicine, Masaryk University, Kamenice 5/A6, 625 00 Brno, Czech Republic.

*Affiliations continued on next page.*

[8]Max Planck Institute for the Science of Human History, Jena, Germany.

[9]Genetics and Conservation, University of Cumbria, UK.

[10]Department of Surgery and Theriogenology, Faculty of Veterinary Medicine, Sokoine University of Agriculture, Morogoro, Tanzania.

[11]Ecology Monitoring Department, Tanzania National Parks (TANAPA), P.O.Box 3134, Arusha, Tanzania.

[12]Research Unit of Emerging Infectious and Tropical Diseases (URMITE) UM63, CNRS 7278, IRD 198, INSERM 1095, Aix-Marseille University, 27, Bd Jean Moulin, 13385 Marseille cedex 05, Marseille, France.

[13]Research Unit of Emerging Infectious and Tropical Diseases (URMITE) IRD 198, Campus IRD/UCAD, Hann Les Maristes, Dakar, Senegal.

[14]Tanzania Wildlife Research Institute (TAWIRI), P.O. Box 661, Arusha, Tanzania.

[15]Department of Public Health, Faculty of Veterinary Medicine, Sokoine University of Agriculture, Morogoro, Tanzania.

[16]National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Diseases Control and Prevention, Atlanta, GA, USA.

[17]Department of Science, Natural Resources and Outdoor Studies, Centre for National Parks and Protected Areas, University of Cumbria, Carlisle, UK, CA1 2HH.

[18]Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, BP 1303, Abidjan 01, Côte d'Ivoire.

[19]Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany.

[20]Center for Bioinformatics, University of Tübingen, Sand 14, 72076 Tübingen, Germany.

[^]Current affiliation: Unité Mixte Internationale 233, Institut de Recherche pour le Développement, INSERM U1175, and University of Montpellier, Montpellier, France.

[*]Correspondence to: Dr. Sebastien Calvignac-Spencer, Robert Koch Institute, Seestrasse 10, 13353 Berlin, Germany, tel.: +49187542501, calvignacs@rki.de; Prof. David Šmajs, Faculty of Medicine, Masaryk University, 625 00 Brno, Czech Republic, tel.: +420 549 497 496, dsmajs@med.muni.cz; Prof. Dr. rer. nat. Johannes Krause, Max Planck Institute for the Science of Human History, Jena, Germany, tel.: +49 3641 686600, krause@shh.mpg.de.

## 6.1 ABSTRACT

A number of African non-human primate (NHP) species exhibit yaws- and syphilis-like symptoms in the wild, yet it is still unclear which treponemes are involved in these infections. We present phylogenomic evidence that in four African countries, The Gambia, Senegal, Côte d'Ivoire and Tanzania, wild NHP populations harbor bacterial strains closely related to the human yaws-causing *Treponema pallidum* ssp. *pertenue*. These pathogens radiated quickly across humans and NHPs and likely jumped repeatedly between their primate hosts. NHPs therefore represent a high-diversity reservoir for this pathogen, which may impact ongoing yaws eradication efforts.

## 6.2 INTRODUCTION

Human-pathogenic treponemes (*Treponema pallidum*; *TP*) causing syphilis (ssp. *pallidum*: *TPA*), yaws (ssp. *pertenue*: *TPE*) and bejel (ssp. *endemicum*: *TEN*) are responsible for a large global disease burden and are considered exclusively human-adapted pathogens (World Health Organization 2012; Mitjà *et al.* 2015). Yet non-human primates (NHPs) can be experimentally infected with *TPE* and *TPA*, and syphilis- and yaws-like symptoms, along with detection of anti-*TP* antibodies, have been reported for a number of NHP species across Africa (Nichols 1910; Elsas *et al.* 1968; Fribourg-Blanc & Mollaret 1969; Knauf *et al.* 2013). Genetic evidence of *TP* in wild NHPs are however very scarce and confined to a baboon population in Guinea (Zobaníková *et al.* 2013) and another in Tanzania (Harper *et al.* 2012). A single genome allowing conclusive determination of the *TP* subspecies is available from a strain isolated from Guinean baboon (*Papio papio*) in 1966 and is highly similar to human-infecting *TPE* genomes (Zobaníková *et al.* 2013). Nonetheless, it remains unclear which treponemes are generally involved in NHP infections. Such information is critical for public health initiatives, which are currently built on the assumption that human infections stem only from other humans.

## 6.3 METHODS

Here we investigated four wild NHP populations that displayed ulcerative skin lesions in the Taï National Park (TNP), Côte d'Ivoire; the Bijilo Forest Park (BFP), the Gambia; the Niokolo-Koba National Park (NKNP), Senegal; as well as at the Lake Manyara National Park (LMNP), Tanzania. Additional Materials and methods are available as supplementary materials at the Science website. These populations either presented with syphilis-like anogenital lesions (Knauf *et al.* 2012) or yaws-like orofacial and limb lesions. While *TP* infections in olive baboons (*Papio anubis*) at LMNP were already confirmed (*8*), we collected skin lesion biopsies or swabs from affected individuals in TNP, BFP, and NKNP to determine their *TP* infection status. We detected *TP* DNA in samples from these three populations and selected promising candidates for whole genome sequencing from all four populations, based on high *TP* copy numbers or the sequencing of longer PCR fragments (**table S6.1**). To overcome the large background of host DNA we used targeted DNA capture coupled with next generation sequencing (Arora *et al.* 2016; Gogarten *et al.* 2016).

## 6.4 RESULTS

After mapping against the same reference *TP* genome, removing all PCR duplicates, and merging different sequencing runs from the same sample, we achieved a range of 22,886-470,303 DNA sequencing reads per sample resulting in an 6.1 to 121-fold average genome coverage for 2 samples per population (n=8; **table S6.2**). After removal of four potentially recombinant genes, we reconstructed the phylogeny of these genomes and all other available *TP* genomes using maximum parsimony (**Fig. 6.1, table S6.3** and **Fig. S6.1**). All simian *TP* strains were closely related to *TPE* strains infecting humans (Cejkova *et al.* 2012). Human yaws-causing *TPE* strains (Samoa D, CDC-2, and Gauthier) spanned a broad geographic and temporal range but were less divergent from each other than the two strains infecting sooty mangabeys (*Cercocebus atys*) from a single social group at TNP. For the two African green monkey (*Chlorocebus sabaeus*) and the olive baboon social groups, intra-group strain divergence was low, though intra-species strain divergence for African green monkeys was again almost as high as the divergence observed between human strains. The *TPE* clade exhibited a star-like phylogenetic branching pattern, with short basal branches receiving low statistical support, suggesting a rapid initial radiation of *TPE* across humans and NHPs. Neither human-infecting *TPE* nor NHP-infecting *TPE* strains formed monophyletic groups, which may indicate that ancestral *TPE* strains were not host species specific.

Using long-range PCR amplification coupled with next generation sequencing (**tables S6.4**), we determined the complete sequence of one of the new NHP-infecting *TPE* (LMNP; primers available from corresponding authors). This genome showed the same structure as the three human-infecting and the single NHP-infecting *TPE* strain genomes already available, including gene synteny in two alternative rRNA operons (Cejkova *et al.* 2012). When comparing the two NHP-infecting *TPE* strains (Fribourg-Blanc and LMNP) to the closest human-infecting *TPE* strains (CDC-2 and Gauthier, respectively), we found that only 7.2% and 9.1% of all coding sequences (77 and 97 coding sequences out of 1065 coding sequences) comprised amino acid substitutions (**tables S6.5, S6.6 and S6.7**). There is, therefore, a strong indication of limited functional divergence between human-infecting and NHP-infecting *TPE* strains. Since the

Fribourg-Blanc strain causes sustainable infection when inoculated into human skin (Smith *et al.* 1971), it seems likely that a number of these *TPE* strains infecting NHPs have zoonotic potential.

## 6.5 DISCUSSION

Antibiotherapy in combination with mass drug applications led to a marked decrease in the incidence of *TP* diseases in the mid-20$^{th}$ century. Unfortunately yaws incidence has increased again in West Africa, Southeast Asia, and the Pacific region (World Health Organization 2012). Efforts aiming to eradicate yaws globally by 2020 are currently underway (Maurice 2012). Our findings provide convincing evidence that NHPs likely represent a large and widespread *TPE* reservoir that could hinder future eradication efforts targeting African human populations. Even if yaws eradication is achieved in humans by 2020, post-treatment surveillance will be an important consideration for sustainable eradication in Africa, where continued high-contact rates to NHPs may allow for disease re-emergence (Mossoun *et al.* 2015).

## 6.6 ACKNOWLEDGEMENTS

JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336), the Canadian Institutes of Health Research's Strategic Training Initiative in Health Research's Systems Biology Training Program, an NSERC Vanier Canada Graduate Scholarship (CGS), and a long-term Research Grant from the German Academic Exchange Service (DAAD-91525837-57048249). HMD was supported by Max Planck Institute for Evolutionary Anthropology and Gent University. KIB was supported by a Social Science and Humanities Research Council of Canada postdoctoral fellowship. The PSGS sequencing was supported the Grant Agency of the Czech Republic to DS and MS (P302/12/0574, GP14-29596P). This project was supported by the German Research Foundation through grants FHL (LE1135/2), SK (KN1097/3-1 and KN1097/4-1), and CR (RO3055/2-1). This study was also supported by the AMIDEX project (No. ANR-11-IDEX-0001-02), funded by the 'Investissements d'Avenir' French Government program, managed by the French National Research Agency (ANR) and the Fondation Méditerranée Infection. All raw read files have been deposited in NCBI as part of the BioProject PRJNA343706.

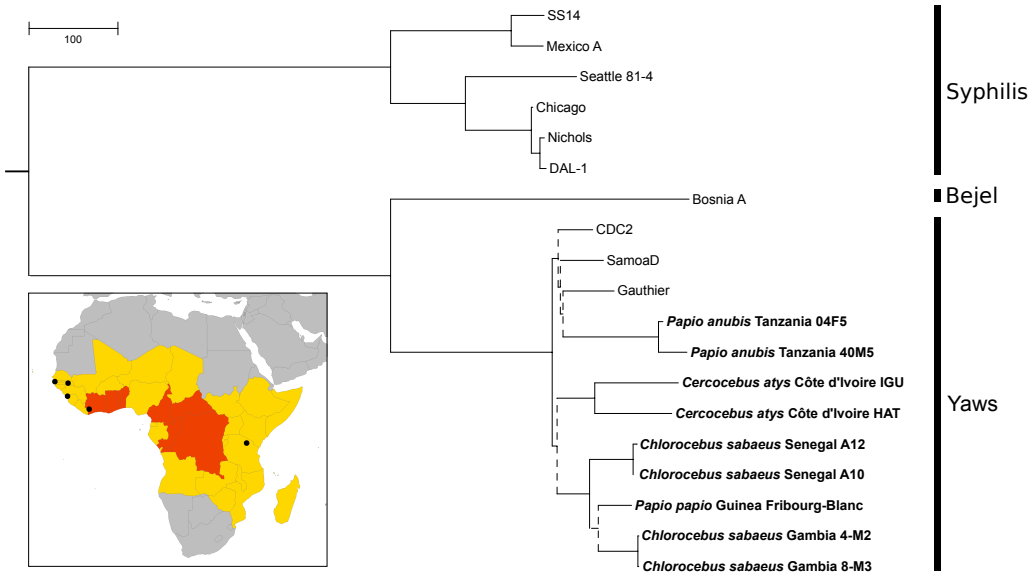



**Figure 6.1**: **Phylogenomic analysis of NHP (in bold) and human-infecting treponemes.** In this maximum parsimony tree, nodes that had less than 90% bootstrap support are indicated with dashed lines. Tip labels indicate the NHP species sampled, the country of origin, and the sample ID. The scale is in nucleotide substitution. The inset is a map of Africa where sites of origin of the NHP samples from which a *TP* genome was determined are indicated with black circles. A country's 2013 yaws status based on the World Health Organization's Global Health Observatory is indicated by its color: grey indicates no previous history of yaws infections in humans, yellow indicates a country previously endemic for yaws though the current status is unknown, and countries in red indicate countries which are currently considered endemic for yaws.

# CHAPTER 7

## CONCLUDING REMARKS AND FUTURE DIRECTIONS

There is a saying in German that has rung true throughout my life; “man lernt nie aus”, which translates roughly to; one is never done learning. I was born in Germany, but am a dual US-German citizen and my family first came to the United States shortly after the Chernobyl nuclear disaster. My father was working in a lab at the time and had access to a Geiger counter, and when we went outside to the grass where my sister was playing we found that it was more radioactive than the lab he was working in. Shortly thereafter, my mother left Germany with me and my sister, heading to the United States for the first time. Since a young age I have been motivated to understand how these massive environmental changes interact with wildlife and human populations.

My studies in Kibale National Park, Uganda, presented in part in Chapters 2 and 3 of this thesis, examined how environmental change impacts non-human primate’s (NHPs) food availability and quality (Gogarten *et al. 2012b*), gastrointestinal parasites (Chapman *et al.* 2012), group sizes (Gogarten *et al.* 2015), behaviour (Gogarten *et al.* 2014b), group genetic structure and dispersal (Miyamoto *et al.* 2013), and the cascading impact of these changes on primate fitness, abundance, and population sizes (Gogarten *et al.* 2014b; Chapman *et al.* 2015a; Chapman *et al.* 2015b). These in turn impacted the broader ecosystem, as NHP species serve as ecosystem engineers (Chapman *et al.* 2013). As part of this dissertation and through my work at the Robert Koch Institute in Berlin, Germany and working at their field site in Taï National Park in the Côte d’Ivoire, I also made my first forays into understanding how these changes may impact populations at the human-wildlife interface. This shift necessitated a strong emphasis on the molecular; In chapter 4, I generated amplicons covering the 16S V4 hypervariable region to examine factors impacting bacterial gut microbiome community composition in a wild NHP community; examining the importance of species, social group and position in the social network and also mother-offspring effects. In the Appendix Chapter, I began studying the ecology and evolution of retroviruses that have spilled over into the surrounding human population and gained insights into the importance of virus-host interactions in determining cross-species transmission risk (Gogarten *et al.* 2014a). In chapter 5, I developed hybridization capture methods on bone samples to generate *Treponema pallidum pertenue* sequences and demonstrate that this pathogen is distributed in several wild NHPs in Taï National Park. In Chapter 6, I expand this work and use this method to test multiple NHP species and find that *T. pallidum* made a recent radiation across NHPs and humans, with major implications for global yaws

eradication. That insights into a reservoir for this pathogen that infects thousands of people annually and has been the focus of huge global eradication efforts is only now being described, highlights the infancy of our understanding of wildlife pathogens.

As it became clear to the world that there was an Ebola virus disease outbreak in West Africa, I happened to be about 150 km from the index village; as part of an interdisciplinary team, I worked as an equally contributing first-author on the study that largely ruled out the possibility of a concurrent large-scale outbreak in NHPs in the region as the source of the current epidemic (Marí Saéz *et al. 2015*). Rather, environmental DNA from a burned tree where the index-case reportedly played, suggested contact with an insectivorous bat species known to survive experimental infections may have been responsible for the spillover (Marí Saéz *et al. 2015*). We subsequently revaluated the evidence supporting fruit bats as the sole wildlife reservoir of Ebolaviruses (Leendertz *et al.* 2015), and have been broadening our search for potential Ebolavirus reservoirs and developing education materials and emergence mitigation strategies. Again, our knowledge about the ecology of this pathogen is rudimentary, making prevention efforts daunting. These efforts, in part, motivated my shifting focus to studying potentially zoonotic pathogens in Chapters 5 and 6.

My thesis has examined how red colobus populations may be adjusting to environmental change by changing their group sizes in response to major changes in the availability of food (Chapter 2). These changes in red colobus group size sizes across the park have required individuals to make behavioural shifts to cope with the changing competitive regimes in theirs groups, including increased travel, decreased time spent feeding, and increased time spent in polyspecific associations (Chapter 3). Interestingly these changes appear to allow animals to successfully compensate for the changing environment, as we saw no major shift in individual fitness. To understand how these changes might be impacting the distribution of microorganisms, I examined the gut microbiome of a community of NHPs at a number of scales; from the stability of the individual, to the impact of social groups and social networks, to between species differences and showing a link between chimpanzee hunting of colobines and their gut microbiomes (Chapter 4). These results suggest that the changing group sizes observed in Kibale National Park might be changing the diversity of gut microbiomes in these populations; individuals acquire similar microbiomes to their group mates and thus larger social groups likely mean that there will be less standing bacterial diversity at the landscape scale. Similarly, large

groups might be subject to less successful predation rates by chimpanzees; the changing red colobus group sizes might ultimately be changing the gut microbiome communities found in chimpanzee populations. To understand how these changes might be impacting pathogens in this ecosystem, I first developed a new method to non-invasively sequence the genome of an elusive pathogen, *Treponema pallidum*. Using this approach on bones, I show that chimpanzees and red colobus are infected with this pathogen that causes yaws and syphilis in humans (Chapter 5). I then scale this approach up to study this pathogen in multiple NHP populations across sub-Saharan Africa. I found that a number of African NHPs that exhibit yaws- and syphilis-like symptoms in the wild, are infected with treponemes. Surprisingly, full genomes show that all NHP populations sampled are infected with a pathogen closely related to human yaws and that these pathogens radiated quickly across humans and NHPs and likely jumped repeatedly between their primate hosts (Chapter 6). NHPs represent a high-diversity reservoir for this pathogen, which likely impacts ongoing yaws eradication efforts. The changing behaviour observed in Chapters 2 and 3, which in turn likely impact the distribution of microorganisms might have major impacts on the prevalence of this pathogen. This in turn could greatly influence disease risk for humans living in proximity to these NHP populations, suggesting that projects are need to minimize this risk and understand changing disease emergence processes. Collectively, these studies suggest that changing environments might have major cascading impacts not only on wildlife populations, but also for humans living in proximity to these populations.

These experiences prompted me to become involved in a small scale Health Education Project in the Taï Region, where we raised funds and supplies and created basic education materials about the hygiene and disease prevention for K-12 students; I became convinced that promoting human healthcare can change conservation outcomes (Chapman *et al.* 2015c), but frustrated at our ignorance involving the distribution and ecology of wildlife pathogens and only being involved in small-scale mitigation efforts. These experiences during the time writing my dissertation have reinvigorated my interest in collaborating with large-scale human health projects to enable an understanding of factors influencing transmission risk and conservation at the human-wildlife interface. Rather than simply documenting areas of transmission or areas with high disease prevalence, I am excited to begin exploring solutions. The complicated relationships between humans, their livestock, wildlife, and their environments have given rise to the Planetary Health paradigm (Whitmee *et al.* 2015). This paradigm calls for interdisciplinary collaborative efforts to

optimize health for people, animals and the environment. While intriguing, it has proven difficult to operationalize these ideas; a quantitative framework for testing whether intervention strategies facilitate the health of all of these parties is currently lacking and these complex systems often react in unexpected ways. The ways in which anthropogenic changes impact disease dynamics of a landscape thus represents a critical area of future research, especially in light of current and predicted global population growth and climate change. This highlights the need for broader monitoring and research to enable the development of informed public health and conservation strategies incorporating disease mitigation (Jones *et al.* 2008; Lafferty 2009). The results of this thesis suggest that monitoring wildlife and humans at this interface is critical for preventing further zoonotic emergence events and for conservation efforts.

Efforts to alleviate extreme poverty and improve global human health are rooted in a philosophy of human rights (London 2008). These efforts have made tremendous progress in the last several decades and it has been argued that more people are living longer, have more access to education, clean water, and basic sanitation; this may in part be driven by higher per capita incomes (United Nations Development Programme 2015); ambitious efforts are underway to scale up these projects to fully eliminate extreme poverty and increase lifespans globally. As the global human population rises above 7.3 billion, the cost of these improvements has shifted disproportionately to wildlife and ecosystems (Whitmee *et al.* 2015). While difficult to measure, there is growing consensus that humans have driven increased rates of wildlife extinction well beyond historic background rates (Pereira *et al.* 2010). This raises the important question: how should animal and ecosystem rights factor into our efforts to alleviate extreme poverty and improve global human health? While beyond the scope of this dissertation, I believe this represents a critical discussion we as a species must undertake as we move forward into the future. This thesis suggests that small changes to the environment can have cascading impacts on wildlife populations, reiterating the need for caution.

Determining how human healthcare interventions and improving livelihoods interact with increasing anthropogenic disturbance to impact ecosystem health and services is a critical area of future research. Reducing human and livestock disease burdens is predicted to reduce transmission risk to wildlife populations, though this has never been tested on a large scale. Increasing access to clean water sources and latrines should reduce microorganism spillover rates. Increasing socioeconomic wellbeing might decrease the need for bushmeat hunting as a

source of nutrition (Golden *et al.* 2011), thus further decreasing spillover. However, in urban areas bushmeat consumption increases with wealth, so it is difficult to predict how practices change (Brashares *et al.* 2011). Reforestation projects and multi-use zones might mitigate pressure on protected areas, but could also serve as areas of spillover. Lastly, as NHP populations decline, pathogens might go extinct or become more virulent or prevalent in high-density fragments. As the world population grows, addressing these questions is critical for mitigating impacts of humans and wildlife on one another. Large-scale public health interventions provide a test case to observe how such initiatives impact the microorganism community and health of NHPs.

Since my first experiences conducting fieldwork in the tropics examining the impacts of environmental changes, disease has remained an omnipresent topic; seemingly weekly funerals of youth in remote communities, and co-workers and I falling ill for days, not being able to walk around, let alone work. My animal study subjects suffered too; slow days and unpleasant symptoms were common, but disease also represented the largest known source of mortality at all of my field sites. The close evolutionary relationship and similar physiology of NHPs and humans make them particularly important for understanding disease emergence and transmission in both directions. This combined with the availability of high-quality long-term data from multiple ecosystems drove me to begin studying wild NHP populations and ultimately, to pursue this dissertation. This thesis provided the opportunity to link changes in the environment to changes in animal behaviour, and then to examine how these changes might be influencing microorganisms in NHP populations and ultimately the abundance of potential zoonotic pathogens. I hope the reader has enjoyed this thesis to even a small fraction of the extent that I myself have. As the German quote starting this concluding chapter highlights, I will never be done learning, and this dissertation has provided a critical stepping stone in my development as a scientist and researcher for which I will be eternally grateful.

**COMPLETE LIST OF REFERENCES**


Abdolrasouli, A., Croucher, A., Hemmati, Y. & Mabey, D. (2013). A case of endemic syphilis, Iran. *Emerging Infectious Diseases*, 19, 162.

Agresti, A. (2002). *Categorical data analysis*. Wiley-Interscience, Hoboken, NJ.

Aivelo, T., Laakkonen, J. & Jernvall, J. (2016). Population-and Individual-Level Dynamics of the Intestinal Microbiota of a Small Primate. *Applied and Environmental Microbiology*, 82, 3537-3545.

Alexander, R.D. (1974). The evolution of social behaviour. *Annual Review of Ecology and Systematics*, 5, 325-382.

Ali, M., Taylor, G.P., Pitman, R.J., Parker, D., Rethwilm, A., Cheingsong-Popov, R., Weber, J.N., Bieniasz, P.D., Bradley, J. & McClure, M.O. (1996). No evidence of antibody to human foamy virus in widespread human populations. *AIDS Research and Human Retroviruses*, 12, 1473-1483.

Altizer, S., Nunn, C.L. & Lindenfors, P. (2007). Do threatened hosts have fewer parasites? A comparative study in primates. *Journal of Animal Ecology*, 76, 304-314.

Altizer, S., Nunn, C.L., Thrall, P.H., Gittleman, J.L., Antonovics, J., Cunningham, A.A., Dobson, A.P., Ezenwa, V. & Jones, K.E. (2003). Social organization and parasite risk in mammals: integrating theory and empirical studies. *Annual Review of Ecology, Evolution, and Systematics*, 34, 517-547.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.

Amato, K.R., Leigh, S.R., Kent, A., Mackie, R.I., Yeoman, C.J., Stumpf, R.M., Wilson, B.A., Nelson, K.E., White, B.A. & Garber, P.A. (2014). The gut microbiota appears to compensate for seasonal diet variation in the wild black howler monkey (*Alouatta pigra*). *Microbial Ecology*, 69, 434-443.

Amato, K.R., Martinez-Mota, R., Righini, N., Raguet-Schofield, M., Corcione, F.P., Marini, E., Humphrey, G., Gogul, G., Gaffney, J. & Lovelace, E. (2015). Phylogenetic and ecological factors impact the gut microbiota of two Neotropical primate species. *Oecologia*, 1-17.

Amato, K.R., Yeoman, C.J., Kent, A., Righini, N., Carbonero, F., Estrada, A., Gaskins, H.R., Stumpf, R.M., Yildirim, S. & Torralba, M. (2013). Habitat degradation impacts black

howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. *The ISME Journal*, 7, 1344-1353.

Anderson, R.M. (1978). The regulation of host population growth by parasitic species. *Parasitology*, 76, 119-157.

Angedakin, S. & Lwanga, J.S. (2011). Changes in group size and composition of the blue monkeys (*Cercopithecus mitis stuhlmanni*) between 1984 and 2009 at Ngogo, Kibale National Park, Uganda. *African Journal of Ecology*, 49, 362-366.

Anthony, S.J., Islam, A., Johnson, C., Navarrete-Macias, I., Liang, E., Jain, K., Hitchens, P.L., Che, X., Soloyvov, A. & Hicks, A.L. (2015). Non-random patterns in viral diversity. *Nature Communications*, 6, 8147.

Arnold, C., Matthews, L.J. & Nunn, C.L. (2010). The 10kTrees website: a new online resource for primate phylogeny. *Evolutionary Anthropology*, 19, 114-118.

Arora, N., Schuenemann, V.J.J., Jäger, G., Peltzer, A., Seitz, A., Herbig, A., Strouhal, M., Grillová, L., Sánchez-Busó, L., Kühnert, D., Bos, K.I.I., Rivero Davis, L.R., Mikalová, L., Bruisten, S., Komericki, P., French, P., Grant, P.R.R., Pando, M.A., Gallo Vaulet, L., Rodríguez-Fermepin, M., Martinez, A., Centurión-Lara, A., Giacani, L., Norris, S.J.J., Šmajs, D., Bosshard, P.P.P., González-Candelas, F., Nieselt, K., Krause, J. & Bagheri, H.C.C. (2016). Origin of modern syphilis and emergence of a contemporary pandemic cluster. *bioRxiv*, 051037.

Ayouba, A., Akoua-Koffi, C., Calvignac-Spencer, S., Esteban, A., Locatelli, S., Li, H., Li, Y., Hahn, B.H., Delaporte, E., Leendertz, F.H. & Peeters, M. (2013). Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Cote d'Ivoire. *AIDS*, 27, 2488-2491.

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Bailey, A.L., Lauck, M., Weiler, A., Sibley, S.D., Dinis, J.M., Bergman, Z., Nelson, C.W., Correll, M., Gleicher, M. & Hyeroba, D. (2014). High genetic diversity and adaptive potential of two simian hemorrhagic fever viruses in a wild primate population. *PLoS ONE*, 9, e90714.

Baker, B. & Armelagos, G. (1988). The origin and antiquity of syphilis: paleopathological diagnosis and interpretation. *Current Anthropology*, 29, 703.

Barelli, C., Albanese, D., Donati, C., Pindo, M., Dallago, C., Rovero, F., Cavalieri, D., Tuohy, K.M., Hauffe, H.C. & De Filippo, C. (2015). Habitat fragmentation is associated to gut microbiota diversity of an endangered primate: implications for conservation. *Scientific Reports*, 5, 14862.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4 *Journal of Statistical Software*, 67, 1-48.

Baylet, R., Thivolet, J., Sepetjian, M., Nouhouay, Y. & Baylet, M. (1970). Natural open treponematosis in the *Papio papio* baboon in Casamance. *Bulletin de la Societe de Pathologie Exotique et de ses Filiales*, 64, 842-846.

Beehner, J.C., Onderdonk, D.A., Alberts, S.C. & Altmann, J. (2006). The ecology of conception and pregnancy failure in wild baboons. *Behavioral Ecology*, 17, 741-750.

Betsem, E., Rua, R., Tortevoye, P., Froment, A. & Gessain, A. (2011). Frequent and recent human acquisition of simian foamy viruses through apes' bites in central Africa. *PLoS Pathogens*, 7, e1002306.

Bittar, F., Keita, M.B., Lagier, J.-C., Peeters, M., Delaporte, E. & Raoult, D. (2014). *Gorilla gorilla gorilla* gut: a potential reservoir of pathogenic bacteria as revealed using culturomics and molecular tools. *Scientific Reports*, 4, 1-5.

Blasse, A., Calvignac-Spencer, S., Merkel, K., Goffe, A.S., Boesch, C., Mundry, R. & Leendertz, F.H. (2013). Mother-offspring transmission and age-dependent accumulation of simian foamy virus in wild chimpanzees. *Journal of Virology*, 87, 5193-5204.

Boesch, C. (1994). Chimpanzees-red colobus monkeys: a predator-prey system. *Animal Behaviour*, 47, 1135-1148.

Boesch, C. & Achermann, H.B. (2000). *The chimpanzees of the Taï forest: behavioural ecology and evolution*. Oxford University Press, Oxford.

Boesch, C. & Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Tai National Park. *American Journal of Physical Anthropology*, 78, 547-573.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A. & Caporaso, J.G. (2012). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10, 57-59.

Bolzoni, L. & De Leo, G.A. (2013). Unexpected consequences of culling on the eradication of wildlife diseases: the role of virulence evolution. *The American Naturalist*, 181, 301-313.

Bongers, F., Poorter, L., Hawthorne, W.D. & Sheil, D. (2009). The intermediate disturbance hypothesis applies to tropical forests, but disturbance contributes little to tree diversity. *Ecology Letters*, 12, 798-805.

Bonnell, T.R., Sengupta, R.R., Chapman, C.A. & Goldberg, T.L. (2010). An agent-based model of red colobus resources and disease dynamics implicates key resource sites as hot spots of disease transmission. *Ecological Modelling*, 221, 2491-2500.

Bordenstein, S.R. & Theis, K.R. (2015). Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biology*, 13, e1002226.

Borries, C., Larney, E., Lu, A., Ossi, K. & Koenig, A. (2008). Costs of group size: lower developmental and reproductive rates in larger groups of leaf monkeys. *Behavioral Ecology*, 19, 1186-1194.

Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., Wood, J., Earn, D.J.D., Herring, D.A., Bauer, P., Poinar, H.N. & Krause, J. (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature*, 478, 506-510.

Bouwman, A.S., Kennedy, S.L., Müller, R., Stephens, R.H., Holst, M., Caffell, A.C., Roberts, C.A. & Brown, T.A. (2012). Genotype of a historic strain of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 109, 18511-18516.

Brashares, J.S., Golden, C.D., Weinbaum, K.Z., Barrett, C.B. & Okello, G.V. (2011). Economic and geographic drivers of wildlife consumption in rural Africa. *Proceedings of the National Academy of Sciences*, 108, 13931-13936.

Bray, J.R. & Curtis, J.T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27, 325-349.

Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T. & Lachmann, M. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, 104, 14616-14621.

Brncic, T.M., Willis, K.J., Harris, D.J. & Washington, R. (2007). Culture or climate? The relative influences of past processes on the composition of the lowland Congo rainforest. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 229-242.

Brown, C.R. & Brown, M.B. (1996). *Coloniality in the cliff swallow: the effect of group size on social behavior*. University of Chicago Press, Chicago.

Butynski, T.M. (1990). Comparative Ecology of Blue Monkeys (*Cercopithecus mitis*) in High-Density and Low-Density Subpopulations. *Ecological Monographs*, 60, 1-26.

Caillaud, D., Craft, M.E. & Meyers, L.A. (2013). Epidemiological effects of group size variation in social species. *Journal of the Royal Society Interface*, 10, 20130206.

Calattini, S., Nerrienet, E., Mauclère, P., Georges-Courbot, M.C., Saïb, A. & Gessain, A. (2004). Natural simian foamy virus infection in wild-caught gorillas, mandrills and drills from Cameroon and Gabon. *Journal of General Virology*, 85, 3313-3317.

Calvignac-Spencer, S., Adjogoua, E.V., Akoua-Koffi, C., Hedemann, C., Schubert, G., Ellerbrok, H., Leendertz, S.A.J., Pauli, G. & Leendertz, F.H. (2012a). Origin of human T-lymphotropic virus type 1 in rural Côte d'Ivoire. *Emerging Infectious Diseases*, 18, 830-833.

Calvignac-Spencer, S., Leendertz, S.A.J., Gillespie, T.R. & Leendertz, F.H. (2012b). Wild great apes as sentinels and sources of infectious disease. *Clinical Microbiology and Infection*, 18, 521-527.

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P.M., Metzger, S., Schubert, G. & Leendertz, F.H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, 22, 915-924.

Cameron, C.E., Lukehart, S.A., Castro, C., Molini, B., Godornes, C. & Van Voorhis, W.C. (2000). Opsonic potential, protective capacity, and sequence conservation of the *Treponema pallidum* subspecies *pallidum* Tp92. *Journal of Infectious Diseases*, 181, 1401-1413.

Campbell, G., Kuehl, H., Diarrassouba, A., N'Goran, P.K. & Boesch, C. (2011). Long-term research sites as refugia for threatened and over-harvested species. *Biology Letters*, 7, 723-726.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G. & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, doi:10.1038/ismej.2012.1038.

Cardiff, R.D., Ward, J.M. & Barthold, S.W. (2007). 'One medicine—one pathology': are veterinary and human pathology prepared? *Laboratory Investigation*, 88, 18-26.

Cardillo, M., Mace, G.M., Jones, K.E., Bielby, J., Bininda-Emonds, O.R.P., Sechrest, W., Orme, C.D.L. & Purvis, A. (2005). Multiple causes of high extinction risk in large mammal species. *Science*, 309, 1239-1241.

Carpenter, J.P. (2007). Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60, 31-51.

Catchpole, W. & Wheeler, J. (1992). Estimating plant biomass: a review of techniques. *Australian Journal of Ecology*, 17, 121-131.

Cejkova, D., Zobanikova, M., Chen, L., Pospisilova, P., Strouhal, M., Qin, X., Mikalova, L., Norris, S.J., Muzny, D.M., Gibbs, R.A., Fulton, L.L., Sodergren, E., Weinstock, G.M. & Smajs, D. (2012). Whole genome sequences of three Treponema pallidum ssp. pertenue strains: yaws and syphilis treponemes differ in less than 0.2% of the genome sequence. *PLOS Neglected Tropical Diseases*, 6, e1471.

Chapman, C.A., Balcomb, S.R., Gillespie, T., Skorupa, J. & Struhsaker, T.T. (2000). Long-term effects of logging on African primate communities: a 28 year comparison from Kibale National Park, Uganda. *Conservation Biology*, 14, 207-217.

Chapman, C.A., Bonnell, T.R., Gogarten, J.F., Lambert, J.E., Omeja, P.A., Twinomugisha, D., Wasserman, M.D. & Rothman, J.M. (2013). Are Primates Ecosystem Engineers? *International Journal of Primatology*, 34, 1-14.

Chapman, C.A., Bonnell, T.R., Schoof, V.A.M., Gogarten, J.F. & Calmé, S. (2015a). Competing pressures on populations: long-term dynamics of food availability, food quality, disease, stress, and animal abundance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140112.

Chapman, C.A., Bowman, D.D., Ghai, R.R., Gogarten, J.F., Goldberg, T.L., Rothman, J.M., Twinomugisha, D. & Walsh, C. (2012). Protozoan parasites in group-living primates: testing the biological island hypothesis. *American Journal of Primatology*, 74, 510-517.

Chapman, C.A. & Chapman, L.J. (1997). Forest regeneration in logged and unlogged forests of Kibale National Park, Uganda. *Biotropica*, 29, 396-412.

Chapman, C.A. & Chapman, L.J. (2000). Determinants of group size in primates: the importance of travel costs. In: *On the move: how and why animals travel in groups* (eds. Boinski, S & Garber, PA). University of Chicago Press Chicago, pp. 24-41.

Chapman, C.A. & Chapman, L.J. (2002). Foraging challenges of red colobus monkeys: influence of nutrients and secondary compounds. *Comparative Biochemistry and Physiology. Part A, Physiology*, 133, 861-875.

Chapman, C.A., Chapman, L.J., Bjorndal, K.A. & Onderdonk, D.A. (2002). Application of protein-to-fiber ratios to predict colobine abundance on different spatial scales. *International Journal of Primatology*, 23, 283-310.

Chapman, C.A., Chapman, L.J., Jacob, A.L., Rothman, J.M., Omeja, P., Reyna-Hurtado, R., Hartter, J. & Lawes, M.J. (2010a). Tropical tree community shifts: implications for wildlife conservation. *Biological Conservation*, 143, 366-374.

Chapman, C.A., Chapman, L.J., Naughton-Treves, L., Lawes, M.J. & McDowell, L.R. (2004). Predicting folivorous primate abundance: validation of a nutritional model. *American Journal of Primatology*, 62, 55-69.

Chapman, C.A., Chapman, L.J., Rode, K.D., Hauck, E.M. & McDowell, L.R. (2003). Variation in the nutritional value of primate foods: among trees, time periods, and areas. *International Journal of Primatology*, 24, 317-333.

Chapman, C.A., Chapman, L.J., Struhsaker, T.T., Zanne, A.E., Clark, C.J. & Poulsen, J.R. (2005). A long-term evaluation of fruiting phenology: importance of climate change. *Journal of Tropical Ecology*, 21, 31-45.

Chapman, C.A., Chapman, L.J. & Wrangham, R. (1995). Ecological constraints on group size: an analysis of spider monkey and chimpanzee subgroups. *Behavioral Ecology and Sociobiology*, 36, 59-70.

Chapman, C.A. & Lambert, J.E. (2000). Habitat alteration and the conservation of African primates: case study of Kibale National Park, Uganda. *American Journal of Primatology*, 50, 169-185.

Chapman, C.A. & Pavelka, M.S.M. (2005). Group size in folivorous primates: ecological constraints and the possible influence of social factors. *Primates*, 46, 1-9.

Chapman, C.A., Rothman, J.M. & Hodder, S.A.M. (2009). Can parasites limit primate group size? A test with red colobus. In: *Primate parasite ecology: the dynamics and study of host-parasite relationships.* (eds. Huffman, MA & Chapman, CA). Cambridge University Press Cambridge, pp. 423-440.

Chapman, C.A., Schoof, V.A., Bonnell, T.R., Gogarten, J.F. & Calmé, S. (2015b). Competing pressures on populations: long-term dynamics of food availability, food quality, disease, stress and animal abundance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140112.

Chapman, C.A., Struhsaker, T.T., Skorupa, J.P., Snaith, T.V. & Rothman, J.M. (2010b). Understanding long-term primate community dynamics: implications of forest change. *Ecological Applications*, 20, 179-191.

Chapman, C.A., van Bavel, B., Boodman, C., Ghai, R.R., Gogarten, J.F., Hartter, J., Mechak, L.E., Omeja, P.A., Poonawala, S., Tuli, D. & Goldberg, T.L. (2015c). Providing health care to improve community perceptions of protected areas. *Oryx*, 49, 636-642.

Chapman, C.A., Wasserman, M.D., Gillespie, T.R., Speirs, M.L., Lawes, M.J., Saj, T.L. & Ziegler, T.E. (2006). Do nutrition, parasitism, and stress have synergistic effects on red colobus populations living in forest fragments? *American Journal of Physical Anthropology*, 131, 525-534.

Chapman, C.A., Wrangham, R.W. & Chapman, L.J. (1994). Indices of habitat-wide fruit abundance in tropical forest. *Biotropica*, 26, 160-171.

Charleston, M.A. & Robertson, D.L. (2002). Preferential Host Switching by Primate Lentiviruses Can Account for Phylogenetic Similarity with the Primate Phylogeny. *Systematic Biology*, 51, 528-535.

Cheney, D.L. & Wrangham, R.W. (1987). Predation. In: *Primate societies* (eds. Smuts, BB, Cheney, DL, Seyfarth, RM, Wrangham, RW & Struhsaker, TT). University of Chicago Press Chicago, pp. 227-239.

Chevalier, C., Stojanović, O., Colin, D.J., Suarez-Zamorano, N., Tarallo, V., Veyrat-Durebex, C., Rigo, D., Fabbiano, S., Stevanović, A. & Hagemann, S. (2015). Gut microbiota orchestrates energy homeostasis during cold. *Cell*, 163, 1360-1374.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*.

Clayton, J.B., Vangay, P., Huang, H., Ward, T., Hillmann, B.M., Al-Ghalith, G.A., Travis, D.A., Long, H.T., Van Tuan, B. & Van Minh, V. (2016). Captivity humanizes the primate microbiome. *Proceedings of the National Academy of Sciences*, 201521835.

Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., Magris, M., Hidalgo, G., Contreras, M. & Noya-Alarcón, Ó. (2015). The microbiome of uncontacted Amerindians. *Science Advances*, 1, e1500183.

Clutton-Brock, T.H. & Harvey, P.H. (1977). Primate ecology and social organization. *Journal of Zoological Society of London*, 183, 1-39.

Collias, N. & Southwick, C. (1952). A field study of population density and social organization in howling monkeys. *Proceedings of the National Academy of Sciences*, 96, 143-156.

Collins, S.M., Surette, M. & Bercik, P. (2012). The interplay between the intestinal microbiota and the brain. *Nature Reviews Microbiology*, 10, 735-742.

Compton, A.A. & Emerman, M. (2013). Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. *PLoS Pathogens*, 9, e1003135.

Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J. & Relman, D.A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science*, 336, 1255-1262.

Cote, I.M. & Poulin, R. (1995). Parasitism and group size in social animals: a meta-analysis. *Behavioral Ecology*, 6, 159-165.

Courgnaud, V., Formenty, P., Akoua-Koffi, C., Noe, R., Boesch, C., Delaporte, E. & Peeters, M. (2003). Partial molecular characterization of two simian immunodeficiency viruses (SIV) from African colobids: SIVwrc from Western red colobus (*Piliocolobus badius*) and SIVolc from olive colobus (*Procolobus verus*). *Journal of Virology*, 77, 744-748.

Cowlishaw, G. (1999). Predicting the pattern of decline of African primate diversity: an extinction debt from historical deforestation. *Conservation Biology*, 13, 1183-1193.

Creel, S. & MacDonald, D. (1995). Sociality, group size, and reproductive suppression among carnivores. *Advances in the Study of Behavior*, 34, 203-257.

Croft, D.P., James, R. & Krause, J. (2008). *Exploring animal social networks*. Princeton University Press, Princeton.

Danish, L., Chapman, C.A., Hall, M.B., Rode, K.D. & Worman, C. (2006). The role of sugar in diet selection in redtail and red colobus monkeys. In: *Feeding ecology in apes and other primates: ecological, physical, and behavioural aspects* (eds. Hohmann, G, Robbins, MM & Boesch, C). Cambridge University Press Cambridge, pp. 471-487.

Darling, A.C., Mau, B., Blattner, F.R. & Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14, 1394-1403.

Daszak, P., Cunningham, A.A. & Hyatt, A.D. (2000). Emerging infectious diseases of wildlife--threats to biodiversity and human health. *Science*, 287, 443.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J. & Turnbaugh, P.J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505, 559-563.

Davies, C.R., Ayres, J.M., Dye, C. & Deane, L.M. (1991). Malaria infection rate of Amazonian primates increases with body weight and group size. *Functional Ecology*, 5, 655-662.

De Nys, H.M., Madinda, N.F., Merkel, K., Robbins, M., Boesch, C., Leendertz, F.H. & Calvignac-Spencer, S. (2015). A cautionary note on fecal sampling and molecular epidemiology in predatory wild great apes. *American Journal of Primatology*, 77, 833-840.

Degnan, P.H., Pusey, A.E., Lonsdorf, E.V., Goodall, J., Wroblewski, E.E., Wilson, M.L., Rudicell, R.S., Hahn, B.H. & Ochman, H. (2012). Factors associated with the diversification of the gut microbial communities within chimpanzees from Gombe National Park. *Proceedings of the National Academy of Sciences*, 109, 13034-13039.

Delm, M.M. (1990). Vigilance for predators: detection and dilution effects. *Behavioral Ecology and Sociobiology*, 26, 337-342.

Demment, M.W. & Van Soest, P.J. (1985). A nutritional explanation for body-size patterns of ruminants and nonruminant herbivores. *The American Naturalist*, 125, 641-672.

Dent, D.H., DeWalt, S.J. & Denslow, J.S. (2013). Secondary forests of central Panama increase in similarity to old-growth forest over time in shade tolerance but not species composition. *Journal of Vegetation Science*, 24, 530-542.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72, 5069-5072.

Devault, A.M., Golding, G.B., Waglechner, N., Enk, J.M., Kuch, M., Tien, J.H., Shi, M., Fisman, D.N., Dhody, A.N. & Forrest, S. (2014). Second-pandemic strain of *Vibrio*

*cholerae* from the Philadelphia cholera outbreak of 1849. *New England Journal of Medicine*, 370, 334-340.

Dheilly, N.M. (2014). Holobiont-holobiont interactions: redefining host-parasite interactions. *PLoS Pathogens*, 10, e1004093.

Dobson, A.J. & Barnett, A. (2008). *An introduction to generalized linear models*. CRC press.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences*, 107, 11971-11975.

Domínguez-Rodrigo, M. (2002). Hunting and scavenging by early humans: the state of the debate. *Journal of World Prehistory*, 16, 1-54.

Dowle, M., Srinivasan, A., Short, T. & Lianoglou, S. (2015). data. table: Extension of data.frame R package version 1.9.6. https://cran.R-project.or/package=data.table.

Eggeling, W.J. (1947). Observations on the ecology of the Budongo rain forest, Uganda. *Journal of Ecology*, 34, 20-87.

Elgar, M.A. (1989). Predator vigilance and group size in mammals and birds: a critical review of the empirical evidence. *Biological Reviews*, 64, 13-33.

Elsas, F.J., Smith, J.L., Israel, C.W. & Gager, W.E. (1968). Late syphilis in the primate. *The British Journal of Venereal Diseases*, 44, 267-273.

Enquist, B.J. & Niklas, K.J. (2001). Invariant scaling relations across tree-dominated communities. *Nature*, 410, 655-660.

Enquist, B.J. & Niklas, K.J. (2002). Global allocation rules for patterns of biomass partitioning in seed plants. *Science*, 295, 1517-1520.

Eppinger, M., Baar, C., Linz, B., Raddatz, G., Lanz, C., Keller, H., Morelli, G., Gressmann, H., Achtman, M. & Schuster, S.C. (2006). Who ate whom? adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet*, 2, e120.

Erickson, D.L., Jones, F.A., Swenson, N.G., Pei, N., Bourg, N.A., Chen, W., Davies, S.J., Ge, X.-j., Hao, Z., Howe, R.W., Huang, C.-L., Larson, A.J., Lum, S.K.Y., Lutz, J.A., Ma, K., Meegaskumbura, M., Mi, X., Parker, J.D., Fang-Sun, I., Wright, S.J., Wolf, A.T., Ye, W., Xing, D., Zimmerman, J.K. & Kress, W.J. (2014). Comparative evolutionary diversity

and phylogenetic structure across multiple forest dynamics plots: a mega-phylogeny approach. *Frontiers in Genetics*, 5, 358.

Eubank, S., Guclu, H., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Toroczkai, Z. & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429, 180-184.

Ezenwa, V.O. (2004). Host social behavior and parasitic infection: a multifactorial approach. *Behavioral Ecology*, 15, 446-454.

Ezenwa, V.O. & Jolles, A.E. (2015). Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science*, 347, 175-177.

Ezenwa, V.O. & Williams, A.E. (2014). Microbes and animal olfactory communication: Where do we go from here? *BioEssays*, 36, 847-854.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L., Clemente, J.C., Knight, R., Heath, A.C. & Leibel, R.L. (2013). The long-term stability of the human gut microbiota. *Science*, 341, 1237439.

FAO (1997). Estimating biomass and biomass change of tropical forests: a primer. In: *Forest Resource Assessment Publication*. Food & Agriculture Org. Rome, p. 134.

Fashing, P.J. (2001). Male and female strategies during intergroup encounters in guerezas (*Colobus guereza*): evidence for resource defense mediated through males and a comparison with other primates *Behavioral Ecology and Sociobiology*, 50, 219-230.

Fashing, P.J., Mulindahabi, F., Gakima, J.-B., Masozera, M., Mununura, I., Plumptre, A.J. & Nguyen, N. (2007). Activity and ranging patterns of *Colobus angolensis ruwenzorii* in Nyungwe Forest, Rwanda: possible costs of large group size. *International Journal of Primatology*, 28, 529-550.

Fedigan, L. (1986). Demographic trends in the *Alouatta palliata* and *Cebus capucinus* populations of Santa Rosa National Park, Costa Rica. In: *Primate ecology and conservation* (eds. Else, JG & Lee, PC). Press Syndicate of the University of Cambridge New York, pp. 287-293.

Felsenstein, J. (1985). Confidence-limits on phylogenies - an approach using the bootstrap. *Evolution*, 39, 783-791.

Felsenstein, J. & Felenstein, J. (2004). *Inferring phylogenies*. Sinauer Associates Sunderland.

Felton, A.M., Felton, A., Raubenheimer, D., Simpson, S.J., Foley, W.J., Wood, J.T., Wallis, I.R. & Lindenmayer, D.B. (2009). Protein content of diets dictates the daily energy intake of a free-ranging primate. *Behavioral Ecology*, 20, 685-690.

Fogel, A.T. (2015). The gut microbiome of wild lemurs: a comparison of sympatric *Lemur catta* and *Propithecus verreauxi*. *Folia Primatologica*, 86, 85-95.

Forstmeier, W. & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, 65, 47-55.

Freeland, W.J. (1976). Pathogens and the evolution of primate sociality. *Biotropica*, 8, 12-24.

Fribourg-Blanc, A. & Mollaret, H.H. (1969). Natural treponematosis of the African primate. *Primates in Medicine*, 3, 113-121.

Fribourg-Blanc, A., Niel, G. & Mollaret, H. (1966). Confirmation serologique et microscopique de la treponemose du cynocephale de Guinee. *Bulletin de la Societe de Pathologie Exotique et de ses Filiales*, 59, 54-59.

Funkhouser, L.J. & Bordenstein, S.R. (2013). Mom Knows Best: The Universality of Maternal Microbial Transmission. *PLoS Biology*, 11, e1001631.

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A., Bradley, D.G. & Orlando, L. (2015). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*.

Ganzhorn, J.U. (2002). Distribution of a folivorous lemur in relation to seasonally varying food resources: integrating quantitative and qualitative aspects of food characteristics. *Oecologia*, 131, 427-435.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M. & Hahn, B.H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, 397, 436-441.

Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52, 203-223.

Gessain, A., Rua, R., Betsem, E., Turpin, J. & Mahieux, R. (2013). HTLV-3/4 and simian foamy retroviruses in humans: discovery, epidemiology, cross-species transmission and molecular virology. *Virology*, 435, 187-199.

Ghai, R.R., Chapman, C.A., Omeja, P.A., Davies, T.J. & Goldberg, T.L. (2014). Nodule worm infection in humans and wild primates in Uganda: cryptic species in a newly identified region of human transmission. *PLoS Neglected Tropical Diseases*, 8, e2641.

Giacani, L., Brandt, S.L., Ke, W., Reid, T.B., Molini, B.J., Iverson-Cabral, S., Ciccarese, G., Drago, F., Lukehart, S.A. & Centurion-Lara, A. (2015). Transcription of TP0126, *Treponema pallidum* putative OmpW homolog, is regulated by the length of a homopolymeric guanosine repeat. *Infection and Immunity*, 83, 2275-2289.

Giacani, L., Lukehart, S. & Centurion-Lara, A. (2007). Length of guanosine homopolymeric repeats modulates promoter activity of subfamily II tpr genes of *Treponema pallidum* ssp. *pallidum*. *FEMS Immunology and Medical Microbiology*, 51, 289-301.

Giacani, L. & Lukehart, S.A. (2014). The endemic treponematoses. *Clinical Microbiology Reviews*, 27, 89-115.

Gillespie, T.R., Greiner, E.C. & Chapman, C.A. (2005). Gastrointestinal parasites of the colobus monkeys of Uganda. *Journal of Parasitology*, 91, 569-573.

Gillespie, T.R., Nunn, C.L. & Leendertz, F.H. (2008). Integrative approaches to the study of primate infectious disease: implications for biodiversity conservation and global health. *American Journal of Physical Anthropology*, 137, 53-69.

Gillings, M.R., Paulsen, I.T. & Tetu, S.G. (2015). Ecology and Evolution of the Human Microbiota: Fire, Farming and Antibiotics. *Genes*, 6, 841-857.

Giraldeau, L.A. (1988). The stable group and the determinants of foraging group size. In: *The ecology of social behavior* (ed. Slobodchikoff, CH). Academic Press New York, pp. 33-53.

Godfrey, L.R. & Irwin, M.T. (2007). The evolution of extinction risk: past and present anthropogenic impacts on the primate communities of Madagascar. *Folia Primatologica*, 78, 405-419.

Goffe, A.S., Blasse, A., Mundry, R., Leendertz, F.H. & Calvignac-Spencer, S. (2012). Detection of retroviral super-infection from non-invasive samples. *PLoS ONE*, 7, e36570.

Gogarten, J.F., Akoua-Koffi, C., Calvignac-Spencer, S., Leendertz, S.A.J., Weiss, S., Couacy-Hymann, E., Koné, I., Peeters, M., Wittig, R.M., Boesch, C., Hahn, B.H. & Leendertz, F.H. (2014a). The ecology of primate retroviruses–an assessment of 12 years of retroviral studies in the Taï National Park area, Côte d′Ivoire. *Virology*, 460, 147-153.

Gogarten, J.F., Bonnell, T.R., Brown, L.M., Campenni, M., Wasserman, M.D. & Chapman, C.A. (2014b). Increasing group size alters behavior of a folivorous primate. *International Journal of Primatology*, 35, 590-608.

Gogarten, J.F., Brown, L.M., Chapman, C.A., Marina, C., Doran-Sheehy, D., Fedigan, L.M., Grine, F.E., Perry, S., Pusey, A.E., Sterck, E.H.M., Wich, S.A. & Wright, P.C. (2012a). Seasonal mortality patterns in non-human primates: implications for variation in selection pressures across environments. *Evolution*, 66, 3252–3266.

Gogarten, J.F., Düx, A., Schuenemann, V.J., Nowak, K., Boesch, C., Wittig, R.M., Krause, J., Calvignac-Spencer, S. & Leendertz, F.H. (2016). Tools for opening new chapters in the book of *Treponema pallidum* evolutionary history. *Clinical Microbiology and Infection*, 22, 916-921.

Gogarten, J.F., Guzman, M., Chapman, C.A., Jacob, A.L., Omeja, P.A. & Rothman, J.M. (*2012b*). What is the predictive power of the colobine protein-to-fiber model and its conservation value? *Tropical Conservation Science*, 5, 381-393.

Gogarten, J.F., Jacob, A.L., Ghai, R.R., Rothman, J.M., Twinomugisha, D., Wasserman, M.D. & Chapman, C.A. (2015). Group size dynamics over 15+ years in an African forest primate community. *Biotropica*, 47, 101-112.

Gogarten, J.P., Doolittle, W.F. & Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution*, 19, 2226-2238.

Goldberg, T.L., Chapman, C.A., Cameron, K., Saj, T., Karesh, W.B., Wolfe, N.D., Wong, S.W., Dubois, M.E. & Slifka, M.K. (2008). Serologic evidence for novel poxvirus in endangered red colobus monkeys, western Uganda. *Emerging Infectious Diseases*, 14, 801-803.

Goldberg, T.L., Sintasath, D.M., Chapman, C.A., Cameron, K.M., Karesh, W.B., Tang, S., Wolfe, N.D., Rwego, I.B., Ting, N. & Switzer, W.M. (2009). Coinfection of Ugandan red colobus (*Procolobus* [*Piliocolobus*] *rufomitratus tephrosceles*) with novel, divergent delta-,lenti-, and spumaretroviruses. *Journal of Virology*, 83, 11318-11329.

Golden, C.D., Fernald, L.C.H., Brashares, J.S., Rasolofoniaina, B.J.R. & Kremen, C. (2011). Benefits of wildlife consumption to child nutrition in a biodiversity hotspot. *Proceedings of the National Academy of Sciences*, 108, 19653-19656.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R. & Bell, J.T. (2014). Human genetics shape the gut microbiome. *Cell*, 159, 789-799.

Graber, J.R., Leadbetter, J.R. & Breznak, J.A. (2004). Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Applied and Environmental Microbiology*, 70, 1315-1320.

Gray, R., Mulligan, C., Molini, B., Sun, E., Giacani, L., Godornes, C., Kitchen, A., Lukehart, S. & Centurion-Lara, A. (2006). Molecular evolution of the tprC, D, I, K, G, and J genes in the pathogenic genus Treponema. *Molecular Biology and Evolution*, 23, 2220-2233.

Griffen, A.L., Beall, C.J., Campbell, J.H., Firestone, N.D., Kumar, P.S., Yang, Z.K., Podar, M. & Leys, E.J. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *The ISME journal*, 6, 1176-1185.

Griffin, R.H. & Nunn, C.L. (2012). Community structure and the spread of infectious disease in primate social networks. *Evolutionary Ecology*, 26, 779-800.

Griffiths, S.W. & Magurran, A.E. (1997). Schooling preferences for familiar fish vary with group size in a wild guppy population. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264, 547-551.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307-321.

Gulland, F.M.D. (1992). The role of nematode parasites in Soay sheep (*Ovis aries* L.) mortality during a population crash. *Parasitology Research*, 105, 493-503.

Hackett, C.J. (1963). On the origin of the human treponematoses (pinta, yaws, endemic syphilis and venereal syphilis). *Bulletin of the World Health Organization*, 29, 7-41.

Hahn, B.H., Shaw, G.M., De Cock, K.M. & Sharp, P.M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science*, 287, 607-617.

Hairston, N.G., Smith, F.E. & Slobodkin, L.B. (1960). Community structure, population control, and competition. *The American Naturalist*, 94, 421-424.

Hamilton, A., Taylor, D. & Vogel, J.C. (1986). Early forest clearance and environmental degradation in south-west Uganda. *Nature*, 320, 164-167.

Han, G.Z. & Worobey, M. (2012). An endogenous foamy-like viral element in the coelacanth genome. *PLoS Pathogens*, 8, e1002790.

Harper, K.N., Fyumagwa, R.D., Hoare, R., Wambura, P.N., Coppenhaver, D.H., Sapolsky, R.M., Alberts, S.C., Tung, J., Rogers, J. & Kilewo, M. (2012). *Treponema pallidum* infection in the wild baboons of East Africa: distribution and genetic characterization of the strains responsible. *PLoS ONE*, 7, e50882.

Harper, K.N., Liu, H., Ocampo, P.S., Steiner, B.M., Martin, A., Levert, K., Wang, D., Sutton, M. & Armelagos, G.J. (2008a). The sequence of the acidic repeat protein (arp) gene differentiates venereal from nonvenereal *Treponema pallidum* subspecies, and the gene has evolved under strong positive selection in the subspecies that causes syphilis. *FEMS Immunology and Medical Microbiology*, 53, 322-332.

Harper, K.N., Ocampo, P.S., Steiner, B.M., George, R.W., Silverman, M.S., Bolotin, S., Pillay, A., Saunders, N.J. & Armelagos, G.J. (2008b). On the origin of the treponematoses: a phylogenetic approach. *PLoS Negl Trop Dis*, 2, e148.

Harris, T.R. & Chapman, C.A. (2007). Variation in the diet and ranging behavior of black-and-white colobus monkeys: implications for theory and conservation. *Primates*, 28, 208-221.

Harris, T.R., Chapman, C.A. & Monfort, S.L. (2010). Small folivorous primate groups exhibit behavioral and physiological effects of food scarcity. *Behavioral Ecology*, 21, 46-56.

Hawthorne, W.D. (1996). Holes and the sums of parts in Ghanaian forest: regeneration, scale and sustainable use. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 104, 75-176.

Heg, D., Brouwer, L., Bachar, Z. & Taborsky, M. (2005). Large group size yields group stability in the cooperatively breeding cichlid *Neolamprologus pulcher*. *Behaviour*, 142, 11-12.

Hehemann, J.-H., Kelly, A.G., Pudlo, N.A., Martens, E.C. & Boraston, A.B. (2012). Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences*, 109, 19786-19791.

Hino, S., Katamine, S., Miyata, H., Tsuji, Y., Yamabe, T. & Miyamoto, T. (1996). Primary Prevention of HTLV-I in Japan. *Journal of Acquired Immune Deficiency Syndromes & Human Retrovirology*, 13, S199-S203.

Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Gordon, D.B., Brizuela, L., McCombie, W.R. & Hannon, G.J. (2009). Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature protocols*, 4, 960-974.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M. & Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics*, 2, 353-359.

Holling, C.S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4, 1-23.

Holmes, J.C. & Price, P.W. (1986). Communities of parasites. In: *Community ecology: pattern and process* (eds. Kikkawa, J & Anderson, DJ). Blackwell Scientific Oxford, pp. 187–213.

Hooper, L.V., Littman, D.R. & Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. *Science*, 336, 1268-1273.

Horner-Devine, M.C. & Bohannan, B.J.M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology*, 87, S100-S108.

Hudson, E.H. (1965). Treponematosis and man's social evolution. *American Anthropologist*, 67, 885-901.

Huson, D.H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254-267.

Irwin, M.T., Junge, R.E., Raharison, J.-L. & Samonds, K.E. (2010). Variation in physiological health of diademed sifakas across intact and fragmented forest at Tsinjoarivo, eastern Madagascar. *American Journal of Primatology*, 72, 1013-1025.

Isbell, L.A. (1991). Contest and scramble competition: patterns of female aggression and ranging behaviour among primates. *Behavioral Ecology*, 2, 143-155.

Isbell, L.A. (1994). Predation on primates: Ecological patterns and evolutionary consequences. *Evolutionary Anthropology: Issues, News, and Reviews*, 3, 61-71.

Isbell, L.A. (2012). Re-evaluating the Ecological Constraints model with red colobus monkeys (*Procolobus rufomitratus tephrosceles*). *Behaviour*, 149, 493-529.

Ives, A.R. & Carpenter, S.R. (2007). Stability and diversity of ecosystems. *Science*, 317, 58-62.

Janson, C.H. & Goldsmith, M.L. (1995). Predicting group size in primates: foraging costs and predation risks. *Behavioral Ecology*, 6, 326-336.

Janson, C.H. & van Schaik, C.P. (1988). Recognizing the many faces of primate food competition: methods. *Behaviour*, 105, 165-186.

Jones, C.G., Lawton, J.H. & Shachak, M. (1994). Organisms as ecosystem engineers. In: *Ecosystem management*. Springer, pp. 130-147.

Jones, C.G., Lawton, J.H. & Shachak, M. (1997). Positive and negative effects of organisms as physical ecosystem engineers. *Ecology*, 78, 1946-1957.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451, 990-994.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L. & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, btt193.

Junglen, S., Hedemann, C., Ellerbrok, H., Pauli, G., Boesch, C. & Leendertz, F.H. (2010). Diversity of STLV-1 strains in wild chimpanzees (*Pan troglodytes verus*) from Côte d'Ivoire. *Virus Research*, 150, 143-147.

Kappeler, P.M., Cremer, S. & Nunn, C.L. (2015). Sociality and health: impacts of sociality on disease susceptibility and transmission in animal and human societies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140116.

Kazadi, W.M., Asiedu, K.B., Agana, N., Mitjà, O. & Guinea, P.N. (2014). Epidemiology of yaws: an update. *Clinical Epidemiology*, 6, 119–128.

Keele, B.F., Jones, J.H., Terio, K.A., Estes, J.D., Rudicell, R.S., Wilson, M.L., Li, Y., Learn, G.H., Beasley, T.M., Schumacher-Stankey, J., Wroblewski, E., Mosser, A., Raphael, J., Kamenya, S., Lonsdorf, E.V., Travis, D.A., Mlengeya, T., Kinsel, M.J., Else, J.G., Silvestri, G., Goodall, J., Sharp, P.M., Shaw, G.M., Pusey, A.E. & Hahn, B.H. (2009). Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature*, 460, 515-519.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Peeters, M. & Hahn, B.H. (2006). Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 313, 523-526.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P. & Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463-1464.

Kembel, S.W. & Hubbell, S.P. (2006). The phylogenetic structure of a neotropical forest tree community. *Ecology*, 87.

Kembel, S.W., O'Connor, T.K., Arnold, H.K., Hubbell, S.P., Wright, S.J. & Green, J.L. (2014). Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proceedings of the National Academy of Sciences*, 111, 13715-13720.

Kircher, M., Sawyer, S. & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*, 40, e3.

Kirchhoff, F. (2010). Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host & Microbe*, 8, 55-67.

Knauf, S., Batamuzi, E.K., Mlengeya, T., Kilewo, M., Lejora, I.A., Nordhoff, M., Ehlers, B., Harper, K.N., Fyumagwa, R., Hoare, R., Failing, K., Wehrend, A., Kaup, F.J., Leendertz, F.H. & Matz-Rensing, K. (2012). Treponema infection associated with genital ulceration in wild baboons. *Veterinary Pathology*, 49, 292-303.

Knauf, S., Liu, H. & Harper, K.N. (2013). Treponemal infection in nonhuman primates as possible reservoir for human yaws. *Emerging Infectious Diseases*, 19, 2058.

Koenig, A. (2002). Competition for resources and its behavioral consequences among female primates. *International Journal of Primatology*, 23, 759-783.

Kolman, C.J., Centurion-Lara, A., Lukehart, S.A., Owsley, D.W. & Tuross, N. (1999). Identification of *Treponema pallidum* subspecies *pallidum* in a 200-year-old skeletal specimen. *Journal of Infectious Diseases*, 180, 2060-2063.

Köndgen, S., Kühl, H., N'Goran, P.K., Walsh, P.D., Schenk, S., Ernst, N., Biek, R., Formenty, P., Mätz-Rensing, K. & Schweiger, B. (2008). Pandemic human viruses cause decline of endangered great apes. *Current Biology*, 18, 260-264.

Krivatkin, S. & Krivatkina, E. (1997). Secondary syphilis transmitted through non-sexual contact. Three cases descriptions. *Journal of the European Academy of Dermatology and Venereology*, 9, S224.

Krüger, O., Affeldt, E., Brackmann, M. & Milhahn, K. (1998). Group size and composition of *Colobus guereza* in Kyambura Gorge, southwest Uganda, in relation to chimpanzee activity. *International Journal of Primatology*, 19, 287-297.

Krützen, M., Willems, E.P. & van Schaik, C.P. (2011). Culture and geographic variation in orangutan behavior. *Current Biology*, 21, 1808-1812.

Kuehl, H.S., Elzner, C., Moebius, Y., Boesch, C. & Walsh, P.D. (2008). The price of play: self-organized infant mortality cycles in chimpanzees. *PLoS ONE*, 3, e2440.

Lafferty, K. (2009). The ecology of climate change and infectious diseases. *Ecology*, 90, 888-900.

Lagier, J.-C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M. & Raoult, D. (2015). Current and past strategies for bacterial culture in clinical microbiology. *Clinical Microbiology Reviews*, 28, 208-236.

Lauck, M., Hyeroba, D., Tumukunde, A., Weny, G., Lank, S.M., Chapman, C.A., O'Connor, D.H., Friedrich, T.C. & Goldberg, T.L. (2011). Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan Red Colobus monkey discovered using direct pyrosequencing. *PLoS ONE*, 6, e19056.

Laws, R.M. (1970). Elephants as agents of habitat and landscape change in East Africa. *Oikos*, 21, 1-15.

Leendertz, F.H., Boesch, C., Ellerbrok, H., Rietschel, W., Couacy-Hymann, E. & Pauli, G. (2004). Non-invasive testing reveals a high prevalence of simian T-lymphotropic virus type 1 antibodies in wild adult chimpanzees of the Taï National Park, Côte d'Ivoire. *Journal of General Virology*, 85, 3305-3312.

Leendertz, F.H., Boesch, C., Junglen, S., Pauli, G. & Ellerbrok, H. (2003). Characterization of a new simian T-lymphotropic virus type 1 (STLV-1) in a wild living chimpanzee (*Pan troglodytes verus*) from Ivory Coast: evidence of a new STLV-1 group? *AIDS Research and Human Retroviruses*, 19, 255-258.

Leendertz, F.H., Pauli, G., Maetz-Rensing, K., Boardman, W., Nunn, C., Ellerbrok, H., Jensen, S.A., Junglen, S. & Christophe, B. (2006). Pathogens as drivers of population declines: the importance of systematic monitoring in great apes and other threatened mammals. *Biological Conservation*, 131, 325-337.

Leendertz, F.H., Zirkel, F., Couacy-Hymann, E., Ellerbrok, H., Morozov, V.A., Pauli, G., Hedemann, C., Formenty, P., Jensen, S.A., Boesch, C. & Junglen, S. (2008). Interspecies transmission of simian foamy virus in a natural predator-prey system. *Journal of Virology*, 82, 7741-7744.

Leendertz, S.A., Locatelli, S., Boesch, C., Kücherer, C., Formenty, P., Liegeois, F., Ayouba, A., Peeters, M. & Leendertz, F.H. (2011). No evidence for transmission of SIVwrc from western red colobus monkeys (*Piliocolobus badius badius*) to wild west african chimpanzees (*Pan troglodytes verus*) despite high exposure through hunting. *BMC Microbiology*, 11, 24.

Leendertz, S.A.J., Gogarten, J.F., Düx, A., Calvignac-Spencer, S. & Leendertz, F.H. (2015). Assessing the evidence supporting fruit bats as the primary reservoirs for ebola viruses. *EcoHealth*, Online first, 1-8.

Leendertz, S.A.J., Junglen, S., Hedemann, C., Goffe, A., Calvignac, S., Boesch, C. & Leendertz, F.H. (2010). High prevalence, coinfection rate, and genetic diversity of retroviruses in wild red colobus monkeys (*Piliocolobus badius badius*) in Tai National Park, Cote d'Ivoire. *Journal of Virology*, 84, 7427-7436.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R. & Tilman, D. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601-613.

Leslie, D.E., Azzato, F., Karapanagiotidis, T., Leydon, J. & Fyfe, J. (2007). Development of a real-time PCR assay to detect Treponema pallidum in clinical specimens and assessment of the assay's performance by comparison with serological testing. *J Clin Microbiol*, 45, 93-96.

Levins, R. (1968). *Evolution in changing environments: some theoretical explorations*. Princeton University Press, Princeton.

Levréro, F., Gatti, S., Gautier-Hion, A. & Ménard, N. (2007). Yaws disease in a wild gorilla population and its impact on the reproductive status of males. *American Journal of Physical Anthropology*, 132, 568-575.

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

Lieberman, D. & Lieberman, M. (1987). Forest tree growth and dynamics at La Selva, Costa Rica (1969-1982). *Journal of Tropical Ecology*, 3, 347-358.

Liégeois, F., Lafay, B., Formenty, P., Locatelli, S., Courgnaud, V., Delaporte, E. & Peeters, M. (2009). Full-length genome characterization of a novel simian immunodeficiency virus lineage (SIVolc) from olive colobus (*Procolobus verus*) and new SIVwrcPbb strains from western red colobus (*Piliocolobus badius badius*) from the Tai Forest in Ivory Coast. *Journal of Virology*, 83, 428-439.

Liu, W., Worobey, M., Li, Y., Keele, B.F., Bibollet-Ruche, F., Guo, Y., Goepfert, P.A., Santiago, M.L., Ndjango, J.B.N., Neel, C., Clifford, S.L., Sanz, C., Kamenya, S., Wilson, M.L., Pusey, A.E., Gross-Camp, N., Boesch, C., Smith, V., Zamma, K., Huffman, M.A., Mitani, J.C., Watts, D.P., Peeters, M., Shaw, G.M., Switzer, W.M., Sharp, P.M. & Hahn, B.H. (2008). Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. *PLoS Pathogens*, 4, e1000097.

Locatelli, S., Roeder, A.D., Bruford, M.W., Noë, R., Delaporte, E. & Peeters, M. (2011). Lack of evidence of simian immunodeficiency virus infection among nonhuman primates in Tai National Park, Cote d'Ivoire: limitations of noninvasive methods and SIV diagnostic tools for studies of primate retroviruses. *International Journal of Primatology*, 32, 288-307.

Loehle, C. (1995). Social barriers to pathogen transmission in wild animal populations. *Ecology*, 76, 326-335.

London, L. (2008). What is a human-rights based approach to health and does it matter? *Health and Human Rights*, 10, 65-80.

Lovell, N., Jurmain, R. & Kilgore, L. (2000). Skeletal evidence of probable treponemal infection in free-ranging African apes. *Primates*, 41, 275-290.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J. & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *The ISME journal*, 5, 169.

Luger, A. (1993). The origin of syphilis: clinical and epidemiologic considerations on the Columbian theory. *Sexually Transmitted Diseases*, 20, 110-117.

Lukehart, S.A., Godornes, C., Molini, B.J., Sonnett, P., Hopkins, S., Mulcahy, F., Engelman, J., Mitchell, S.J., Rompalo, A.M., Marra, C.M. & Klausner, J.D. (2004). Macrolide

resistance in *Treponema pallidum* in the United States and Ireland. *New England Journal of Medicine*, 351, 154-158.

Lwanga, J.S., Struhsaker, T.T., Struhsaker, P.J., Butynski, T.M. & Mitani, J.C. (2011). Primate population dynamics over 32.9 years at Ngogo, Kibale National Park, Uganda. *American Journal of Primatology*, 73, 997-1011.

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M.R., Hallows, J.L., Kusebauch, U., Vigl, E.E., Malfertheiner, P. & Megraud, F. (2016). The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science*, 351, 162-165.

Malim, M.H. & Bieniasz, P.D. (2012). HIV restriction factors and mechanisms of evasion. *Cold Spring Harbor Perspectives in Medicine*, 2, a006940.

Mann, J., Connor, R.C., Barre, L.M. & Heithaus, M.R. (2000). Female reproductive success in bottlenose dolphins (*Tursiops sp.*): life history, habitat, provisioning, and group-size effects. *Behavioral Ecology*, 11, 210-219.

Manokaran, N. & Kochummen, K. (1987). Recruitment, growth and mortality of tree species in a lowland dipterocarp forest in Peninsular Malaysia. *Journal of Tropical Ecology*, 3, 315-330.

Marcogliese, D.J. (2005). Parasites of the superorganism: are they indicators of ecosystem health? *International Journal for Parasitology*, 35, 705-716.

Marí Saéz, A., Weiss, S., Nowak, K., Lapeyre, V., Zimmermann, F., Düx, A., Kühl, H.S., Kaba, M., Regnaut, S., Merkel, K., Sachse, A., Thiesen, U., Villányi, L., Boesch, C., Dabrowski, P.W., Radonić, A., Nitsche, A., Leendertz, S.A.J., Petterson, S., Becker, S., Krähling, V., Couacy-Hymann, E., Akoua-Koffi, C., Weber, N., Schaade, L., Fahr, J., Borchert, M., Gogarten, J.F., Calvignac-Spencer, S. & Leendertz, F.H. (*2015*). Zoonotic origin of the West African Ebola epidemic. *EMBO Molecular Medicine*, 7 17-23.

Maurice, J. (2012). WHO plans new yaws eradication campaign. *The Lancet*, 379, 1377-1378.

McCord, A.I., Chapman, C.A., Weny, G., Tumukunde, A., Hyeroba, D., Klotz, K., Koblings, A.S., Mbora, D.N., Cregger, M. & White, B.A. (2014). Fecal microbiomes of non-human primates in western Uganda reveal species-specific communities largely resistant to habitat perturbation. *American Journal of Primatology*, 76, 347-354.

McCullagh, P. & Nelder, J.A. (1996). *Generalized linear models*. Chapman and Hall, London.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. & Hugenholtz, P. (2011). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6, 610-618.

McGraw, W.S. & Bshary, R. (2002). Association of terrestrial mangabeys (*Cercocebus atys*) with arboreal monkeys: experimental evidence for the effects of reduced ground predator pressure on habitat use. *International Journal of Primatology*, 23, 311-325.

McGraw, W.S., Zuberbühler, K. & Noë, R. (2007). Monkeys of the Taï Forest: an African primate community. Cambridge University Press Cambridge, p. 342.

McMurdie, P.J. & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8, e61217.

Meyer, M. & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010, pdb prot5448.

Milton, K. (1979). Factors influencing leaf choice by howler monkeys: a test of some hypotheses of food selection by generalist herbivores. *The American Naturalist*, 114, 363-378.

Milton, K. (1984). Habitat, diet and activity patterns of free-ranging woolly spider monkeys (*Brachyteles arachnoides* E. Geoffroy 1806). *International Journal of Primatology*, 5, 491-514.

Milton, K. (1996). Effects of bot fly (*Alouattamyia baeri*) parasitism on a free-ranging howler (*Alouatta palliata*) population in Panama. *Journal of Zoological Society of London*, 239, 39-63.

Miner, B.G., Sultan, S.E., Morgan, S.G., Padilla, D.K. & Relyea, R.A. (2005). Ecological consequences of phenotypic plasticity. *Trends in Ecology & Evolution*, 20, 685-692.

Minhós, T., Nixon, E., Sousa, C., Vicente, L.M., da Silva, M.F., Sá, R. & Bruford, M.W. (2013). Genetic evidence for spatio-temporal changes in the dispersal patterns of two sympatric African colobine monkeys. *American Journal of Physical Anthropology*, 150, 464-474.

Mitani, J.C., Sanders, W.J., Lwanga, J.S. & Windfelder, T.L. (2001). Predatory behavior of crowned hawk-eagles (*Stephanoaetus coronatus*) in Kibale National Park, Uganda. *Behavioral Ecology and Sociobiology*, 49, 187-195.

Mitani, J.C., Struhsaker, T.T. & Lwanga, J.S. (2000). Primate community dynamics in old growth forest over 23.5 years at Ngogo, Kibale National Park, Uganda: Implications for conservation and census methods. *International Journal of Primatology*, 21, 269-286.

Mitani, J.C. & Watts, D.P. (1999). Demographic influences on the hunting behavior of chimpanzees. *American Journal of Physical Anthropology*, 109, 439-454.

Mitani, J.C. & Watts, D.P. (2001). Why do chimpanzees hunt and share meat? *Animal Behaviour*, 61, 915-924.

Mitjà, O., Marks, M., Konan, D.J.P., Ayelo, G., Gonzalez-Beiras, C., Boua, B., Houinei, W., Kobara, Y., Tabah, E.N., Nsiire, A., Obvala, D., Taleo, F., Djupuri, R., Zaixing, Z., Utzinger, J., Vestergaard, L.S., Bassat, Q. & Asiedu, K. (2015). Global epidemiology of yaws: a systematic review. *The Lancet Global Health*, 3, e324-e331.

Miyamoto, M.M., Allen, J.M., Gogarten, J.F. & Chapman, C.A. (2013). Microsatellite DNA suggests that group size affects sex-biased dispersal patterns in red colobus monkeys. *American Journal of Primatology*, 75, 478–490.

Moeller, A.H., Caro-Quintero, A., Mjungu, D., Georgiev, A.V., Lonsdorf, E.V., Muller, M.N., Pusey, A.E., Peeters, M., Hahn, B.H. & Ochman, H. (2016a). Cospeciation of gut microbiota with hominids. *Science*, 353, 380-382.

Moeller, A.H., Foerster, S., Wilson, M.L., Pusey, A.E., Hahn, B.H. & Ochman, H. (2016b). Social behavior shapes the chimpanzee pan-microbiome. *Science Advances*, 2, e1500997.

Moeller, A.H., Li, Y., Ngole, E.M., Ahuka-Mundeke, S., Lonsdorf, E.V., Pusey, A.E., Peeters, M., Hahn, B.H. & Ochman, H. (2014). Rapid changes in the gut microbiome during human evolution. *Proceedings of the National Academy of Sciences*, 111, 16431-16435.

Moeller, A.H., Peeters, M., Ayouba, A., Ngole, E.M., Esteban, A., Hahn, B.H. & Ochman, H. (2015). Stability of the gorilla microbiome despite simian immunodeficiency virus infection. *Molecular Ecology*, 24, 690–697.

Moeller, A.H., Peeters, M., Ndjango, J.-B., Li, Y., Hahn, B.H. & Ochman, H. (2013a). Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. *Genome Research*, 23, 1715-1720.

Moeller, A.H., Shilts, M., Li, Y., Rudicell, R.S., Lonsdorf, E.V., Pusey, A.E., Wilson, M.L., Hahn, B.H. & Ochman, H. (2013b). SIV-induced instability of the chimpanzee gut microbiome. *Cell Host & Microbe*, 14, 340-345.

Montiel, R., Solórzano, E., Díaz, N., Álvarez-Sandoval, B.A., González-Ruiz, M., Cañadas, M.P., Simões, N., Isidro, A. & Malgosa, A. (2012). Neonate human remains: a window of opportunity to the molecular study of ancient syphilis. *PLoS ONE*, 7, e36371.

Mori, A.S. (2011). Ecosystem management based on natural disturbances: hierarchical context and non-equilibrium paradigm. *Journal of Applied Ecology*, 48, 280-292.

Morozov, V.A., Leendertz, F.H., Junglen, S., Boesch, C., Pauli, G. & Ellerbrok, H. (2009). Frequent foamy virus infection in free-living chimpanzees of the Taï National Park (Côte d'Ivoire). *Journal of General Virology*, 90, 500-506.

Mossoun, A., Pauly, M., Akoua-Koffi, C., Couacy-Hymann, E., Leendertz, S.A.J., Anoh, A.E., Gnoukpoho, A.H., Leendertz, F.H. & Schubert, G. (2015). Contact to non-human primates and risk factors for zoonotic disease emergence in the Taï region, Côte d'Ivoire. *EcoHealth*, 12, 580-591.

Mouinga-Ondémé, A., Caron, M., Nkoghé, D., Telfer, P., Marx, P., Saïb, A., Leroy, E., Gonzalez, J.P., Gessain, A. & Kazanji, M. (2012). Cross-species transmission of simian foamy virus to humans in rural Gabon, central Africa. *Journal of Virology*, 86, 1255-1260.

Mulligan, C.J., Norris, S.J. & Lukehart, S.A. (2008). Molecular studies in *Treponema pallidum* evolution: toward clarity? *PLoS Neglected Tropical Diseases*, 2, e184.

Murcia, C. (1995). Edge effects in fragmented forests: implications for conservation. *Trends in Ecology & Evolution*, 10, 58-62.

Murray, S. & Linial, M. (2006). Foamy virus infection in primates. *Journal of Medical Primatology*, 35, 225-235.

N'Goran, P.K., Boesch, C., Mundry, R., N'Goran, E.K., Herbinger, I., Yapi, F.A. & Kuehl, H.S. (2012). Hunting, law enforcement, and African primate conservation. *Conservation Biology*, 26, 565-571.

Nakazawa, N., Hanamura, S., Inoue, E., Nakatsukasa, M. & Nakamura, M. (2013). A leopard ate a chimpanzee: first evidence from East Africa. *Journal of Human Evolution*, 65, 334-337.

Nei, M. & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford University Press, Oxford.

Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L. & Graf, J. (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS One*, 9, e94249.

Newman, L., Kamb, M., Hawkes, S., Gomez, G., Say, L., Seuc, A. & Broutet, N. (2013). Global estimates of syphilis in pregnancy and associated adverse outcomes: analysis of multinational antenatal surveillance data. *PLoS Medicine*, 10, e1001396.

Nichols, H.J. (1910). Experimental yaws in the monkey and rabbit. *The Journal of Experimental Medicine*, 12, 616-622.

Noë, R. & Bshary, R. (1997). The formation of red colobus–Diana monkey associations under predation pressure from chimpanzees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264, 253-259.

Norden, N., Chazdon, R.L., Chao, A., Jiang, Y.-H. & Vílchez-Alvarado, B. (2009). Resilience of tropical rain forests: tree community reassembly in secondary forests. *Ecology Letters*, 12, 385-394.

Norris, S.J., Paster, B.J., Moter, A. & Göbel, U.B. (2006). The genus *Treponema*. In: *The Prokaryotes*. Springer New York, pp. 211-234.

Nunn, C.L. & Altizer, S. (2006). *Infectious diseases in primates: behavior, ecology and evolution*. Oxford University Press, Oxford.

Nunn, C.L., Craft, M.E., Gillespie, T.R., Schaller, M. & Kappeler, P.M. (2015). The sociality–health–fitness nexus: synthesis, conclusions and future directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140115.

Nunn, C.L. & Heymann, E.W. (2005). Malaria infection and host behavior: a comparative study of Neotropical primates. *Behavioral Ecology and Sociobiology*, 59, 30-37.

Nunn, C.L., Thrall, P.H., Bartz, K., Dasgupta, T. & Boesch, C. (2009). Do transmission mechanisms or social systems drive cultural dynamics in socially structured populations? *Animal Behaviour*, 77, 1515-1524.

Oates, J.F. (1977). The guereza and its food. In: *Primate ecology* (ed. Clutton-Brock, TH). Academic Press New York, pp. 275-321.

Oates, J.F., Whitesides, G.H., Davies, A.G., Waterman, P.G., S.M., G., DaSilva, G.L. & Mole, S. (1990). Determinants of variation in tropical forest primate biomass: new evidence from West Africa. *Ecology*, 71, 328-343.

Ochman, H., Worobey, M., Kuo, C.-H., Ndjango, J.-B.N., Peeters, M., Hahn, B.H. & Hugenholtz, P. (2010). Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. *PLoS Biology*, 8, e1000546.

Oksanen, J.F., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2016). Vegan: Community Ecology Package. R package version 2.4-0. https://CRAN.R-project.org/package=vegan.

Oksanen, J.F., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2012). Vegan: Community Ecology Package. R package version 2.0-3 http://CRAN. *R-project. org/package= vegan*.

Olupot, W. (1994). Ranging patterns of the grey-cheeked mangabey *Cercocebus albigena* with special reference to food finding and food availability in Kibale National Park. In: *Zoology*. Makerere University Kampala, p. 108.

Omeja, P.A., Jacob, A.L., Lawes, M.J., Lwanga, J.S., Rothman, J.M., Tumwesigye, C. & Chapman, C.A. (2014). Changes in Elephant Abundance Affect Forest Composition or Regeneration? *Biotropica*, 46, 704-711.

Opdam, P. & Wascher, D. (2004). Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. *Biological Conservation*, 117, 285-297.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N. & Pearse, W. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5. 2. 2013.

Ostfeld, R.S., Glass, G.E. & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20, 328-336.

Parker, J., Rambaut, A. & Pybus, O.G. (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infection, Genetics and Evolution*, 8, 239-246.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, 37, 637-669.

Pavelka, M.S.M., Brusselers, O.T., Nowak, D. & Behie, A.M. (2003). Population reduction and social disorganization in *Alouatta pigra* following a hurricane. *International Journal of Primatology*, 24, 1037-1055.

Pearse, W.D., Cadotte, M.W., Cavender-Bares, J., Ives, A.R., Tucker, C.M., Walker, S.C. & Helmus, M.R. (2015). pez: phylogenetics for the environmental sciences. *Bioinformatics*, 31, 2888-2890.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J. & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biology*, 17, 1-14.

Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R. & Walpole, M. (2010). Scenarios for global biodiversity in the 21st century. *Science*, 330, 1496-1501.

Perkins, S.E., Cagnacci, F., Stradiotto, A., Arnoldi, D. & Hudson, P.J. (2009). Comparison of social networks derived from ecological data: implications for inferring infectious disease dynamics. *Journal of Animal Ecology*, 78, 1015-1022.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & the R Development Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models. *R package version 3.1-106.*

Poinar, H.N. & Cooper, A. (2000). Ancient DNA: do it right or not at all. *Science*, 5482, 1139.

Poulin, R. (1995). Phylogeny, ecology, and the richness of parasite communities in vertebrates. *Ecological Monographs*, 65, 283-302.

Pride, R.E. (2005). Optimal group size and seasonal stress in ring-tailed lemurs (*Lemur catta*). *Behavioral Ecology*, 16, 550-560.

Prince, A.M., Brotman, B., Lee, D.H., Andrus, L., Valinsky, J. & Marx, P. (2002). Lack of evidence for HIV type 1-related SIVcpz infection in captive and wild chimpanzees (*Pan troglodytes verus*) in West Africa. *AIDS Research and Human Retroviruses*, 18, 657-660.

Pulliam, H.R. & Caraco, T. (1984). Living in groups: is there an optimal group size. In: *Behavioural ecology: an evolutionary approach* (eds. Krebs, JR & Davies, NB). Sinauer Associates Sunderland, England, pp. 122-147.

Purvis, A., Agapow, P.M., Gittleman, J.L. & Mace, G.M. (2000). Nonrandom extinction and the loss of evolutionary history. *Science*, 288, 328-330.

R Core Team (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*, URL https://www.R-project.org/.

R Development Core Team (2012). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria*, ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rambaut, A., Lam, T.T., Max Carvalho, L. & Pybus, O.G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, 2.

Range, F. & Noë, R. (2002). Familiarity and dominance relations among female sooty mangabeys in the Taï National Park. *American Journal of Primatology*, 56, 137-153.

Refisch, J. & Koné, I. (2005). Impact of commercial hunting on monkey populations in the Taï region, Côte d'Ivoire. *Biotropica*, 37, 136-144.

Ren, T., Grieneisen, L.E., Alberts, S.C., Archie, E.A. & Wu, M. (2015). Development, diet and dynamism: longitudinal and cross-sectional predictors of gut microbial communities in wild baboons. *Environmental Microbiology*.

Reynolds, V., Lloyd, A.W., Babweteera, F. & English, C.J. (2009). Decaying *Raphia farinifera* palm trees provide sodium for wild chimpanzees in the Budongo forest, Uganda. *PLoS Biology*, 4, e6194.

Rifkin, J.L., Nunn, C.L. & Garamszegi, L.Z. (2012). Do animals living in larger groups experience greater parasitism? A meta-analysis. *The American Naturalist*, 180, 70.

Ripley, B. (2011). nnet: feed-forward neural networks and multinomial log-linear models. *R package version 7,3-1*.

Rode, K.D., Chapman, C.A., Chapman, L.J. & McDowell, L.R. (2003). Mineral resource availability and consumption by colobus in Kibale National Park, Uganda. *International Journal of Primatology*, 24, 541-573.

Rode, K.D., Chapman, C.A., McDowell, L.R. & Stickler, C. (2006). Nutritional correlates of population density across habitats and logging intensities in redtail monkeys (*Cercopithecus ascanius*). *Biotropica*, 38, 625-634.

Rohland, N. & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols*, 2, 1756-1762.

Román, G.C. & Román, L.N. (1986). Occurrence of congenital, cardiovascular, visceral, neurologic, and neuro-ophthalmologic complications in late yaws: a theme for future research. *Review of Infectious Diseases*, 8, 760-770.

Rosenbaum, B., O'Brien, T.G., Kinnaird, M.F. & Supriatna, J. (1998). Population densities of Sulawesi crested black macaques (*Macaca nigra*) on Bacan and Sulawesi, Indonesia: effects of habitat disturbance and hunting. *American Journal of Primatology*, 44, 89-106.

Rothman, J., Chapman, C. & Soest, P. (2012). Methods in primate nutritional ecology: a user's guide. *International Journal of Primatology*, 33, 542-566.

Rothman, J.M., Chapman, C.A., Struhsaker, T.T., Raubenheimer, D., Twinomugisha, D. & Waterman, P.G. (2015). Long-term declines in nutritional quality of tropical leaves. *Ecology*, 96, 873-878.

Rothman, J.M., Dusinberre, K. & Pell, A.N. (2009). Condensed tannins in the diets of primates: a matter of methods? *American Journal of Primatology*, 71, 70-76.

Rothman, J.M., Raubenheimer, D. & Chapman, C.A. (2011). Nutritional geometry: protein and energy in mountain gorilla diets. *American Journal of Physical Anthropology*, 144, 257-258.

Rothman, J.M., Van Soest, P.J. & Pell, A.N. (2006). Decaying wood is a sodium source for mountain gorillas. *Biology Letters*, 2, 321-324.

Rudran, R. (1978). *Socioecology of the Blue Monkeys (Cercopithecus mitis stuhlmanni) of the Kibale Forest, Uganda*. Smithsonian Contributions to Zoology: no. 249, Washington.

Russon, S.E., Davies, S.J., King, D.A. & Tan, S. (2005). Soil-related performance variation and distributions of tree species in a Bornean rain forest. *Journal of Ecology*, 93, 879-889.

Rwego, I.B., Isabirye-Basuta, G., Gillespie, T.R. & Goldberg, T.L. (2008). Gastrointestinal bacterial transmission among humans, mountain gorillas, and livestock in Bwindi Impenetrable National Park, Uganda. *Conservation Biology*, 22, 1600-1607.

Ryan, A.M., Chapman, C.A. & Rothman, J.M. (2013). How do differences in species and part consumption affect diet nutrient concentrations? A test with red colobus monkeys in Kibale National Park, Uganda. *African Journal of Ecology*, 51, 1-10.

Ryan, S.J. & Walsh, P.D. (2011). Consequences of non-Intervention for infectious disease in African great apes. *PLoS ONE*, 6, e29030.

Salyer, S.J., Gillespie, T.R., Rwego, I.B., Chapman, C.A. & Goldberg, T.L. (2012). Epidemiology and molecular relationships of *Cryptosporidium spp.* in people, primates, and livestock from Western Uganda. *PLoS Neglected Tropical Diseases*, 6, e1597.

Samson, M., Libert, F., Doranz, B.J., Rucker, J., Liesnard, C., Farber, C.M., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C., Muyldermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R.J., Collman, R.G., Doms, R.W., Vassart, G. & Parmentier, M. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature*, 382, 722-725.

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M. & Brookfield, J.F. (2005). Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Tai Forest, Cote d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *Journal of Virology*, 79, 12515-12527.

Santiago, M.L., Rodenburg, C.M., Kamenya, S., Bibollet-Ruche, F., Gao, F., Bailes, E., Meleth, S., Soong, S.J., Kilby, J.M., Moldoveanu, Z., Fahey, B., Muller, M.N., Ayouba, A., Nerrienet, E., McClure, H.M., Heeney, J.L., Pusey, A.E., Collins, D.A., Boesch, C., Wrangham, R.W., Goodall, J., Sharp, P.M., Shaw, G.M. & Hahn, B.H. (2002). SIVcpz in wild chimpanzees. *Science*, 295, 465.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C. & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, 413, 591-596.

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., Fiori, J., Gotti, R., De Bellis, G., Luiselli, D., Brigidi, P., Mabulla, A., Marlowe, F., Henry, A.G. & Crittenden, A.N. (2014). Gut microbiome of the Hadza hunter-gatherers. *Nature Communications*, 5.

Schofield, P., Mbugua, D.M. & Pell, A.N. (2001). Analysis of condensed tannins: a review. *Animal Feed Science and Technology*, 91, 21-40.

Schubert, G., Stockhausen, M., Hoffmann, C., Merkel, K., Vigilant, L., Leendertz, F.H. & Calvignac-Spencer, S. (2015). Targeted detection of mammalian species using carrion fly-derived DNA. *Molecular Ecology Resources*, 15, 285-294.

Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A. & Busso, P. (2013). Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science*, 341, 179-183.

Sender, R., Fuchs, S. & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *BioRxiv*.

Sharp, P.M. & Hahn, B.H. (2011). Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine*, 1, a006841.

Sharp, P.M., Shaw, G.M. & Hahn, B.H. (2005). Simian immunodeficiency virus infection of chimpanzees. *Journal of Virology*, 79, 3891-3902.

Sheil, D., Jennings, S. & Savill, P. (2000). Long-term permanent plot observations of vegetation dynamics in Budongo, a Ugandan rain forest. *Journal of Tropical Ecology*, 16, 765-800.

Shumway, R.H. & Stoffer, D.S. (2000). *Time series analysis and its applications*. Springer-Verlag, New York.

Sibley, R.M. (1983). Optimal group size is unstable. *The American Naturalist*, 31, 947–948.

Sih, A., Ferrari, M.C. & Harris, D.J. (2011). Evolution and behavioural responses to human-induced rapid environmental change. *Evolutionary Applications*, 4, 367-387.

Simpson, E.H. (1949). Measurement of diversity. *Nature*, 163, 688-688.

Skorupa, J.P. (1989). Crowned Eagles *Strephanoaetus coronatus* in rainforest: observations on breeding chronology and diet at a nest in Uganda. *Ibis*, 131, 294-298.

Šmajs, D., Norris, S.J. & Weinstock, G.M. (2012). Genetic diversity in *Treponema pallidum*: implications for pathogenesis, evolution and molecular diagnostics of syphilis and yaws. *Infection, Genetics and Evolution*, 12, 191-202.

Šmajs, D., Zobaníková, M., Strouhal, M., Čejková, D., Dugan-Rocha, S., Pospíšilová, P., Norris, S.J., Albert, T., Qin, X. & Hallsworth-Pepin, K. (2011). Complete genome sequence of *Treponema paraluiscuniculi*, strain Cuniculi A: the loss of infectivity to humans is associated with genome decay. *PLoS ONE*, 6, e20415.

Smith, J.L., David, N., Indgin, S., Israel, C., Levine, B., Justice Jr, J., McCrary 3rd, J., Medina, R., Paez, P. & Santana, E. (1971). Neuro-ophthalmological study of late yaws and pinta. II. The Caracas project. *British Journal of Venereal Diseases*, 47, 226.

Snaith, T.V. & Chapman, C.A. (2005). Towards an ecological solution to the folivore paradox: patch depletion as an indicator of within-group scramble competition in red colobus. *Behavioral Ecology and Sociobiology*, 59, 185-190.

Snaith, T.V. & Chapman, C.A. (2007). Primate group size and socioecological models: do folivores really play by different rules? *Evolutionary Anthropology*, 16, 94-106.

Snaith, T.V. & Chapman, C.A. (2008). Red colobus monkeys display alternative behavioural responses to the costs of scramble competition. *Behavioral Ecology*, 19 1289-1296.

Snaith, T.V., Chapman, C.A., Rothman, J.M. & Wasserman, M.D. (2008). Bigger groups have fewer parasites and similar cortisol levels: a multi-group analysis in red colobus monkeys. *American Journal of Primatology*, 70, 1-9.

Soetaert, K. & Heip, C. (1990). Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Marine Ecology Progress Series*, 59, 305-307.

Sokal, R.R. & Rohlf, F.J. (1995). *Biometry: the principles and practice of statistics in biological research*. 3rd edn. W. H. Freeman and Company, New York.

Sparks, T.H. & Carey, P.D. (1995). The responses of species to climate over two centuries: an analysis of the marsham phenological record, 1736-1947. *Journal of Ecology*, 83, 321-329.

Stamm, L.V. & Bergen, H.L. (2000). A point mutation associated with bacterial macrolide resistance is present in both 23S rRNA genes of an erythromycin-resistant *Treponema pallidum* clinical isolate. *Antimicrobial Agents and Chemotherapy*, 44, 806-807.

Stanford, C.B. (1995). The influence of chimpanzee predation on group size and anti-predator behaviour in red colobus monkeys. *Animal Behaviour*, 49, 577-587.

Steenbeek, R. & van Schaik, C.P. (2001). Competition and group size in Thomas's langurs (*Presbytis thomasi*): the folivore paradox revisited. *Behavioral Ecology and Sociobiology*, 49, 100-110.

Sterck, E.H., Watts, D.P. & van Schaik, C.P. (1997). The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology*, 41, 291-309.

Stickler, C.M. (2004). The effects of logging on primate-habitat interactions: a case study of redtail monkeys (*Cercopithecus ascanius*) in Kibale National Park, Uganda. M.Sc. Thesis, University of Florida Gainesville, p. 94.

Stockwell, C.A., Hendry, A.P. & Kinnison, M.T. (2003). Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, 18, 94-101.

Stoneking, M. & Krause, J. (2011). Learning about human population history from ancient and modern genomes. *Nature Reviews Genetics*, 12, 603-614.

Strier, K.B. & Mendes, S.L. (2012). The Northern Muriqui (*Brachyteles hypoxanthus*): lessons on behavioral plasticity and population dynamics from a critically endangered species. In: *Long-term field studies of primates* (eds. Kappeler, PM & Watts, DP). Springer-Verlag Berlin, pp. 125-140.

Struhsaker, T.T. (1975). *The red colobus monkey*. University of Chicago Press, Chicago.

Struhsaker, T.T. (1978). Food habits of five monkey species in the Kibale Forest, Uganda. In: *Recent advances in primatology* (eds. Chivers, DJ & Herbert, J). Academic Press London, pp. 225-248.

Struhsaker, T.T. (1997). *Ecology of an African rain forest: logging in Kibale and the conflict between conservation and exploitation*. University of Florida Press, Gainesville.

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

Struhsaker, T.T. & Leakey, M. (1990). Prey selectivity by crowned hawk-eagles on monkeys in the Kibale Forest, Uganda. *Behavioral Ecology and Sociobiology*, 26, 435-443.

Struhsaker, T.T. & Leland, L. (1985). Infanticide in a patrilineal society of red colobus monkeys. *Zeitschrift für Tierpsychologie*, 69, 89-132.

Struhsaker, T.T. & Leland, L. (1987). Colobines: infanticide by adult males. In: *Primate societies* (eds. Smuts, BB, Cheney, DL, Seyfarth, RM, Wrangham, RW & Struhsaker, TT). University of Chicago Press Chicago, pp. 83-97.

Switzer, W.M., Bhullar, V., Shanmugam, V., Cong, M.E., Parekh, B., Lerche, N.W., Yee, J.L., Ely, J.J., Boneva, R., Chapman, L.E., Folks, T.M. & Heneine, W. (2004). Frequent simian foamy virus infection in persons occupationally exposed to nonhuman primates. *Journal of Virology*, 78, 2780-2789.

Switzer, W.M., Garcia, A.D., Yang, C., Wright, A., Kalish, M.L., Folks, T.M. & Heneine, W. (2008). Coinfection with HIV-1 and simian foamy virus in West Central Africans. *Journal of Infectious Diseases*, 197, 1389-1393.

Switzer, W.M., Salemi, M., Shanmugam, V., Gao, F., Cong, M.E., Kuiken, C., Bhullar, V., Beer, B.E., Vallet, D., Gautier-Hion, A., Tooye, Z., Villinger, F., Holmes, E.C. & Heneine, W. (2005). Ancient co-speciation of simian foamy viruses and primates. *Nature*, 434, 376-380.

Swofford, D.L. (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Symington, M.M. (1988). Demography, ranging patterns, and activity budgets of black spider monkeys (*Ateles paniscus chamek*) in the Manu National Park, Peru. *American Journal of Primatology*, 15, 45-67.

Synnott, T. (1971). Annotated list of the perennial woody vegetation of the West Ankole forests. *The Uganda Journal*, 35, 1-21.

Takeda, K., Sato, S. & Sugawara, K. (2003). Familiarity and group size affect emotional stress in Japanese Black heifers. *Applied Animal Behaviour Science*, 82, 1-11.

Tamura, K., Stecher, G., Peterson, D. & Filipski, A. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*.

Teelen, S. (2007). Primate abundance along five transect lines at Ngogo, Kibale National Park, Uganda. *American Journal of Primatology*, 69, 1030-1044.

Teelen, S. (2008). Influence of chimpanzee predation on the red colobus population at Ngogo, Kibale National Park, Uganda. *Primates*, 49, 41-49.

Theis, K.R., Dheilly, N.M., Klassen, J.L., Brucker, R.M., Baines, J.F., Bosch, T.C.G., Cryan, J.F., Gilbert, S.F., Goodnight, C.J., Lloyd, E.A., Sapp, J., Vandenkoornhuyse, P., Zilber-Rosenberg, I., Rosenberg, E. & Bordenstein, S.R. (2016). Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. *mSystems*, 1.

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M. & Michel, G. (2012). Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. *Environmental Microbiology*, 14, 2379-2394.

Thornburg, R.W. & Baseman, J.B. (1983). Comparison of major protein antigens and protein profiles of *Treponema pallidum* and *Treponema pertenue*. *Infection and Immunity*, 42, 623-627.

Traina-Dorge, V.L., Lorino, R., Gormus, B.J., Metzger, M., Telfer, P., Richardson, D., Robertson, D.L., Marx, P.A. & Apetrei, C. (2005). Molecular epidemiology of simian T-cell lymphotropic virus type 1 in wild and captive sooty mangabeys. *Journal of Virology*, 79, 2541-2548.

Tremaroli, V. & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, 489, 242-249.

Tung, J., Barreiro, L.B., Burns, M.B., Grenier, J.-C., Lynch, J., Grieneisen, L.E., Altmann, J., Alberts, S.C., Blekhman, R. & Archie, E.A. (2015). Social networks predict gut microbiome composition in wild baboons. *Elife*, 4, e05224.

Turkington, R. (2009). Top-down and bottom-up forces in mammalian herbivore - vegetation systems: An essay review. . *Botany*, 87, 723-739.

Turner, M.G., Romme, W.H., Gardner, R.H., O'Neill, R.V. & Kratz, T.K. (1993). A revised concept of landscape equilibrium: Disturbance and stability on scaled landscapes. *Landscape Ecology*, 8, 213-227.

Tutin, C.E.G. & Oslisly, R. (1995). *Homo, Pan*, and *Gorilla*: coexistence over 60,000 years at Lope in central Gabon. *Journal of Human Evolution*, 28, 597-602.

Uganda Wildlife Authority (2003). Kibale National Park General Management Plan 2003-2013. Uganda Wildlife Authority Kampala, Uganda, p. 79.

United Nations Development Programme (2015). Human Development Report. p. 288.

van Gemerden, B.S., Olff, H., Parren, M.P.E. & Bongers, F. (2003). The pristine rain forest? Remnants of historical human impacts on current tree species composition and diversity. *Journal of Biogeography*, 30, 1381-1390.

van Schaik, C.P. (1983). Why are diurnal primates living in groups? *Behaviour*, 87, 120-144.

van Schaik, C.P. (1989). The ecology of social relationships amongst female primates. In: *Comparative socioecology: the behavioural ecology of humans and other mammals* (eds. Standen, V & Foley, RA). Blackwell Scientific Publications Boston, pp. 195-218.

van Schaik, C.P. & Kappeler, P.M. (1997). Infanticide risk and the evolution of male–female association in primates. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264, 1687-1694.

van Schaik, C.P., Noordwijk, M.A., Boer, R.J. & Tonkelaar, I. (1983). The effect of group size on time budgets and social behaviour in wild long-tailed macaques (*Macaca fascicularis*). *Behavioral Ecology and Sociobiology*, 13, 173-181.

van Schaik, C.P. & van Hooff, J.A.R.A.M. (1983). On the ultimate causes of primate social systems. *Behaviour*, 85, 91-117.

van Woerden, J.T., Willems, E.P., van Schaik, C.P. & Isler, K. (2011). Large brains buffer energetic effects of seasonal habitats in catarrhine primates. *Evolution*, 66, 191-199.

Venables, W.N. & Ripley, B.D. (2002). *Modern applied statistics with S*. Springer-Verlag, New York.

von Hunnius, T.E., Roberts, C.A., Boylston, A. & Saunders, S.R. (2006). Histological identification of syphilis in pre-Columbian England. *American Journal of Physical Anthropology*, 129, 559-566.

von Hunnius, T.E., Yang, D., Eng, B., Waye, J.S. & Saunders, S.R. (2007). Digging deeper into the limits of ancient DNA research on syphilis. *Journal of Archaeological Science*, 34, 2091-2100.

Warinner, C., Rodrigues, J.F.M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y. & Fiddyment, S. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, 46, 336-344.

Waser, P. (1975). Monthly variations in feeding and activity patterns of the mangabey, *Cercocebus albigena* (Lydekker). *East African Wildlife Journal*, 13, 249-263.

Wasserman, M.D. & Chapman, C.A. (2003). Determinants of colobine monkey abundance: the importance of food energy, protein and fibre content. *Journal of Animal Ecology*, 72, 650-659.

Wasserman, M.D., Chapman, C.A., Milton, K., Gogarten, J.F., Wittwer, D.J. & Ziegler, T.E. (2012). Estrogenic plant consumption predicts red colobus monkey (*Procolobus rufomitratus*) hormonal state and behavior. *Hormones and Behavior*, 62, 553-562.

Waterman, P.G., Ross, J.A.M., Bennet, E.L. & Davies, A.G. (1988). A comparison of the floristics and leaf chemistry of the tree flora in two Malaysian rain forests and the influence of leaf chemistry on populations of colobine monkeys in the old world. *Biological Journal of the Linnean Society*, 34, 1-32.

Watts, D.P., Potts, K.B., Lwanga, J.S. & Mitani, J.C. (2012). Diet of chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, Kibale National Park, Uganda, 2. Temporal variation and fallback foods. *American Journal of Primatology*, 74, 130-144.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 475-505.

Weldon, C., Du Preez, L.H., Hyatt, A.D., Muller, R. & Speare, R. (2004). Origin of amphibian chytrid fungus. *Emerging Infectious Diseases*, 10, 2100-2105.

Whitmee, S., Haines, A., Beyrer, C., Boltz, F., Capon, A.G., de Souza Dias, B.F., Ezeh, A., Frumkin, H., Gong, P., Head, P., Horton, R., Mace, G.M., Marten, R., Myers, S.S., Nishtar, S., Osofsky, S.A., Pattanayak, S.K., Pongsiri, M.J., Romanelli, C., Soucat, A., Vega, J. & Yach, D. (2015). Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health. *The Lancet*, 386, 1973-2028.

Williams, J.M., Lonsdorf, E.V., Wilson, M.L., Schumacher-Stankey, J., Goodall, J. & Pusey, A.E. (2008). Causes of death in the Kasekela chimpanzees of Gombe National Park, Tanzania. *American Journal of Primatology*, 70, 766-777.

Wittig, R.M. & Boesch, C. (2003). Food competition and linear dominance hierarchy among female chimpanzees of the Tai National Park. *International Journal of Primatology*, 24, 847-867.

Wolfe, N.D., Escalante, A.A., Karesh, W.B., Kilbourn, A., Spielman, A. & Lal, A.A. (1998). Wild primate populations in emerging infectious disease research: the missing link? *Emerging Infectious Diseases*, 4, 149–158.

Wolfe, N.D., Switzer, W.M., Carr, J.K., Bhullar, V.B., Shanmugam, V., Tamoufe, U., Prosser, A.T., Torimiro, J.N., Wright, A., Mpoudi-Ngole, E., McCutchan, F.E., Birx, D.L., Folks, T.M., Burke, D.S. & Heneine, W. (2004). Naturally acquired simian retrovirus infections in central African hunters. *The Lancet*, 363, 932-937.

Wood, S. & Wood, M.S. (2011). Package 'mgcv'.

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 73, 3-36.

World Health Organization (2012). *Global incidence and prevalence of selected curable sexually transmitted infections-2008*. World Health Organization, Geneva.

Wrangham, R.W. (1980). An ecological model of female-bonded primate groups. *Behaviour*, 75, 262-300.

Wrangham, R.W., Gittleman, J.L. & Chapman, C.A. (1993). Constraints on group size in primates and carnivores: population density and day-range as assays of exploitation competition. *Behavioral Ecology and Sociobiology*, 32, 199-209.

Wright, J.P. & Jones, C.G. (2006). The concept of organisms as ecosystem engineers ten years on: progress, limitations, and challenges. *BioScience*, 56, 203-209.

Wright, S.J. (2003). The myriad consequences of hunting for vertebrates and plants in tropical forests. *Perspectives in Plant Ecology, Evolution, and Systematics*, 6, 73-86.

Wylie, T.N., Wylie, K.M., Herter, B.N. & Storch, G.A. (2015). Enhanced virome sequencing using targeted sequence capture. *Genome Research*, 25, 1910-1920.

Yeager, C.P. & Kirkpatrick, C.R. (1998). Asian colobine social structure: ecological and evolutionary constraints. *Primates*, 39, 147-155.

Yeager, C.P. & Kool, K. (2000). The behavioral ecology of Asian colobines. In: *Old world monkeys* (eds. Whitehead, PF & Jolly, CJ). Cambridge University Press Cambridge, pp. 496-521.

Zaiss, Mario M., Rapin, A., Lebon, L., Dubey, Lalit K., Mosconi, I., Sarter, K., Piersigilli, A., Menin, L., Walker, Alan W., Rougemont, J., Paerewijck, O., Geldhof, P., McCoy, Kathleen D., Macpherson, Andrew J., Croese, J., Giacomin, Paul R., Loukas, A., Junt, T., Marsland, Benjamin J. & Harris, Nicola L. (2015). The intestinal microbiota contributes to the ability of helminths to modulate allergic inflammation. *Immunity*, 43, 998-1010.

Zilber-Rosenberg, I. & Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, 32, 723-735.

Zobaníková, M., Strouhal, M., Mikalová, L., Ambrozová, L., Pospísilová, P., Fulton, L.L., Chen, L., Sodergren, E., Weinstock, G.M. & Smajs, D. (2013). Whole genome sequence of the *Treponema Fribourg-Blanc:* unspecified simian isolate is highly similar to the yaws subspecies. *PLoS Neglected Tropical Diseases*, 7.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009). *Mixed effects models and extensions in ecology with R*. Springer Verlag, New York.

# APPENDIX LINKING STATEMENT

In the Chapter 4, I began making the link between grouping behaviour, within- and between-species interactions, and the microbiome. The retroviral studies pursued at Taï National Park represent a novel integration of primate ecology and behavioural findings for the understanding the ecology and evolution of one aspect of the microbiome, viruses. Attempting to address some of the within- and between-species transmission questions highlighted in Chapter four, I sought to examine how these changes might be impacting microorganism transmission. I began examing retroviruses in Taï National Park, a field site discussed in detail in the following chapter, because they are quickly evolving, allowing the accumulated mutations over a host's lifetime, to be used to understand the transmission process. Unfortunately my own efforts to use non-invasive sampling to screen samples for retroviruses were unsuccessful for SIVsmm, while SFVsmm, a selective sweep meant that there was not enough diversity to examine transmission processes in my study group. In this Appendix Chapter, I draw upon the available data from this ecosystem and my own data to gather insights into the prevalence, within- and cross-species transmission of primate retroviruses in this community. These were used to generate insights into the ecology and evolution of primate retroviruses in a NHP community and I discuss how these processes might be shifting in changing environments.

# APPENDIX CHAPTER

## THE ECOLOGY OF PRIMATE RETROVIRUSES – AN ASSESSMENT OF 12 YEARS OF RETROVIRAL STUDIES IN THE TAÏ NATIONAL PARK AREA, CÔTE D'IVOIRE

Jan F. Gogarten[1,2,5*], Chantal Akoua-Koffi[4*], Sebastien Calvignac-Spencer[1*], Siv Aina J. Leendertz[1*], Sabrina Weiss[1*], Emmanuel Couacy-Hymann[5], Inza Koné[6], Martine Peeters[7], Roman M. Wittig[2], Christophe Boesch[2], Beatrice H. Hahn[8], Fabian H. Leendertz[1]

A version of this chapter is published in *Virology*, Volumes 460-461, pp. 147-153

### A.1 ABSTRACT

*Authors contributed equally to this work.

[1]Research group Epidemiology of Highly Pathogenic Microorganisms RKI, Berlin, Germany.

[2]Primatology department, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

[3]Department of Biology, McGill University, Montreal, Quebec, Canada.

[4]Reseach Center for the Development and Teaching Hospital, Université Alassane Ouattara de Bouake, Cote d'Ivoire; [5]Central Laboratory for Animal Diseases, Bingerville, Côte d'Ivoire.

[6]Taï Monkey Project, Centre Suisse de Recherches Scientifiques, B.P. 1303 Abidjan, Côte d'Ivoire and Laboratory of Zoology, University of Cocody, 22 B.P. 582, Abidjan 22, Côte d'Ivoire.

[7]UMI 233, TransVIHMI, Institute for Research and Development (IRD) and University of Montpellier 1, Montpellier, France.

[8]Department of Microbiology and Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America.

The existence and genetic make-up of most primate retroviruses was revealed by studies of bushmeat and fecal samples from unhabituated primate communities. For these, detailed data on intra- and within-species contact rates are generally missing, which makes identification of factors influencing transmission a challenging task. Here we present an assessment of 12 years of research on primate retroviruses in the Taï National Park area, Côte d'Ivoire. We discuss insights gained into the prevalence, within- and cross-species transmission of primate retroviruses (including towards local human populations) and the importance of virus-host interactions in determining cross-species transmission risk. Finally we discuss how retroviruses ecology and evolution may change in a shifting environment and identify avenues for future research.

## A.2 INTRODUCTION

The close evolutionary relationship and similar physiology of non-human primates (NHPs) and humans make NHPs a likely source for the zoonotic transmission of viruses (reviewed in: Wolfe *et al.* 1998; Gillespie *et al.* 2008; Calvignac-Spencer *et al.* 2012b; Gessain *et al.* 2013). NHP retroviruses are arguably the best illustration of this prediction (e.g., simian origins of human immunodeficiency viruses HIV-1 and 2; reviewed in: Sharp & Hahn 2011) and among the best-characterized NHP viruses. However, despite decades of research, behavioural and ecological factors affecting within- and between- species transmission of retroviruses in NHPs remain poorly understood. A full understanding of the ecology of primate retroviruses requires knowledge of the virus itself, the host-virus interface, and the host's ecology. This necessitates multidisciplinary research efforts that are only possible where primatology research projects have been run on a long-term basis in collaboration with veterinarians and molecular biologists.

Taï National Park in the Côte d'Ivoire hosts one of the world's best-studied wild primate populations (Fig. A.1). Studies on the chimpanzees and monkeys of Taï National Park were initiated in 1979 and 1989 respectively (Boesch & Achermann 2000; McGraw *et al.* 2007). Taï National Park harbours 11 different NHP species, many of which regularly interact (Figure 2; McGraw *et al.* 2007). For example, many spend much of their time in polyspecific associations (e.g. red colobus - *Piliocolobus badius badius*, sooty mangabeys - *Cercocebus atys*, and Diana monkeys - *Cercopithecus diana*; Noë & Bshary 1997; McGraw & Bshary 2002). Hunter–prey relationships also exist in this community as chimpanzees (*Pan troglodytes verus*) regularly hunt other NHPs (Figure 2; Boesch & Boesch 1989). Local human populations in the area also interact with the primate community, mostly through hunting of NHPs (Figure 2; Refisch & Koné 2005). This complex set of interactions offers ample opportunities for microorganism transmission: through biting, grooming, mating and hunting-related activities such as butchering of carcasses and meat consumption.

In 2002, a perennial veterinary program began its association with the primatology research program. The latter has provided data and access to three groups of well-habituated chimpanzees (Boesch & Achermann 2000) and well-habituated monkey species (McGraw *et al.* 2007). This has helped make the primate community in Taï National Park one of the rare instances where thorough data are generated regarding primate ecology and the microorganisms that infect these

same individuals (Calvignac-Spencer *et al.* 2012b). This is particularly true for retroviruses, whose diversity and transmission patterns have been scrutinized since the very beginning of the veterinary studies of this community (although with some bias). These insights are augmented by human health projects that have spawned in the area around Taï National Park (Calvignac-Spencer *et al.* 2012a; Ayouba *et al.* 2013) allowing the study of zoonotic retroviral transmission stemming from this primate community. Here we review 12 years of research on primate retroviruses in the Taï National Park area and highlight how the long-term primatology research program (Boesch & Achermann 2000; McGraw *et al.* 2007) allowed virological results to be embedded in a relevant ecological context. We focus on results relevant to the ecology of the three retroviruses that have been extensively studied in Taï National Park; namely the simian immunodeficiency viruses (SIV) and their human counterparts (HIV), the simian T-lymphotropic viruses type 1 (STLV-1) and their human counterparts (HTLV-1), and the simian foamy viruses (SFV). We summarize insights gained into their prevalence, virus-host interactions, within- and between-species transmission, within and between NHP species and into the surrounding human population. Finally, we discuss the likely impact of shifting primate community dynamics in the framework of ongoing "natural" and anthropogenic changes.

### A.3 Prevalence and diversity of retroviruses in Taï National Park NHPs

To date in the NHPs of Taï National Park, SIV has been detected in sooty mangabeys (SIVsmm), western red colobus (SIVwrc) and olive colobus (SIVolc). The prevalence of SIVsmm in sooty mangabeys is high (apparent prevalence = 59% (95% CI = 0.35 to 0.88); Santiago *et al.* 2005). Similarly, the prevalence of SIVwrc in the red colobus population is reported to be one of the highest observed in wild NHPs (estimated prevelance = 82% (95% CI = 0.66 to 0.98); Leendertz *et al.* 2010), whereas SIVolc has so far only been reported in a single olive colobus monkey (Courgnaud *et al.* 2003). Strict host specificity was observed for the SIV strains found in these NHPs (Courgnaud *et al.* 2003; Liégeois *et al.* 2009; Leendertz *et al.* 2010). A collection of fecal samples obtained from the other six monkey species living in the park were screened both serologically and by PCR without any positive results (Locatelli *et al.* 2011). However, since sampling sizes were very small ($n \leq 10$) except for two species: black and white colobus (*Colobus polykomos, n*=27 distinct individuals) and Diana monkeys (n=23 distinct individuals), low

prevalences cannot be ruled out. Finally, the Taï National Park chimpanzee population, like other *Pan troglodytes verus* populations (Gao *et al.* 1999; Prince *et al.* 2002; Santiago *et al.* 2002), is not infected with SIVcpz (n=32, out of a total of 300 chimpanzees living in the park; Leendertz *et al.* 2011), the chimpanzee specific virus which infects the central and east African chimpanzee subspecies *P .t. troglodytes* and *P. t. schweinfurthii* (Keele *et al.* 2006)

STLV-1 has been detected in sooty mangabeys, red colobus monkeys and chimpanzees. Three of five sooty mangabeys tested in Taï National Park were infected with STLV-1, which does not allow for an accurate estimation of the prevalence in this species (Traina-Dorge *et al.* 2005; Calvignac-Spencer *et al.* 2012a). The prevalence of STLV-1 could be estimated in red colobus monkeys (apparent prevalence = 50% (95% CI = 0.29-0.71); Leendertz *et al.* 2010). It is markedly higher than at Kibale National Park in Uganda (Goldberg *et al.* 2009). Behavioural differences between these red colobus populations might play a role in these different prevalences; higher seasonality in Taï National Park seems to lead to a distinct breeding season with higher competition, promiscuity and aggression rates when compared to the Kibale community where births occur year round (Leendertz *et al.* 2010). Interestingly, all red colobus individuals in Tai (and Kibale) that tested positive for STLV-1 were co-infected with either SIV or SFV or both (Goldberg *et al.* 2009; Leendertz *et al.* 2010). In chimpanzees, the prevalence of STLV-1 is also high (apparent prevalence = 46% (95% CI = 0.28-0.65); Leendertz *et al.* 2003; Leendertz *et al.* 2004; Junglen *et al.* 2010). The STLV-1 strains circulating in these NHPs are not strictly species-specific, STLV-1 infecting sooty mangabeys on the one hand and red colobus and chimpanzees on the other, form two relatively homogeneous clades (Calvignac-Spencer *et al.* 2012a). This suggests that the generally assumed lack of host specificity of STLV-1 may not hold true at this small geographic scale.

SFV has been found in sooty mangabeys, red colobus and chimpanzees. SFV prevalence in sooty mangabeys seems high, but the sample size is still too small to derive a meaningful prevalence estimate (nine positive individuals out of twelve tested; J Gogarten and F Leendertz unpublished data). SFV in red colobus has one of the highest prevalences of all retroviruses so far tested for at Taï National Park (apparent prevalence = 86% (95% CI 72-100); Leendertz *et al.* 2010). This is a similarly high prevalence as found in red colobus in East Africa and NHP populations in general (Calattini *et al.* 2004; Liu *et al.* 2008; Goldberg *et al.* 2009). SFV also infects Taï National Park chimpanzees at very high prevalence (apparent prevalence = 90% (95% CI = 0.80-0.95); Liu *et*

*al.* 2008; Morozov *et al.* 2009; Blasse *et al.* 2013), which is similar to prevalence estimates at other chimpanzee study sites (44% to 100%; Liu *et al.* 2008). SFV from Taï National Park NHPs conform to the strong pattern of host-parasite co-divergence observed in other vertebrates (Switzer *et al.* 2005; Murray & Linial 2006; Leendertz *et al.* 2008; Morozov *et al.* 2009; Han & Worobey 2012).

In summary, sooty mangabeys and red colobus are infected at high prevalence by all three retroviruses while chimpanzees are not infected by SIV but frequently infected with STLV-1 and SFV. For the other primate species found in Taï National Park the occurrence of retroviruses is currently not known, but their genetic relationship to other retrovirus-infected NHPs in other parts of Africa suggests they could serve as hosts for these viruses.

### A.4 WITHIN-SPECIES TRANSMISSION OF RETROVIRUSES IN TAÏ NATIONAL PARK NHPS

Understanding retroviral transmission within a host species is a necessary first step for understanding retroviral ecology. Although within-species transmission dynamics seem an obvious area where retrovirology and primatology could synergistically collaborate, it remains essentially unexplored. This is true at Taï National Park as well, where only a handful of studies have investigated within-species retroviral transmission patterns.

The modalities of transmission have been partially addressed for a number of host-virus combinations. Both SIVsmm and SFVcpz were shown to be regularly transmitted both vertically (mother-to-infant) and horizontally (Santiago *et al.* 2005; Blasse *et al.* 2013). STLV-1 in chimpanzees is more rarely transmitted from mother to offspring (only 2/17 infants and juveniles born from STLV-1 positives mothers tested positive thus far; Leendertz *et al.* 2004). This rate is similar to the vertical transmission rate described for humans that breastfed over 12 months (15.7%; Hino *et al.* 1996) but it should be noted that here chimpanzee mother status was only determined at the time of this cross-sectional study implying this transmission rate can only be an overestimate (one of the mothers might have been STLV-1 negative throughout their breastfeeding). The importance of STLV-1 horizontal transmission during adulthood is even more equivocal, as many infections appear to arise due to transmission from red colobus, which may occur during frequent hunting of red colobus by chimpanzees (hunting preferences described in detail below); the slow evolutionary rate and lack of depth (in terms of sampling

viral diversity) has prevented an understanding of STLV-1 epidemiological processes (Leendertz *et al.* 2004; Junglen *et al.* 2010).

As retroviruses reach high prevalences in Taï National Park, an important question is whether individuals accumulate multiple-strains over their lifetime (Goffe *et al.* 2012). SIVsmm positivity is age-structured, with infections detected more often in adults than juveniles. This could be a function of more frequent aggressive interactions and increased sexual activity associated with adulthood (Santiago *et al.* 2005). Distinguishable SFVcpz strains, likely the result of multiple independent infections, were shown to accumulate with age in chimpanzees. SFVcpz infection appears to first occur via vertical transmission but is followed in adult life by the acquisition of further infections, possibly stemming from aggressive interactions (Blasse *et al.* 2013). Interestingly, sex does not seem to influence the accumulation of SFVcpz, which may reflect the involvement of both sexes in aggressive interactions (Blasse *et al.* 2013). STLV-1 infection in chimpanzees shows a comparable trend, with seropositivity/PCR positivity increasing with age (Leendertz *et al.* 2004; Junglen *et al.* 2010), which might be a result of age related hunting activity (see below).

Most NHPs live in complex social systems, whose organization likely influences retroviral transmission (Griffin & Nunn 2012). Santiago et al. (Santiago *et al.* 2005) examined the distribution of SIVsmm positivity with respect to dominance rank and demonstrated an excess of high ranking females among SIVsmm positive females. A number of possible explanations for the observed pattern were proposed (e.g., increased mating or grooming behaviour associated with higher rank) but the exact mechanism driving this pattern remains unknown (Santiago *et al.* 2005). While intragroup relationships play a role in the epidemiology of retroviruses, interactions between groups are also expected to play a major role in shaping retroviral transmission patterns. NHPs exhibit a huge degree of variation in intergroup dynamics and social structure, and understanding this variation has been a focus of primatologists for several decades (Wrangham 1980; van Schaik & van Hooff 1983; Sterck *et al.* 1997). For example, while female philopatry (i.e., that females stay in their natal group their entire lives while males typically disperse) is common among many NHP species, red colobus exhibit a strong dispersal bias whereby sub-adult females leave their natal groups (Struhsaker 2010). In this species, the fact that dispersal generally occurs before or concomitantly with sexual maturity and that extra-group copulations are rare, may explain the finding that SIVwrc strains circulating in two red colobus groups were

found to segregate according to a particular individual's group membership (Locatelli *et al.* 2011). Other important grouping behaviours (e.g., fission, fusion, intergroup aggression) are also expected to influence patterns of retroviral genetic diversity, although evidence remains weak and largely circumstantial (Locatelli *et al.* 2011). The aforementioned studies demonstrate how behavioural observations can aid in the interpretation of retroviral genetic diversity, while highlighting that successful examples of such collaboration are limited to date..

### A.5 CROSS-SPECIES TRANSMISSION OF RETROVIRUSES AMONG TAÏ NATIONAL PARK NHPS

Many NHPs at Taï National Park interact regularly with one another. The position of chimpanzees within this interaction network closely mimics that of humans (Fig. A.2). As a result of this likeness and the physiological and genetic similarity of chimpanzees to humans, the chimpanzee/NHP prey system has been used to provide insight into zoonotic transmission risk in human/NHP prey systems (Boesch 1994; Leendertz *et al.* 2004; Leendertz *et al.* 2008). However, major differences exist between chimpanzee and human hunting behaviour: the former species concentrates largely on a single prey, the red colobus, while humans exhibit much less specificity in hunting preference. Bushmeat market analyses reveal that nearly all NHP species found in Taï National Park are hunted and sold around the park. In markets located on the western edge of Taï National Park, no species accounts for less than 5% of the overall NHP bushmeat biomass, with red colobus and black and white colobus almost equally present (respectively 24.7 and 22.3%; Refisch & Koné 2005). In contrast, chimpanzees were observed to capture 215 red colobus over a 12 year period, while only successfully capturing six olive colobus and a single sooty mangabey (Boesch & Achermann 2000). This means that over their lifespan, chimpanzees will be exposed to several hundred kilograms of meat infected with retroviruses, the vast majority of which will be coming from red colobus (Leendertz *et al.* 2011), while humans will be exposed to retroviruses from a much broader range of species.

Results at Taï National Park have confirmed the importance of red colobus as a source of retroviruses for chimpanzees. Most STLV-1 sequences obtained from chimpanzees all belong to clades consisting of one of their prey's STLV-1, with a majority being closely related to STLV-1 identified in red colobus (Leendertz *et al.* 2004). Two SFVcpz-infected individuals were also found to harbour SFV from their prey, in both cases SFVwrc from red colobus (Leendertz *et al.*

2008). To date, no evidence of a transmission chain in chimpanzees showing the spread of newly acquired retrovirus has been documented at Taï. SIVcpz, which arose in Central West African chimpanzees after their split from the Western sub-species, consists of a mosaic genome of SIVs from two of its prey species (red-capped mangabeys - *Cercocebus torquatus* - and either mustached guenons - *Cercopithecus cephus*, mona monkeys - *Cercopithecus mona*, greater spot-nosed monkeys - *Cercopithecus nictitans* - or an ancestor of these cercopithecines; Courgnaud *et al.* 2003; Sharp *et al.* 2005). The Taï National Park chimpanzees also regularly consume SIV infected monkeys, especially red colobus infected with SIVwrc. This triggered a targeted search for SIVwrc infections in the Taï National Park chimpanzees, but despite extensive efforts (about 10% of the Taï National Park chimpanzee population), no seropositive (n=23) or PCR-positive (n=30) individual could be detected (Leendertz *et al.* 2011).

A number of recent studies have provided insights into host mechanisms for SIV/HIV resistance and how these are evaded by different viral genes; these studies have provided evidence for the role of a number host receptors (e.g., CCR5; Samson *et al.* 1996) and restriction factors (e.g., APOBEC3G, TRIM5a, tetherin, reviewed in; Malim & Bieniasz 2012), as well as a number of viral genes that antagonize these antiviral host restriction factors (e.g., Vif, Vpr, Vpu, Nef, reviewed in: Kirchhoff 2010). With respect to the lack of SIVwrc infections in the Taï National Park chimpanzees, a seemingly important interaction is that of host APOBEC3G and lentiviral viral integration factor (Vif). APOBEC3G is a intracellular cytidine deaminase that restrict retroviruses by hypermutating their genomes. Retroviruses counteract APOBEC3G with the activity of Vif, which promotes APOBEC3G degradation by the proteasome. The binding site of Vif onto NHP APOBEC3G was recently shown to be well conserved among cercopithecines, although the sequence itself was under strong positive selection (Compton & Emerman 2013). By contrast, the site of binding of Vif in the colobine APOBEC3G has shifted, most likely as a way for colobine SIVs to cope with a unique insertion in their host's APOBEC3G sequence that might have concealed the "classical" binding site (Compton & Emerman 2013). A by-side of this adaptation is that SIVolc (the only colobine SIV tested for sensitivity to APOBEC3G in: Compton & Emerman 2013) is fully sensitive to the activity of all other NHP APOBEC3G against which it was tested. While further studies are needed to specifically verify that SIVwrc is efficiently restricted by chimpanzee APOBEC3G,, it is restricted by human APOBEC3G, which is strictly identical to chimpanzee APOBEC3G at the Vif binding site (Compton & Emerman

2013). This suggests that SIVwrc is also efficiently restricted by chimpanzee APOBEC3G, preventing infections from propagating, despite extremely high levels of exposure. Conversely, given that SIVsmm efficiently inhibits human APOBEC3G, the lack of SIVsmm transmission into the Taï National Park chimpanzee population (which contrasts with the frequent transmission of SIVsmm to humans; see below) may either be a result of the rarity with which chimpanzees hunt this species or another, as yet undescribed, restriction factor.

### A.6 ZOONOTIC TRANSMISSION OF RETROVIRUSES IN THE TAÏ NATIONAL PARK AREA

HTLV-1 strains found in West Africa mostly belong to the HTLV-1 subtype A, often referred to as the Cosmopolitan subtype, a clade restricted to humans for which there is no evidence of recent zoonotic transmission events. This makes humans inhabiting the Taï National Park area an ideal test case for local STLV-1 transmission, as zoonotic strains would be immediately identifiable through phylogenetic analyses. Out of ten HTLV-1 strains identified in villages bordering Taï National Park, six belonged to the subtype A while four were most closely related to local NHP STLV-1 strains (Calvignac-Spencer *et al.* 2012a). Interestingly, three were likely the results of cross-species transmission from sooty mangabeys while another apparently stemmed from a red colobus monkey, strongly suggesting multiple zoonotic transmission events have occurred in this hunter/prey system.

SIVsmm appears to have crossed the species barrier several times in the Taï National Park area and it is hypothesized that the West African HIV-2 epidemics started here (Santiago *et al.* 2005). In fact, SIVsmm transfer seems to happen relatively frequently as evidenced by a recent study documenting ongoing transmission of SIVsmm through the identification of a ninth HIV-2 lineage (Ayouba *et al.* 2013). Mirroring what is observed in the chimpanzee population, SIVwrc and SIVolc have not been detected in the local human population. As discussed above, this might be due in part to the action of human APOBEC3G, which efficiently restricts SIVolc and, maybe, SIVwrc (Compton & Emerman 2013).

While zoonotic transmission of SFV to humans has been documented in numerous studies, whether secondary transmission occurs remains an open question (Switzer *et al.* 2004; Wolfe *et al.* 2004; Switzer *et al.* 2008; Betsem *et al.* 2011; Mouinga-Ondémé *et al.* 2012). At the moment these transmission events, thought to occur during bushmeat hunting, appear to remain isolated

cases that occur infrequently and SFV has not become established in a human population (reviewed in: Gessain *et al.* 2013). Around Taï National Park there are no conclusive data on SFV infections in humans; serological evidence suggested that infections may be present (M. Peeters unpublished data; Ali *et al.* 1996) but no virus genetic material could be amplified from these samples (F Leendertz unpublished data).

### A.7 RETROVIRUSES IN A CHANGING ENVIRONMENT

The ongoing massive climate and habitat changes induced by human activities have led many primate communities into non-equilibrium states (Chapman *et al.* 2010a). This dynamism in the backdrop in which retroviruses circulate creates new evolutionary pressures and epidemiological processes, which might result in new patterns of cross-species and zoonotic transmission. For example, red colobus are also the most frequently hunted prey of the eastern chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo, in Kibale National Park, Uganda (Mitani & Watts 1999; Watts *et al.* 2012), where unsustainable levels of chimpanzee hunting pressure have caused the red colobus population to undergo a major decline (Teelen 2008). At Kyambura Gorge in Queen Elizabeth National Park in Uganda, no red colobus are present and black-and-white colobus (*Colobus guereza*) are the most frequently hunted prey (Krüger *et al.* 1998). This suggests that chimpanzees shift their prey in the face of changing availability, which will likely have major impacts on zoonotic retroviral transmission patterns; especially if resistance is strain specific as discussed above. For example, if sooty mangabeys become a regular prey item of the chimpanzees in Taï, SIVsmm transmission to chimpanzees could be facilitated (assuming present low exposure is the main explanation for the absence of SIVsmm infection in chimpanzees). Human food preferences and hence exposure may also shift in light of changing prey availability. At Taï National Park, current human hunting levels of all monkeys species have led to their extirpation in many areas within the park (N'Goran *et al.* 2012). The dramatic changes in monkey population distributions at Taï National Park will likely have major impacts on their retroviruses.

### A.8 CONCLUSIONS

The retroviral studies pursued at Taï National Park highlight the importance of the ongoing integration of primate ecology and behaviour findings for the interpretation of viral data. The primatology research programs will be indispensable partners for improving our understanding of retroviral ecology in the coming years. Collaborative efforts at other long-term field sites have already yielded results as striking as the discovery of increased mortality, decreased fitness and AIDS-like disease symptoms caused by SIVcpz in the chimpanzees of Gombe, Tanzania (Keele *et al.* 2009). Findings such as these would not have been possible without dense longitudinal sampling in combination with detailed primate behaviour and life history data. Yet the infection status of many NHP species at Taï National Park remains unknown and even in those species that have been carefully scrutinized, within-species retroviral transmission dynamics are poorly understood. Further, the fitness costs incurred by hosts, genetic counter-strategies deployed in the arms race between viruses and hosts, and the interplay with concurrent co-infections and a host's resident microbiome remain completely unexplored at Tai National Park. Fortunately, the use of non-invasive samples is coming of age, greatly expanding the number of samples and species available for study (Santiago *et al.* 2002). Ultimately, an increased knowledge about the viral ecology within the park will hopefully help provide a better understanding and predictive framework for the emergence of retroviruses in this area, particularly in the face of a changing environment.

## A.9 ACKNOWLEDGEMENTS

We thank the Max Planck Institute for Evolutionary Anthropology, the Ivorian Ministry of Environment and Forests, the Ministry of Research, the directorship of the Tai National Park, the Office Ivoirien des Parcs et Réserves , the Centre Suisse de Recherche Scientifique, and the Taï Chimpanzee Project and its team of field assistants for their continuous support. This work was supported through the Deutsche Forschungsgemeinschaft (DFG LE1813/4-1). JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336), the Canadian Institutes of Health Research's Strategic Training Initiative in Health Research's Systems Biology Training Program, an NSERC Vanier Canada Graduate Scholarship (CGS), and a Quebec Centre for Biodiversity Science Excellence Award. For discussions and suggestions on this manuscript,

JFG thanks Colin Chapman and Leone Brown. Finally we thank Genevieve Campbell and Hjalmar Kühl for providing the Fig. A.1 of this manuscript.

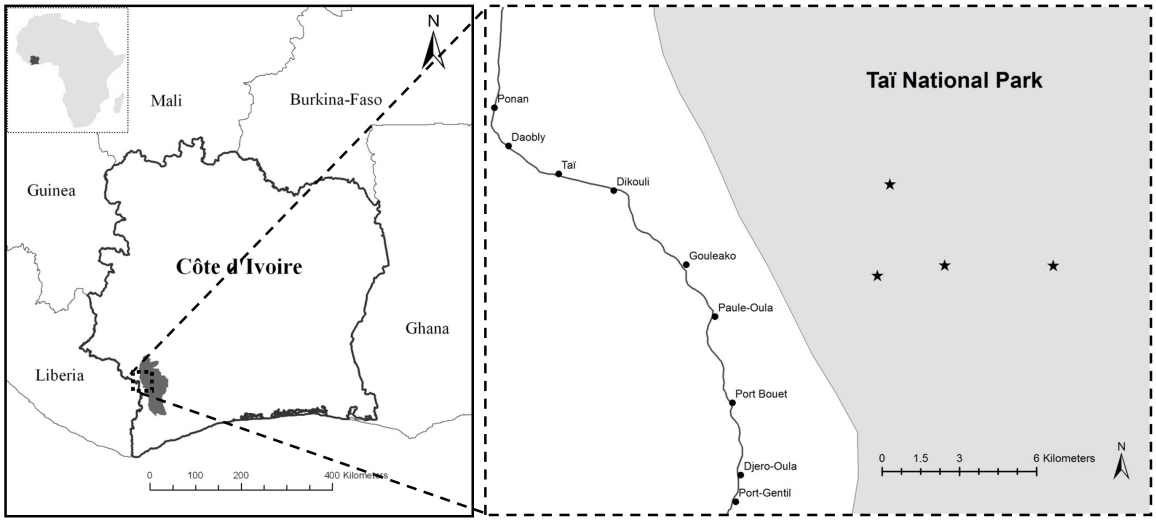



**Figure A.1: Location of the main study area.** a) Location of the Taï National Park in Africa and Côte d'Ivoire; b) Close-up of the research area. Stars indicate primatology research camps. This figure was provided by Genevieve Campbell and Hjalmar Kühl and is derived from (Campbell *et al.* 2011).

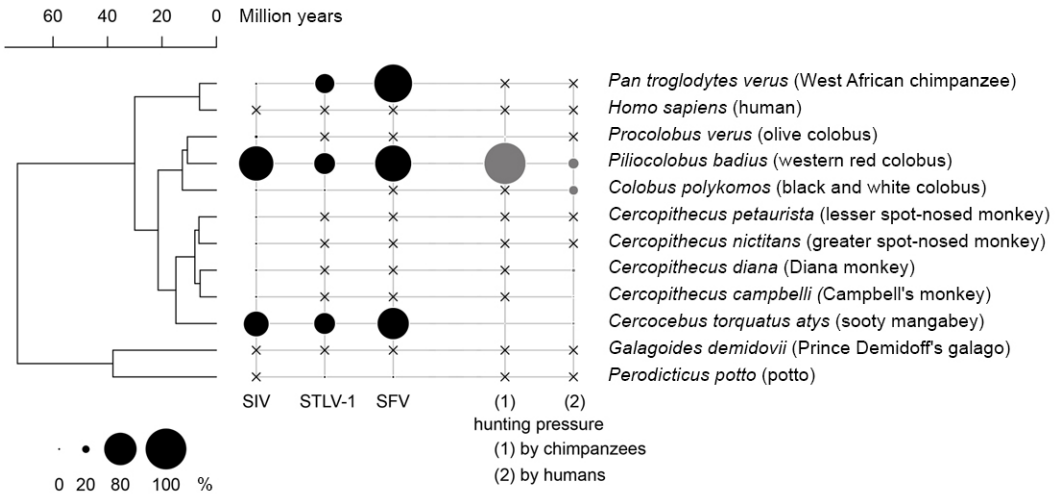



**Figure A.2: Phylogenetic relationships, retroviral prevalence and hunting pressure on primates in Taï.** This maximum clade credibility tree was identified from a sample of 1,000 posterior trees obtained from the 10k project (version 3; Based on eleven mitochondrial, six autosomal genes, and two Y-chromosome genes using a Bayesian MCMC performed with MrBayes under a calibrated, relaxed clock model; Arnold *et al.* 2010). Time axis is in millions of years. All branches are supported by maximal posterior probability values (1.0) but the two supporting the cercopithecine and the spot-nosed monkey clades (0.86 for both). Circle sizes are proportional to prevalence (black circles) or hunting pressure (grey circles).

# SUPPLEMENTARY MATRIALS FOR CHAPTER 4

## HUNTING CAPTURES MORE THAN PREY: EVIDENCE FOR MICROBIOME ASSIMILATION BY CHIMPANZEES

## S.4.1 SUPPLEMENTARY METHODS

### S.4.1.1 Intraspecific examination of beta diversity

#### *4.1.1.2 Intraspecific examination of beta diversity using Bray-Curtis*

We also used the Bray-Curtis index to examine pairwise dissimilarity between the bacterial communities of chimpanzee and sooty mangabey microbiomes, conducting all comparisons as conducted using UniFrac. The Bray-Curtis index incorporates abundance information and does not increase based on shared absences (Bray & Curtis 1957). Bray-Curtis dissimilarities were calculated using an implementation in the R package *Vegan* (Oksanen *et al.* 2016), and range from 0, when communities are completely similar, to 1, when communities are completely dissimilar.

#### *4.1.1.2 Testing for microbiome community differences by species, social group, and individual*

We followed Ochman et al. (2010) to evaluate whether similarity between primate fecal microbiome mirrors the phylogenetic relatedness of the hosts. In addition to the threshold based analysis, we also considered each bacterial OTU as a character assigned to one of two states, presence or absence (hereafter referred to as the presence-absence based approach) within each NHP host.

## S.4.2 Supplementary results

#### *S.4.2.1 Bray-Curtis*

Using both sampling periods together, we also found that samples from the same individual were more similar than samples from different individuals ($n_{samples}$ = 229, $n_{individuals}$ = 59, $\bar{x}_{different\ individuals}$ = 0.563, $\bar{x}_{same\ individual}$ = 0.512, Mantel test: permutation $P < 0.001$; Fig. S4.2A). This effect was also more pronounced when we used only the 2014 samples ($n_{samples}$ = 154, $n_{individuals}$ = 58, $\bar{x}_{different\ individuals}$ = 0.549, $\bar{x}_{same\ individual}$ = 0.446, Mantel test: permutation $P < 0.001$) or the 2015 samples ($n_{samples}$ = 74, $n_{individuals}$ = 27, $\bar{x}_{different\ individuals}$ = 0.544, $\bar{x}_{same\ individual}$ = 0.497, Mantel test: permutation $P < 0.001$). The microbiomes of samples collected between years were more similar when they originated from the same individual than from different individuals, also suggesting the similarity of microbiomes from an individual persisted between years ($n_{samples}$ = 229, $n_{individuals}$ = 59, $\bar{x}_{different\ individuals\ different\ years}$ = 0.582, $\bar{x}_{same\ individual\ different\ years}$ = 0.561, Mantel

test: permutation $P = 0.003$; Fig. S4.2B). For the 26 individuals sampled in both years, gut microbiome similarity was also greater within than between sampling years ($n_{samples} = 146$, $\bar{x}_{between\ sampling\ years} = 0.565$, $\bar{x}_{with\ sampling\ years} = 0.475$, Wilcoxon test: $T^{+} = 348$, $N = 26$, $P < 0.001$; Fig. S4.2C). For sooty mangabeys, samples from the same study group were more similar to one another than samples from different groups ($n_{samples} = 258$, $n_{individuals} = 87$, $\bar{x}_{different\ group} = 0.593$, $\bar{x}_{same\ group} = 0.563$, Mantel test: permutation $P = 0.004$). Bacterial communities from mother-offspring pairs were not significantly different from non-mother offspring-pairs when measured using the Bray-Curtis dissimilarity index ($n_{samples} = 117$, $n_{mother\text{-}offspring\ pairs} = 18$, $\bar{x}_{non\text{-}mother\text{-}offspring\ pair} = 0.543$, $\bar{x}_{mother\text{-}offspring\ pair} = 0.519$, Mantel test: permutation $P = 0.106$; Fig. S4.2D).

For the South chimpanzee study group, samples from the same individual were more similar to one another than samples from different individuals ($n_{samples} = 47$, $n_{individuals} = 24$, $\bar{x}_{different\ individuals} = 0.638$, $\bar{x}_{same\ individual} = 0.545$, Mantel test: permutation $P < 0.001$; S2E). For the North group samples from the same individual were also more similar to one another than samples from different individuals ($n_{samples} = 23$, $n_{individuals} = 12$, $\bar{x}_{different\ individuals} = 0.653$, $\bar{x}_{same\ individual} = 0.554$, Mantel test: permutation $P = 0.003$; S2F). Samples from the three study groups were more similar within groups than between groups ($n_{samples} = 98$, $n_{individuals} = 64$, $\bar{x}_{Bray\text{-}Curtis\ different\ group} = 0.653$, $\bar{x}_{Bray\text{-}Curtis\ same\ group} = 0.633$, Mantel test: permutation $P = 0.006$).

### *S.4.2.2 Microbiome community differences by species, social group, and individual: presence/absence*

The presence-absence maximum parsimony analysis suggested distinct chimpanzee, colobine, and cercopithecine clades, though it was common to observe a sooty mangabey or colobine sample occasionally appearing in another clade (Fig. S4.5). The average branch length distance between samples of the different primate species to chimpanzee samples, also supported the proximity of chimpanzees to colobines (Fig. S4.6A). Similarly, the average branch length distance between samples of the different primate species to samples of three colobine species provided support for a colobine clade, with chimpanzees as a sister clade (Fig. S4.6 B-C).

For the rooting of presence-absence based maximum parsimony phylogeny; none of 100 bootstrap replicates supported a chimpanzee outgroup to all other primate samples. The rooting of 100 bootstrap replicates was consistent with a chimpanzee-colobine clade, though for 47 of these one colobine sample was placed in the cercopithecine clade, while for another 13, one

sooty mangabey sample appeared in the colobine clade. The BaTS analysis of the presence-absence maximum parsimony phylogenies revealed strong phylogenetic structure of samples on these trees based on the host species (Table S4.2). A BaTS analysis of the reduced presence-absence phylogenies including only chimpanzees or sooty mangabeys, found that social group membership and individual identity were also structured on these phylogenies (Table S4.3-7).

The UPGMA clustering found that samples from a primate species tended to cluster together, though several of the cercopithecine species clustered in the sooty mangabey clade (Fig. 4.4A). Similarly, one chimpanzee sample was found closer to the colobines than to other chimpanzee samples. Unfortunately such an approach does not allow for anwith empirical rooting of this topology, though placement of the root on the branch leading from the chimpanzee clade to the cercopithcine clade revealed a phylogeny extremely similar to that obtained from the maximum parsimony analysis (Fig.4B). These results again, are generally inconsistent with the primate phylogeny and suggest a colobine-chimpanzee clade (Fig. 4.2B). Using UniFrac dissimilarity scores, we compared each sooty mangabey, chimpanzee, and red colobus sample to every other sample collected and grouped these dissimilarity scores by species. This analysis suggested that sooty mangabeys had the most shared branch length with the other cercopithecines (Fig. S4.7A), suggesting functionally similar communities mirroring their close phylogenetic relationship. Chimpanzee (Fig. S4.7B), and red colobus (Fig. S4.7C) microbiomes tended to have the most similar communities with conspecifics, also supporting the findings from the GLMM and PAUP based analysis and suggesting a strong effect of a hosts species on the composition of the microbiome. While there was a great deal of variation in UniFrac dissimilarity by species, sooty mangabeys appeared to have microbiomes most similar to other cercopithecines and much more dissimilar communities when compared with chimpanzees and colobines, which was highlighted by the UPGMA analysis.

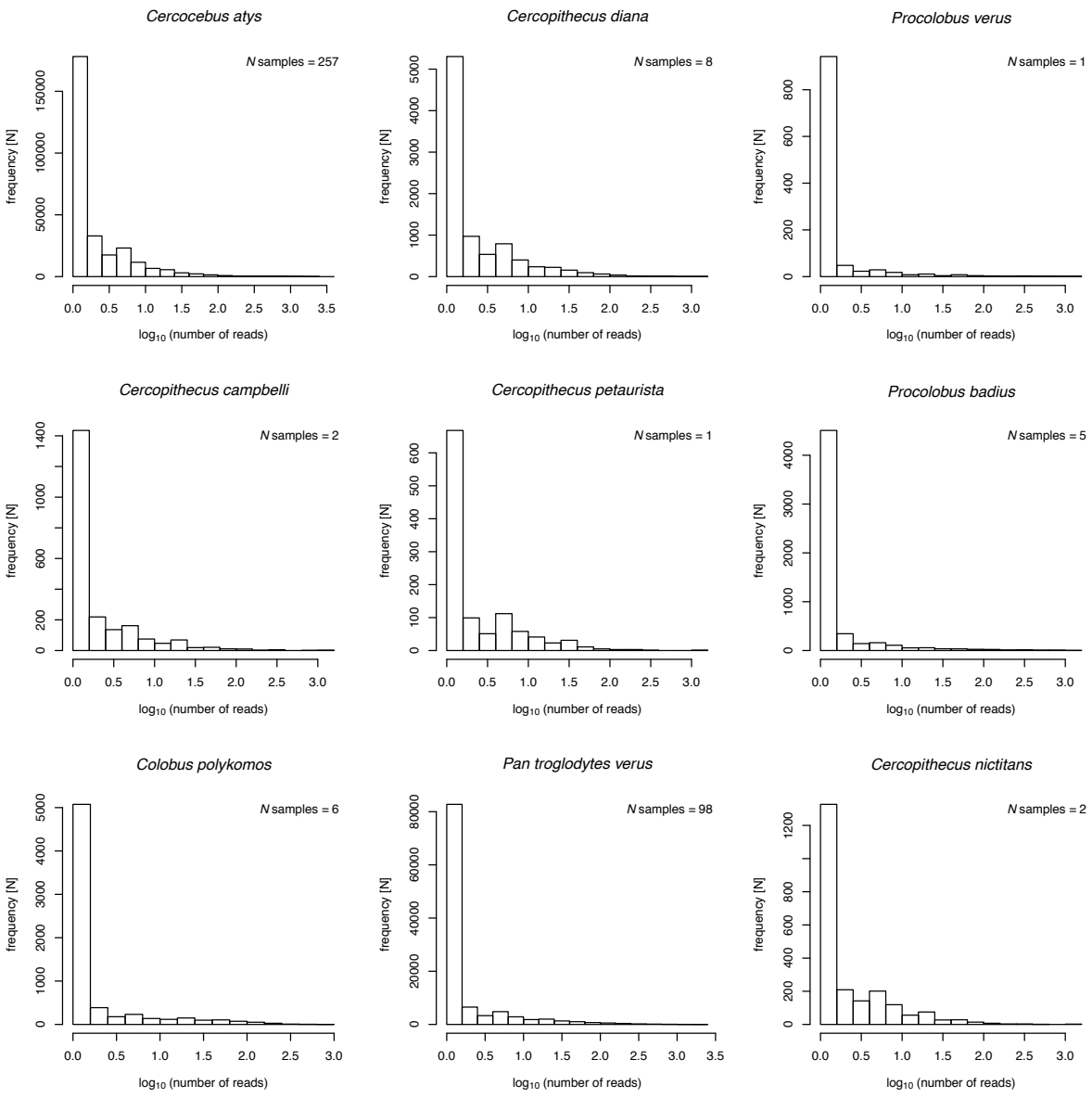



**Figure S4.1: Number of reads assigned to each bacterial OTU by host species.** Sample size for each NHP host is indicated in the upper right of each group.

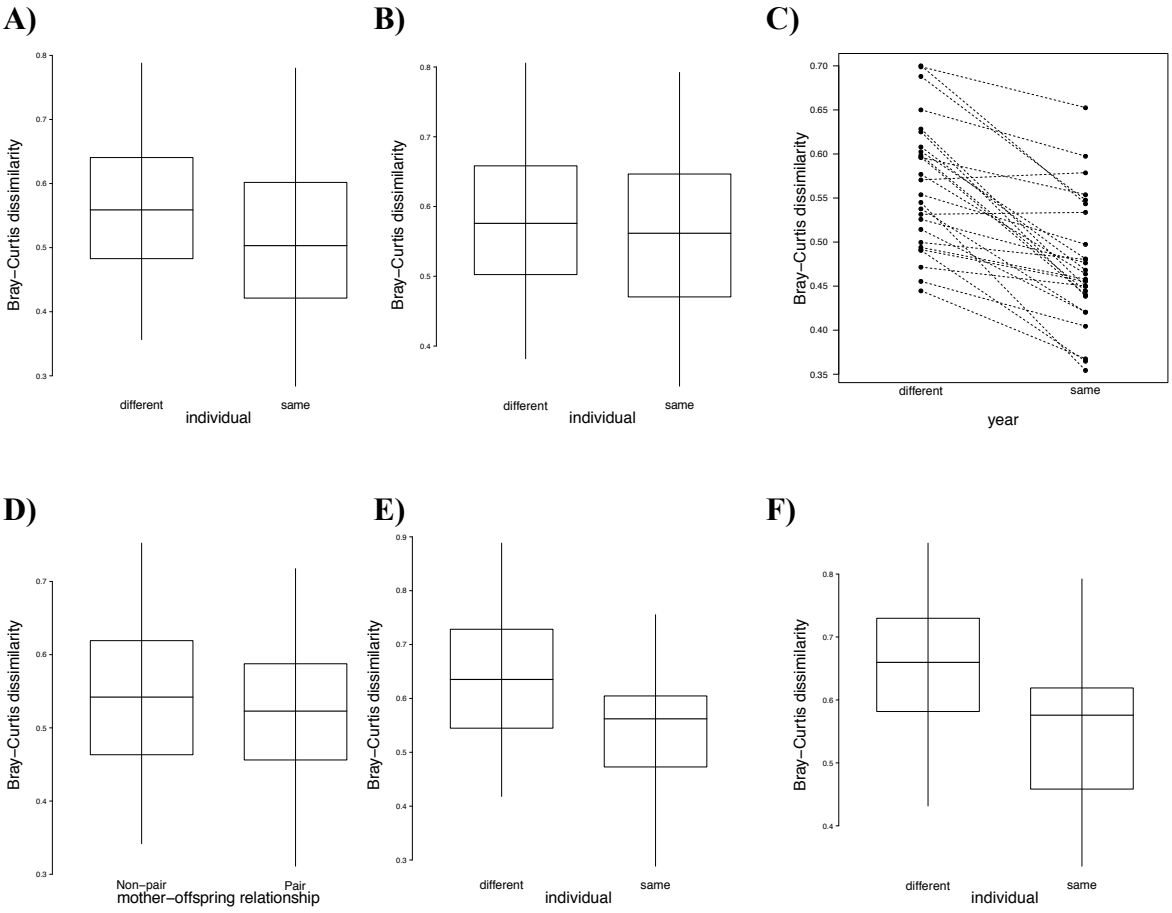
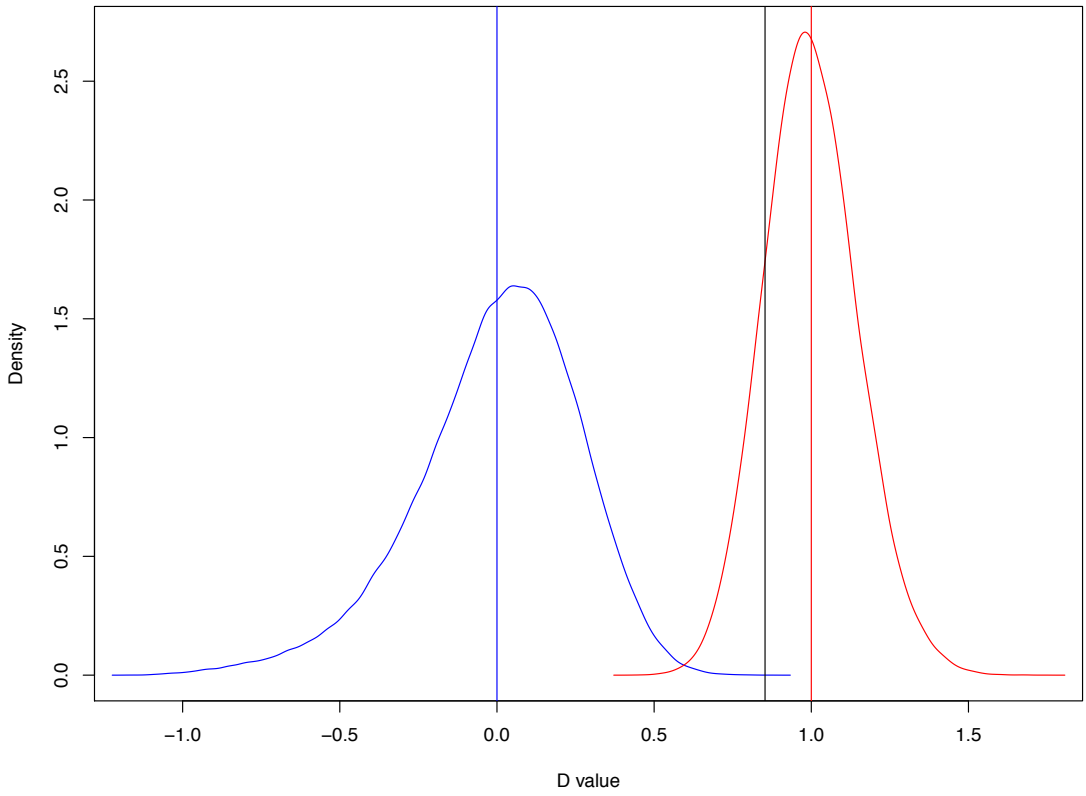



**Figure S4.2**: **Bray-Curtis dissimilarities between samples** from (**A**) sooty mangebeys in the Audrenissrou group, when stemming from the same or different individuals, (**B**) between samples from different sampling years when stemming from the same or different individuals (**C**) for individuals sampled in the same or different years, (**D**) when stemming from a mother offspring-pair or a non-mother-offspring pair, (**E**) between samples from the same or different individuals for chimpanzees in the South group and (**F**) North group. The middle horizontal line represent the median while the rectangle shows the quartiles and the vertical line represents the 2.5 and 97.5% percentiles. Dashed lines in (**C**) indicate the paired nature of the dataset, connecting the dissimilarities for samples from each individual from the same or different sampling year.

**A)**



**B)**

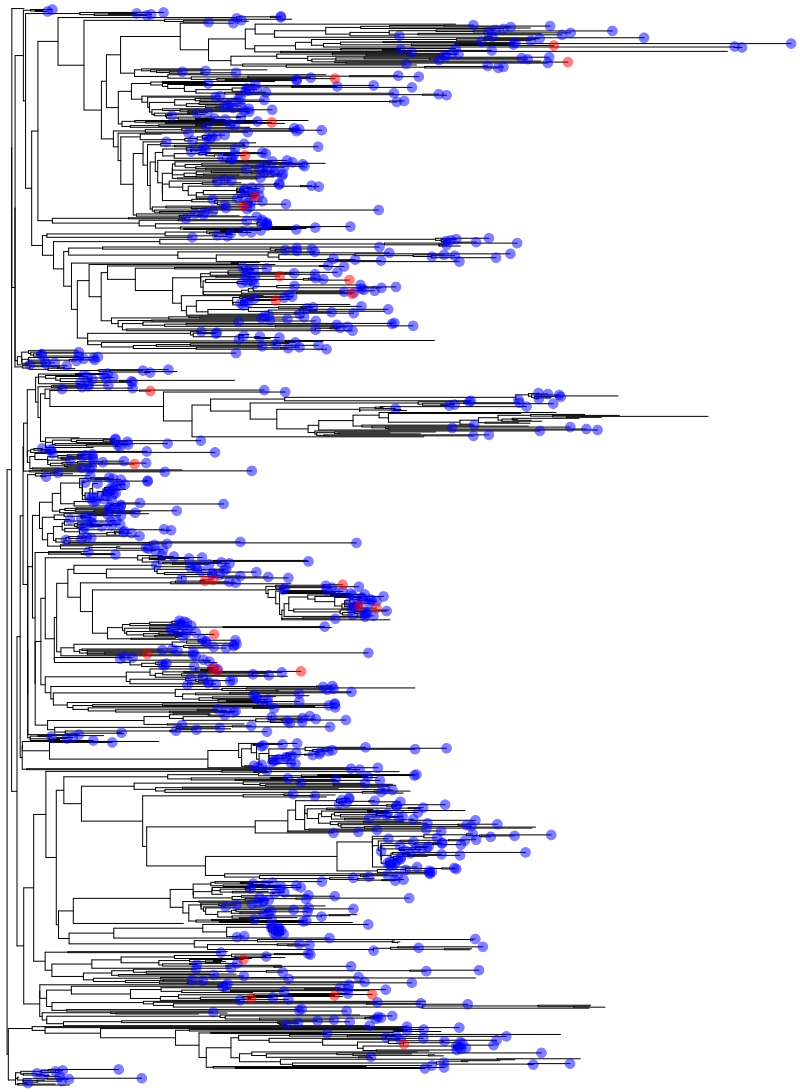


**Figure S4.3: A) Testing for phylogenetic signal in bacterial taxa for which host species had a major effect on the abundance**. Density plots of the distributions of the two simulations relative to the observed D value (solid vertical black line); a value of 0 indicates that the trait was as phylogenetically conserved as expected under a Brownian threshold model (shown in blue) while a value of 1 indicates a random distribution across the phylogeny (shown in red). B) Phylogeny of bacterial OTUs; those with a significant species effect are indicated with a blue circle at the tip, while those with a non-significant species effect are shown in red.


**Figure S4.4: A) Phylogeny of samples estimated with a heuristic maximum parsimony using the threshold based scores of bacterial abundance as characters.** Colors indicate the host's species. Names indicate the species, the social group, individual host, and a unique sample identifier separated by an underscore; these are illegible in a printed version though readers are invited to zoom in on the electronic version to see these details. Bootstrap support is shown for the species clades and the root position was supported by our TempEst analysis. **B)** Phylogeny of the primate hosts based on eleven mitochondrial and six autosomal genes made available through the 10kTrees project (Arnold *et al.* 2010).

Click on the image to be taken to a high resolution version available at the end of this document.

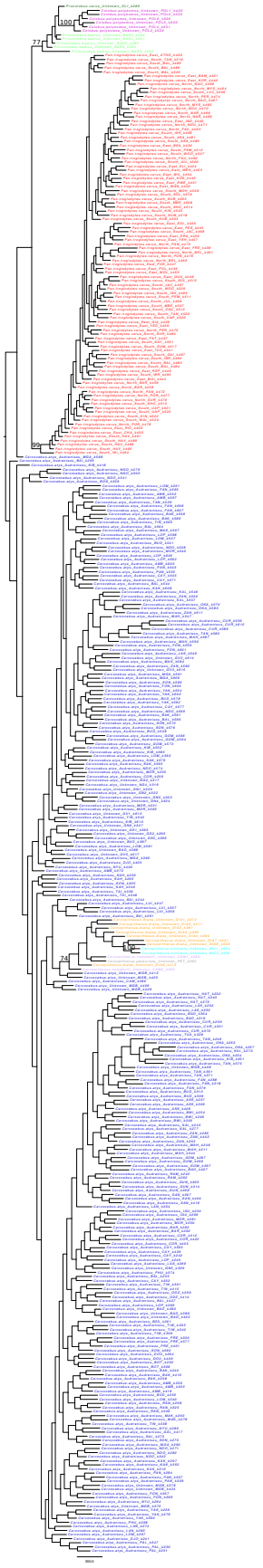


**Figure S4.5: Phylogeny of samples estimated with a heuristic maximum parsimony considering the presence or absence of bacteria as characters.** Colors indicate the host's species, while taxa names indicate the species, the social group, individual host, and a unique sample identifier separated by an underscore; these are illegible in a printed version though readers are invited to zoom in on the electronic version to see these details. Bootstrap support is shown for the species clades and the root position was supported by our TempEst analysis.

Click on the image to be taken to a high resolution version available at the end of this document.

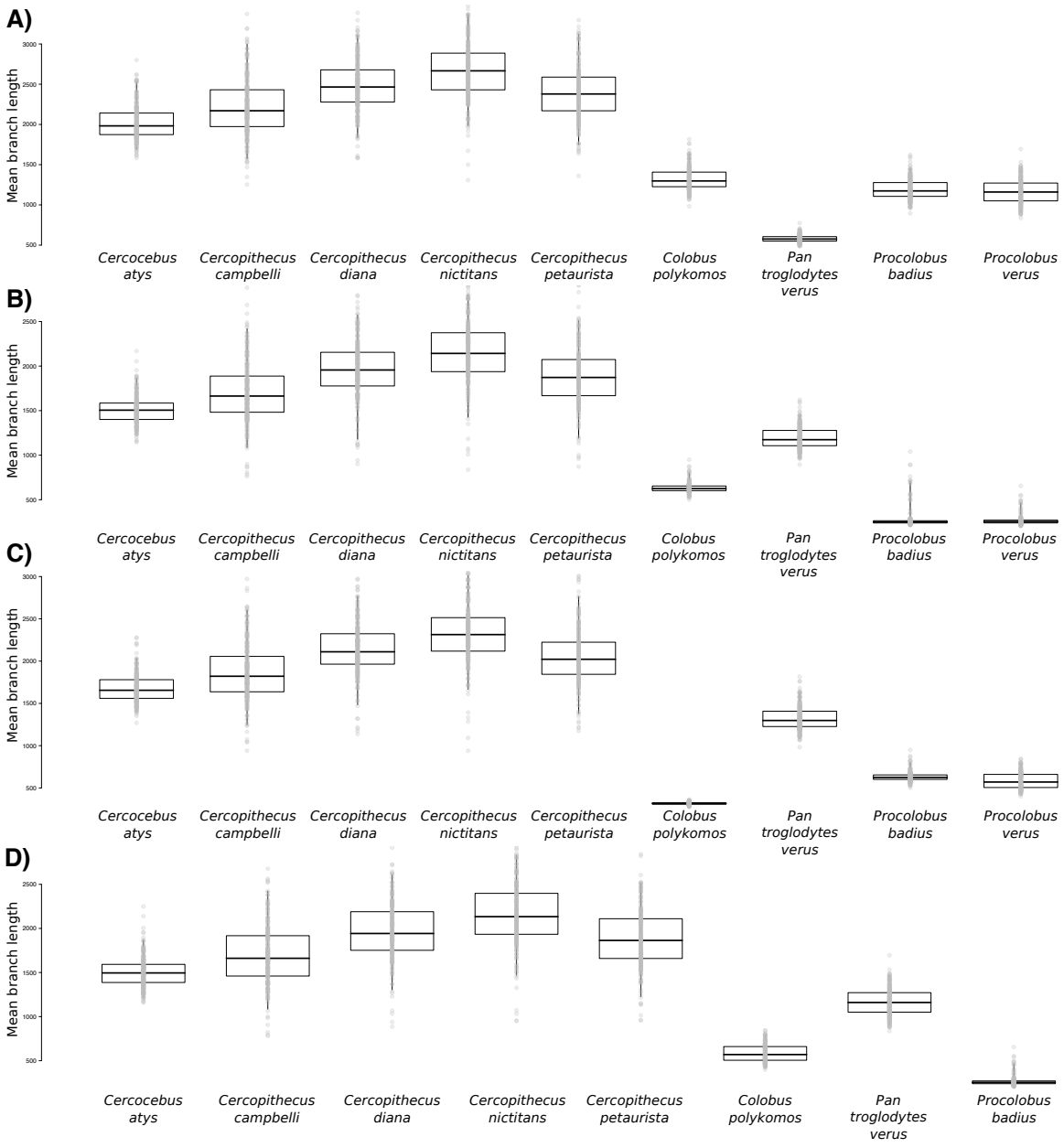



**Figure S4.6: Mean branch lengths between species for presence-absence based maximum parsimony analysis.** Based on comparisons of mean branch lengths of all samples to samples of A) chimpanzees, B) red colobus, C) king colobus, and D) olive colobus.

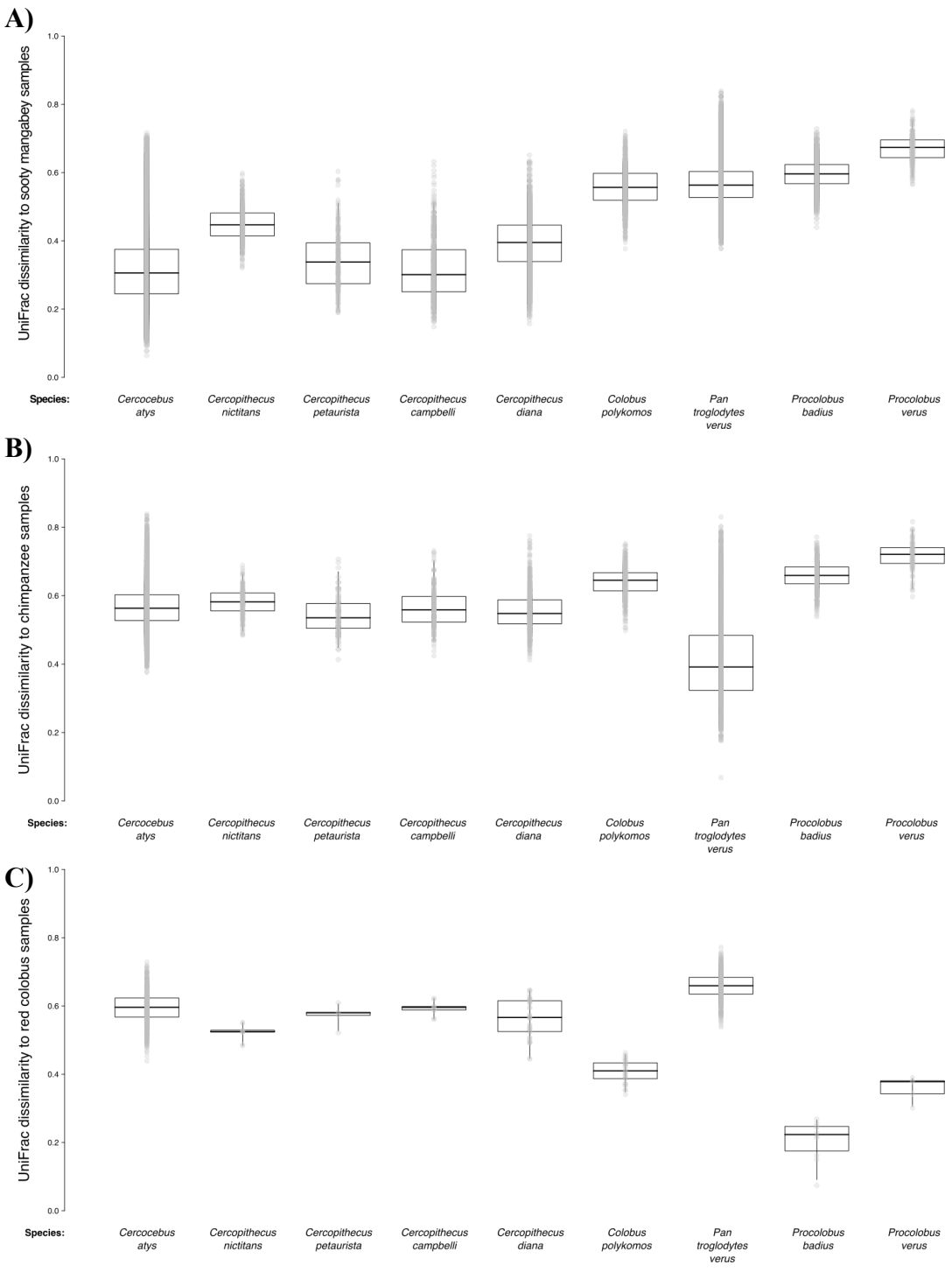



**Figure S4.7: Unifrac dissimilarity scores** between every **A)** sooty mangabeys, **B)** chimpanzees, and **C)** red colobus, sample and all other samples, grouped here by species.

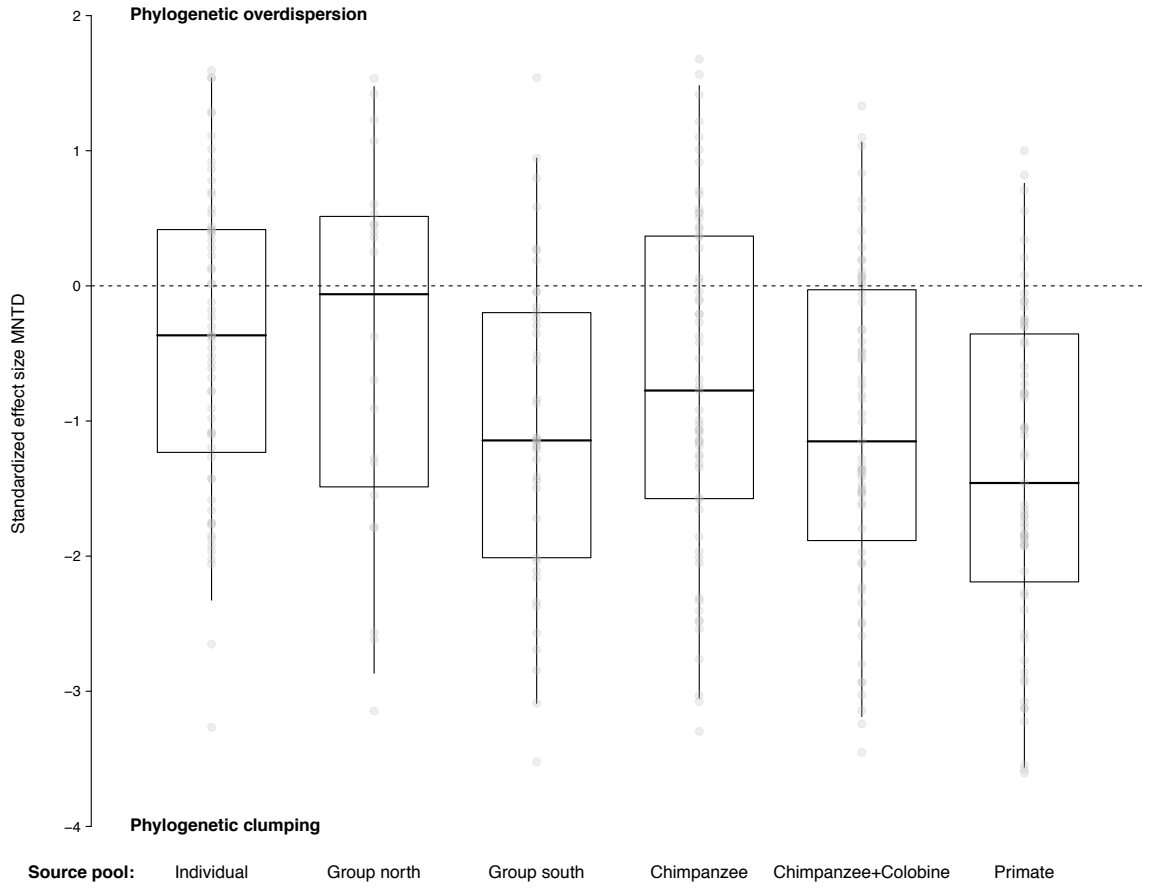



**Figure S4.8: Chimpanzee SES MNTD based on null model simulations of the bacterial community in each fecal sample using different bacterial source pools**. We ran the group analysis for each of the two groups for which we had repeated sampling of individuals. In addition we included a source pool level that consisted of bacteria found in any colobine or chimpanzee sample. The solid middle horizontal line of the rectangles represents the median, the rectangle shows the quartiles and the vertical line represents the 2.5 and 97.5% percentiles, while the values for each sample are indicated by overlapping gray circles. Values above the dashed line are those exhibiting phylogenetic overdispersion while those below the line exhibit phylogenetic clumping.

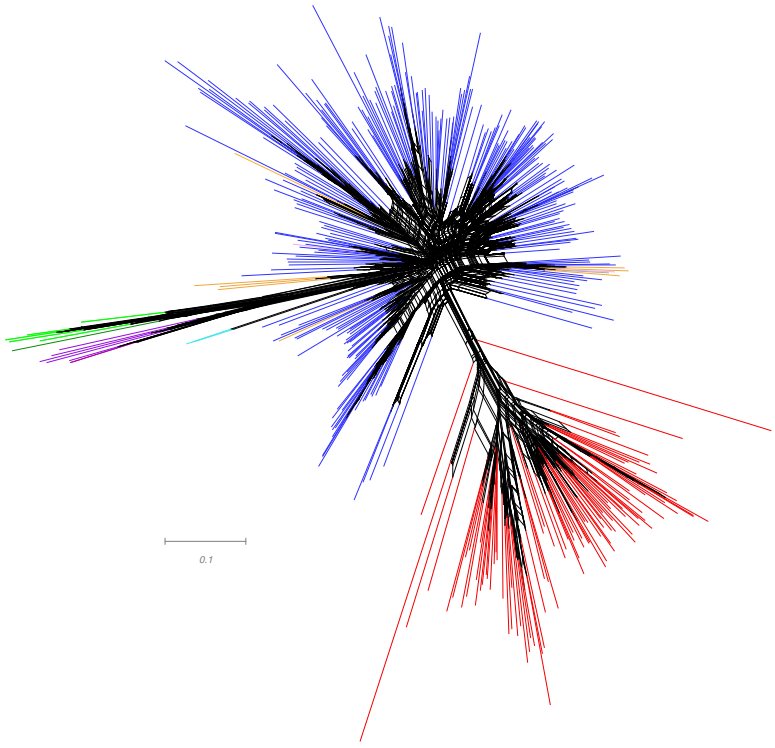



**Figure S4.9: Unrooted phylogenetic network built using the unifrac dissimiarlity matrix as input and the SplitsTree4.** Tips represent samples and are colored by the host species as in Figure S4.4.

**Supplementary Table 4.1: Spearman rank correlations between behavioral associations and bacterial community dissimilarity measured with Bray-Curtis**

| | 2014 | | 2015 | |
|---|---|---|---|---|
| Behavior | $r_{sb}$ | *P* | $r_{sb}$ | *P* |
| Grooming | -0.086 | 0.921 | -0.037 | 0.758 |
| Aggression | -0.013 | 0.637 | -0.061 | 0.792 |
| Proximity | -0.097 | 0.966 | -0.024 | 0.679 |

**Supplementary Table 4.2: Results of BaTS testing of host species relationships with phylogenetic position, using presence absence based maximum parsimony phylogeny**

| Test statistic | Host species | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| AI | | 0.461 | 20.044 | 17.940 | 22.288 | **<0.001** |
| PS | | 9.000 | 111.500 | 106.000 | 117.000 | **<0.001** |
| MC | *Cercocebus atys* | 97.000 | 8.270 | 6.000 | 13.000 | **0.010** |
| MC | *Cercopithecus diana* | 2.000 | 1.070 | 1.000 | 2.000 | 0.070 |
| MC | *Cercopithecus petaurista* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | *Cercopithecus nictitans* | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | *Cercopithecus campbelli* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | *Procolobus verus* | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | *Colobus polykomos* | 6.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | *Procolobus badius* | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | *Pan troglodytes verus* | 98.000 | 2.690 | 2.000 | 4.000 | **0.010** |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Supplementary Table 4.3: Results of BaTS testing for a correlation between group membership and phylogenetic position, using presence absence based maximum parsimony phylogeny**

| Test statistic | Social group | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| *Cercocebus atys* | | | | | | |
| Group: AI | | 2.585 | 5.948 | 4.864 | 7.195 | **<0.001** |
| Group: PS | | 17.000 | 27.130 | 26.000 | 28.000 | **<0.001** |
| MC | Audrenissrou | 14.000 | 15.600 | 11.000 | 23.000 | 0.640 |
| MC | Neighbor | 4.000 | 1.600 | 1.000 | 2.000 | **0.010** |
| *Pan troglodytes verus* | | | | | | |
| Group: AI | | 4.244 | 6.499 | 5.317 | 7.500 | **<0.001** |
| Group: PS | | 31.000 | 42.440 | 39.000 | 46.000 | **<0.001** |
| MC | East | 2.000 | 2.440 | 2.000 | 4.000 | 0.960 |
| MC | North | 5.000 | 3.570 | 2.000 | 6.000 | 0.120 |
| MC | South | 2.000 | 2.080 | 1.000 | 3.000 | 0.840 |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Supplementary Table 4.4: Full results of BaTS testing for a correlation between chimpanzee individual identity and phylogenetic position, using threshold based maximum parsimony phylogeny**

| Test statistic | Individual | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| AI | | 8.987 | 11.512 | 11.215 | 11.618 | **<0.001** |
| PS | | 81.000 | 90.110 | 88.000 | 92.000 | **<0.001** |
| MC | ATHO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | POL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WOL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BEA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WIL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IND | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | FAU | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BEL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MYS | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | NAO | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | NAR | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | NOU | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | LUC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ASA | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | ISH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ERA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PES | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | JAC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PEM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | EOL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | YEH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PAN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | TAN | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | JUL | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | SUM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MBE | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | WAL | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | KIN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KUB | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | SHO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BAM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KOR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | GIA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BAL | 2.000 | 1.070 | 1.000 | 2.000 | 0.080 |
| MC | UAP | 2.000 | 1.040 | 1.000 | 1.000 | **0.050** |
| MC | RIC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | POR | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | SUR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | OSC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | POS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | QUA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | YED | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WAN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WOO | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | PON | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PER | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | GAI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | FAT | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | SOL | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | KAY | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IVO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IBR | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | MAI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | REP | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BAR | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | RWE | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | CHA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | HAV | 4.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | WEH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ELI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KOS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MOH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | FRE | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | INI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Supplementary Table 4.5: Full results of BaTS testing for a correlation between mangabey individual identity and phylogenetic position, using threshold based maximum parsimony phylogeny**

| Test statistic | Individual | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| AI | | 12.708 | 29.210 | 28.640 | 29.723 | **<0.001** |
| PS | | 132.000 | 205.040 | 201.000 | 209.000 | **<0.001** |
| MC | DJO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PAL | 2.000 | 1.050 | 1.000 | 1.000 | 0.060 |
| MC | LOM | 1.000 | 1.060 | 1.000 | 2.000 | 1.000 |
| MC | BAK | 1.000 | 1.030 | 1.000 | 1.000 | 1.000 |
| MC | MOR | 2.000 | 1.090 | 1.000 | 2.000 | **0.100** |
| MC | COR | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | BOS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | FAN | 2.000 | 1.050 | 1.000 | 1.000 | 0.060 |
| MC | TIN | 1.000 | 1.030 | 1.000 | 1.000 | 1.000 |
| MC | SON | 2.000 | 1.050 | 1.000 | 1.000 | 0.060 |
| MC | AMB | 3.000 | 1.060 | 1.000 | 2.000 | **0.010** |
| MC | NGO | 3.000 | 1.020 | 1.000 | 1.000 | **0.010** |
| MC | NYU | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | MGA | 2.000 | 1.020 | 1.000 | 1.000 | **0.030** |
| MC | MAH | 3.000 | 1.070 | 1.000 | 2.000 | **0.010** |
| MC | TAN | 1.000 | 1.030 | 1.000 | 1.000 | 1.000 |
| MC | BUD | 2.000 | 1.060 | 1.000 | 2.000 | 0.070 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | YAK | 3.000 | 1.040 | 1.000 | 1.000 | **0.010** |
| MC | GOM | 2.000 | 1.080 | 1.000 | 2.000 | 0.090 |
| MC | BAL | 2.000 | 1.020 | 1.000 | 1.000 | **0.030** |
| MC | KAL | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | ZAN | 3.000 | 1.070 | 1.000 | 2.000 | **0.010** |
| MC | OKA | 2.000 | 1.060 | 1.000 | 2.000 | 0.070 |
| MC | FON | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | LAS | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | MAS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | LOP | 1.000 | 1.050 | 1.000 | 1.000 | 1.000 |
| MC | KIB | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | CAY | 2.000 | 1.050 | 1.000 | 1.000 | 0.060 |
| MC | NDO | 3.000 | 1.050 | 1.000 | 1.000 | **0.010** |
| MC | KAK | 3.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | PAN | 3.000 | 1.070 | 1.000 | 2.000 | **0.010** |
| MC | PHU | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | KAH | 3.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | BZA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BAI | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | RAN | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | SAL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PRE | 3.000 | 1.020 | 1.000 | 1.000 | **0.010** |
| MC | HAT | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | CUR | 5.000 | 1.040 | 1.000 | 1.000 | **0.010** |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | BAO | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | SAN | 3.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | RAM | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | ASS | 3.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | BWI | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | TUA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | GAL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IGU | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | BAR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | GUN | 3.000 | 1.010 | 1.000 | 1.000 | **0.010** |
| MC | MAK | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | DOU | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | LUI | 3.000 | 1.010 | 1.000 | 1.000 | **0.010** |
| MC | LAN | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | ODZ | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | BOT | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | CAM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | TIW | 2.000 | 1.020 | 1.000 | 1.000 | 0.030 |
| MC | TAI | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Supplementary Table 4.6: Full results of BaTS testing for a correlation between chimpanzee individual and phylogenetic position, using presence absence based maximum parsimony phylogeny**

| Test statistic | Individual | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| AI | | 7.990 | 10.653 | 10.322 | 10.767 | **<0.001** |
| PS | | 79.000 | 90.200 | 88.000 | 92.000 | **<0.001** |
| MC | ATHO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KUB | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BAM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KOR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | FAU | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | POL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WOL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | CHA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | HAV | 4.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | RIC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | POR | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | BAR | 2.000 | 1.020 | 1.000 | 1.000 | **0.030** |
| MC | WAL | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | SUR | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | GIA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | UAP | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | PAN | 1.000 | 1.030 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | YED | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KIN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | SHO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MBE | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | POS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | JAC | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | WOO | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | JUL | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | PEM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | TAN | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | OSC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ISH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IVO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | GAI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IBR | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | REP | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MAI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KAY | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | SOL | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | SUM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BAL | 2.000 | 1.020 | 1.000 | 1.000 | **0.030** |
| MC | FAT | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PER | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | PON | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | ERA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | YEH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | PES | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | IND | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | NAO | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | BEL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | LUC | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | NOU | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MYS | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | NAR | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | QUA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WAN | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | RWE | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BEA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | EOL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ASA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WIL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ELI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KOS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | WEH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | FRE | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MOH | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | INI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Supplementary Table 4.7: Full results of BaTS testing for a correlation between sooty mangabey individual identity and phylogenetic position, using threshold based maximum parsimony phylogeny**

| Test statistic | Individual | Observed Mean | Null mean | Null lower 95% CI | Null upper 95% CI | *P* |
|---|---|---|---|---|---|---|
| AI | | 14.333 | 28.221 | 27.604 | 28.698 | **<0.001** |
| PS | | 149.000 | 205.560 | 202.000 | 209.000 | **<0.001** |
| MC | PAL | 3.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | LOM | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | TAN | 2.000 | 1.070 | 1.000 | 2.000 | 0.080 |
| MC | SON | 1.000 | 1.080 | 1.000 | 2.000 | 1.000 |
| MC | MAS | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | FAN | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | BUD | 2.000 | 1.110 | 1.000 | 2.000 | 0.120 |
| MC | AMB | 2.000 | 1.070 | 1.000 | 2.000 | 0.080 |
| MC | PAN | 3.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | BAK | 2.000 | 1.060 | 1.000 | 2.000 | 0.070 |
| MC | TIN | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | LOP | 1.000 | 1.020 | 1.000 | 1.000 | 1.000 |
| MC | BAL | 2.000 | 1.040 | 1.000 | 1.000 | **0.050** |
| MC | NDO | 3.000 | 1.040 | 1.000 | 1.000 | **0.010** |
| MC | MOR | 2.000 | 1.050 | 1.000 | 1.000 | 0.060 |
| MC | FON | 1.000 | 1.040 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | CAY | 4.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | KIB | 1.000 | 1.070 | 1.000 | 2.000 | 1.000 |
| MC | KAK | 3.000 | 1.070 | 1.000 | 2.000 | **0.010** |
| MC | MGA | 2.000 | 1.040 | 1.000 | 1.000 | **0.050** |
| MC | BAR | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | YAK | 2.000 | 1.040 | 1.000 | 1.000 | **0.050** |
| MC | CUR | 3.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | GOM | 3.000 | 1.020 | 1.000 | 1.000 | **0.010** |
| MC | KAL | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | OKA | 1.000 | 1.030 | 1.000 | 1.000 | 1.000 |
| MC | ZAN | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | LAS | 2.000 | 1.040 | 1.000 | 1.000 | **0.050** |
| MC | MAH | 3.000 | 1.030 | 1.000 | 1.000 | **0.010** |
| MC | IGU | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | COR | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | HAT | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | GAL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | ASS | 3.000 | 1.020 | 1.000 | 1.000 | **0.010** |
| MC | GUN | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | BWI | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | BAO | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | SAN | 3.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | TUA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | NYU | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| MC | PHU | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | RAM | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | PRE | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | DOU | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | BOT | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | LAN | 2.000 | 1.030 | 1.000 | 1.000 | **0.040** |
| MC | ODZ | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | TIW | 1.000 | 1.010 | 1.000 | 1.000 | 1.000 |
| MC | CAM | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | KAH | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | BAI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | NGO | 2.000 | 1.000 | 1.000 | 1.000 | **0.010** |
| MC | BOS | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | DJO | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | TAI | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | BZA | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | RAN | 3.000 | 1.020 | 1.000 | 1.000 | **0.010** |
| MC | SAL | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| MC | MAK | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |
| MC | LUI | 2.000 | 1.010 | 1.000 | 1.000 | **0.020** |

PS - Fitch parsimony score; AI - association index MC - monophyletic clade statistic indicating the maximum observed exclusive single-state clade size.

**Supplementary Table 4.8: Model stability estimates for SES-MPD obtained from a model based on the full data set with maximum and minimum effect size estimates from models excluding one individual at a time.**

| | Original estimate | Minimum | Maximum |
|---|---|---|---|
| Intercept - Species: *Cercopithecus atys* | -3.713 | -4.029 | -3.035 |
| Species: *Cercopithecus diana* | -1.545 | -2.223 | -1.229 |
| Species: *Colobus polykomos* | -2.513 | -3.191 | -2.123 |
| Species: *Pan troglodytes verus* | 5.055 | 4.400 | 5.385 |
| Species: *Procolobus badius* | -0.591 | -1.269 | -0.275 |
| subject@(Intercept) | 0.764 | 0.000 | 0.786 |
| Group.Member@(Intercept) | 0.465 | 0.000 | 0.510 |

**Supplementary Table 4.9: Model estimates for GLMM of SES-MPD predicted by species.**

| | Estimate | Std. Error | *t* value | Lower CL | Upper CL |
|---|---|---|---|---|---|
| (Intercept): Species: *Cercopithecus atys* | -3.713 | 0.202 | -18.367 | -4.127 | -3.125 |
| Species: *Cercopithecus.diana* | -1.545 | 0.557 | -2.773 | -2.756 | -0.445 |
| Species: *Colobus.polykomos* | -2.513 | 0.620 | -4.051 | -3.826 | -1.292 |
| Species: *Pan.troglodytes.verus* | 5.055 | 0.279 | 18.113 | 4.300 | 5.625 |
| Species: *Procolobus.badius* | -0.591 | 0.667 | -0.887 | -1.982 | 0.719 |

CL: Confidence limit.

# SUPPLEMENTARY MATRIALS FOR CHAPTER 5

## TOOLS FOR OPENING NEW CHAPTERS IN THE BOOK OF *TREPONEMA PALLIDUM* EVOLUTIONARY HISTORY

## S.5.1 SUPPLEMENTARY MATERIAL

### S.5.1.1 Extraction Protocol

DNA was extracted using a silica-based method, following a modified protocol of Rohland and Hofreiter (2007) and Gamba et al. 2015 (Rohland & Hofreiter 2007; Gamba *et al.* 2015). For each bone 100-150mg of fine powder drilled at low speed were extracted using a 48h digestion in 5 ml of extraction Buffer A (0.5 M EDTA, 0.5% N-lauryl-Sarcosyl, 1 mg/mL Proteinase K at pH 8.0). The supernatant, recovered after spinning the solution at 2000 rpm for 5 min, was transferred into 20 ml of Buffer B (GuSCN 5 M, Tris 50 mM, NaCl 25 mM, EDTA 20 mM, TritonX-100 1x) with 80 µL of a fresh silica pellet prepared as described in Rohland and Hofreiter (2007). The final pH was adjusted to 4.0–5.0 using 37% HCl and pH paper. DNA binding to silica surfaces was performed for 3 h at room temperature with agitation. The supernatant was then removed except for 1ml after centrifugation at 2000 rpm for 2 min. The silica pellet was resuspended, transferred to a 1.5-mL tube and centrifuged at 1 2000 rpm for 2 min. The supernatant was removed and washed twice with 1 mL 80% ethanol, centrifuging again at 12000 rpm for 2 min. The silica pellet was dried for 20min at room temperature in the laminar flow-hood and resuspended in 90 µl of nuclease free water. The suspension was incubated for 20-30 min at 37°C and centrifuged at 12000rpm for 2 min. The supernatant was transferred to a new tube and stored at -20°C.

### S.5.1.2 PCR diagnostics for choosing samples for enrichment by hybridisation capture

Extracts were tested using a standard PCR amplifying a 67 bp DNA fragment, including primers from the DNA polA gene previously described in Leslie et al. (2007) and used by Knauf et al. to screen samples from olive baboons (2012). Amplifications were performed in 25 µl reactions containing 0.2 µM of each primer, 200µM dUTPs, 4 mM MgCl2, 2.5 µl 10X PCR buffer, 0.25 µl Platinum® Taq polymerase (Invitrogen, Carlsbad, USA), and 0.3 µl AmpErase® Uracil N-glycosylase (UNG) to degrade any potential contamination from PCR products generated in the laboratory. Assays were run under the following conditions: 7 min at 45°C, 5 min at 95°C, 40 cycles [15 s at 95°C, 45 s at 60°C, 60 s at 72°C], 7 min at 72°C. To enable sequencing of this short region, fusion primers were appended to those reactions positive on a gel; these reactions were run without AmpErase UNG as this would have degrade the product and with the following conditions: 5 min at 95°C, 5 cycles [15 s at 95°C, 45 s at 60°C, 60 s at 72°C], 7 min at 72°C. To

confirm the presence of *Treponema pallidum* and maximize the probability of success in a hybridisation capture experiments, the 12 samples that were sequence positive for *polA* were submitted to two nested PCRs amplifying longer sequences of the *Treponema pallidum* genome (Harper *et al.* 2008b). The first amplification round for GDP was performed in 25 µl reactions as for the first *polA* screening, though without AmpErase UNG, and under the following conditions: 5 min at 95°C, 35 cycles [30 s at 95°C, 60s at 55°C, 60 s at 72°C], 7 min at 72°C. The nested round was performed using 2µl of amplified product in a 25 µl reaction as for the first round. To reduce the possibility of contamination, the *GDP* and *cfpA* PCRs were also performed using simply the 2$^{nd}$ round of the nest PCR but including the AmpErase UNG step. Amplified products were sequenced using Sanger's sequencing and sequences were compared to publicly available sequences in EMBL through BLAST (Altschul *et al.* 1990).

### S.5.1.3 Library preparation protocols at Robert Koch Institute

The extracts from the three PCR positive bone samples were analysed using a Bioanalyzer: two samples (11787 and 22-52), were further fragmented using a Covaris S220 Focused-ultrasonicator® in a volume of 130ul low EDTA TE buffer) using settings to generate a 400 bp fragment side (Intensity = 4, Duty cycle = 10%, Cycles per burst = 200, Treatment time = 55 s, Temperature = 7°C). Those fragmented extracts were then concentrated using the MinElute PCR purification kit and eluted into 2x10ul low EDTA TE buffer. DNA concentrations of all extracts were then measured using a Qubit dsDNA High Sensitivity kit. For each extract a target of 1µg DNA or all available remaining DNA extract were used as input into the following library preparation procedure. Libraries were prepared using the Accel-NGS 2S DNA library kit following the standard protocol, with unique single indices. Libraries were quantified using the KAPA HiFI library quantification kit and subsequently amplified using the KAPA Hot Start Library Amplification Kit, using Illumina adaptor specific primers (5'-3': AATGATACGGCGACCACCGA and 5'-3': CAAGCAGAAGACGGCATACGA, 65°C Annealing Temperature and 45 s elongation) and then requantified. Extracts underwent variable number of cycles to reach the desired amount of starting material (15028 = 9 cycles, 11786 = 8 cycles, 22-52 = 8 cycles). Samples were pooled to equal contribute to a total of 258 ng DNA of starting material for input into the in-solution hybridisation capture, following the concentration with a MinElute PCR Purification Kit.

### S.5.1.4 In-Solution capture at the Robert Koch Institute using MYbaits Custom Target Enrichment Kit

RNA baits covering the entire *T. p. pertenue* Fribourg-Blanc genome were designed with 2-fold tiling of 120mer baits. We followed the Mybaits Sequence Enrichment for Targeted Sequencing protocol (Version 2.3.1) using a hybridisation time of 16 hours. Following this first round of capture, the surviving DNA was reamplified using the KAPA Hot Start Library Amplification Kit for 8 cycles. The pool was then quantified using the KAPA HiFi Library Quantification Kit and further amplified for 11 cycles to reach 100-500 ng starting material for the 2$^{nd}$ round of capture. DNA was concentrated using a MinElute PCR Purification Kit and 264 ng of DNA were input into the second round of hybridisation capture, performed using the same conditions as for the first round of capture described above. The surviving DNA was reamplified using the KAPA Hot Start Library Amplification Kit for 8 cycles, purified using the MinElute PCR Purification Kit, quantified using the KAPA HiFi Library Quantification Kit, and diluted to 4nM as input for the Illumina MiSeq (v3 2x300 Chemistry). Following the successful MiSeq run, a technical error occurred while writing the last 100bp of the 2$^{nd}$ read, but all reads could be still be assigned to their proper index and quality appeared stable for available reads.

### S.5.1.5 Library preparation protocols at Tübingen University

In advance of the library preparation the extracted DNA was sheared using the Covaris S220 machine in order to produce fragments between 300 and 500 bp with the following conditions: 5 intensity, 200 cycles per burst, 45 s. Afterwards the samples were concentrated to 30 µl each using Amicon ultra centrifugal filters (Millipore) according to manufacturer's instructions. Then all sheared and concentrated DNA extracts were converted into double-stranded Illumina libraries as described in Meyer and Kircher, 2010. Via amplification sample specific indexes were added to both library adapters to obtain double indexed libraries (Kircher et al. 2012). Library blanks were treated accordingly. Adaptor ligation and addition of the barcodes were both monitored with a quantification assay using the primer sets IS7, IS8 and IS5, IS6 (Meyer & Kircher 2010), the DyNAmo Flash SYBR Green qPCR Kit (Biozym) and the Lightcycler 96 (Roche).

A second amplification was performed for all indexed libraries in 100 µl reactions containing 5 µl library template, 4 units AccuPrime Taq DNA Polymerase High Fidelity (Invitrogen), 1 unit

10X AccuPrime buffer (containing dNTPs) and 0.3 µM IS5 and IS6 primers (Meyer & Kircher 2010). The thermal profile used was the following: 2-min initial denaturation at 94°C, followed by 5 to 17 cycles consisting of 30-s denaturation at 94°C, a 30-s annealing at 60°C and a 2-min elongation at 68°C and a 5-min final elongation at 68°C. The amplified products were purified with MinElute spin columns (Qiagen) according to the manufacturer's protocol. The amplified indexed libraries were quantified using an Agilent Bioanalyzer DNA 1000 Chip and pooled in equimolar amounts for hybridisation capture.

### S.5.1.6 Hybridisation capture and sequencing at the University of Tübingen

For the genome wide enrichment of the equimolar pooled libraries two rounds of capture hybridisation were performed using 1 million Agilent SureSelect arrays with designed probes (60 bp length and 4 bp tiling density) spanning the *T. p. pallidum* genome (Arora *et al.* 2016) following the protocol previously described by Hodges et al. (2009). After the first round of hybridisation, captured products were eluted in 490 µl $H_2O$, quantified via qPCR as previously described and amplified in 100 µl reactions using 24 µl template and the reagents and thermo profile detailed previously. The amplified library pools were purified using MinElute columns (Qiagen), the concentrations were determined using an Agilent 2100 Bioanalyzer DNA 1000 chip and used in the second round of capture. After the second hybridisation the captured products were eluted in 490 µl $H_2O$ and processed as detailed previously using 48 µl template for amplification. After the final quantification the pools were diluted to 10 nM for high-throughput sequencing. Paired-end dual index sequencing was performed on an Illumina Hiseq 2500 platform using 2*100+7+7 cycles and the manufacturer's protocols for multiplex sequencing (TruSeq PE Cluster Kit v3-cBot-HS).

### S.5.1.7 Precautions taken to avoid contamination

Capture analysis was performed in two separate laboratories, using distinct methods in both of the spaces. Both approaches converged on a similar finding, covering distinct regions of the genome. Both these laboratories are set up to avoid contamination and a take number of precautions including a dedicated DNA extraction room where only bones were extracted inside of sealed glove box, with UV sterilization and bleaching of surfaces following drilling and extraction. All PCR reactions were prepared in in glove boxes with pipets and laboratory equipment only used for this purpose, in rooms physically separated from the room where DNA

extracts were added to the mix, again in glove boxes with equipment only used for this step of the procedure. PCRs were conducted in another laboratory with separate lab coats and equipment and on a given day, reseachers never went from the DNA and PCR product rooms to the extraction or PCR reaction set up rooms. All PCRs included a negative water control that remained negative. We confirmed PCRs with Amperase UNG to rule out contamination with any PCR products and all PCRs in this laboratory are conducted with uracil in place of thymine, suggesting PCR products would be digested by this enzyme. For all steps we used filtered pipet tips, clean gloves, and regular bleaching of lab spaces. Extraction of bones and sending of an aliquot of extract to Tübingen was done before any libraries were constructed for *Treponema pallidum* in this laboratory space.

### S.5.1.8 Details on generation of phylogeny of full genomes in Figure 5.1

Analyses suggest that TPR genes have been under strong positive selection and may be recombining, suggesting these genes may be inappropriate for inferring phylogenetic history (Gray *et al.* 2006). We removed these genes from these genomes based on their annotations and then genomes were aligned with Mauve v2.3.1 (Darling *et al.* 2004), using an implementation in Geneious 8.1.6 with standard settings, retaining only regions present in all 12 genomes. Genome alignments were then stripped of identical sites in Geneious, also removing any site where any of the genomes had a gap, leaving 10,441 variable sites. The phylogeny was estimated from non-identical sites using PhyML (Build 20140926 (Guindon *et al.* 2010)) using a GTR+G4 model of nucleotide evolution. Equilibrium frequencies, topology and branch lengths were all optimized and the gamma distribution parameter was estimated from the data. The tree search used a combination of nearest neighbour interchange and subtree pruning and regrafting (BEST approach). Support values were calculated using Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-like aLRT).

**Table S5.1. Diagnostic primers and probe used to preliminary assess *Treponema pallidum* infection**

| Genetic region | Forward primer (5'-3')<br>Reverse primer (5'-3') | Size of amplicon | Reference |
|---|---|---|---|
| *polA* | AGGATCCGGCATATGTCCAA<br>GTGAGCGTCTCATCATTCCAAA | 67bp | (Leslie et al. 2007) |
| *polA* fusion primer for sequencing | GTAAAACGACGGCCAGAGGATCCGGCA TATGTCCAA<br>CAGGAAACAGCTATGACGTGAGCGTCT CATCATTCCAAA | 100bp | Current study |
| *cfpA* Nested | GAGCGTCTGGACGTAATGG<br>TAGGATGGCAATCTCCTTCG | 189bp | (Harper *et al.* 2008b) |
| *gpd* | AAGAACTTTCCCTCCTCCGTGC<br>CGTTTGATACGCTTCAGCTCG | 331bp | (Harper *et al.* 2008b) |
| *gpd* Nested | GTGGGTTGGAACAGACAACC<br>CGTTTGCACATACACTAGATCC | 161bp | (Harper *et al.* 2008b) |

**Table S5.2: Summary of bones extracted and tested using three diagnostic PCR systems.**

| **Country** | **Site** | **Species** | **Death date** | ***PolA* gel band** | ***PolA* seq** | ***GPD* band** | ***GPD* just second round with UNG** | ***GDP* sequence** | ***CFA* Band** | ***CFA* sequence** |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Cercocebus atys* | 2000 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Cercocebus atys* | 2004 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Cercocebus atys* | 2006 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Cercocebus atys* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Cercopithecus campbelli* | 2000 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Cercopithecus diana* | 2000 | + | + | - | Multiple bands: no sequence | Multiple bands: no sequence | - | NA |
| CI | Tai National Park | *Cercopithecus diana* | 2002 | + | + | - | + | - | - | - |
| CI | Tai National Park | *Cercopithecus diana* | 2007 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Colobus polykomos* | 1994 | + | + | - | - | *NA* | - | *NA* |
| CI | Tai National Park | *Colobus polykomos* | 2003 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1992/ 93 | + | + | - | - | *NA* | - | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1991 | + | + | - | - | *NA* | - | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | + | + | - | Multiple bands: no sequence | Multiple bands: no sequence | - | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1992/ 93 | + | + | - | - | *NA* | - | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1993 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1992 | + | + | - | + | - | + | + |
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1992 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | + | - | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1989 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1992/ 93 | + | + | - | - | *NA* | - | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1996 | + | + | - | - | *NA* | - | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Pan troglodytes verus* | 1996 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1999 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 2000 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | ~2000 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1999 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | unknown | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 1999 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Pan troglodytes verus* | 2004 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 2004 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 2002 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 2005 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | ~2006 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 2001 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Pan troglodytes verus* | 2002 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Pan troglodytes verus* | 1998 | + | + | + | + | + | - | NA |
| CI | Tai National Park | *Procolobus badius* | 2010 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 2001 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 2002 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 2002 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 2002 | + | + | - | + | + | - | NA |
| CI | Tai National Park | *Procolobus badius* | 2003 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CI | Tai National Park | *Procolobus badius* | 2004 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 1994 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |
| CI | Tai National Park | *Procolobus badius* | 1997 | - | *NA* | *NA* | *NA* | *NA* | *NA* | *NA* |

CI = Côte d'Ivoire; *NA* = Not applicable; + indicates sequence matching *Treponema pallidum*; - indicates no sequence or sequence not matching *Treponema pallidum*

**Table S5.3: Names and accession numbers of *Treponema* genomes used blastn comparison to filter sequences.**

| *Treponema pallidum* genomes for Blast comparision | | Selected non- *Treponema pallidum* genomes for Blast comparision | |
|---|---|---|---|
| Organism | Accession # | *Organism* | Accession number |
| *Treponema pallidum subsp. pertenue* str. SamoaD | NC_016842 | *Treponema azotonutricium* ZAS-9 | NC_015577 |
| *Treponema pallidum subsp. pertenue* str. Gauthier | NC_016843 | *Treponema brennaborense* DSM 12168 | NC_015500 |
| *Treponema pallidum subsp. pallidum* DAL-1 | NC_016844 | *Treponema bryantii* NK4A124 | NZ_ATWV00000000 |
| *Treponema pallidum subsp. pertenue* str. CDC2 | NC_016848 | *Treponema caldarium* DSM 7334 | NC_015732 |
| *Treponema pallidum subsp. pallidum* str. Chicago | NC_017268 | *Treponema denticola* AL-2 | NZ_AGDQ00000000 |
| *Treponema pallidum subsp. pallidum* str. Mexico A | NC_018722 | *Treponema denticola* AL-2 | NZ_CM001798 |
| *Treponema pallidum subsp pertenue* str. Fribourg-Blanc | NC_021179 | *Treponema denticola* ASLM | NZ_AGDR00000000 |
| *Treponema pallidum subsp. pallidum* str. | NC_021490 | *Treponema denticola* ATCC 33520 | NZ_AGDS00000000 |

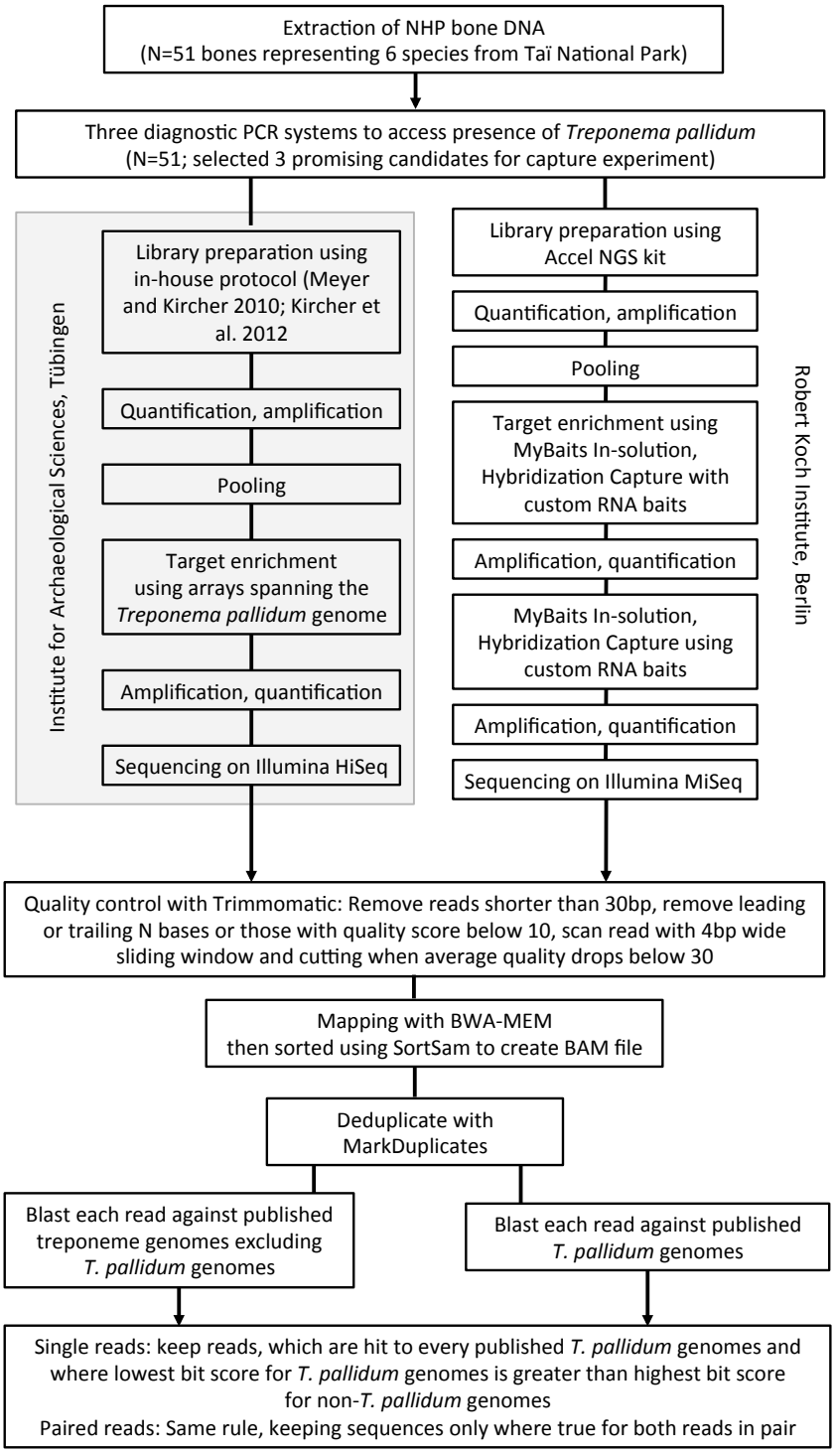
| | | | |
|---|---|---|---|
| Nichols | | | |
| *Treponema pallidum subsp. pallidum* SS14 | NC_021508 | *Treponema denticola* ATCC 33521 | NZ_AGDT00000000 |
| *Treponema pallidum subsp. pallidum* str. Sea 81-4 | NZ_CP003679 | *Treponema denticola* ATCC 35404 | NZ_AGDU00000000 |
| *Treponema pallidum subsp. endemicum* str. BosniaA | NZ_CP007548 | *Treponema denticola* ATCC 35404 | NZ_CM001796 |
| | | *Treponema denticola* ATCC 35405 | NC_002967 |
| | | *Treponema denticola* F0402 | NZ_ADEC00000000 |
| | | *Treponema denticola* H-22 | NZ_AGDV00000000 |
| | | *Treponema denticola* H-22 | NZ_CM001795 |
| | | *Treponema denticola* H1-T | NZ_AGDW00000000 |
| | | *Treponema denticola* H1-T | NZ_CM001794 |
| | | *Treponema denticola* MYR-T | NZ_AGDX00000000 |
| | | *Treponema denticola* OTK | NZ_AGDY00000000 |
| | | *Treponema denticola* OTK | NZ_CM001797 |

| | |
|---|---|
| *Treponema denticola* SP23 | NZ_AHAB00000000 |
| *Treponema denticola* SP32 | NZ_AHAC00000000 |
| *Treponema denticola* SP33 | NZ_AGDZ00000000 |
| *Treponema denticola* SP37 | NZ_AGEA00000000 |
| *Treponema denticola* SP44 | NZ_AHAD00000000 |
| *Treponema denticola* US-Trep | NZ_AGEB00000000 |
| *Treponema lecithinolyticum* ATCC 700332 | NZ_AWVH00000000 |
| *Treponema maltophilum* ATCC 51939 | NZ_ATFF00000000 |
| *Treponema medium* ATCC 700293 | NZ_ATFE00000000 |
| *Treponema paraluiscuniculi* Cuniculi A | NC_015714 |
| *Treponema pedis* isoE1186 | NZ_AOTP00000000 |
| *Treponema pedis* isoM1220 | NZ_AOTQ00000000 |

| | |
|---|---|
| *Treponema pedis* isoM1224 | NZ_AOTR00000000 |
| *Treponema pedis* str. B 683 | NZ_AOTN00000000 |
| *Treponema pedis* str. T A4 | NC_022097 |
| *Treponema pedis* str. T M1 | NZ_AOTM00000000 |
| *Treponema phagedenis* | NZ_CDNC00000000 |
| *Treponema phagedenis* 4A | NZ_AQCF00000000 |
| *Treponema phagedenis* F0421 | NZ_AEFH00000000 |
| *Treponema primitia* ZAS-1 | NZ_AEEA00000000 |
| *Treponema primitia* ZAS-2 | NC_015578 |
| *Treponema putidum* | NZ_CP009228 |
| *Treponema saccharophilum* DSM 2985 | NZ_AGRW00000000 |
| *Treponema socranskii subsp. paredis* ATCC 35535 | NZ_ATFD00000000 |
| *Treponema socranskii subsp. socranskii* VPI | NZ_AUZJ00000000 |

| | |
|---|---|
| DR56BR1116 = ATCC 35536 | |
| *Treponema socranskii subsp. socranskii* VPI DR56BR1116 = ATCC 35536 | NZ_AVQI00000000 |
| *Treponema sp. C6A8* | NZ_JHVB00000000 |
| *Treponema sp. JC4* | NZ_AJGU00000000 |
| *Treponema sp. OMZ 838* | NZ_CP009227 |
| *Treponema succinifaciens* DSM 2489 | NC_015385 |
| *Treponema vincentii* ATCC 35580 | NZ_ACYH00000000 |
| *Treponema vincentii* F0403 | NZ_ATFC00000000 |
| *Treponema vincentii* F0403 | NZ_ATFC01000000 |

**Figure S5.1: Visualization of workflow for hybridisation capture experiment.**



# SUPPLEMENTARY MATRIALS FOR CHAPTER 6

## NON-HUMAN PRIMATES AS RESERVOIRS FOR YAWS

## S.6.1 SUPPLEMENTARY MATERIALS

### S6.1.1 Ethical statement

All procedures performed on sooty mangabeys in Taï National Park (TNP) were approved by the Ministry of Environment and Forests as well as the Ministry of Research, the Office Ivoirien des Parcs et Réserves, and the director of TNP. Baboon samples from Lake Manyara National Park (LMNP) in Tanzania were taken in accordance with the Tanzania Wildlife Research Institute's Guidelines for Conducting Wildlife Research (2001; 2012) and with permission of Tanzania National Parks (TNP/HQ/E.20/08B) as well as the Commission for Science and Technology in Tanzania (2007-56-NA-2006-176). The Joint Management Research Committee (JMRC) of the TAWIRI Board and TANAPA approved sample collection. The Animal Welfare and Ethics Committee of the German Primate Center approved the use of samples for this study. Procedures on green monkeys in Bijilo Forest Park (BFP) were conducted with ethical approval from the University of Cumbria and permission from the Forestry Department and Department of Parks and Wildlife Management in The Gambia. Green monkeys were captured in the wild in Niokolo Koba National Park (NKNP), Senegal, under license No. 1302/DPN/MEDD (10.16.2015) granted by the Republic of Senegal. Good veterinary practice and animal welfare were considered in all procedures carried out at all field sites. Anesthetized animals were monitored for vital functions and remained under close supervision from the time of induction until full recovery, and until the animals were able to reunite with their social group.

## S6.2 MATERIALS AND METHODS

### S6.2.1 Laboratory contributions

The study brought together NHP samples from four field sites and involved analysis conducted in multiple different laboratories. As a consequence of this, several analysis steps (e.g., DNA extraction, PCR testing, DNA capture, sequencing and genome assembly) were performed with different methods. Briefly, samples from West Africa were collected by the University of Cumbria (Carlisle, UK), the Research Unit of Emerging Infectious and Tropical Diseases (Marseilles, France and Dakar, Senegal: URMITE) and the group Epidemiology of Highly Pathogenic Microorganisms (Robert Koch Institute, Berlin, Germany: RKI) and processed by URMITE and RKI. Samples from East Africa were collected and processed by the Work Group

Neglected Tropical Diseases (German Primate Center, Göttingen, Germany: DPZ). While DNA samples from West Africa were analyzed by in-solution capture at RKI and by microarray capture at the Department of Archaeological Sciences (Eberhard-Karls-University Tuebingen, Germany: EKU), samples from East Africa were only microarray captured at EKU. In addition, East African samples were used to run long-range PCR followed by next-generation sequencing at the Department of Biology (Masaryk University, Brno, Czech Republic: MU). Obtained sequence data from all samples were collected and shared between collaborators and jointly analyzed at EKU.

### S6.2.2 Study sites and sample size

The five sooty mangabeys sampled for this project come from a study group that was habituated to human observers in 2012 as part of the Taï Chimpanzee Project, in TNP, Côte d'Ivoire, representing one of the many interacting NHP species under long-term disease surveillance in the Park (Gogarten *et al.* 2014a). All individuals in the study group were individually identifiable as of June 2013. The five samples from African green monkeys from BFP, The Gambia, come from a group that has been habituated to human presence as a result of several years of feeding by tourists and guides. No group composition and epidemiological data is currently available, as most of the monkeys have not yet been individually identified. The three samples from African green monkeys from NKNP, Senegal, come from monkeys captured with a wire mesh trap installed in the park (13°04'N-12°43'W) in proximity to a National Park guardhouse and along the Niokolo Koba River, about 55m south of the national road N°7. The two baboon samples from Tanzania were extracted from a study that aimed to investigate the pathogen that causes genital ulceration in olive baboons at LMNP (Knauf *et al.* 2012). Samples originate from a female (4F5230307) and a male (40M5160407) baboon from a non-habituated group in the northern part of the national park.

### S6.2.3 Clinical manifestation

In TNP, we observed that many juveniles in a group of habituated sooty mangabeys (*Cercocebus atys*) presented with ulcerative lesions on the face or lower extremities. A single adult exhibited extensive facial tissue destruction, damage to bone and cartilage, and a poor physical condition. In BFP, we observed facial and anogenital lesions in a number of juveniles and adults belonging to a group of habituated African green monkeys (*Chlorocebus sabaeus*). The most severe

manifestations were also observed in an adult with extensive destruction and scarring of facial and anogenital tissues. In NKNP, genital lesions were only observed during the clinical examination of adult males belonging to an unhabituated group of African green monkeys. In LMNP, we had previously described frequent and severe anogenital lesions in juvenile and adult baboons (*Papio anubis*) (Knauf *et al.* 2012).

### S6.2.4 Sampling procedures in NHPs

Seven clinically affected and four unaffected individuals from the TNP and BFP groups were anesthetized using a combination of ketamine (10mg/kg)/xylazine (1mg/kg) (TCP) or ketamine (5mg/kg)/medetomidine (50µg/kg) (BFP), which was administered intramuscularly via blowpipe (Telinject GmbH; TCP group) or hand-injection after trapping (BFP group). An intramuscular injection of atipamezole (1mg/10 mg xylazine or 5mg/1mg medetomidine) was administered after sampling for reversal of anesthesia. Skin biopsies and blood samples were collected from all individuals. Skin biopsies were taken from the edge of a lesion in clinically affected individuals and from normal unaffected tissue in both clinically affected and healthy individuals. Skin biopsies were preserved in a 10 % formalin solution, in a formaldehyde/glutaraldehyde mixture and/or frozen (here and subsequently this refers to freezing in liquid nitrogen in the field and at -80°C in the laboratory). Where refrigeration was not immediately available (BFP group), samples were preserved in RNAlater (Life technologies, NY) and kept at room temperature for three weeks, then transferred to -80°C. Blood was collected in EDTA tubes from the femoral vein. Whole blood was preserved frozen (TCP group) or in RNAlater (BFP group). For the TCP group, swabs from lesions preserved in STGG transport medium were also collected and refrigerated until analysis. The two baboons at LMNP were chemically short-term immobilized by remote-distance intramuscular injection of ketamine (10mg/kg)/xylazine (0.2 mg/kg). Skin samples were taken from the margin of ulcerated genital tissue, using a sterile 6 mm biopsy punch (Heiland VET Vertriebsgesellschaft mbH & Co. KG, Hamburg, Germany). Skin samples were immediately transferred into RNAlater and were frozen 12 hours later at -20°C until export to Germany. Swabs of penile lesions from African green monkeys in Senegal were transported without culture medium in liquid nitrogen.

### S6.2.5 DNA extraction

DNA extraction from LMNP baboon samples was performed twice (Knauf *et al.* 2012). Tissue was cut into small pieces and ground with Precillys-Keramik beads (peQlab Biotechnologie GmbH, Erlangen, Germany), followed by 4-6 hours of digestion with proteinase Kat 56°C. The first extraction (in 2007) was performed using the NucleoSpin Tissue extraction kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany), the second extraction (in 2014) using the First-DNA-All-Tissue kit (Gen-ial GmbH, Troisdorf, Germany). In both cases, DNA was eluted in molecular grade water. Aliquots were kept frozen at -80°C. At the RKI, DNA was extracted from tissue, blood and swabs using DNA/RNA purification kits (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany), QIAamp DNA Blood Mini Kits (QIAGEN GmbH, Hilden, Germany) and QIAamp Viral RNA Mini Kits (QIAGEN GmbH, Hilden, Germany) respectively.

### S6.2.6 PCR testing

Tissue and blood samples (TCP, BFP and NKNP groups) were screened for *TP* infections using PCR and subsequent sequencing (**table S6.1**). Extracts were tested using a standard PCR amplifying a 67 bp DNA fragment (includes primers) of the DNA polymerase I gene (polA) as previously described (Knauf *et al.* 2012). Amplifications were performed in 25 µl reactions containing 0.2 µM of each primer, 200 µM dNTPs, 4 mM $MgCl_2$, 2,5 µl 10X PCR buffer and 0.25 µl Platinum® Taq polymerase (Invitrogen, Carlsbad, USA). Assays were run under the following conditions: 5 min at 95°C, 40 cycles [15 s at 95°C, 45 s at 60°C, 60 s at 72°C], 7 min at 72°C.

### S6.2.7 qPCR

*TP* copy numbers in the LMNP baboon samples were measured using a TaqMan PCR run on a 7500 Real Time PCR System (Applied Biosystems Inc., Foster City, CA, USA) targeting the same 67-bp fragment of the DNA polymerase I gene (polA) of *T. pallidum* (2007). A plasmid dilution series containing the amplicon was used as a standard. Cloning was performed using the TOPO TA cloning kit (Invitrogen, Karlsruhe, Germany) and plasmid preparation was achieved using the NucleoSpin Plasmid kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). Samples were tested as duplicates in 2007 and triplets in 2014. For the individual 4F5230307, copy numbers were 4,696±1,824/100 ng genomic DNA (gDNA; mean±SD; DNA extraction in 2007) and 2±1/100 ng gDNA in DNA extracted in 2014. Likewise, DNA extracted in 2007 from

animal 40M5160407 revealed a copy number of 7,250/100 ng gDNA and 102±2/100 ng gDNA copies (mean±SD) for the DNA extracted in 2014. The low copy number found in DNA extracted in 2014 could be an effect of the different extraction methods used or of the unequal distribution of spirochetes in skin samples. In addition, the first DNA extraction was performed immediately after samples have been imported to Germany in 2007.

### S6.2.8 DNA amplification, capture, and sequencing

Three different methods were used to amplify and sequence the whole genome of the simian *TP* strains; methods used depended on availability of DNA extracts in specific laboratories. West African *TP* isolates were whole genome sequenced using in-solution and microarray based capture techniques, whereas the LMNP strain was sequenced by microarray capture and long-range PCR. Long-range PCR was able to generate high genome coverage in multiple repeat regions and paralogous regions where the hybridization capture enrichment approaches were less effective.

### S6.2.9 Amplification of target DNA (long-range PCR)

The purified genomic DNA of East African baboon 40M5160407 was amplified using the multiple displacement amplification approach (REPLI-g kit, QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. Resulting DNA was 100x diluted and used for the pooled segment genome sequencing (PSGS) as described previously (Cejkova *et al.* 2012) (Zobaníková *et al.* 2013). Briefly, DNA was amplified with 278 pairs of specific primers to obtain overlapping PCR products (**table S6.4)** covering the entire genome of the LMNP isolate (40M5160407). PCR products were amplified with PrimeSTAR GXL DNA Polymerase (Takara Bio Inc., Otsu, Japan) using touchdown PCR. The cycling conditions were: denaturation at 94°C for 1 min; 8 cycles: 98°C for 10 s, 68-60°C for 15 s (annealing temperature gradually reduced by 1°C/every cycle), and 68°C for 6 min; 35 cycles: 98°C for 10 s, 61°C for 15 s, and 68°C for 6 min (43 cycles in total); followed by the final extension at 68°C for 7 min. The PCR products were purified using a QIAquick PCR Purification Kit (QIAGEN, Valencia, CA, USA) according to the manufacturer's instructions. To facilitate sequencing of RNA operons, *tpr* genes and other paralogous regions, PCR products were split equimolarly into four distinct pools. Prior to next-generation sequencing on a MiSeq platform (Illumina), these pools were labeled with multiplex identifier (MID) adapters and sequenced as four different samples.

### S6.2.10 Microarray capture

DNA extracts were sheared with a Covaris S220 Focused-ultrasonicator® to produce fragments between 300 and 500 bp using the following conditions: 5 intensity, 200 cycles per burst, 45 seconds. Then the samples were concentrated to 30 µl each with Amicon® Ultra Centrifugal Filters (Merck KGaA, Darmstadt, Germany) using the manufacturer's instructions and were converted into double-stranded Illumina libraries as described by Meyer and Kircher (2010). Sample specific barcodes were added to both library adapters to obtain double indexed libraries (Kircher *et al.* 2012). Library blanks were treated accordingly. The efficiency of the reaction was tested using a quantification assay with the primer set IS5 and IS6 (Meyer & Kircher 2010), the DyNAmo Flash SYBR Green qPCR Kit (Biozym Scientific GmbH, Hessisch Oldendorf, Germany) and the Lightcycler® 96 (Roche Life Science). For all indexed libraries a second amplification was performed in 100 µl reactions containing 5 µl library template, 4 units AccuPrime Taq DNA Polymerase High Fidelity (Invitrogen), 1 unit 10X AccuPrime buffer (containing dNTPs) and 0.3 µM IS5 and IS6 primers (Meyer & Kircher 2010), with the following thermal profile: 2-min initial denaturation at 94°C, followed by 5 to 18 cycles consisting of 30-sec denaturation at 94°C, a 30-sec annealing at 60°C and a 2-min elongation at 68°C and a 5-min final elongation at 68°C. The purification of the amplified products was performed using MinElute spin columns (QIAGEN GmbH, Hilden, Germany) following the manufacturer's protocol. For hybridization capture the amplified indexed libraries were quantified with an Agilent Bioanalyzer DNA 1000 Chip and pooled in equimolar amounts. Treponemal DNA was enriched from equimolarly pooled libraries via two rounds of hybridization capture using 1 million Agilent SureSelect arrays with designed probes (60 bp length and 4 bp tiling density) that span the *T. pallidum* genome (Arora *et al.* 2016) following the protocol described by Hodges et al., 2009 (2009). The first round of hybridization was followed by an elution of the capture products in 490 µl H2O, a quantification via qPCR (see previous section) and amplifications in 100 µl reactions using 24 µl template (reagents and thermo profile as described previously). The amplified products were purified using MinElute columns (Qiagen) and the concentrations were determined using an Agilent 2100 Bioanalyzer DNA 1000 chip. These amplified products were then used for the second round of capture, after which the captured products were again eluted in 490 µl H2O and processed as described above with 48 µl template for amplification. Afterwards the captured products were quantified and diluted to 10

nM for high-throughput sequencing. Paired-end dual index sequencing was conducted on an Illumina Hiseq 2500 platform using 2*100+7+7 cycles and the manufacturer's protocols for multiplex sequencing (TruSeq PE Cluster Kit v3-cBot-HS).

**S6.2.11 In-solution capture**

Selected PCR positive extracts were fragmented using a Covaris S220 Focused-ultrasonicator® in a total volume of 130 µL (filled with low EDTA TE buffer), using settings aiming to generate ~400 bp fragments (Intensity = 4, Duty cycle = 10%, Cycles per burst = 200, Treatment time = 55 seconds, Temperature = 7°C). Fragmented extracts were then concentrated using a MinElute PCR purification kit and each fragmented extract was eluted into 2x10 µL low EDTA TE buffer. DNA concentration was measured using a Qubit dsDNA High Sensitivity kit. 1µg DNA or all available remaining DNA extract were used for subsequent library preparation using the Accel-NGS 2S DNA library kit following the standard protocol and a sample specific unique index. Quantification was conducted using a KAPA HiFi library quantification kit and libraries were then amplified using a KAPA Hot Start Library Amplification Kit and Illumina adapter specific primers (5'-3': AATGATACGGCGACCACCGA and 5'-3': CAAGCAGAAGACGGCATACGA, 45 s at 98°C, variable number of cycles [15 s at 98°C, 30 s at 65°C, 45 s at 72°C], 1 min at 72°. Following amplification libraries were requantified to ensure the desired amount of starting material for capture (total number of cycles before 1$^{st}$ capture: HAT = 3 cycles, M3 and IGU = 4 cycles, A10 and A12 = 5 cycles, A9 and M2 = 6 cycles). HAT, IGU, M2 and M3 were pooled to equally contribute to a total of 240 ng DNA of starting material for input into the in-solution hybridization capture, following concentration with a MinElute PCR Purification Kit (pool A). A9, A10 and A12 were pooled to contribute equally to 500 ng DNA after concentration with a MinElute PCR Purification Kit (pool B). These two pools underwent separate hybridization capture and sequencing at RKI. We designed RNA baits to span the *T. p. pertenue* Fribourg-Blanc genome (accession number NC_021179) with 2-fold tiling and 120mer baits, and used these for hybridization capture enrichment of *TP* DNA as described previously (Gogarten *et al.* 2016). Briefly, we followed the Mybaits Sequence Enrichment for Targeted Sequencing protocol (Version 2.3.1) using the recommended hybridization time of 16 hours. Following an initial round of capture, pools of surviving libraries were reamplified using a KAPA Hot Start Library Amplification Kit (Pool A = 18 cycles, Pool B = 15 cycles) to generate 100-500 ng starting material for a second round of capture. Pools were again quantified using the KAPA HiFi

Library Quantification Kit to determine concentration of libraries. DNA was concentrated using a MinElute PCR Purification Kit prior to commencing with a second round of hybridization capture applying the same conditions as for the first round of capture described above. Surviving DNA was reamplified using the KAPA Hot Start Library Amplification Kit (Pool A=10 cycles, Pool B =13 cycles), purified using the MinElute PCR Purification Kit, and quantified using the KAPA HiFi Library Quantification Kit. The two pools were diluted to 4nM as input for two separate sequencing runs on an Illumina MiSeq (v3 2x300 Chemistry).

### S6.2.12 DNA sequencing and genome assembly

The sequencing results are summarized in **table S6.2**. We applied EAGER (Peltzer *et al.* 2016), a comprehensive pipeline for read pre-processing, mapping, variant identification, and genome reconstruction on all sequenced samples. Each of the steps performed using EAGER is described below.

***Read preprocessing of sequenced genome samples***: The sequenced products for all samples were paired-end reads with a varying number of overlapping nucleotides between corresponding forward and reverse reads. Several pre-processing steps were necessary such as adapter clipping, merging of corresponding paired-end reads in the overlapping regions and finally quality trimming of the resulting reads. To remove sequencing adapters from the paired-end reads an overlap alignment of the respective adapter with the 3' end of each forward and reverse read was produced. Regions at the 3' end of each read that were contained in the alignment were clipped. Reads that were shorter than 30 nucleotides after adapter clipping were removed. This procedure resulted in three kinds of remaining reads: forward reads that did not have a corresponding reverse read, reverse reads that did not have a corresponding forward read, and matching forward and reverse reads that could be used in the merging process. Merging was performed for all paired-end reads with a minimum overlap of 10 nucleotides and at most 5% mismatches in the overlap region with the Clip&Merge tool implemented in EAGER. On average about 60% of all paired-end reads were merged in each sample. All reads that could not be merged were first trimmed at the 3' end such that all bases have a phred quality score of at least 20 and then mapped individually.

***Mapping assembly***: After adapter clipping, merging and quality trimming, the resulting reads for all samples were mapped using the F-B genome as a reference. All reads (merged and unmerged) were treated as single-end reads and mapping was performed using the BWA-MEM algorithm

(Li 2013) with default parameters. After mapping the Genome Analysis Toolkit (GATK) was used to generate a mapping assembly for each strain that had at least 80% coverage of the F-B genome with a minimum of 3 reads. For this procedure, the UnifiedGenotyper module of GATK was applied to call reference bases and variants from the mapping. The reference base was called if the genotype quality of the call was at least 30 and the position was covered by at least 3 reads. A variant position (SNP) was called if the following criteria were met: i) the position was covered by at least 3 reads; ii) the genotype quality of the call was at least 30 and iii) the minimum SNP allele frequency was 90%. If the requirements for a variant call were not fulfilled, the reference base was called instead but only if at least 3 reads confirmed the reference base and the quality threshold was reached. If neither of the requirements for a reference base call nor the requirements for a variant call were met, the character 'N' was inserted at the respective position. For the generation of draft genome sequences we used the tool VCF2Genome of the EAGER pipeline, which reads a VCF file and outputs a genome in fasta format.

***Long range PCR-based genome assembly:*** The Illumina sequencing reads obtained from 4 distinct pools (sequenced as 4 different samples – see **table S6.2**) were separately assembled *de novo* using SeqMan NGen v4.1.0 software (DNASTAR, Madison, WI, USA). A total of 99, 81, 62, and 138 contigs (obtained for each pool 1, 2, 3, and 4, respectively) were aligned to the corresponding sequences (representing each pool sequence) of the reference F-B genome (Zobaníková *et al.* 2013) (GenBank CP003902.1) using Lasergene software (DNASTAR, Madison, WI, USA). In addition, the Illumina sequencing reads were also mapped to the F-B genome and processed as mentioned above. All gaps in the genome sequence and all discrepancies between contig sequences and reference-guided consensus were resolved using Sanger sequencing. Altogether, 20 genomic regions of the baboon isolate genome from East Africa (40M5160407) were amplified and Sanger sequenced. The final overlapping pool sequences were joined to obtain complete genome sequence of the baboon isolate. The sequences of genes containing tandem repeats, i.e. *arp* (*tp0433*) and *tp0470* genes, were also resolved using Sanger sequencing. The number of tandem repeats in these genes was estimated based on the gel electrophoresis. Gene *tprK* (*tp0897*) showed intra-strain variability and therefore nucleotides in variable regions were replaced with 'N's in the complete genome sequence. In addition, the G/C-homopolymeric stretches revealed intra-strain variability throughout the genome. The prevailing number of G/Cs in these regions was used in the final genome sequence.

### S6.2.13 SNP effect analysis

The genetic effect of each of the SNPs occurring in at least one strain was analyzed using SnpEff (Cingolani *et al.* 2012). We used an annotation database of *T. p. pallidum* (Fribourg-Blanc with RefSeq acc. ID NC_021179.1) built from the genomic annotation (the respective gff file was also retrieved from NCBI). These annotations include protein-coding genes as well as non-coding RNAs and pseudogenes. The up-/downstream region size parameter for reporting SNPs that are located upstream or downstream of protein-coding genes was set to 100 nt. For all other parameters default values were used. The results were used to compile a table providing information on the genetic effect for each occurring SNP. This table is available as a separate supplementary file. High-throughput sequencing produced between 0.36 and 43.3 million reads per sample. After duplicate removal between 1,310 and 470,303 reads mapped to the Fribourg-Blanc reference genome resulting in mean coverage estimates of 0.12 to more than 120-fold average. For the samples of 4F5230307 and 40M5160407 we pooled the respective mapped reads into a common file for each sample before the subsequent genotyping steps. After this pooling step, all 8 samples had a coverage of at least 80% of the Fribourg-Blanc genome with a minimum of 3 reads. After SNP calling using GATK between 101 and 243 SNPs in comparison to F-B could be identified. The sequencing and genotyping results are summarized in **table S6.2**.

### S6.2.14 Gene identification, annotation and classification

Both protein-coding genes and genes for noncoding RNA were annotated in the genome sequence of East African baboon isolate 40M5160407 based on the annotation of previously published *TPE* strain Gauthier (GenBank CP002376.1). Lasergene software (DNASTAR, Madison, WI, USA) was used for Gauthier orthologous gene alignment and recalculation of gene coordinates to East African baboon isolate (40M5160407). A gene size limit of 150 bp was applied. Genes were tagged with the TPE40M5-prefix and the locus tag numbering corresponds to the tag numbering of orthologous genes annotated in the TPE strain Gauthier genome.

### S6.2.15 Processing of published genomes

In order to also apply the EAGER analysis pipeline to the complete genomic sequences already available in GenBank, we generated artificial reads using the tool Genome2Reads (also part of the EAGER software). Genome2Reads uses a tiling approach with an offset of 1, to artificially generate reads of length 100 nucleotides, resulting in an average coverage of 100X. For the

resulting samples we applied the same mapping, SNP calling and genome reconstruction procedure as for the sequenced samples in order to obtain consistent and comparable results for phylogeny reconstruction.

### S6.2.16 Phylogenetic analysis

MEGA6 was used to generate Maximum Parsimony trees. Bootstrap values were inferred from 100 replicates. The analysis involved 19 nucleotide sequences, 8 monkey (see **table S6.2**) and 11 human strains (see **table S6.3**). All positions with less than 85% site coverage were eliminated; that is, fewer than 15% alignment gaps, missing data, and ambiguous bases were allowed at any position. We computed one phylogenetic tree using all 2,317 informative positions (**fig. S6.1 A**) and one phylogenetic tree removing all positions from putative recombinant genes (Arora *et al.* 2016), resulting in 2,012 informative positions (**fig. S6.1 B**). Both trees are nearly identical, with the only differences are the respective branch lengths, which can be attributed to the different number of informative positions.

## S6.3 Supplementary Text: Comparison of whole genome sequences

### S6.3.1 General overview

The overall genome structure of East African baboon isolate 40M5160407 (obtained from *Papio anubis* from Lake Manyara National Park) is highly similar to other genomes of human *T. pallidum* subsp. *pertenue* strains (Gauthier, CDC-2, and Samoa D) and the simian F-B isolate. As in other genomes, different numbers in the 60 bp-long repetitions in the *arp* gene and different number of 24 nt-long repetitions in the *tp0470* gene were found. The number of repetitions in the *tp0470* is identical to the published *TPE* strain Gauthier and the number of repetitions in the *arp* gene (9) is close to the number of repetitions in the published *TPE* strain Gauthier (10). All 60 bp-repeats in *arp* gene of the baboon LMNP genome were of Type II and were identical to other *TPE* strains previously described (Harper *et al.* 2008a). In contrast to all characterized *TPE* and *T. pallidum* subsp. *endemicum* (*TEN*) strains (with the exception of *TPE* strain Gauthier), *tprC* and *tprD* genes are identical in the analyzed 40M5160407 genome. Similarly, *tprF* and *tprI* genes are identical in the analyzed genome of baboon isolate 40M5160407. However, this is a common feature of all known *TPE* strains. The *tprH* gene of the LMNP strain from East Africa

has a frameshift mutation leading to protein shortening. The *tprK* gene only has 3 variable regions, V5-V7, when compared to other *TPE* strains. Interestingly, we found the genome of baboon isolate 40M5160407 to be more related to the genome of human *TPE* Gauthier strain than to simian isolate F-B. The number of nucleotide differences of various lengths between the genome of the baboon isolate 40M5160407 and the published *TPE* Gauthier and *TPE* F-B genomes (CP002376.1 and CP003902.1, respectively) is shown in **table S6.6** indicating that the majority of differences are single nucleotide substitutions or indels. Altogether, differences were found in 266 and 325 chromosomal positions when the baboon isolate was compared to Gauthier and F-B genomes, respectively. In addition to differences in *tp0433*, *tp0470*, and *tprK* genes, larger differences were determined in TPEGAU_0136 (33 nt-long deletion; specific for str. Gauthier and Samoa D), in TPFB_0548 (42 nt-long deletion; specific for str. F-B), in TPEGAU_0858 (79 nt-long deletion; specific for str. Gauthier), in IGR between TPEGAU_0628 and TPEGAU_0629 (302 nt-long deletion; specific for str. Gauthier), and in IGR between TPFB_0696 and TPFB_0697 (430 nt-long insertion; specific for str. F-B); the length of other nucleotide changes ranged between 1-15 nts.

### S6.3.2 RNA operons

The structure of RNA operons (231,180-236,139; 279,584-284,533; according to *TPE* strain Gauthier: CP002376.1) is similar to strains Gauthier, CDC-2, and F-B, but different to what is found in the Samoa D, Samoa F, and CDC-1strains. The sequence of 16S-5S-23S is identical in both operons and the 23S RNA sequences were identical to other *TPE* strains (except for str. F-B). There are no mutations associated with macrolide resistance (A2058G, A2059G) (Stamm & Bergen 2000; Lukehart *et al.* 2004).

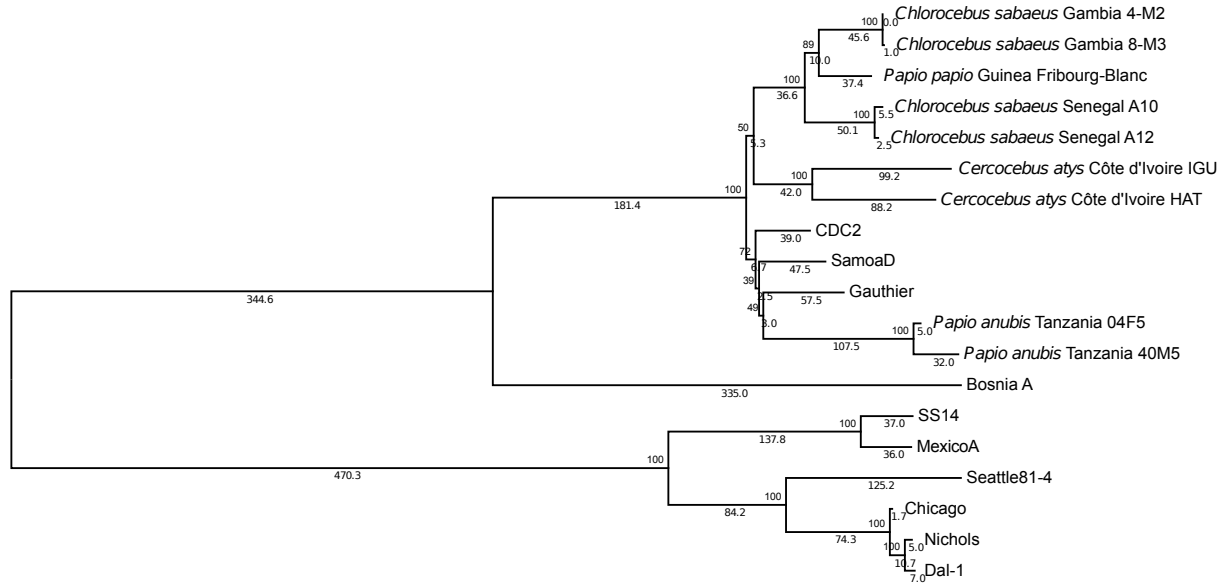
### S6.3.3 Intergenic regions

Out of a total of 266 nucleotide changes different from the sequence of Gauthier strain (1-302 in length), 28 (10.53 %) were located in intergenic regions. The length of intergenic regions in the *TPE* strains (Gauthier, CDC2, Samoa D, and F-B) is about 4.63-4.68% (Cejkova *et al.* 2012; Zobaníková *et al.* 2013). Indel changes detected in the 40M5160407 *T. pallidum* str.-intergenic regions were located upstream of genes coding for Tpr proteins (C, D, G, I, and J), chemotaxis proteins (Mcp and CheB), proteins involved in transport and metabolism (TmpC, SecA, EmrE, and TP0925), ribosomal protein S19 (RpsS), RNA polymerase (RpoB), and hypothetical proteins

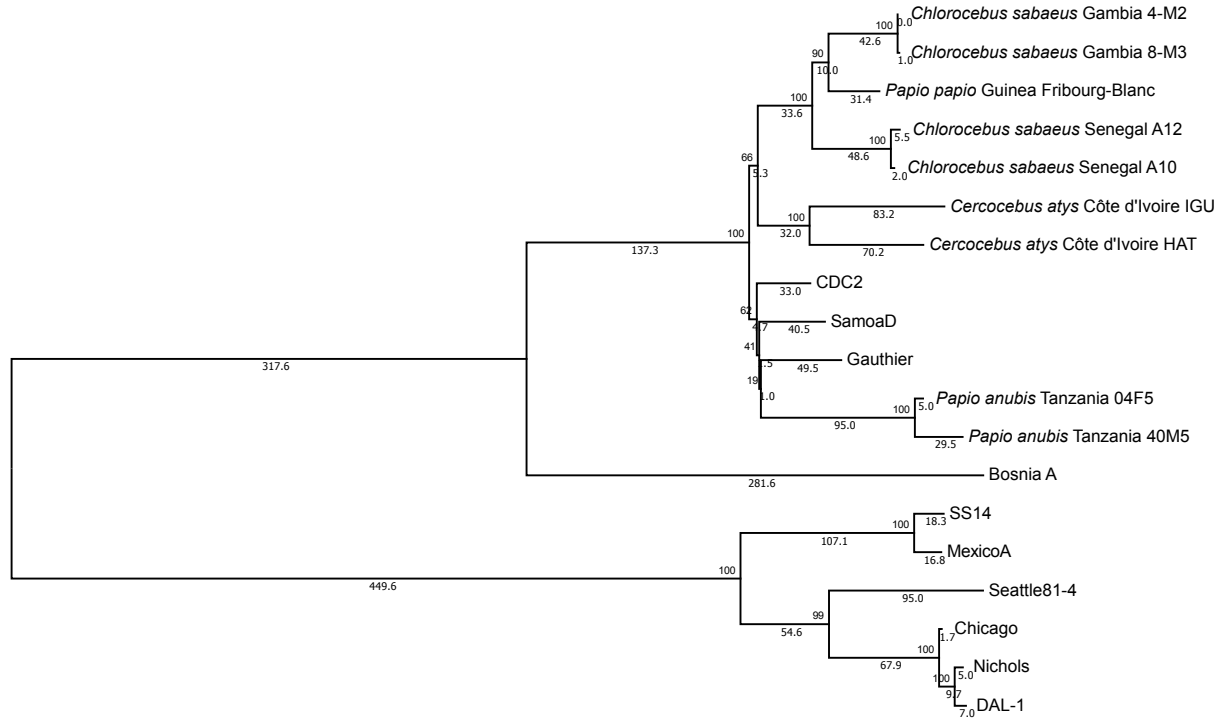
(TP0381, TP0383, TP0479, TP0480, and TP0629). These indel changes were predominantly identified in G/C-homopolymeric tracts and were previously shown to affect transcription rate of the downstream genes (Giacani *et al.* 2007; Giacani *et al.* 2015).

### S6.4 SUPPLEMENTARY ACKNOWLEDGMENTS:

This article represents a chapter in the doctoral dissertation of JFG, and has thus benefited greatly from the input of his supervisor committee and advisors Jonathan Davies, David Marcogliese, Charles Nunn, and Louis Lefebvre. Simone Lüert and Ulla Thiesen are thanked for discussions and technical support. We thank the Ivorian Ministry of Environment and Forests, the Ministry of Research, the directorship of the Taï National Park, the Office Ivoirien des Parcs et Réserves, the Centre Suisse de Recherche Scientifique, the Taï Chimpanzee Project and the Taï Monkey project and their teams of field assistants for their support. We thank Jonathan Müller-Tiburtius, Therese Löhrich, Sylvain Lemoine, Simon T. Kannieu, Daniel Gnimion, Richard Peho, and Martina Magris for field assistance at TNP. For assistance and support in BFP we thank the Department of Parks and Wildlife Management and the Forestry Department. For assistance and support with sequencing and read processing at the RKI we are grateful to Wojciech Dabrowski, Julia Hinzmann, Andreas Nitsche and Julia Tesch. We would like to thank the Senegalese authorities, in particular the officials of Niokolo-Koba, who helped our work in Senegal immensely. We also thank Annick Abeille, who performed the initial PCR screening and sequencing of the Senegalese green monkey samples. For the work on East African simian strains, we would like to thank TAWIRI, TANAPA, Lake Manyara National Park staff, especially the chief park warden, ecology monitoring unit and park rangers, as well as Tanzania Commission for Science and Technology (COSTECH).




**A**




**B**

**Figure S6.1: Phylogenetic trees of *Treponema pallidum* whole genome sequences.** The evolutionary history was inferred using the Maximum Parsimony method. The percentage of

replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the nodes (Felsenstein 1985). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei & Kumar 2000) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). Branch lengths were calculated using the average pathway method (Nei & Kumar 2000) and are in the units of the number of changes over the whole sequence. They are shown next to the branches. All positions with less than 85% site coverage were eliminated. That is, fewer than 15% alignment gaps, missing data, and ambiguous bases were allowed at any position. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.* 2013). **A**. **Maximum parsimony tree of alignment with all positions in SNP alignment**. There were a total of 2,317 positions in the final dataset. **B**. **Maximum parsimony tree of alignment without positions of recombinant genes** (TPFB_0136, TPFB_0326, TPFB_0488, TPFB_0865). There were a total of 2,046 positions in the final dataset.

**Table S6.1**: **Molecular analysis (PCR and sequencing) performed on blood samples, tissue samples and lesion swabs (RKI)**.

| Species | Group | ID | Sample type | *Treponema polA* |
|---|---|---|---|---|
| *Cercocebus atys* | TCP | Hat*,$ | face lesion biopsy | P |
| | | | normal skin biopsy | N |
| | | | blood | N |
| | | Igu*,$ | face lesion biopsy | P |
| | | | arm lesion biopsy | P |
| | | | normal skin biopsy | N |
| | | Kah | normal skin biopsy | N |
| | | | blood | N |
| | | Ran | normal skin biopsy | N |
| | | | blood | N |
| | | Pha* | face lesion biopsy | N |
| | | | blood | N |
| *Chlorocebus sabaeus* | BFP | M2*,$ | face lesion biopsy | P |
| | | | blood | N |
| | | M3*,$ | face lesion biopsy | P |
| | | | normal skin biopsy | N |
| | | | blood | N |
| | | M7* | face lesion biopsy | P |
| | | | genital lesion biopsy | N |
| | | | blood | N |
| | | M6 | normal skin biopsy | N |
| | | | blood | N |
| | | M4 | normal skin biopsy | N |
| | | | blood | N |

P: positive PCR result or sequences generated but too short to be uploaded into EMBL; N: negative PCR result. * clinically affected individual. $ samples from which genome sequencing was successful.

**Table S6.2: Read mapping and genotyping results (EKU).** Read mapping and genotyping results using EAGER. Here, details on raw reads, mapped reads, mean coverage of F-B genome, percentage of genome that is covered by at least 3 reads and number of SNPs are shown. Note that EAGER outputs a more extensive result table, including percentage of endogenous DNA, duplication factors and many more. The full report table is available as supplementary material.

| Sample Name | **# of raw reads** | **Mapped reads after duplicate removal** | **Mean coverage** | **Coverage ≥ 3X [%]** | **# of SNPs** |
|---|---|---|---|---|---|
| *Cercocebus atys* IGU | 355672 | **22886** | 6.12 | 82.39 | 207 |
| *Cercocebus atys* HATO | 1279702 | **108150** | 29.64 | 99.90 | 228 |
| *Chlorocebus sabaeus* M2 | 6618162 | **418774** | 113.12 | 99.99 | 101 |
| *Chlorocebus sabaeus* M3 | 5741494 | **447331** | 121.26 | 100.00 | 102 |
| *Chlorocebus sabaeus* A10 | 7491960 | **470303** | 114.71 | 99.96 | 115 |
| *Chlorocebus sabaeus* A12 | 3169682 | **207318** | 46.71 | 96.53 | 109 |
| BS1-40M5 | 4.33E+07 | 50,746 | 5.27 | 75.08 | 158 |
| BS2-40M5 | 1.19E+08 | 142490 | 15.07 | 98.18 | 208 |
| BS3-40M5 | 2.75E+07 | 36505 | 3.98 | 56.58 | 127 |
| BS9-40M5 | 3.44E+07 | 8296 | 0.94 | 8.86 | 29 |

| | | | | | |
|---|---|---|---|---|---|
| *Papio anubis* 40M5-pooled | - | **240262** | 25.85 | 99.80 | 244 |
| BS4-04-F5 | 2.83E+07 | 21422 | 2.20 | 23.84 | 46 |
| BS5-04-F5 | 2.99E+07 | 26246 | 2.67 | 29.99 | 71 |
| BS6-04-F5 | 1.85E+07 | 17257 | 1.76 | 16.83 | 36 |
| BS7-04-F5 | 3.99E+07 | 37511 | 3.85 | 48.12 | 103 |
| BS8-04-F5 | 3.49E+07 | 1310 | 0.12 | 0.45 | 0 |
| *Papio anubis* 04-F5-pooled | - | **77809** | **7.97** | 88.70 | 174 |

**Table S6.3: Published genomes used for phylogenetic analyses.**

| Sample | RefSeq ID |
|---|---|
| Nichols | NC_021490.2 |
| SS14 | NC_021508.1 |
| Chicago | NC_017268.1 |
| Mexico A | NC_018722.1 |
| DAL-1 | NC_016844.1 |
| SEA81-4 | CP003679.1 |
| Fribourg-Blanc | NC_021179.1 |
| SamoaD | NC_016842.1 |
| CDC2 | NC_016848.1 |
| Gauthier | NC_016843.1 |
| BosniaA | CP007548.1 |

**Table S6.4: Summary of the PSGS sequencing results of 4 genomic DNA (gDNA) pools of the East African baboon isolate (40M5160407; MU, EKU, DPZ). Each gDNA pool resulted in overlapping PCR products. Processing was done on an Illumina Nextera NX library preparation and sequencing on MiSeq nano v2 2x250 bp. Raw reads in fastq format (past filter). *R1: forward read; R2: reverse read; Mean Qx%: percent bases with a Phred score of at least x; Mean Q: mean Phred score; Masked: reads smaller 10 bases get masked with 35 consecutive "N". ** length according to the published Fribourg-Blanc (F-B) genome (Zobaníková et al. 2013) (GenBank CP003902.1); nts=nucleotides.**

| Sample* | Pool** (nts) | No. of reads | No. of bases | Mean read length | Mean Q20%* | Mean Q30%* | Mean Q* | Masked* |
|---|---|---|---|---|---|---|---|---|
| M1_S19_L001_R1_001 | 1 (259,918) | 216,572 | 38,354,135 | 178 | 97.43 | 96.32 | 37 | 1,153 |
| M1_S19_L001_R2_001 | | 216,572 | 38,449,943 | 178 | 93.89 | 91.58 | 36 | 1,151 |
| M2_S31_L001_R1_001 | 2 (253,932) | 293,796 | 48,039,997 | 165 | 95.40 | 93.58 | 36 | 2,503 |
| M2_S31_L001_R2_001 | | 293,796 | 48,446,711 | 166 | 90.34 | 87.07 | 35 | 2,500 |
| M3_S43_L001_R1_001 | 3 (254,562) | 313,125 | 52,076,572 | 167 | 97.24 | 96.09 | 37 | 2,138 |
| M3_S43_L001 | | 313,125 | 52,293,004 | 168 | 93.04 | 90.60 | 36 | 2,123 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| _R2_001 | | | | | | | | |
| M4_S55_L001_R1_001 | 4 (376,699) | 315,732 | 53,669,732 | 171 | 96.80 | 95.53 | 37 | 2,175 |
| M4_S55_L001_R2_001 | | 315,732 | 53,901,788 | 172 | 92.78 | 90.29 | 36 | 2,176 |

**Table S6.5: Number of nucleotide differences (i.e. indels and SNVs) of various lengths between the genome of the baboon isolate 40M5160407 and the published *TPE* Gauthier and TPE F-B genomes. *tprD* and *tprK* genes as well as the differences in the number of repeats in *arp* and *tp0470* genes were excluded from the analysis.**

| Length of nucleotide differences | M5-40 vs. Gauthier | M5-40 vs. F-B |
|---|---|---|
| 1 nt | 252 | 307 |
| 2 nt | 3 | 5 |
| 3 nt | 1 | 5 |
| 4 nt | 1 | 1 |
| 5 nt | 1 | 0 |
| 6 nt | 0 | 2 |
| 9 nt | 4 | 2 |
| 15 nt | 1 | 1 |
| 33 nt | 1 | 0 |
| 42 nt | 0 | 1 |
| 79 nt | 1 | 0 |
| 302 nt | 1 | 0 |
| 430 nt | 0 | 1 |
| **Total number** | **266** | **325** |

**Table S6.6: Proteins encoded by the *TPE* Fribourg-Blanc baboon isolate with 1 and more amino acid changes when compared to the *TPE* strain CDC-2 proteome.**

| Gene*<br>Protein / Functional group | *TPE* F-B whole genome coordinates$^{\Delta}$ | Type of change in comparison with *TPE* Gauthier | Number of aa changes | Result of the frameshift mutation |
|---|---|---|---|---|
| **TPFB_0005** GyrA, cDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase / DNA replication, repair, recombination | 6654 | 1 SNV | 1 | |
| **TPFB_0012**$^{\circ}$ | 12487-12488 | 1 bp deletion resulting in frameshift mutation | | considered as pseudogene, (deletion in position 114 out of 177 bp) |
| **TPFB_0018** GreA, transcription elongation factor / transcription | 20705 | 1 SNV | 1 | |
| **TPFB_0033** hypothetical protein / unknown | 41224 | 1 SNV | 1 | |
| **TPFB_0040 M**cp, putative methyl-accepting chemotaxis protein / cell rocesses | 49369-49373 | 5 bp insertion resulting in frameshift mutation | 6 | protein shortening on C-terminus from 814 to 810 aa (808 aa similar to CDC-2) |

| | | | | |
|---|---|---|---|---|
| **TPFB_0092** RpoE, DNA-directed RNA polymerase sigma subunit / transcription | 101955 | 1 SNV | 1 | |
| **TPFB_0098** DnaJ1, chaperone / cell processes | 106884, 107247 | 2 SNV | 2 | |
| **TPFB_0117** TprC / unknown | within region 134966-136550 | 19 SNV | 16 | |
| **TPFB_0126a** hypothetical protein / unknown | within region 148979-148985 | 6 SNV | 2 | |
| **TPFB_0126b** hypothetical protein / unknown | 148982-148985 | 4 SNV | 14 | SNV in START codon, protein shortening on N-terminus from 135 to 121 aa |
| **TPFB_0131** TprD / unknown | 153989 | 1 SNV | 1 | |
| **TPFB_0152a** hypothetical protein / unknown | 176235 | 1 SNV | 1 | |
| **TPFB_0179** hypothetical protein / unknown | 199635-199637 | 3 bp insertion | 1 | |
| **TPFB_0196** RplP, ribosomal protein L16 / translation | 210334 | 1 SNV | 1 | |
| **TPFB_0200** RplX, ribosomal protein L24 / translation | 211401 | 1 SNV | 1 | |

| | | | |
|---|---|---|---|
| **TPFB_0236** NusG, transcription antitermination protein / regulation | 243855 | 1 SNV | 1 |
| **TPFB_0242** RpoC, DNA-directed RNA polymerase subunit beta prime / transcrition | 254357, 254472 | 2 SNV | 1 |
| **TPFB_0245** hypothetical protein / unknown | 258570 | 1 SNV | 1 |
| **TPFB_0279** bifunctional cytidylate kinase/ribosomal protein / translation | 295340-295342 | 3 bp insertion | 1 |
| **TPFB_0303** MutL, DNA mismatch repair protein / DNA replication, repair, recombination | 319012, 319401, 321132 | 3 SNV | 3 |
| **TPFB_0316** TprF / unknown | within region 332557-333448 | 7 SNV | 6 |
| **TPFB_0322** sugar ABC superfamily ATP binding cassette transporter, membrane protein / transport | 340498-340499, 340529 | 1 bp deletion and 1 bp insertion in a close proximity | 9 |
| **TPFB_0324** putative outer membrane protein / unknown | within region 342126-345775 | 5 SNV | 5 |
| **TPFB_0326** Tp92, outer membrane protein / virulence | 346413, 348114, 348115 | 3 SNV | 2 |

| | | | | |
|---|---|---|---|---|
| **TPFB_0344** TrcF, transcription-repair coupling factor / transcription | 368400 | 1 SNV | 1 | |
| **TPFB_0345a** hypothetical protein / unknown | 372913, 372920 | 2 SNV | 1 | |
| **TPFB_0346** putative lipoprotein / unknown | 373273, 373352, 373484 | 3 SNV | 3 | |
| **TPFB_0347** putative membrane protein / unknown | 373760, 373761 | 2 bp insertion leading to frameshift mutation | 40 | protein shortening on N-terminus from 276 to 236 aa |
| **TPFB_0370** hypothetical protein / unknown | 396159-396160 | 6 bp deletion | 2 | |
| **TPFB_0433** Arp, acidic repeat protein / unknown | 463017-463676 | 11x60 bp insertion | 220 | |
| **TPFB_0457** hypothetical protein / unknown | 488933 | 1 SNV | 1 | |
| **TPFB_461a** hypothetical protein / unknown | 493022-493023 | 1 bp deletion resulting in frameshift mutation | 46 | protein elongation on C-terminus from 60 to 81 aa (35 aa similar to TPE CDC-2) |
| **TPFB_0462** putative lipoprotein / unknown | 493401-493403 | 3 bp insertion | 1 | |

| | | | | |
|---|---|---|---|---|
| **TPFB_0463** | 494190 | 1 SNV | 1 | |
| **TPFB_0470** TPR domain protein / unknown | 499964-499965 | 15x24 bp deletion | 120 | |
| **TPFB_0484** hypothetical protein / unknown | 517708-517709 | 1 bp deletion resulting in frameshift mutation | 103 | protein shortening on N-terminus from 671 to 568 aa |
| **TPFB_0488** Mcp, methyl-accepting chemotaxis protein / cell processes | within region 523706-525356 | 10 SNV | 10 | |
| **TPFB_0529** NtpA2, two-sector ATPase, V(1) subunit A / transport | 575936 | 1 SNV | 1 | |
| **TPFB_0548** hypothetical protein / unknown | 594092-594093, 594408, 594420, 594421, 595033 | 42 bp deletion and 4 SNV | 17 | |
| **TPFB_0552** hypothetical protein / unknown | 600656 | 1 SNV | 1 | |
| **TPFB_0577** putative membrane protein / unknown | 630702, 630712 | 2 SNV | 2 | |
| **TPFB_0584** hypothetical protein / unknown | 636174 | 1 SNV | 1 | |
| **TPFB_0596** PcnB2, tRNA polynucleotide | 650042 | 1 SNV | 1 | |

| | | | |
|---|---|---|---|
| adenylyltransferase / translation | | | |
| **TPFB_0620** TprI / unknown | within region 673711-674602 | 8 SNV | 6 |
| **TPFB_0622** putative membrane protein / unknown | 679037 | 1 SNV | 1 |
| **TPFB_0639** Mcp, methyl-accepting chemotaxis protein / cell processes | 700639 | 1 SNV | 1 |
| **TPFB_0640** Mcp, methyl-accepting chemotaxis protein / cell processes | 703788 | 1 SNV | 1 |
| **TPFB_0652** PotA, spermidine/putrescine ABC superfamily ATP binding cassette transporter, ABC protein / transport | 718955 | 1 SNV | 1 |
| **TPFB_0675** hypothetical protein / unknown | 743072 | 1 SNV | 1 |
| **TPFB_0690** putative lipoprotein / unknown | 760088 | 1 SNV | 1 |
| **TPFB_0730** PgsA2, CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase / general metabolism | 798522 | 1 SNV | 1 |
| **TPFB_0733** hypothetical protein / unknown | 801328, 801349 | 2 SNV | 2 |

| | | | |
|---|---|---|---|
| **TPFB_0747** hypothetical protein / unknown | 814737 | 1 SNV | 1 |
| **TPFB_0817** Eno, phosphopyruvate hydratase / general metabolism | 888576 | 1 SNV | 1 |
| **TPFB_0854** putative lipoprotein / unknown | 928417 | 1 SNV | 1 |
| **TPFB_0856** putative lipoprotein / unknown | 936307, 936308, 936310 | 3 SNV | 2 |
| **TPFB_0856a** hypothetical protein / unknown | 937211, 937317, 937445, 938301 | 4 SNV | 4 |
| **TPFB_0858** putative lipoprotein / unknown | 937211, 937317, 937445, 938301 | 4 SNV | 4 |
| **TPFB_0859** hypothetical protein / unknown | 938535, 939155-939156 | 1 SNV and 3 bp deletion | 2 |
| **TPFB_0861** GlmS, glutamine--fructose-6-phosphate transaminase / general metabolism | 946841, 946842 | 2 SNV | 1 |
| **TPFB_0865** putative outer membrane protein / unknown | within region 946841-946848 | 7 SNV | 3 |

| | | | | |
|---|---|---|---|---|
| **TPFB_0891** InfB, initiation factor IF2 / translation | 972710 | 1 SNV | 1 | |
| **TPFB_0896**° | within region 976990-977041 | 6 SNV | | SNV leading to STOP codon, considered as pseudogene |
| **TPFB_0901** NorM, MATE family multi antimicrobial extrusion protein / transport | 985974 | 1 SNV | 1 | |
| **TPFB_0949a** hypothetical protein / unknown | 1034780 | 1 SNV | 1 | |
| **TPFB_0952** putative lipase/esterase / general metabolism | 1036094 | 1 SNV | 1 | |
| **TPFB_0957** TRAP-T family tripartite ATP-independent periplasmic transporter / transport | 1041824 | 1 SNV | 1 | |
| **TPFB_0960** FlgG1, flagellar basal body rod protein / cell structure | 1044647 | 1 SNV | 1 | |
| **TPFB_0966** putative lipoprotein / unknown | 1051534 | 1 SNV | 1 | |
| **TPFB_0967** hypothetical protein / unknown | 1053139, 1053293 | 2 SNV | 1 | |
| **TPFB_0968** hypothetical protein / unknown | 1053676, 1053893, 1054025, | 4 SNV | 4 | |

| | | | |
|---|---|---|---|
| | 1055233 | | |
| **TPFB_0973** PheS, phenylalanine--tRNA ligase alpha subunit / translation | 1059810 | 1 SNV | 1 |
| **TPFB_0976** putative membrane protein / unknown | 1062865 | 1 SNV | 1 |
| **TPFB_0993a** hypothetical protein / unknown | 1081313 | 1 SNV | 1 |
| **TPFB_0998** AlsT2, sodium/alanine symporter family protein / transport | 1088215 | 1 SNV | 1 |
| **TPFB_1007** ThyX, thymidylate synthase / general metabolism | 1100256 | 1 SNV | 1 |

***** Gene *tprK* (TPE40M5_0897) was omitted from the analysis due to its intra-strain variability.

$^{\Delta}$ Coordinates correspond to the positions leading to amino acid changes.

$^{\circ}$ Gene was not annotated in the TPE F-B genome.

**Table S6.7: Proteins encoded by the LMNP baboon isolate 40M5160407 with 1 and more amino acid changes when compared to the *TPE* strain Gauthier proteome.**

| Gene*<br><br>Protein / Functional group | 40M5160407 whole genome coordinates$^{\Delta}$ | Type of change in comparison with *TPE* Gauthier | Number of aa changes | Result of the frameshift mutation |
|---|---|---|---|---|
| **TPE40M5_0001** DnaA, DNA-directed DNA replication initiator protein / DNA replication, repair, recombination | 263 | 1 SNV | 1 | |
| **TPE40M5_0006** putative lipoprotein / unknown | 7547-7548, 8213 | 1 bp deletion resulting in frameshift mutation, 1 SNV | 237 | protein shortening on C-terminus from 415 to 195 aa (178 aa similar to TPE Gauthier) |
| **TPE40M5_0012** hypothetical protein / unknown | 12486-12487 | 1 bp deletion resulting in frameshift mutation | 20 | protein shortening on C-terminus from 58 to 42 aa (38 aa similar to TPE Gauthier) |
| **TPE40M5_0023** NSS family putative amino acid:sodium (Na+) symporter / transport | 27869 | 1 SNV | 1 | |
| **TPE40M5_0040** Mcp, putative methyl-accepting chemotaxis protein / cell processes | 49368-49371 | 5 bp insertion resulting in frameshift mutation | 6 | protein shortening on C-terminus from 771 to 767 aa (765 aa similar to TPE Gauthier) |

| | | | | |
|---|---|---|---|---|
| **TPE40M5_0042** hypothetical protein / unknown | 49826 | 1 SNV | 1 | |
| **TPE40M5_0067** TPR domain protein / unknown | 72691-72694 | 4 bp insertion resulting in frameshift mutation | 11 | protein shortening on N-terminus from 390 to 376 aa |
| | 72732-72733 | 9 bp deletion | 3 | |
| | 72862, 72952 | 2 SNV | 2 | |
| **TPE40M5_0083** hypothetical protein / unknown | 93789 | 1 SNV | 1 | |
| **TPE40M5_0098** DnaJ chaperone / cell processes | 107102 | 1 SNV | 1 | |
| **TPE40M5_0117** TprC / unknown | 136694-136695 | 1 bp deletion | | deletion in START codon, instead of GTG gene starts with ATG |
| | within region 134911-136545 | 26 SNV | 15 | |
| **TPE40M5_0119** methionine ABC superfamily ATP binding cassette transporter / transport | 138272 | 1 SNV | 1 | |
| **TPE40M5_0127a** hypothetical protein / unknown | 150210-150211 | 1 bp insertion resulting in frameshift | 106 | protein elongation on C-terminus from 126 to 222 aa (116 aa similar |

| | | | | |
|---|---|---|---|---|
| | | mutation | | to TPE Gauthier) |
| **TPE40M5_0127b** putative lipoprotein / unknown | 150210-150211, 150463 | 1 bp insertion resulting in frameshift mutation and 1 SNV | | considered as pseudogene (insertion in position 74-75 out of 414 bp) |
| **TPE40M5_0131** TprD / unknown | 154140-154141 | 1 bp deletion | | deletion in START codon, instead of GTG gene starts with ATG |
| | within region 152357-153408 | 25 SNV | 14 | |
| **TPE40M5_0134** putative outer membrane protein / unknown | 156213-156215, 156253 | 3 bp insertion and 1 SNV | 2 | |
| **TPE40M5_0136** putative outer membrane protein / virulence | 158538-158570 | 33 bp insertion | 11 | |
| | within region 158132-158511 | 7 SNV | 3 | |
| **TPE40M5_0143** ABC superfamily ATP binding cassette transporter, membrane protein / transport | 164599 | 1 SNV | 1 | |

| | | | | |
|---|---|---|---|---|
| **TPE40M5_0219**<br>probable sigma factor regulatory protein / regulation | 224855, 226013 | 2 SNV | 2 | |
| **TPE40M5_0230**<br>PriA, DNA replication factor Y / DNA replication, repair, recombination | 239958 | 1 SNV | 1 | |
| **TPE40M5_0249**<br>FlaA, flagellar filament outer layer protein / cell structure | 262681 | 1 SNV | 1 | |
| **TPE40M5_0259** LysM domain protein / unknown | 271476-271484 | 9 bp insertion | 3 | |
| **TPE40M5_0279**<br>bifunctional cytidylate kinase/ribosomal protein / translation | 296015, 296193 | 2 SNV | 1 | |
| **TPE40M5_0286** hypothetical protein / unknown | 301572 | 1 SNV | 1 | |
| **TPE40M5_0312a** hypothetical protein / unknown | 329051-329052 | 1 bp deletion resulting in frameshift mutation | | considered as pseudogene (deletion in position 103-104 out of 162 bp) |
| **TPE40M5_0313** TprE / unknown | 330487, 331148, 331175, | 4 SNV | 2 | |

| | | | | |
|---|---|---|---|---|
| | 331206 | | | |
| **TPE40M5_0316** TprF / unknown | 334039-334040 | 1 bp deletion | 1 | deletion in START codon, instead of GTG gene starts with ATG |
| | within region 332289-334033 | 24 SNV | 18 | |
| **TPE40M5_0317** TprG / unknown | 334965 | 1 SNV | 1 | |
| **TPE40M5_0319** TmpC, sugar ABC superfamily ATP binding cassette transporter, membrane protein / transport | 336858, 336869, 336963 | 3 SNV | 3 | |
| | 337590 | 1 SNV | | read through STOP codon, fusion to TPE40M5_0320 |
| **TPE40M5_0321** sugar ABC superfamily ATP binding cassette transporter / transport | 337867 | 1 SNV | 1 | |
| **TPE40M5_0326** Tp92, outer membrane protein / virulence | within region 347585-348193 | 7 SNV | 7 | |
| **TPE40M5_0334** HTH domain protein / unknown | 359096 | 1 SNV | 1 | |

| | | | | |
|---|---|---|---|---|
| **TPE40M5_0346** putative lipoprotein / unknown | 373022, 373479, 373488 | 3 SNV | 2 | |
| **TPE40M5_0347** putative membrane protein / unknown | 373756-373757 | 2 bp insertion resulting in frameshift mutation | 40 | protein shortening on N-terminus from 276 to 236 aa |
| **TPE40M5_0398** FliE flagellar hook-basal body protein / cell structure | 424529 | 1 SNV | 1 | |
| **TPE40M5_0401** FliH, IIISP family Type III (virulence-related) secretory pathway protein / virulence | 428180 | 1 SNV | 1 | |
| **TPE40M5_0433** Arp, acidic repeat protein / unknown | 462316 | 1 SNP | 1 | |
| | 463318-463319 | 1x60 bp deletion | 20 | |
| **TPE40M5_0444** LysM domain protein / unknown | 472228 | 1 bp insertion resulting in frameshift mutation | 42 | protein shortening on N-terminus from 342 to 300 aa |
| **TPE40M5_0462** putative lipoprotein / unknown | 493896-493897 | 9 bp deletion | 3 | |
| | 493749 | 1 SNP | 1 | |

| | | | |
|---|---|---|---|
| **TPE40M5_0483** hypothetical protein / unknown | 514807 | 1 SNV | 1 |
| **TPE40M5_0488** Mcp, methyl-accepting chemotaxis protein / cell processes | within region 523457-524502 | 11 SNV | 11 |
| **TPE40M5_0496** TPR domain protein / unknown | 533718 | 1 SNV | 1 |
| **TPE40M5_0505** hexokinase / general metabolism | 541183 | 1 SNV | 1 |
| **TPE40M5_0512** IspDF, bifunctional 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase/2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase / general metabolism | 550259, 550820 | 2 SNV | 2 |
| **TPE40M5_0514** UvrA, excision endonuclease subunit / DNA Replication, Repair, Recombination | 552914 | 1 SNV | 1 |
| **TPE40M5_0526** HrpA, ATP-dependent helicase / DNA Replication, Repair, Recombination | 571132 | 1 SNV | 1 |
| **TPE40M5_0528** | 573806 | 1 SNV | 1 |

| | | | | |
|---|---|---|---|---|
| NtpB2, two-sector ATPase, V(1) subunit B / transport | | | | |
| **TPE40M5_0534** hypothetical protein / unknown | 579146, 579341 | 2 SNV | 2 | |
| **TPE40M5_0546** S1B subfamily peptidase / general metabolism | 591706 | 1 SNV | 1 | |
| **TPE40M5_0548** outer membrane protein / unknown | within region 593836-594613 | 7 SNV | 7 | |
| **TPE40M5_0549** ClpA, S14 family endopeptidase / cell processes | 595254 | 1 SNV | 1 | |
| **TPE40M5_0569** Xaa-Pro aminopeptidase / general metabolism | 620521 | 1 SNV | 1 | |
| **TPE40M5_0584** hypothetical protein / unknown | 636180 | 1 SNV | 1 | |
| **TPE40M5_0610** TprH / unknown | 664542-664543 | 1 bp deletion resulting in frameshift mutation | 280 | protein shortening from 693 to 475 aa (413 aa similar to TPE Gauthier) |
| **TPE40M5_0611** ABC superfamily ATP binding cassette transporter, ABC | 666233 | 1 SNV | 1 | |

| | | | | |
|---|---|---|---|---|
| protein / transport | | | | |
| **TPE40M5_0620** TprI / unknown | 674940-674941 | 1 bp deletion | 1 | deletion in START codon, instead of GTG gene starts with ATG |
| | within region 673190-674934 | 24 SNV | 18 | |
| **TPE40M5_0621** TprJ / unknown | 675865 | 1 SNV | 1 | |
| **TPE40M5_0622** putative membrane protein / unknown | 677330 | 1 SNV | 2 | SNV leading to STOP codon, protein truncation on C-terminus from 595 to 593 aa |
| **TPE40M5_0629** hypothetical protein / unknown | 689822 | 1 SNV | 1 | |
| **TPE40M5_0639** methyl-accepting chemotaxis protein / cell processes | 700841, 701394 | 2 SNV | 2 | |
| **TPE40M5_0640** Mcp, methyl-accepting chemotaxis protein / cell processes | 702455, 702637, 702679 | 3 SNV | 3 | |
| **TPE40M5_0652** PotA, spermidine/putrescine ABC superfamily ATP binding cassette transporter / transport | 717776 | 1 SNV | 1 | |

| | | | | |
|---|---|---|---|---|
| **TPE40M5_0654** PotC, spermidine/putrescine ABC superfamily ATP binding cassette transporter / transport | 719981, 720341, 720540 | 3 SNV | 3 | |
| **TPE40M5_0678** hypothetical protein / unknown | 744875 | 1 SNV | 1 | |
| **TPE40M5_0684** MglB, galactose ABC superfamily ATP binding cassette transporter, binding protein / transport | 750550 | 1 SNV | 1 | |
| **TPE40M5_0693** putative lipoprotein / unknown | 762579 | 1 SNV | 1 | |
| **TPE40M5_0696** putative nicotinamidase / general metabolism | 767109 | 1 bp insertion resulting in frameshift mutation | 20 | protein shortening on N-terminus from 278 to 258 aa |
| **TPE40M5_0698** hypothetical membrane protein / unknown | 768282 | 1 SNV | 1 | |
| **TPE40M5_0722** FliL2, flagellar basal body-associated protein / cell structure | 792057 | 1 SNV | 1 | |
| **TPE40M5_0733** hypothetical protein / unknown | 800658, 800673, 800717 | 3 SNV | 3 | |
| **TPE40M5_0741** NadD, putative nicotinate- | 808690 | 1 SNV | 1 | |

| | | | | |
|---|---|---|---|---|
| nucleotide adenylyltransferase / general metabolism | | | | |
| **TPE40M5_0746** PpdK, pyruvate, phosphate dikinase / general metabolism | 812279, 813232 | 2 SNV | 2 | |
| **TPE40M5_0748** CfpA, cytoplasmic filament protein A / cell processes | 816355 | 1 SNV | 1 | |
| **TPE40M5_0758** RpsU, ribosomal protein S21 / translation | 824067 | 1 SNV | 1 | |
| **TPE40M5_0761** hypothetical protein / unknown | 827081-827082 | 2 bp deletion resulting in frameshift mutation | | fusion to TPE40M5_0762 |
| **TPE40M5_0764** HD-GYP domain protein / regulation | 830136 | 1 SNV | 1 | |
| **TPE40M5_0796** putative ApbE family protein / general metabolism | 864688, 865092 | 2 SNV | 1 | |
| **TPE40M5_0804** sugar ABC superfamily ATP binding cassette transporter / transport | 874803 | 1 SNV | 1 | |
| **TPE40M5_0842** Map, methionyl aminopeptidase / translation | 917314 | 1 SNV | 1 | |

| | | | | |
|---|---|---|---|---|
| **TPE40M5_0856a** hypothetical protein / unknown | within region 936623-936821 | 6 SNV | 6 | |
| **TPE40M5_0858** putative lipoprotein / unknown | 937531-937609 | 79 bp insertion resulting in frameshift mutation | 49 | protein elongation on C-terminus from 385 to 409 aa (360 aa similar to TPE Gauthier) |
| | within region 936623-937432 | 8 SNV | 6 | |
| **TPE40M5_0859** hypothetical protein / unknown | 938460-938461 | 9 bp deletion | 3 | |
| | 938313 | 1 SNP | 0 | |
| **TPE40M5_0865** putative outer membrane protein / unknown | 945781, 946149 | 2 SNP | 2 | |
| **TPE40M5_0896** hypothetical protein / unknown | 976339, 976341, 976342, 976344 | 4 SNP | | SNV leading to STOP codon, considered as pseudogene |
| **TPE40M5_0898**<br>RecB,<br>exodeoxyribonuclease V beta subunit / DNA replication, repair, recombination | 977954 | 1 bp insertion resulting in frameshift mutation | 7 | protein truncation from 1239 to 1238 aa (1232 aa similar to TPE Gauthier) |
| **TPE40M5_0919** | 998374, 998555 | 2 SNV | 2 | |

| | | | |
|---|---|---|---|
| thioredoxin group 1 family protein / general metabolism | | | |
| **TPE40M5_0931** hypothetical protein / unknown | 1015467 | 1 SNV | 1 |
| **TPE40M5_0939** pyruvate synthase / general metabolism | 1025096 | 1 SNV | 1 |
| **TPE40M5_0949** Oxa1, family cytochrome oxidase biogenesis protein / transport | 1032124 | 1 SNV | 1 |
| **TPE40M5_0952** putative lipase/esterase / general metabolism | 1035289, 1035290 | 2 SNV | 1 |
| **TPE40M5_0966** thioredoxin group 1 family protein / general metabolism | 1049816 | 1 SNV | 1 |
| **TPE40M5_0967** hypothetical protein / unknown | 1052396, 1052398 | 2 SNV | 1 |
| **TPE40M5_0968** hypothetical protein / unknown | 1053293 | 1 SNV | 1 |
| **TPE40M5_0969** putative outer membrane protein / unknown | 1055256 | 1 SNV | 1 |
| **TPE40M5_0976** hypothetical protein / unknown | 1062454-1062468 | 15 bp insertion | 5 |

| | | | |
|---|---|---|---|
| **TPE40M5_1012** RpoD, DNA-directed RNA polymerase sigma subunit / transcription | 1105470 | 1 SNV | 1 |
| **TPE40M5_1035** valine--tRNA ligase / translation | 1130609, 1132018 | 2 SNV | 1 |

* Gene *tprK* (TPE40M5_0897) was omitted from the analysis due to its intra-strain variability.

$^{\Delta}$ Coordinates correspond to the positions leading to amino acid changes.

Figure 5.2: High resolution version

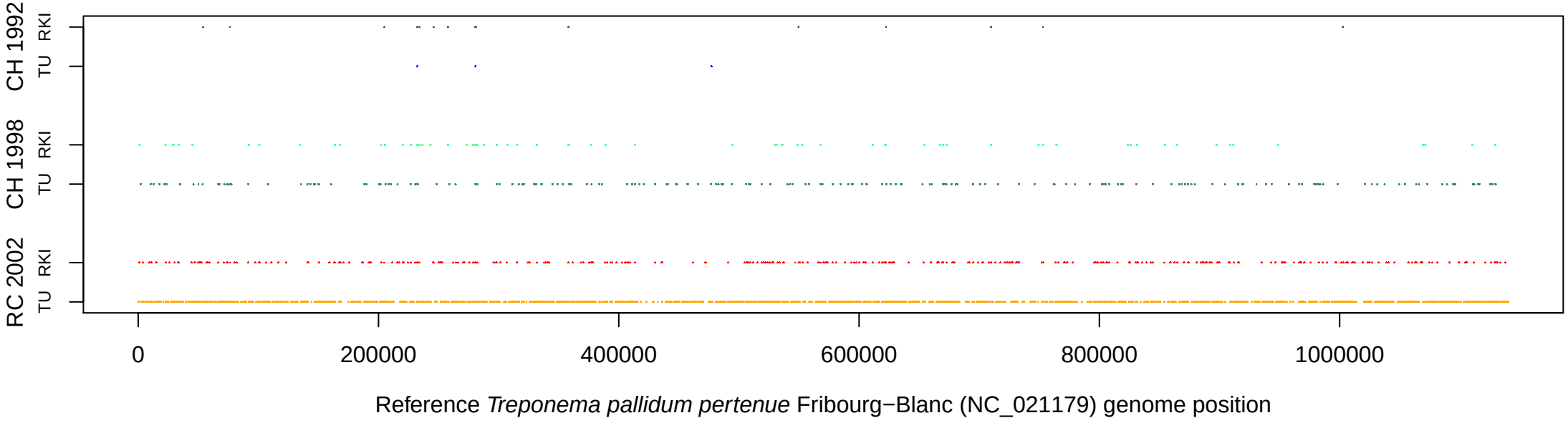


Figure S4.4: High resolution version

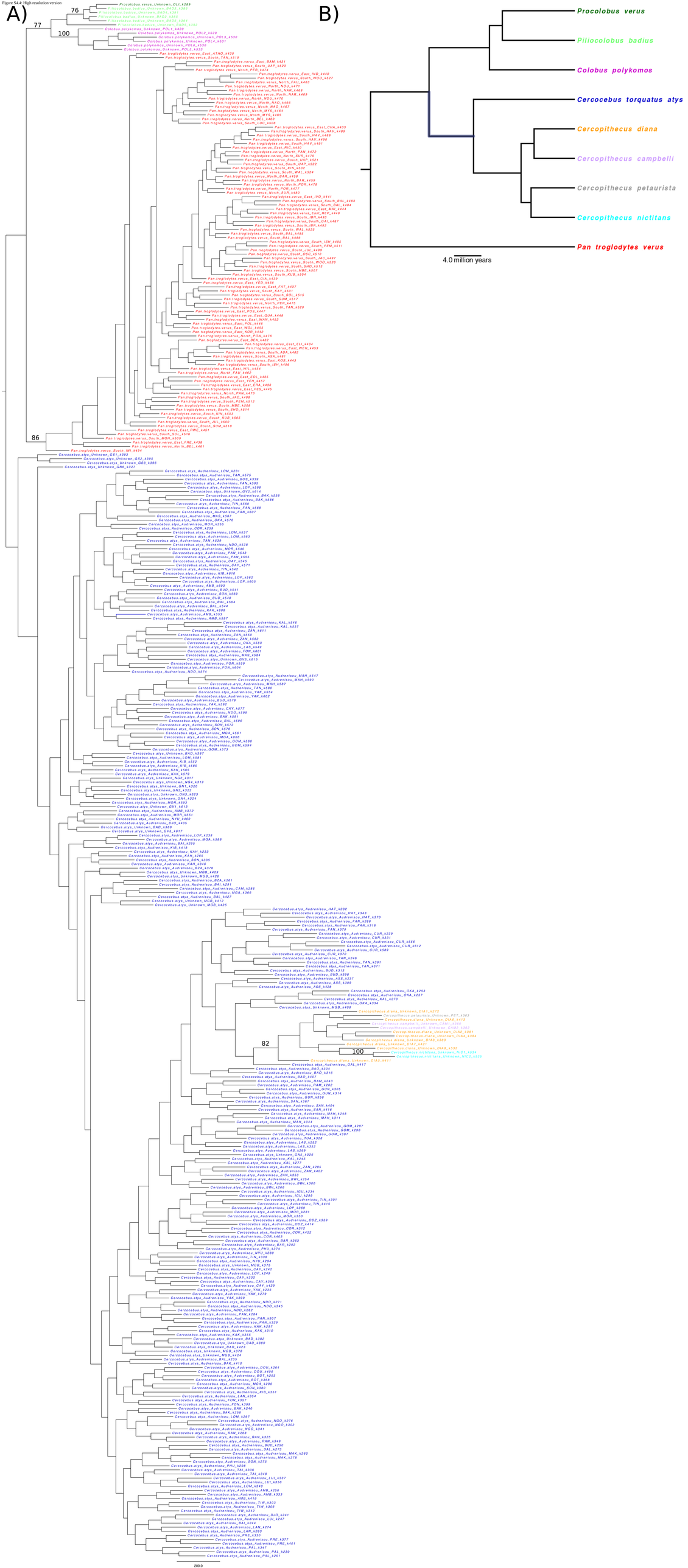


Figure S4.5: High resolution version