Spring 2019

A Novel SNP Genotyping Technique to Determine Orangutan Relatedness and Genetic Diversity at Camp Leakey in Tanjung Puting, Central Kalimantan

Ruth Ella Linsky
Central Washington University, ruth.linsky@cwu.edu

A NOVEL SNP GENOTYPING TECHNIQUE TO DETERMINE ORANGUTAN RELATEDNESS AND GENETIC DIVERSITY AT CAMP LEAKEY IN TANJUNG PUTING, CENTRAL KALIMANTAN

A Thesis

Presented to

The Graduate Faculty

Central Washington University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Primate Behavior

---

by
Ruth Ella Linsky
May 2019

CENTRAL WASHINGTON UNIVERSITY

Graduate Studies

We hereby approve the thesis of

Ruth Ella Linsky

Candidate for the degree of Master of Science

APPROVED FOR THE GRADUATE FACULTY

_____ _____
Dr. Lori Sheeran, Committee Chair

_____ _____
Dr. Biruté Mary Galdikas

_____ _____
Dr. Joseph Lorenz

_____ _____
Dean of Graduate Studies

ABSTRACT

A NOVEL SNP GENOTYPING TECHNIQUE TO DETERMINE ORANGUTAN RELATEDNESS AND GENETIC DIVERSITY AT CAMP LEAKEY IN TANJUNG PUTING, CENTRAL KALIMANTAN

by

Ruth Ella Linsky

May 2019

Genetic studies of dispersal patterns in wild populations of orangutans (*Pongo* spp.) have sought to confirm behavioral observations that female orangutans tend to stay near their natal range while males disperse. In order to genotype a previously unsampled wild population of endangered orangutans at Tanjung Puting National Park I developed novel application of a methyl based magnetic capture for enrichment of fecal DNA and commercial human targeted single nucleotide polymorphism (SNP) microarray technology. I confirmed results of this new genotyping technique through standard microsatellite short tandem repeat (STR) micro-capillary genotyping. I estimate genetic diversity and relatedness (*r*) for 32 (21 female and 9 male) wild orangutans at the Camp Leakey Study Site. I successfully isolated 125 known human SNP loci (0.08% of those targeted) which hybridized orangutan DNA on the human targeted *Illumina Infinium QC* array. Average relatedness within the population, estimated from our combo SNP/STR dataset using *TrioML* estimator, is at a level between half and first cousins (r = .082), and I found no significant difference of *r* between males and females. All males and females had relatives within the study site but paternity was not assigned to any potential fathers sampled. Results indicate all sampled males and females are from the local population. High and near equal relatedness for both sexes in this group, combined with a low number of males sampled, suggests conditions for the potential for deep inbreeding. This is a particular concern because the population is at risk from further isolation through habitat fragmentation.

ACKNOWLEDGMENTS

I first and foremost wish to thank Dr. Biruté Mary Galdikas for her mentorship, inspiration, and for the truly unparalleled efforts she continues to put towards protecting, conserving, and understanding orangutans. I wish to thank Dr. Steven Wagner, Dr. Lori Sheeran, Dr. Joe Lorenz, and Dr. Jessica Mayhew for making this work possible, inspiring my ideas, and getting me through the tough times during this research. This project would not have been mine to continue without the initial work and concepts of Cherri McGarvie, so thank you for your work and passion for orangutans. I wish to thank Ibu Waliyati, Ibu Renie, Ibu Sumi, and Pak Yandi for your friendship, for helping me navigate the administration and bureaucracy surrounding this research and for your lifetimes of work helping save and protect orangutans. Without you OFI would not exist. Thank you to all the people of OFI who helped me along the way to understand the world of the orangutan, Pak Tumin, Pak Sehat, Pak Fajar, Pak Robert, Dr. Ketut, Dr. Prima, Dr. Popo, Ibu Tuti, Ibu Lupi, Ibu Mila, Ibu Maryanti and so many more. Special thanks to those who helped me collect samples, Pak Ranto, Pak Kikis, Pak Tono, Pak Rahman, almarhum Pak Rocky, Pak Budi, Pak Asmaja, Pak Maat, Pak Engkot, Pak Umar, Pak Faisal, Ibu Iyum, Ibu Siti, Ibu Imelda. Thank you to those who helped me with samples along the way Miss Morgan Pettersson, Miss Emily Patton, Miss Celine Vincart, Miss Edythe Guice, Miss Casey Murtagh, Miss Divya Rao, Miss Sarah Godan-Blouin. Many thanks to almarhum Bapak Suwido Limin and Bapak Ici Kulu, Ibu Siska and Ibu Elvi at CIMTROP for your partnership and support. Absolute gratitude and respect to all of RISTEKDIKTI for allowing me to conduct research in Indonesia and for all you do for

your country. Without the Indonesian Ministry of Forestry there would be no protected forests in Indonesia, and I would not have been able to study orangutans or collect or export samples, so thank you for all you do. I extend thanks to LIPI for supporting my research and granting me permits to collect these samples. I wish to thank the CWU Graduate and Research department for financial support for this research as well as Pete and Sandra Barlow and Arlen and Debra Prentice for their Award and Scholarship funds. I wish to thank Kelley Quirk and all involved with the CWU Gear-Up program and for your support and understanding during my research.

I had incredible support from my family, Bernard Linsky, Elizabeth Millar, Andrea and Jean Linsky, and Peter von Hahn throughout this process and would not be who I am without them. My friends in Ellensburg (especially Kaitlin Wright and Krista Banda) and beyond were crucial for my sanity and soul and should be credited for part of this work.

I want to thank the orangutans of Camp Leakey for their patience and for allowing me to follow them and collect their refuse, may you all live long and happy lives.

I wish to acknowledge this research was conducted under the approved CWU IACUC protocol A011523 and all samples were collected with permit from the Indonesian government and exported from Indonesia with CITES permit 01152/IV/SATS-LN/2017.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In 2016 the International Union for Conservation of Nature (IUCN) upgraded the classification of Bornean orangutans (*Pongo pygmaeus*) from "endangered" to "critically endangered," and most orangutan populations are believed to be in sharp decline (Utami-Atmoko et al., 2017). Scientists have recently estimated a loss of over 100,000 individuals between 1999 and 2015 in Borneo alone, with more than 50% of populations being affected by habitat destruction, logging, and industrialized plantations (Voigt et al., 2018). Current population numbers are difficult to estimate directly; but decline has been exacerbated over the last several decades with increased widespread human driven deforestation and development. The most recent orangutan population and habitat viability analysis (Utami-Atmoko et al., 2017) identified the wild orangutan population found in Tanjung Puting National Park (TPNP) in Central Kalimantan (Indonesian Borneo) as essential to the long-term survival of the Bornean species because it is one of the largest remaining populations with approximately 4,000-6,000 individuals. The following research is based on the wild orangutan population at the Camp Leakey study area in TPNP, Central Kalimantan. This is the site of ongoing long-term research begun in 1971 by Dr. Biruté Mary Galdikas alongside local staff associated with Yayasan Orangutan Internasional Kalimantan (YOIK), Orangutan Foundation International (OFI), and students from the biology department at Universitas Nasional in Jakarta. This long-term research mainly focuses on wild behavior and ecology (Galdikas, 1985b, 1985c; Galdikas & Ashbury, 2013; Galdikas & Wood, 1990), and detailed data has been kept on

behavior and presence-absence within the study site over the last four decades. Matrilines have been recorded over several generations during this time. However, paternal and sibling relationships are difficult if not impossible to establish based on behavioral observations alone because of the orangutans' unique semi-solitary social organization and associated promiscuous mating. Here I attempt to identify the kinship and pairwise relationships, along with the overall genetic relatedness of the local wild orangutans at Camp Leakey using non-invasive DNA sample collection methods and a new human-based commercially available single nucleotide polymorphism (SNP) microarray genotyping method.

The objective of our study is to conduct a detailed investigation into the genetic relationships among the wild population of Bornean orangutans (*Pongo pygmaeus*) at the Camp Leakey study site. This novel investigation provides estimates of genetic variability and relatedness overall among those sampled as well as within sex groups in this population. Genetic variability or diversity parameters such as genotype and allele frequencies can be used to estimate relatedness and assess inbreeding in a small or decreasing population (Wright, 1931). These parameters can then help to deduce cryptic or complex social structure and mating patterns. Relatedness estimated from just a small number of microsatellite loci has recently been used to provide evidence for female philopatry and male-biased dispersal among orangutans in another isolated forest also within Central Kalimantan (Arora et al., 2012). The specific objectives of the current study were to assess genetic relatedness and kinship among orangutan individuals at Camp Leakey to determine genetic diversity within the population not yet determined

through behavioral observation. I hope to use estimates of genetic diversity to clarify dispersal patterns, and assess levels of inbreeding within the population, which if high could lead to loss of genetic diversity and can compromise long-term fitness of the population. I seek to use pedigree and parentage analyses to determine parentage and kinship and/or the minimum number of fathers among the sampled individuals. The number of individuals genetically contributing to population offspring can help determine mating system strategies and directly influences effective population size and genetic structure.

The proposed microarray genotyping technique in our study detailed below, can yield thousands of fine scale genetic markers to distinguish individual and population level differences and provide data applicable to estimating many different population characteristics. Estimating these population parameters provides important data for assessing population viability and can directly influence Camp Leakey management decisions and be extended to other orangutan rehabilitation and protected areas.

The following comprises a review of relevant literature regarding the study species, study site, objectives, and methods of the research conducted.

CHAPTER II

LITERATURE REVIEW

Review of Orangutans

Orangutans can be described as semi-solitary, although different age and sex classes exhibit varying degrees and types of sociality (Galdikas, 1985b). The driving force of this solitude is the dispersed nature of orangutan food sources in space and time throughout the forest. Orangutan food density differs by degree between Sumatra and Borneo, the only two islands orangutans currently inhabit, and have resulting consequences for the orangutans on each. However, it has been proposed that climatic changes during the late Miocene and Pliocene, specifically the onset of El Niño Southern Oscillation, brought changes in forest production cycles, creating the mast fruiting phenomenon and the severe periods of low fruit availability currently seen on both islands (Harrison & Chivers, 2007). Orangutans are primarily frugivorous but have been documented in Tanjung Puting National Park (Galdikas, 1988; Galdikas & Vasey, 1992) to eat nearly 400 species of fruit, plant stems, pith, roots, shoots, young leaves, buds, seeds, bark, vines, flowers, saps, as well as invertebrates such as ants, termites, and bees (and their products such as wild forest honey). These food sources are often sparse, with no species being found in large densities within the forest. In most locations, trees' phenology is not synchronized within or between species resulting in only some trees providing fruit, flowers, or young leaves on an isolated and unpredictable basis often dictated by microclimate factors. Orangutans must therefore dedicate a large portion of their time and energy budgets to solitary foraging throughout the forest canopy and lower levels in search of fruit and other high quality preferred food types (Rodman, 1979).

The orangutans' diffuse social organization and complex and somewhat cryptic social structure is a further consequence of this demand on time to meet nutritional and energy requirements. Low food source density means orangutans cannot forage in groups as some other great apes and primates do. The low population densities that result from solitary foraging also result in more relaxed social relationships between and among both sexes when compared to the other great apes. However, this is not the same as strict non-sociality, and these relaxed social relationships have complex behavioral and reproductive consequences for both sexes.

The main orangutan social unit consist of solitary adult males, semi-solitary adult females often accompanied by one or two dependent offspring, and adolescent or sub-adult males and adolescent females who are generally much more gregarious than are adults (Galdikas, 1985b). It has been documented through behavioral studies and supported through several genetic investigations that orangutans exhibit female philopatry with males the predominant dispersing sex (Arora et al., 2012; Goossens et al., 2006; Morrogh-Bernard, Morf, Chivers, & Krützen, 2011; van Noordwijk et al., 2012). Females' range sizes differ geographically, but ranges for *Pongo pygmaeus wurmbii*, the subspecies found in Tanjung Puting averages between 2.5-6km^2 (Singleton, Knott, Morrogh-bernard, Wich, & Schaik, 2009). Orangutan densities within the same subspecies found at other study sites are estimated to be between 2-4.2 individuals/km^2. Male range sizes are less well documented, but it is estimated at all current sites that males, both flanged and unflanged, range 3-5 times further and have larger home ranges than females (Atmoko Utami, Singleton, Noordwijk, Schaik, & Setia, 2009). Male and female orangutans come together in consortships during which they travel and feed

together before and after a period of several days during which mating occurs intensely (Galdikas, 1985a). These consortships can last several days to several weeks, and in Tanjung Puting have been recorded to go on for up to 2 years (Galdikas, 1985a).

Orangutans are one of the most sexually dimorphic primate, and among the great apes, follow closely only the gorilla (*Gorilla gorilla* spp) in size difference. Adult males are estimated to be between 2-2.3 times as heavy as females (Leigh & Shea, 1995). As their solitary lifestyle dictates, male orangutans are unable to rely on male-male collaborative strategies for reproductive access to females as compared to their African ape relatives chimpanzees, or to mate guard females in groups as do gorillas (Biruté M. Galdikas; personal communication, 2009). They must rely instead on sheer size and dominance in one-on one contest competition and have developed a system of delayed maturation corresponding to dual male reproductive strategies with different ways of locating and achieving reproductive success with females (Utami, Goossens, Bruford, De Ruiter, & van Hooff, 2002).

All male orangutans, once fully adult also exhibit extreme secondary sexual characteristics that accompany their large size. These include long hair, large throat sacs, and distinct facial flanges. The throat sac allows them to make the adult male orangutan's characteristic long call, which may announce his presence to male and female orangutans in the area (Galdikas, 1983). The long call may also indicate the direction in which he will be traveling in the following hours or next day (van Schaik, Damerius, & Isler, 2013). Data showing female movement towards dominant calling males (van Schaik et al., 2013; Setia & van Schaik, 2007) further supports the idea that this communication is intended to both attract females as well as deter less dominant males, as was originally

predicted by Galdikas (Galdikas, 1983). Adult (fully flanged mature) males, in Tanjung Puting appear to exhibit two ranging patterns (transient and resident), with residents generally staying within an approximately 6km^2 area, and transients ranging much further (Galdikas, 1985a). While adult males are totally intolerant of each other, resident males have been found to live in overlapping ranges, with evidence that they may actively avoid contact/conflict with the more dominant of males. Dominant males prefer to mate with older parous females rather than adolescents and adult females approach and prefer to initiate contact with dominant adult males (Galdikas, 1985b).

Male orangutans go through a period of subadulthood during which they are larger than females but have yet to develop the full adult characteristics (flanges). At Tanjung Puting juveniles appear to reach subadulthood at 12-15 years of age and are estimated to achieve adulthood generally around 19-20 years (Galdikas, 1985c). This period of subadulthood has been recorded on one instance to last until the age of 30 at one site on the island of Sumatra (Utami et al., 2002). Subadult males are sexually active, engage in forced copulations, and are less selective with mate choice than are adult males. These males do not long call and are believed to range much further than flanged males. Subadult males cannot compete with the much larger, more dominant, flanged males and are easily displaced, although generally much more tolerated compared to other non-dominant flanged males (Galdikas & Vasey, 1992). Because they are tolerated, these subadult males often follow dominant males during consortships with adult parous females and appear to adopt a strategy of waiting until the dominant male is distracted or out of sight before forcibly copulating with unwilling females who often protest but due to size differences often cannot escape (Galdikas, 1985c). Investigations into

reproductive success of each of the male morphs (subadult unflanged and adult flanged) have indicated both groups father offspring, however to varying degree with dominant flanged males (Banes, Galdikas, & Vigilant, 2015; Tajima, Malim, & Inoue, 2018) or multiple flanged males seeming to dominate paternities in some communities (Goossens et al., 2006). However, scientists have found at one site in Sumatra that unflanged males fathered up to half the offspring in over a 15 year time period (Utami et al., 2002).

Adult female orangutans often have overlapping ranges with female relatives and have been observed to associate, travel, and even feed with female relatives when resources permit (van Noordwijk et al., 2012). Female philopatry allows for greater tolerance towards female relatives resulting in more frequent association between adult females and their accompanying offspring(van Noordwijk et al., 2012). This may be due to reduced stress of being in the company of trusted relatives. This tolerance allows offspring opportunities to observe and engage in play with related con-specifics, which is rare for the species.

Wild female orangutans at Camp Leakey on average reproduce for the first time at the age of 15-16 years (Galdikas & Ashbury, 2013) and reproduce on average every 7-8 years after this (Galdikas & Wood, 1990). This lengthy inter-birth interval is accompanied by and due in part to the extensive maternal care and investment made into each offspring by the mother. This investment includes the mother carrying the infant almost full time for the first few years while providing milk and tolerating a close presence and even sharing food with the offspring until it is seven or eight years of age (van Adrichem, Utami, Wich, van Hooff, & Sterck, 2006). It is during this time that the

infant accompanies, observes, shares, and learns the complex feeding strategies required for survival in their rainforest home.

This long inter-birth interval results in females returning into sexually receptive and conceptive cycles up to eight years after pregnancy (Galdikas, 1982b). A thorough comparison between female orangutans and chimpanzees suggests that according to sexual-selection theory this reduction in overall reproductive potential over the orangutan's lifetime should result in an increase in female selection of high-quality males as mating partners (Stumpf, Emery Thompson, & Knott, 2008). The comparison also predicts females of both species to exhibit prosexual behavior outside of the periovulatory period as an anti-infanticide strategy of paternity confusion. This does appear to be the case in orangutans; researchers have documented at several sites that adult females prefer to mate with dominant, flanged males and show proceptivity to non-dominant males during early stages of pregnancy (Galdikas, 1985b; Tajima et al., 2018; Utami-Atmoko, 2000).

Previous Studies on Orangutan Relatedness

To date several estimates of orangutan genetic diversity and relatedness have been conducted at other long-term research sites in both Borneo and Sumatra. The first data published on reproductive success and relatedness of Sumatran orangutans (*Pongo abelii*), from the Ketambe research site in Ache Sumatra, using fecal samples and human microsatellite markers, indicated generally negative mean relatedness vales in both males and females from this population (Utami et al., 2002). The authors concluded that both sexes disperse from the natal area, however, these authors urge caution as their

conclusions are based on a small sample size that also included reintroduced ex-captive orangutans not natal to the area.

Further studies conducted by one of the co-authors of the Ketambe research, B. Goossens and colleagues in Sabah, Malaysia, the northern province of the Island of Borneo, investigated genetic diversity, relatedness, and dispersal within Bornean orangutans (*Pongo pygmaeus*) in the Lower Kinabatangan Wildlife Sanctuary (Goossens et al., 2005, 2006). They also used human microsatellite DNA markers isolated from fecal samples and reported unexpectedly high diversity within the entire population and similarly high relatedness values within males and females, further indicating both sexes as equally dispersing and remaining in natal areas. Kinabatangan is comprised of riparian forest blocks on either side of the Kinabtangan river, creating a ~270km^2 corridor of habitat along the river that is fragmented and surrounded by oil palm plantations. The authors have suggested orangutans in this refuge currently live in "communities of related individuals of both sexes" (p.2585) but admit that this result could be affected by intense habitat fragmentation and destruction, which may reduce orangutans' dispersal abilities.

Interestingly, Morrogh-Bernard and colleagues (2011), arrived at different conclusions in their study at the Sebangau Protected forest in Central Kalimantan, Indonesian Borneo. This study used similar fecal sampling methods as well as similarly human-derived microsatellite and mitochondrial markers; however, the results revealed significantly higher relatedness values among females sampled than among males sampled. Results suggested mitochondrial haplotype diversity was higher for males than females, which additionally supported the authors' conclusion of loose-knit female

philopatric groupings and male dispersal within orangutans at the Sebangau site. These findings were also similarly reported from another Central Kalimantan Site, The Tuanan Orangutan Research area (Arora et al., 2012; Nietlisbach et al., 2012; Nietlisbach, Nater, Greminger, Arora, & Krützen, 2010). This research group investigated genetic diversity and relatedness using autosomal, mitochondrial, and Y-chromosome microsatellite markers isolated from fecal samples from 40 orangutans and discovered significantly higher male mitochondrial haplotype diversity and significantly higher female average pairwise relatedness values. Arora and colleagues (2012) uniquely combined genetic analysis with GPS locational data for individuals sampled and confirmed that 10 of the 15 females sampled shared the study area with another female relative. Nietlisbach's (2012) group also compared mitochondrial and Y-chromosomal single nucleotide polymorphism (SNP) diversity from several Bornean sites. When combined, these data estimated higher Y-chromosome diversity and higher mitochondrial grouping across all sites, which further provides evidence in support of female philopatry and long distance male dispersal within these sites.

Both the Tuanan (7.5km^2) (Arora et al., 2012) and Sebangau (9.0km^2) (Morrogh-Bernard et al., 2011) orangutan research sites are within large contiguous protected and generally intact forests, the Mawas Conservation area (3090 km^2), and the Sebangau National Park (5140 km^2) respectively. The Mawas conservation area can be classified as a lowland peat swamp forest of intact orangutan habitat. The Sebangau research site is also within unfragmented mixed peat swamp habitat, but the forest in this area was selectively logged for more than three decades beginning in the 1970's. These two areas

represent two of the larger remaining orangutan populations within Kalimantan according to the 2016 orangutan habitat population viability analysis (Utami-Atmoko et al., 2017).

Most recent investigations into paternity and reproductive success among the ex-captive community at the Camp Leakey study site were conducted by Galdikas and colleagues (Banes et al., 2015). This study also utilized human-derived microsatellite markers isolated from fecal samples and focused on the ex-captive orangutans and their offspring that remain near to the rehabilitation site. They estimated paternity for many of the second and third generation individuals still present. Results indicated paternity skew in favor of the resident dominant male orangutan at the time, but this study did not estimate overall relative relatedness between the individuals sampled.

The research discussed here expands on the previous sampling conducted at Camp Leakey to the resident wild orangutan population surrounding the rehabilitation site. Significantly, this study uses commercially available high throughput human-based SNP genotyping microarray technology to reveal relatedness and other population parameters of the wild orangutans within Tanjung Puting National Park. This data can provide a wider range of information than traditional microsatellite markers used in most previous studies.

Study Site

Behavioral data on the local wild orangutans at the Camp Leakey (CL) Study site in Tanjung Puting National Park has been collected continuously since the study area was established by Biruté Mary Galdikas and Rod Brindamour in 1971. Galdikas' seminal research into the feeding ecology and behavior of orangutans at CL represents many

foundations of our understanding of the species. This research included descriptions of male and female sociality and reproduction (Galdikas, 1979, 1982b, 1985b, 1985c, 1985a); orangutan tool use (Galdikas, 1982a); and orangutan diet, range, and activity patterns (Galdikas, 1988).

At approximately 4,150km^2 (1886 km^2 of orangutan habitat), Tanjung Puting National Park is one of the largest protected areas in Central Kalimantan (Utami-Atmoko et al., 2017). The CL study area contains a mix of dry ground tropical heath and dipterocarp forests with veins of permanently wet and seasonally flooded peat swamp threaded throughout (Galdikas, 1979). The CL study area was initially established within a 35km^2 area of a protected are which is now Tanjung Puting National Park (Figure 1). The local wild orangutans are behaviorally monitored on an ongoing basis within a core are of 14km^2 (1400ha) which contain maintained trails (Figure 2).

In 1971 orangutan rehabilitation and release began at CL, and it is estimated ~90 individuals were released at this site until 1985 (Galdikas & Ashbury, 2013). These orangutans have also been the subject of various investigations into the species' cognitive abilities such as imitation (Russon & Galdikas, 1993) as well as the adoption and use of sign language (Shapiro, 1982). More recent investigations into ex-captive reproductive parameters have compared ex-captive data to wild orangutans (Galdikas & Ashbury, 2013).In ongoing data collection at CL, a subset of local wild individuals, who are identifiable by local staff and researchers, as well as any unidentified individuals encountered, are behaviorally studied through focal follows from nest to nest each day.

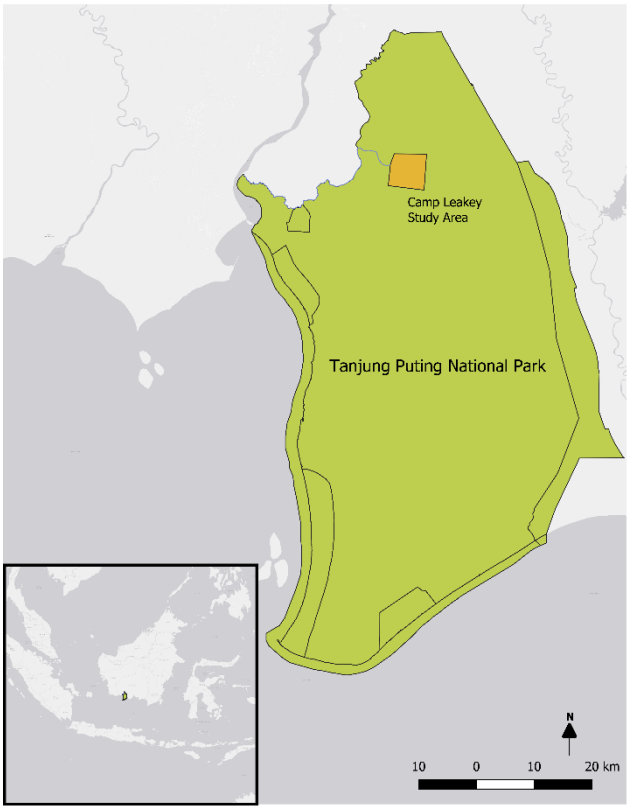


Figure 1 – The Camp Leakey Study Area is located along the Sekonyer River at the northern portion of Tanjung Puting National Park in Central Kalimantan, Indonesian Borneo

Data are collected by focal follows of orangutans for ten day periods. Matrilineal lines have been recorded for up to four generations on females whose home range includes the CL study area site. Thus, some pairwise relatedness between many individuals is known. However, overall relatedness and genetic diversity are unknown for the existing local population and are important factors in understanding existing population structure. These factors help discern behavioral patterns and can be used to assess potential impacts from deforestation and isolation in addition reintroduction and translocations.

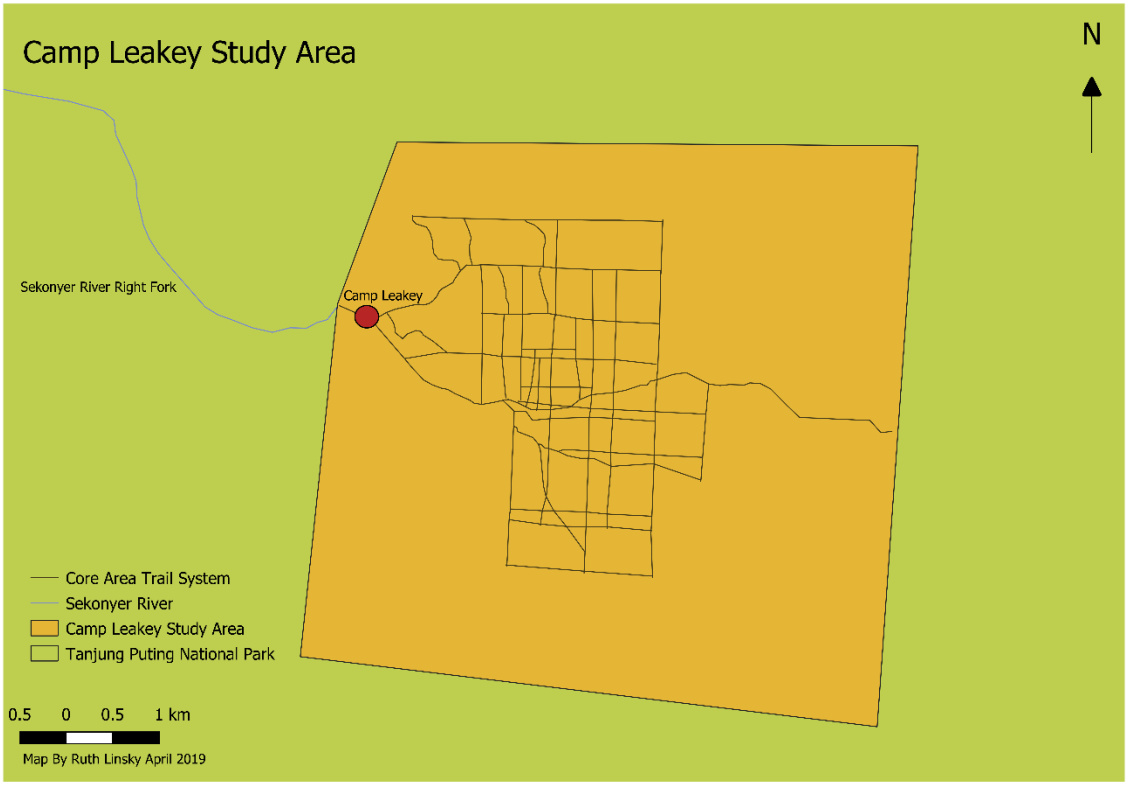


Figure 2 – The core trail system within the Camp Leakey Study area a total of 35km2. The red dot marks the location of Camp Leakey.

Review of Sample Collection Methods

Most previous studies have used similar methods for non-invasively collecting DNA samples from orangutans. A two-step method (Nsubuga et al., 2004) for fecal collection involving an initial 24-36 hour submersion in ethanol followed by immediate desiccation and freezing on silica beads significantly increases average concentration of DNA extracted from samples using standard commercial *Qiagen* DNA extraction kits. As most previous studies have made use of microsatellite or mitochondrial DNA markers, researchers have focused on maximising the production of PCR products post extraction. Goossens et al. (2000) used fecal samples from Sumatran orangutan to produce recommendations of multiple sampling and extractions in multiple tubes of at least three

extracts resulting in at least 12 PCR reactions per sample. These results are based on acquiring high enough PCR positive products to reduce risks of allelic drop out and to ensure positive PCR reaction products (Morin, Chambers, Boesch, & Vigilant, 2001; Taberlet et al., 1996).

Review of Genotyping Techniques

Advances in sequencing technology in the last decade have made possible full genome sequencing at a much more affordable and timelier rate. SNP marker analysis provides a much broader scale of understanding and provides more detailed information about individual and population genetics than ever before possible with microsatellite or mitochondrial gene marker analysis. Many targeted products for SNP genotyping, including microarrays, have been designed for model organisms (including and especially humans) and are available commercially from several companies with *Illumina* being the largest provider currently on the market.

Cross species microarray analysis has shown to be possible within species as genetically distant as oryx (*Oryx* spp.) and modern domesticated bovine (*Bos taurus*) which shared a common ancestor at least 23 million years ago (Ogden, Baird, Senn, & McEwing, 2012) and Antarctic fur seal (*Arctocephalus gazella*) and domestic dogs (*Canis lupus familiaris*) who diverged approximately 44 million years ago (Hoffman, Thorne, McEwing, Forcada, & Ogden, 2013). These studies were able to identify 185 of 54,001 (0.34%) and 173 of 173,662 (0.01%) polymorphic loci in common, respectively. A further study of wild thin horn (*Ovis dalli*) and bighorn (*Ovis canadensis*) sheep genotyped on a chip designed for commercial domestic sheep (*Ovis aries*) (divergent

relatively more recently at approximately 3 million years ago) identified 868 of 49,034 loci (1.7%) to be polymorphic and in common (Miller, Poissant, Kijas, & Coltman, 2011). Early investigations into ancestral alleles among humans and apes also proved ape DNA can be genotyped using human microarrays (Hacia et al., 1999) and confirmed 0.75% and 0.5% common loci between humans and bonobos (*Pan paniscus*), and humans and gorillas respectively.

The *Illumina Infinium Human QC* array contains 15,949 genomic, mitochondrial and sex linked markers to be used for human genotyping purposes. Precedent for using targeted capture of homologous human markers in non-human primates was set by Vallender (2011). Although current *Illumina Infinium* human microarrays have yet to be proven for closely related non-human primates, these arrays offer a low cost, low effort, low input DNA solution for genotyping. In its standard application microbeads on the chips hybridize specific known human SNP locations using targeted probes. UV light illuminates colored probes as specific nucleotides hybridize resulting in light intensity and color data. This data is translated using custom proprietary Illumina software, GenomeStudio2.0. The resulting data are then analyzed by clustering, quality control, and eventual output of genotypes for SNPs selected on numerous qualities.

Microsatellite markers, both autosomal, and Y-chromosome, have been used in the majority of orangutan relatedness studies previously conducted (Nietlisbach et al., 2010). These short-repeated sections are amplified using targeted primers and through polymerase chain reaction (PCR) thermocycling. PCR amplifications are then electrophoresed on standard (agarose or polyacrylamide) gel medium to separate amplicons of varying sizes. Gels are analyzed and coded by machinery or by hand.

Despite much research and improvement in genotyping practices microsatellite (or Short Tandem Repeat, STR) analysis, is time consuming, requires multiple replicates to ensure all loci are amplified during the PCR process, and are plagued with issues around low initial template DNA and inhibitors from co-extracted fecal matter. Resulting DNA fragments require visualization through gel electrophoresis, a multistep process.

New micro-capillary chip-based electrophoresis technologies, such as the *Agilant Bioanalyzer 2100*, have been developed in order to reduce and standardize STR genotyping methods especially in the field of forensics. These small, portable, standard chip-based machines require much lower PCR product input and can be read almost real-time (within minutes) as the amplicons are separated on chip. Electropherogram data and gel like images can be visualized and compared using the *Agilant 2100 Expert* software. Peaks and bands can be coded and genotypes created in much less time with less PCR product required. Reagent and chip products for this equipment are reasonably affordable and commercially available.

Techniques since the initial development of non-invasive sampling have also focussed on targeted capture of desired DNA especially from samples such as feces where large amounts of external DNA may be included. These capture techniques include those of size selection via electrophoreses described by Perry and colleagues (2010), and a method called primer extension capture using biotinylated primers to extend target areas and then capture these synthesized strands using magnetic beads (Briggs, 2012). Most recently a methylation-based enrichment technique was published specifically for use in fecal based genomic studies (Chiou & Bergey, 2018). This technique, based on the *New England Bio-labs NEBnext Microbiome DNA Enrichment Kit*, uses methyl-tagged

magnetic beads that bind selectively to CpG-methylated eukaryotic DNA so that it can then be separated using a magnetic field from the remaining co-extracted microbiome DNA. The resulting host DNA, shown to be enriched in concentration up to ~200 fold in wild baboon feces (Chiou & Bergey, 2018), was then utilized for next-gen sequencing. All these methods help ensure targeted species DNA makes up the bulk of the samples pooled for high throughput sequencing and other genotyping applications.

Review of Relatedness and Pedigree Analysis

Software useful for molecular censusing was reviewed by Vigilant and Guschanski (2009), with specific respect to primate populations, and by Excoffier and Heckel (2006) in a broader context. This software was used for determining allele frequencies, relatedness amongst individuals of a population, and for estimating population size. Programs used to estimate relatedness in the studies of orangutan populations mentioned above included *Kinship 1.3* (formerly *RELATEDNESS*) ((Goodnight & Queller, 1999), as used by (Utami et al., 2002)), *POPASSIGN* ((Goossens et al., 2002), as used by (Goossens et al., 2006), and (Morrogh-Bernard et al., 2011)), *SPAGeDi* (Hardy & Vekemans, 2002), as used by (Morrogh-Bernard et al., 2011)), *POPTOOLS* ((Rohlf & Sokal, 1995), as used by (Morrogh-Bernard et al., 2011), and *COANCESTRY* (Wang, 2011), as used by (Arora et al., 2012).

The *COANCESTRY* software provides estimates using seven different estimators comprising two statistical methods, product moment estimators and maximum-likelihood (ML) estimators. The ML estimator *TRIOML* as compared to the others in this program has since been shown by Wang (2007) to produce the most accurate estimates for large

datasets both SNP and STR. Subsequent publications have shown that in small datasets, where allele frequencies are estimated from the sample for which one is calculating relatedness, that the modified product moment estimators by Wang (2014, 2017) and Ritland (Lynch & Ritland, 1999) give the least biased estimates for high and low related individuals respectively.

Also estimated by the *COANCESTRY* software are Identity by Descent (IBD) coefficients which when combined with relatedness estimates can be used to determine some common relationships assuming no inbreeding in the population. An example of the most common relationships is found in Appendix B as provided by the *COANCESTRY* user guide.

Other various estimators used for pedigree and parentage assignment include *CERVUS* (Marshall TC , Slate, J., Slate, J, Kruuk, 1997) as used by Utami et al. (2002) and Goossens et al. (2006) and *CERVUS* 3.0 (Kalinowski, Taper, & Marshall, 2007), as used by (Arora et al., 2012), and (Banes et al., 2015) and *COLONY2* (Wang & Santure, 2009). These tools use maximum likelihood and exclusion methods for assignment of parentage or paternity, and sibship.

CHAPTER III

STATEMENT OF RESEARCH QUESTION

Despite recent technological advances in genetics and genomics scientists and conservationists still know relatively little about the genetic relationships within wild ape populations, with whom we humans share close ancestry. Orangutans present an extreme example of how difficult genetic sampling can be in the wild. In accordance with regulations surrounding endangered species and in an effort to avoid invasive sampling practices, geneticists have moved towards non-invasive methods by collecting feces, urine, food waste, and other discarded materials (Inoue, Inoue-Murayama, Takenaka, & Nishida, 2007; Rutledge, Holloway, Patterson, & White, 2009). It is now possible to extract viable DNA from these by-products. Primate conservationists have begun to build protocols to investigate populations using these sample types (Goossens et al., 2000; Nsubuga et al., 2004).

Over the past two decades in the hopes of genetically confirming reproductive success and species dispersal patterns, several orangutan populations have been sampled to assess genetic diversity and relatedness (Arora et al., 2012; Goossens et al., 2005; Morrogh-Bernard et al., 2011). Dispersal and reproductive patterns have consequences on species viability in the wild. Knowing these patterns is important in making management decisions for conservation practitioners. Results from studies at several sites on the island of Borneo have varied in overall and sex specific relatedness estimates.

Despite much research and improvement in genotyping practices, microsatellite (or Short Tandem Repeat, STR) analysis, is time consuming, requires multiple replicates

to ensure all loci are amplified during the PCR process, and are plagued with issues around low initial template DNA and inhibitors from co-extracted fecal matter. Resulting DNA fragments require visualization through gel electrophoresis, a multistep process. Once samples are measured through qPCR, this processes requires costly reagents and multiple sets of equipment (Morin et al., 2001).

Recently improvements in next generation sequencing techniques have provided a streamlined and relatively swift genotyping process which is becoming increasingly affordable for often poorly funded primate conservation projects (Vigilant & Guschanski, 2009). Single nucleotide polymorphism (SNPs) loci have been identified and mapped across the human genome, and are being discovered increasingly for model and non-model organisms. The use of commercially designed sequencing kits and microarray genotyping chips across some closely related species has helped the discovery of conserved SNP loci. However, scientists have not yet taken advantage of the relative genetic closeness of orangutans to humans, for which the most commercial products are currently targeted.

In order to streamline the process of non-invasive genotyping to investigate the genetic relatedness of the local wild orangutan population at the Camp Leakey research site, I designed and assessed a new protocol for microarray SNP genotyping of DNA isolated from feces. This paper details a novel protocol which combines fecal DNA extraction with a modified magnetic bead enrichment capture technique, *FecalSeq* (Chiou & Bergey, 2018), and the results of SNP genotyping using human targeted microarray chips. These results are complimented and assessed by comparison to micro-capillary

STR marker analysis, producing two data sets for genetic diversity and relatedness estimate comparisons. This investigation provides a baseline genetic analysis for a population whose behavior and ecology have been studied for almost five decades at the Camp Leakey Study site (Galdikas, 1985b; Galdikas & Ashbury, 2013; Galdikas & Wood, 1990) and offers insight in to the genetic status and health of this important population. These results can then be compared to studies from other orangutan populations.

CHAPTER IV

METHODS USED

Sample Collection

Fecal sampling of wild orangutans at Camp Leakey took place between January and August 2016. Once located and visually identified and confirmed by experienced field assistants as either a known or unknown wild individual, orangutans were followed continuously until defecation was observed. Fecal samples were collected in duplicate and stored as per a two-step method (Nsubuga et al., 2004) where samples (approx. 2-10g) were collected using sterile gloves and a sterile collection spoon to avoid contamination, and initially stored in 30ml of 97% ethanol solution (step one). Ethanol solution was discarded 24 to 36 hours later and ~10g of silica gel beads were placed inside sample container (step two). Samples were then transferred stored in refrigerator at -40C until processing. Samples were exported from Indonesia to Central Washington University under the CITES export permit 01152/IV/SATS-LN/2017.

DNA Extraction

DNA was extracted using the *QIAamp DNA Stool Mini Kit® Qiagen* DNA extraction kit. Initial DNA extraction was conducted by hand and then subsequent extractions were automated using the *QIAcube*™ machine for increased standardization. The standard kit protocol was used except with modification of an extended cell lysis step with an overnight incubation period of 14-18 hours in 23°C heat block after addition of Lysis Buffer and prior to insertion into the *QIAcube*™ machine. An additional extended

incubation hold of 30-120 mins was added before final elution step. For microsatellite (STR) analysis, extracts underwent a double inhibitor cleanse where samples (or existing extracts) were incubated in *Inhibit-Ex* buffer as well as an *Inhibit-Ex* tablet for the 14-18 hours as stated above. Total DNA was then quantified using a *NanoDrop 2000* spectrophotometer reading.

Analysis was conducted on DNA isolated from fecal samples from 32 wild individuals at CL. This included two adult wild males, 14 adult females, two adolescent nulliparous females (with known maternity), eight infant or young juvenile offspring (seven with known maternity among those sampled), four juvenile/adolescent males (one with known maternity, the others unknown but found traveling on their own), and two unknown subadult males. In total 12 of these individuals were known and named, in that they were identified by at least two local field staff and witnessed on multiple occasions in the study area. The others represent either unknown individuals or those whose identity could not be verified.

To gauge initial quantity of orangutan DNA within extractions, quantitative real time PCR was conducted on samples using universal mammalian MYCBP primers (Morin et al., 2001) and a SYBR green Universal Master Mix on the *BioRad iQ5 Optical qPCR* system. Multiple (2-4) DNA extractions were conducted for each individual (except for three individuals only able to be extracted once due to low quantity sample), the initial 69 of which were quantified using real time PCR. Only samples with total orangutan DNA greater than 20ng were used in SNP microarray analysis. Those extracts that were less than 20ng, were pooled for each individual for either enrichment or to be

run directly on the microarray. For those individuals whose samples were pooled for SNP genotyping, further extracts were conducted for STR analysis but were not quantified using qPCR.

Genotyping

For microarray SNP analysis, the *FecalSeq* (Chiou & Bergey, 2018) technique, based on the New England Bio-labs *NEBnext Microbiome DNA Enrichment Kit*, was used on a subset 27 individuals to separate host orangutan DNA from co-extracted microbial DNA. This technique uses methyl-tagged magnetic beads which bind selectively to CpG-methylated eukaryotic DNA which can then be separated from the remaining bacterial sample using a magnetic field. The resulting host enriched DNA was then utilized for microarray SNP analysis. Samples post enrichment were further quantified using the above outline qPCR technique, with the addition of a universal bacterial 16S rRNA primer (Corless et al., 2000), to document observed decrease in bacterial concentration, and thus assess sample enrichment success.

SNP microarray analysis was conducted on 48 samples, 27 of which had undergone the *FecalSeq* enrichment process, one human DNA extract as a positive control, and the rest were un-enriched extraction products including some duplicates of those that underwent enrichment. DNA extracts underwent a quality check and gender confirmation using *Taqman* real time PCR quantification and were run on an *Illumina Infinium Human QC* microarray SNP chip to identify homologous human single nucleotide polymorphism (SNP) loci. Microbeads on the chip hybridize specific known human SNP locations using targeted probes. UV light illuminates colored probes as

specific nucleotides hybridize resulting in light intensity and color data. These data are translated using custom proprietary Illumina software, *GenomeStudio2.0*.

In order to assess the quality of microarray genotypes and resulting relatedness estimates, microsatellite (STR) marker genotyping through targeted amplification and visualization was also performed. Eight STR autosomal markers were selected from those used in several former studies and described by Nietlisbach et al. (2010). These short-repeated sections were amplified using targeted primers and through polymerase chain reaction (PCR) thermocycling. PCR conditions followed Arrora et al. (2010) and Nietlisbach et al. (2010) using *SigmaAldrich* Redtaq mastermix. A subset of samples underwent multiple amplifications for each of the targeted regions. Error rates were calculated from this subset. PCR amplifications were electrophoresed on micro-capillary DNA1000 chips on the *Agilant Bioanalyzer 2100* machine. Resulting electropherogram data were visualized and analyzed using the *Agilant 2100 Expert* software. Fragment variant lengths for eight autosomal tetra-nucleotide loci, five *Pongo* specific (Nietlisbach et al., 2010) and three human specific (Goossens et al., 2005), were coded visually using the gel-like densitometry plot data comparison view. High quality *Pongo* DNA, and human DNA, as positive control was amplified alongside samples to confirm band sizes and intensity. Bands were identified as separate loci when repeatedly amplified or observed (more than once) and when at least four base pairs apart from bands above or below (once corrections between chip runs was done). In total 33 individuals were attempted to be genotyped using this method (22 of the same individuals with SNP genotypes).

Statistical Analysis

Allele frequencies and distinct individual identities were confirmed using *Cervus* (Kalinowski et al., 2007) software for both genotyping techniques. Resulting genotypes from both methods were analyzed statistically to determine accordance with Hardy-Weinberg equilibrium and to assess linkage disequilibrium using online software *GenePop* (Rousset, 2008).

In order to assess quality of SNP genotypes and to compare to the STR dataset, resulting pairwise relatedness values were calculated using the triadic likelihood estimator, *TrioML* (Wang, 2007), and two moment estimators used in past studies (the coefficient of *Wang* (Wang, 2002), and Queller & Goodnight's (Queller & Goodnight, 1989) pairwise relatedness estimator, r_{xy},) for the 22 individuals with both SNP and STR genotypes using the *Colony* (Jones & Wang, 2010) and *COANCESTRY* (Wang, 2011) software. Relatedness values for the subset of 22 individuals with both datasets were calculated using allele frequencies from each entire dataset. However, due to low sample size and the fact that our dataset includes known relatives, one would expect some pairwise relatedness estimates will be underestimates of the true relatedness (Wang, 2007), and overall relatedness estimates using moment estimators will be small and close to zero (Wang, 2014, 2017). Overall group and pairwise relatedness values for each estimator were tested for correlation through paired and unpaired *t*-tests and mantel matrix correlation tests in the R (R Core Team, 2014) statistical environment. In order to test for significant differences in relatedness between male and female groups observed

averages were tested through 1000 bootstrap re-samplings as computed using the *COANCESTRY* Software.

CHAPTER V

RESULTS

Genotyping

Average total orangutan DNA proportion per 100ul extract was 3.31% (0-1729.8ng). After undergoing the *FecalSeq* magnetic bead enrichment process to separate endogenous orangutan DNA from that of contaminating microbiome found in feces, bacterial DNA quantities in extracts decreased from 30-500 fold (median 140 fold decrease). While *Pongo* DNA quantities decreased as well, the average was by about half (median = 0.57, range 0.12-0.958). The 27 post-enrichment samples with highest orangutan concentrations, were then chosen to be run on the *Illumina* SNP microarray.

Initial quality control *Taqman* qPCR testing confirmed the known sexes of all 48 samples run on the microarray. Total DNA concentrations for enriched samples ranged from 0.846ng/ul to 7.39ng/ul and total DNA concentrations for un-enriched samples were 14.3ng/ul to 477ng/ul.

Microarray data from un-enriched samples was poor, with light intensity (Norm R) and color (Norm Theta) results scattered across the spectrum and not clustering as should be expected in comparison to human DNA results. Kinship analysis conducted using genotypes produces with these data did not confirm any known relatedness or duplicate individuals among the data set. This pattern was also seen with enriched samples on the vast majority of the 15, 949 known SNP loci probed for on the microarray. However, through visual inspection of SNP loci with high assignment among the 27 enriched samples, 125 loci (0.78% of the total on the microarray) were identified

as presenting similar clustering patterns to those of human data (examples of this at four SNP loci discovered are shown in Figure 3) and having high enough minor allele frequencies (125 > 0.018, 104 > 0.1, 61 > 0.2) to provide useful variation among individuals to ensure sufficient informativity for relatedness analysis (Krawczak, 1999; Ross et al., 2014).

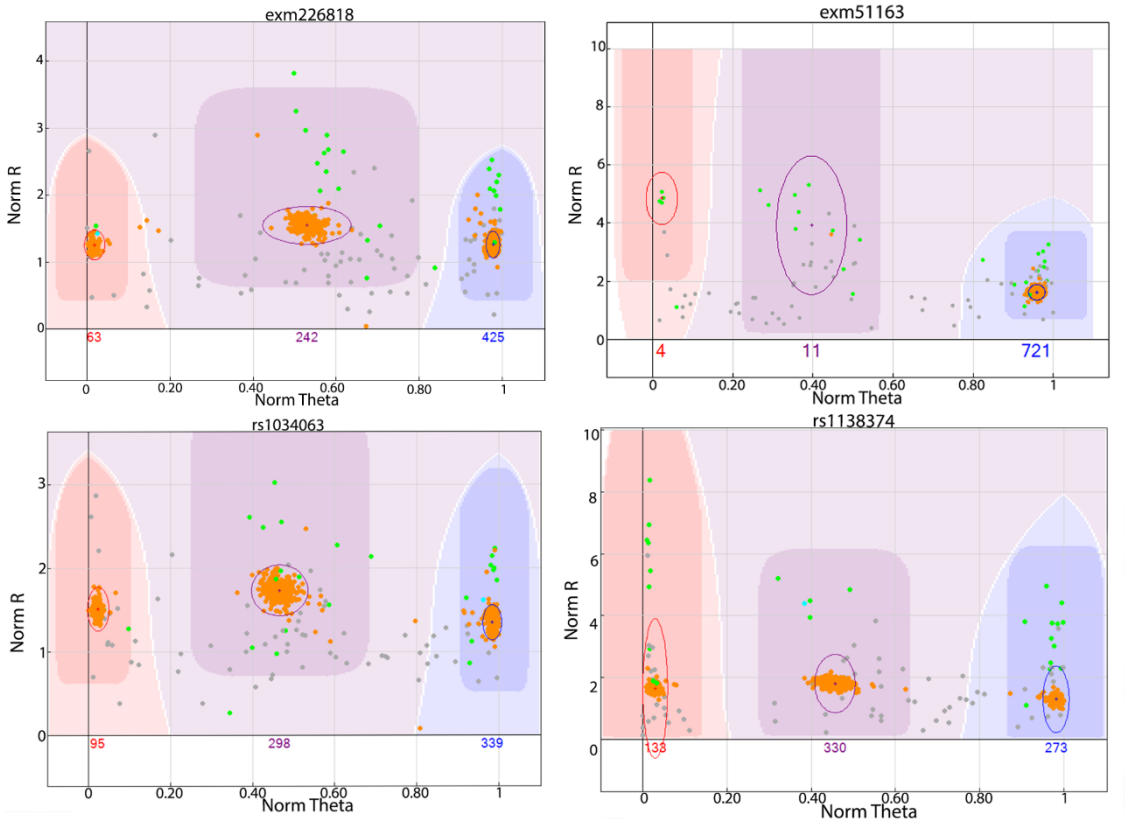


Figure 3 – Example of *GenomeStudio* SNP graph clustering results for four SNP loci. Orange dots are sample human data, green dots are the 27 enriched samples and dark grey dots are unenriched samples. Norm Theta represents the light color read and Norm R is the light intensity. Circles and dark colouration are areas where allele assignments or calls are made. The two exterior red and blue circles represent homozygous calls and center purple is homozygous call.

These 125 homologous bi-allelic SNP loci (Appendix A) were used to create genotypes for 27 individuals at a minimum of 65 (52% of the total 125) loci. Two pairs of

individuals in this SNP dataset were identified as the same individual through identity analysis, matching at 93 and 81 loci in common and mismatching at 0. These two individuals represent suspected duplicates whose identity were not able to be confirmed by multiple field staff. Mean proportion of SNP loci typed was 0.72 and combined non-exclusion probability of identity was $3.07e^{-37}$.

STR genotyping was conducted on 355 PCR amplifications for 33 individuals. Samples from four individuals repeatedly did not amplify so genotypes were not produced for them using this method. Thus, I produced successful genotypes for 29 individuals. Of the successful amplifications 39% were individuals genotyped in duplicate, 14% were individuals genotyped in triplicate, and 7% were individuals genotyped more than three times. Allelic drop out error rates were calculated from this multiple genotyping to be 0.055. Mean proportion of STR loci typed was 0.88, and combined non-exclusion probability of identity was $6.6e^{-7}$.

Genetic Diversity

Overall averaged observed heterozygosity (H_o) for the 125 SNP markers was 0.36, SD = 0.19, average expected heterozygosity (H_e) was 0.34, SD = 0.14 with average polymorphic information content (PIC) of 0.27, SD = 0.092. Average inbreeding co-efficient (F_{IS}) was -0.037, SD = 0.32. Of the 125 individual SNPs, 10 loci differed significantly from H-W equilibrium (α = 0.05), with both lower and higher than expected heterozygosity and corresponding high and low F_{IS} values. These values can be found in Appendix A.

Allele frequency calculations and tests for heterozygosity and deviation from Hardy-Weinberg (H-W) equilibrium for the eight microsatellite STR autosomal markers found average observed heterozygosity (H_o) to be 0.65 with an expected (H_e) value of 0.66 with average *PIC* of 0.591, with one marker D13S765 showing significant deviation from H-W equilibrium ($p = 0.001$). Average F_{IS} inbreeding coefficient for the eight autosomal markers was 0.003 ($SD = 0.14$). These values can be found in Table 1.

Table 1

Autosomal Microsatellite (STR) Markers

Name	Marker Type	A	*n*	H_O	H_E	PIC	P-val	S.E.	Fis W&C
O4_6	*Pongo*	3	25	0.60	0.49	0.42	0.652	0.004	-0.233
O4_B5	*Pongo*	5	26	0.81	0.78	0.72	0.194	0.007	-0.042
O4_A1	*Pongo*	4	29	0.69	0.64	0.55	0.514	0.008	-0.082
O4_B20	*Pongo*	2	25	0.40	0.44	0.34	0.663	0.002	0.101
O4_CHR5	*Pongo*	6	26	0.77	0.69	0.64	0.850	0.007	-0.112
D13S765	Human	6	27	0.59	0.75	0.69	0.001*	0.001	0.213
D6S501	Human	7	25	0.72	0.74	0.69	0.439	0.016	0.020
D13S321	Human	4	21	0.62	0.73	0.66	0.580	0.006	0.160
	Mean	5	26	0.65	0.66	0.59			0.003
	SD	2	2	0.12	0.12	0.13			0.140

Note. Allelic diversity A, number of genotyped individuals, observed heterozygosity H_O, expected heterozygosity H_E, polymorphic information content PIC, average F_{IS} (Weir & Cockerham 1984) and p-value plus standard error S.E. of probability test for deviation from Hardy-Weinberg (HW) equilibrium
*indicates a statistical departure from HW equilibrium, $p < 0.05$

Relatedness

Initial *Cervus* identity analysis of the SNP dataset confirmed the shared identity of two pairs of individuals each sampled twice within the SNP samples and verified all other individuals as unique. These duplicate genotypes were removed and not used for further analysis. Identity analysis of the STR data confirmed all the individuals as unique. This

resulted in 25 individuals with SNP data, 29 with STR data, and 32 unique individuals in total.

Pairwise relatedness estimates produced by *Colony* and parentage analysis using *Cervus* confirmed four of five known mother offspring pairs in the SNP Data (the one non-confirmed pair gave a maximum r estimate of 0.25), and confirmed all seven within the STR data. Average mother-offspring relatedness across estimators was 0.49 (SNP data) and 0.43 (STR data).

Pairwise relatedness values for the 22 individuals with both SNP and STR genotypes were calculated in the program *COANCESTRY* (Wang, 2011) and compared using a Mantel correlation test in the *ade4* (Dray & Dufour, 2007) package in the statistical program R (R Core Team, 2014). Relatedness values between the two data sets were highly positively correlated using three estimators *TrioML* ($r = 0.81$, $p < 0.001$), *Wang* ($r = 0.34$, $p = 0.018$), and the *Queller&Goodnight* ($r = 0.42$, $p = 0.0032$). The two datasets were then combined for all 32 individuals and a new combo dataset was produced.

Overall relatedness within the population and pairwise relatedness within adult female and adult male groups was calculated for each data set using three estimators and the combined dataset and is presented in Table 2.

Overall relatedness in all 32 individuals from the Combo (SNP & STR) dataset using the *TrioML* estimator was 0.082 (*var* = 0.021).

Results of a paired *t*-test to compare the overall *TrioML* averages of the SNP ($r = 0.096$, *var* = 0.023) and the Combo dataset found no significant difference with

conditions ($t(612) = -1.31$, $p = 0.19$). A test to compare the STR mean $r = 0.082$ (*var* = 0.020) and the Combo dataset also showed no significant difference ($t(869) = 0.022$, $p = 0.98$).

Table 2

Mean Relatedness Values for Each Dataset and within Adult Females and Adult Males

Dataset	# of loci	*n*	*TrioML* *(r)*	*var*	*Wang* *(r)*	*var*	*Q&G* *(r)*	*var*
Over All Relatedness								
SNP	125	25	0.096	0.023	-0.014	0.213	-0.062	0.137
STR	8	29	0.082	0.020	0.008	0.097	-0.041	0.072
Combo	**133**	**32**	**0.082**	**0.021**	**0.015**	**0.096**	**-0.071**	**0.101**
Adult Females								
SNP	125	10	0.079	0.014	0.189[a]	0.043	0.054[b]	0.057
STR	8	14	0.074	0.019	-0.031	0.109	-0.065	0.073
Combo	**133**	**15**	**0.074**	**0.018**	**0.029**	**0.086**	**-0.047**	**0.077**
Adult & Sub Adult Males								
SNP	125	4	0.153	0.038	-0.487[a]	0.302	-0.233[b]	0.166
STR	8	3	0.018	0.001	-0.319	0.053	-0.172	0.016
Combo	**133**	**4**	**0.208**	**0.057**	**-0.191**	**0.200**	**-0.070**	**0.148**

Note. Each dataset, number of loci used and how many individuals were included *n*, then each of the three relatedness estimators TrioML, Wang, and Queller & Goodnight Q&G, followed by the variance for each estimate
[a, b] indicates significant difference through bootstrapping between adult male and adult female average *r*

TrioML relatedness within adult females ($n = 15$) using the combo dataset was 0.074 (*var* = 0.018), and relatedness between the two adult fully flanged males ($n = 2$)

was 0. Estimates for *TrioML* relatedness within all potential post dispersal (PPD) males in the combo data set including subadult males (total $n = 4$) was calculated using the combo dataset to be 0.21 (*var* = 0.057). Average pairwise relatedness between adult females in the combo dataset and all PPD males was 0.053 (*var* = 0.011) and between adult females and just the two fully flanged males was 0.055 (*var* = 0.0079). Bootstrap comparisons found no significant difference between average *TrioML* relatedness within adult females or all PPD adult males in the combo dataset ($n = 4$). When just the one known local and one unknown encountered flanged adult males were included there is a significant difference since their pairwise r was 0 and $n = 2$.

Relatedness calculated using two moment estimators (*Wang* and *Q&G*) produced lower and very close to 0 overall relatedness values (with higher variance) using the combo dataset, *Wang* = 0.015 (*var* = 0.096) and *Q&G*= -0.071 (*var* = 0.10). This trend was consistent for all three datasets and is expected since allele frequencies are calculated directly from genotypes sampled, and these numbers represent an average of correlations relative to the population average (Wang, 2017). Bootstrapping comparisons did find a significant difference in the SNP dataset between average *Wang* and *Q&G* relatedness estimates within adult females and adult males with males having lower average relatedness than females relative to the population mean. Bootstrapping comparisons using the STR and combo dataset with all three estimators found no significant difference between males and female average *r*.

Male average relatedness using the *Wang* estimator is lower than females for all three datasets, however this is in reference to the population average rather than absolute

0 relatedness. Using a combination of the modified *Wang* (Wang, 2014) and *TrioML* estimators as well as estimated IDB coefficients of relatedness in combination, all males are determined to be related to two or more individuals in the population at the level of first cousins or higher. A mother was assigned to two unknown males (one subadult and one juvenile) within the study area. The two adult fully flanged males are estimated to have first cousin and higher relatedness with both females and other males within the sample. Each individual analyzed and their first and second degree relatedness is presented in Appendix C.

Average adult female relatedness within the sample is approximately between the levels of half cousin (or first cousin once removed) and first cousins (0.0625-0.125), with all adult and juvenile females having a close (at least half sib, aunt/niece, or first cousin) relative within the study area. One unknown adult female – adult daughter pair observed and sampled within close proximity was identified. This is the same mother of the two unknown males. Two other unknown adult females were identified as full siblings. Using the Full Likelihood method in the *Colony* software to estimate sibship, all females except the two full female siblings cluster into one large half sibship group. Parentage testing was not able to assign paternity to either of the fully flanged adult males or two sub adult males sampled for any of the infant or juvenile individuals tested. Full likelihood parentage testing including using all individuals sampled and known relatedness identified a minimum of three unsampled fathers for the eight infants individuals (age range ~ six years or under) in our dataset (Table 3).

Table 3

Predicted Paternity and Maternity for Eight Infants Sampled Using Colony Likelihood Parentage Assignment

Offspring	Father	Mother
IF_ABD06	1	AF_BD06
IF_AB10	1	#1
IM_ABD07	2	AF_BD07
IM_ABD05	1	#1
IF_ABD01	1	AF_BD01
IF_ABDM12	2	AF_BDM1
IF_ABDR1	3	AF_BDR1
IF_ABDM21	3	#2

CHAPTER VI

DISCUSSION

This is the first known study to combine the use of non-invasive fecal DNA sampling and extraction, methyl based enrichment *FecalSeq*^{TM} (Chiou & Bergey, 2018), and human targeted *Illumina Infinium* SNP microarray genotyping technology for population monitoring of an endangered great ape. The use of fecal DNA sampling has become standard for cryptic and sensitive endangered species. However, difficulties arising from low endogenous DNA quantity within samples remain pervasive. Combining extraction with the *FecalSeq* methyl based magnetic bead capture enrichment technique increased the concentration of orangutan DNA (vs bacterial DNA etc). Of the samples that were tested on the *Illumina* microarray chip only those that underwent the *FecalSeq* enrichment process produced successful genotypes. Initial quantification of DNA samples indicated averages of approximately 3% endogenous DNA in our extractions which is consistent with published numbers (Chiou & Bergey, 2018; Perry et al., 2010). It is likely that this large amount of non-specific exogenous DNA in these samples overwhelmed the small quantity of orangutan DNA in pure un-enriched samples run. By decreasing these non-specific DNA concentrations through enrichment, the orangutan DNA was successfully amplified and bound successfully to the homologous human based SNP tagged beads on the microarray. Despite low initial DNA quantities our results corroborate evidence that reliable results can be produced from "near nanogram" levels (Okitsu, Berg, Lieber, & Hsieh, 2013) on *Illumina Infinium* SNP microarrays.

Cross species microarray analysis has shown to be possible within species as genetically distant as oryx (*Oryx* spp.) and modern domesticated bovine (*Bos taurus*) (divergent at least 23 million years) (Ogden et al., 2012) and Antarctic fur seal (*Arctocephalus gazella*) and domestic dogs (*Canis lupus familiaris*) who diverged approximately 44 million years ago (Hoffman et al., 2013). These studies were able to respectively identify 185 of 54,001 (0.34%) and 173 of 173,662 (0.01%) homologous polymorphic loci in common. A further study of wild thin horn (*Ovis dalli*) and bighorn (*Ovis canadensis*) sheep genotyped on a chip designed for commercial domestic sheep (*Ovis aries*) (divergent relatively more recently than the other examples at approximately 3 million years ago) identified 868 of 49,034 loci (1.7%) to be polymorphic and in common (Miller et al., 2011). Early investigations into ancestral alleles among humans and apes also proved ape DNA can be genotyped using human microarrays (Hacia et al., 1999), and confirmed three and two common homologous polymorphic loci of 397 (0.75% and 0.5%) between bonobos and gorillas and humans respectively. Our results producing 125 common polymorphic loci for *P. pygmaeus* of the 15,949 human loci (0.78%) probed for, fit percentages found by these past cross-species studies. These positive results suggest further attempts at cross species genotyping of *Pongo* DNA on much larger human mapped chips could identify many more common polymorphic SNP loci.

In order to verify the quality of the SNP genotypes produced, microsatellite STR genotyping was conducted for comparison using the *Agilent Bioanalyser 2100* through micro-capillary based electrophoretic chips. Despite the fact that resolution of tetrameric

STR loci can be problematic on this platform (Fraige, Travensolo, & Carrilho, 2013) the use of the 2100 *Expert* software electropherogram overlay and comparison context allowed for calibration of inter-gel and inter-well differences. Using repeated amplifications, visual inspection, and known fragment lengths published by previous authors (Nietlisbach et al., 2010; Utami et al., 2002) variants were identifiable within known ranges and genotypes were successfully assigned.

Allele frequencies calculations revealed one marker for the STR dataset and ten SNP loci showing significant deviation from H-W equilibrium. Average F_{IS} values across all loci was negative and very close to 0. The one STR and six SNP loci with high departure from HW equilibrium and high average individual inbreeding coefficients for all those sampled could be signs of loss of heterozygosity through inbreeding within the population. This could also be due to genotyping error as error rates for the SNP dataset is unknown (although can be assumed to be low because two duplicate individuals were correctly identified mismatching at 0 loci) and for the STR dataset is relatively high.

Several comparisons of pairwise estimators of relatedness have detailed differences between various statistical methods categorized as either moment and likelihood methods. Most recently Wang (2007) compared the *TrioML* maximum likelihood estimator to several moment estimators including his newest moment estimator (referred to as *Wang* in table) as well as a more commonly used moment estimator by Queller and Goodnight (*Q&G*). The three relatedness estimators clearly produce quantitatively different population averages in our study when calculated using the SNP data. Estimates using all three estimators were not significantly different between datasets

and mantel correlation test of matrices of pairwise relatedness estimates for the 22 same individuals showed statistically significant strong positive correlations between r values between any two of the estimators compared. Thus, similar pairwise relatedness estimates have been successfully achieved with both genotyping methods.

Our overall population wide *TrioML* relatedness value of 0.082 using the combined Combo dataset is similar to population values published at other sites in Kalimantan (Arora et al., 2012; Goossens et al., 2006; Morrogh-Bernard et al., 2011) as are our relatively low and slightly negative numbers and published from Sumatra (Utami et al., 2002) using the Queller & Goodnight method. Using each of the three estimators and the combo dataset there was no significant difference in relatedness within males versus females within the CL Study Site. This is similar to those results found from northern Borneo and Sumatra. However, the results are opposite to those found at the two other sites within Central Kalimantan. It is worth noting that there were far fewer males than females encountered and sampled for this study, and analysis indicates that not all fathers of offspring in the area were sampled. However, all males sampled had some male and female relatives within the study site. This suggests that the males that were encountered and sampled are local individuals, either pre-dispersal in the case of sub-adults or adults who have not dispersed. Although males were less related to females in the area, the lack of completely unrelated males in the sample indicates that new males do not appear to be moving into the study area or were not present during sampling. This prediction is corroborated by Galdikas (personal communication) from her behavioral observations over the past 50 years.

High female relatedness does support the behavioral records at CL which indicate female matrilines living within overlapping ranges within the study site. The fact that the majority of the individuals sampled cluster into one group (with the rest in one other smaller group) suggests that two large family groups makes up the majority of the individuals within the area. Without more potential fathers sampled or more detailed mitochondrial data it is extremely difficult to create a concise pedigree for the all of the individuals. I confirmed that two females who nested and fed within the same tree before sampling were a mother-daughter pair and two individuals sampled within the northern side of the study site were a full sibling pair. The high level of half siblings within the population is not surprising since orangutans are known to exhibit both male and female polygyny.

The fact that paternities were not assigned for any individuals is a clear indication that not all potential fathers (sexually mature males) in the area were sampled. It has been shown at this site that the dominant male has fathered the majority of offspring within the local ex-captive community, and there were many other males within this community not included in our study. However, this parentage analysis identified four different fathers of eight of the youngest offspring in our sample, each fathering more than one infant. It is clear that there is not just one dominant male fathering all of the wild offspring at CL. Since I wasn't able to identify these fathers I don't know if these males were flanged or subadult males.

The high average relatedness in our population is unusual for large bodied mammals. For example, in a large multi-generational population of red deer (*Cervus*

elaphus), *r* was estimated to be 0.00687 (Stopher et al., 2012). This number was high in comparison to multiple scenarios modeled by researchers. These scenarios include one of complete random mating within the population, which was estimated to have produced an average *r* of 0.00174 (*SD* = 0.00005). Their observed six fold increase in relatedness from randomness within the population was documented along with two behavioral phenomenon, "female mate fidelity", and "intralineage polygyny". Researchers discovered that females re-mated with the same male in successive mating seasons, and females from the same matriline mated with the same male more often than would be expected under random mating conditions. This resulted in what they called "deep inbreeding" (inbreeding between distant relatives) effects raising average relatedness and inbreeding coefficients.

Our data from the CL study site show that both males and females appear to have high relatedness indicating both sexes stay within their natal area (with more females than males in our sample). One possible case of female mate fidelity was documented within the non-native community in a recent genetic study (Banes et al., 2015) at CL. However, in that study as in our own not all possible fathers were sampled. It is possible that the dominant male was not the father of all the offspring assigned to him because it is probable that he had mature sons in the area who were not sampled (Galdikas unpublished data). Despite the orangutans' long interbirth intervals, it is possible that female mate fidelity could also be occurring within the wild local population at CL resulting in high relatedness and high inbreeding (*F* and F_{IS}) estimates. However, this is not likely given dispersal and ranging patterns of wild adult males. High and unusual

relatedness patterns within the adult females may also be due to females being related both maternally and paternally as in the case of double first cousins or even (paternal) sister –(maternal) cousins. This could easily occur with philopatric groups of females (sisters and aunts and nieces) all mating with a single male in one generation. If such events occur naturally, population isolation and habitat fragmentation causing restriction of dispersal may also be adding to the frequency of these deep inbreeding events.

In comparison to other great ape populations, relatedness calculated for 108 individuals across three chimpanzee communities (known to exhibit female-biased dispersal) found similar relatedness between males and females in all communities with females showing relatively low mean *r* values (*Q&G* = 0.0153 for females across the three communities, and 0.068 for males) (Vigilant, Hofreiter, Siedel, & Boesch, 2002). In one community male relatedness was high, *r* =0.147, close to the male and overall relatedness within the wild orangutan population in the CL study site. This may indicate that further sampling is needed on an ongoing basis both within and outside/adjacent to the core CL population in order to assess the effects of sampling regime on our results and to ensure sampling is capturing all potential fathers/males for the community.

This study shows that new genotyping technologies provide opportunities for understanding ape populations and providing critical genetic data to support conservation efforts. Orangutans possess several recognized characteristics that already put them at a higher risk of extinction than other mammals. The combination of being a large bodied, long lived primate, with the longest inter birth interval recorded in primates (Galdikas & Wood, 1990) coupled with rapid habitat destruction and fragmentation present

tremendous challenges for population survival. These challenges may have already resulted in massive decreases in previous levels of male dispersal. However, our study has demonstrated that there are wild orangutan population clusters with very high relatedness indicative of “deep” inbreeding. Along with female mate fidelity and intralineage polygyny it is probable that there are additional behavioral phenomena contributing to deep inbreeding in orangutans. More attention should be paid to genetic characteristics of wild populations in the course of orangutan conservation.

CHAPTER VII

CONCLUSIONS

The results from this study show that orangutan DNA collected from feces can provide useful SNP genotypes when run on a human targeted micro-array. The small number of samples and number of SNPs that produced results are consistent with past research. Endogenous fecal DNA is often co-extracted with high concentrations of bacterial DNA and inhibitors, thus the *Fecalseq* enrichment technique was critical to capture endogenous orangutan DNA and separate contaminators before conducting SNP analysis. Additionally, research shows that despite the overall genetic similarity between humans and orangutans, the number of polymorphic SNP sites the same between species is a percentage of just below 1%. A study conducted in 1999 comparing results of human and ape DNA (chimpanzee, bonobo, and gorilla) on a microarray of 397 known human polymorphic SNPs also found between 0-0.75% homology (Hacia et al., 1999). The fact I was successful at using this technique provides an exciting new avenue for great ape researchers for discovering SNP loci and genotyping from non-invasive fecal samples from the wild.

High female relatedness found in this study supports the behavioral records at CL which indicate female matrilines living within overlapping ranges within the study site. The fact that the majority of the individuals sampled cluster into two groups suggest two large family groups makes up the majority of the individuals within the area. Without more potential fathers sampled or more detailed mitochondrial data, it is difficult to create a concise pedigree for the all of the individuals. I confirmed that two females who

nested and fed within the same tree before sampling were a mother-daughter pair and two individuals sampled within the northern side of the site were a full sibling pair. The high level of half siblings within the population is not surprising since orangutans are known to exhibit both male and female polygyny. With further sampling and repeat genotyping including sex linked loci however, it may be possible to create a complete pedigree for the wild CL population.

It is clear that many male individuals from the wild local CL study area have yet to be sampled. It is critical that sample collection and analysis be continued by OFI researchers and assistants to expand upon findings from this research. Additionally, there is much sampling and analysis to be conducted on the ex-captive community that lives within the CL area and whom are very possibly interbreeding with the wild individuals. This ongoing research is extremely important for population monitoring into the future and sampling can be expanded to other OFI and Tanjung Puting National Park camps, rehabilitation sites and posts around the park. The more individuals sampled, the better the estimate of underlying allele frequencies, and thus more accurate pairwise and population wide estimates can be made.

Understanding the genetic diversity, structure, and relatedness among CL orangutans will allow more informed management decisions to be made about protected areas, future reintroductions and can be used in population viability analyses required by the Species Survival Plan.

As a long standing and influential partner organization to the Indonesian government, research and data produced by OFI including orangutan rehabilitation data,

local mapping, and the wild orangutan research data can be highly valuable to the Indonesian government and used in decision making and policy formation. The genetic information produced from this study of the CL orangutans is the first step in an exciting new avenue of investigation and source of information for TNTP, OFI, and all stakeholders involved in management of Tanjung Puting. Additionally, any expansion of our knowledge of any of the remaining wild orangutan populations are critical for our overall understanding of the species as a whole and the likelihood of their survival long term. In addition to its significance for protecting the CL orangutans, the genetic information gathered in this study can be used in population viability analysis, species survival plans and other species-wide research endeavours.

REFERENCES

Arora, N., Nater, A., Schaik, C. P. van, Willems, E. P., van Noordwijk, M. A., Goossens, B., … Goodman, M. (2010). Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21376–21381. https://doi.org/10.1073/pnas

Arora, N., van Noordwijk, M. A., Ackermann, C., Willems, E. P., Nater, A., Greminger, M., … Krützen, M. (2012). Parentage-based pedigree reconstruction reveals female matrilineal clusters and male-biased dispersal in nongregarious Asian great apes, the Bornean orang-utans (*Pongo pygmaeus*). *Molecular Ecology*, *21*(13), 3352–3362.

Atmoko Utami, S. . S., Singleton, I., Noordwijk, M. A. Van, Schaik, C. P. Van, & Setia, T. M. (2009). Male – male relationships in orangutans. In S. A. Wich, S. S. Utami-Atmoko, T. M. Setia, & C. P. van Schaik (Eds.), *Orangutans: Geographic variation in behavioral ecology and conservation* (pp. 225–234). Oxford Universtiy Press USA - OSO.

Banes, G. L., Galdikas, B. M. F., & Vigilant, L. (2015). Male orang-utan bimaturism and reproductive success at Camp Leakey in Tanjung Puting National Park, Indonesia. *Behavioral Ecology and Sociobiology*, *69*(11), 1785–1794. https://doi.org/10.1007/s00265-015-1991-0

Briggs, A. W. (2012). Rapid retrieval of DNA target sequences by primer extension capture. *Methods in Molecular Biology*, *772*, 145–154. https://doi.org/10.1007/978-1-61779-228-1_8

Chiou, K. L., & Bergey, C. M. (2018). Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. *Scientific Reports*, *8*(1), 1975. https://doi.org/10.1038/s41598-018-20427-9

Corless, C. E., Guiver, M., Borrow, R., Edwards-Jones, V., Kaczmarski, E. B., & Fox, A. J. (2000). Contamination and sensitivity issues with a real-time universal 16s rRNA PCR. *Journal of Clinical Microbiology*, *38*(5), 1747–1752.

Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04

Excoffier, L., & Heckel, G. (2006). Computer programs for population genetics data analysis: A survival guide. *Nature Reviews Genetics*, *7*(10), 745–758. https://doi.org/10.1038/nrg1904

Fraige, K., Travensolo, R. F., & Carrilho, E. (2013). Analysis of seven STR human loci for paternity testing by microchip electrophoresis. *Brazilian Archives of Biology and Technology*, *56*(2), 213–221. https://doi.org/10.1590/S1516-89132013000200006

Galdikas, B. M. F. (1979). Orangutan adaptation at Tanjung Puting Reserve: Mating and Ecology. In D. A. Hamburg & E. R. McCown (Eds.), *The great apes : Perspectives on human evolution Volume V* (pp. 195–233). Menlo Park, Calif.: The Benjamin/Cummings Publishing Company.

Galdikas, B. M. F. (1982a). Orang-utan tool-use at Tanjung Puting Reserve, Central Indonesian Borneo (Kalimantan Tengah). *Journal of Human Evolution*, *11*(1),

19,IN1,25-24,IN1,33. https://doi.org/10.1016/S0047-2484(82)80028-6

Galdikas, B. M. F. (1982b). Wild orangutan birth at Tanjung Puting Reserve. *Primates*, *23*(4), 500–510. https://doi.org/10.1007/BF02373961

Galdikas, B. M. F. (1983). The orangutan long call and snag crashing at Tanjung Puting Reserve. *Primates*, *24*(3), 371–384. https://doi.org/10.1007/BF02381982

Galdikas, B. M. F. (1985a). Adult male sociality and reproductive tactics among orangutans at Tanjung Puting. *Folia Primatologica*, *45*(1), 9–24. https://doi.org/10.1159/000156188

Galdikas, B. M. F. (1985b). Orangutan sociality at Tanjung Puting. *American Journal of Primatology*, *8*, 87–99.

Galdikas, B. M. F. (1985c). Subadult male orangutan sociality and reproductive behavior at Tanjung Puting. *American Journal of Primatology*, *8*, 87–99.

Galdikas, B. M. F. (1988). Orangutan diet, range, and activity at Tanjung Puting, Central Borneo. *International Journal of Primatology*, *9*(1), 1–35. https://doi.org/10.1007/BF02740195

Galdikas, B. M. F., & Ashbury, A. (2013). Reproductive parameters of female orangutans (*Pongo pygmaeus wurmbii*) 1971-2011, a 40-year study at Tanjung Puting National Park, Central Kalimantan, Indonesia. *Primates*, *54*(1), 61–72. https://doi.org/10.1007/s10329-012-0331-5

Galdikas, B. M. F., & Vasey, P. (1992). Why are orangutans so smart?: Ecological and social hypothesis. In F. D. Burton (Ed.), *Social processes and mental abilities in*

non-human primates, evidences from longitudinal field studies (pp. 183–224). Lewiston/Queenston/Lampeter: The Edwin Mellen Press.

Galdikas, B. M. F., & Wood, J. W. (1990). Birth spacing patterns in humans and apes. *American Journal of Physical Anthropology*, *83*, 185–191.

Goodnight, K. F., & Queller, D. C. (1999). Computer software for performing likelihood tests. *Molecular Ecology*, *8*, 1231–1234.

Goossens, B., Chikhi, L., Jalil, M. F., Ancrenaz, M., Lackman-Ancrenaz, I., Mohamed, M., … Bruford, M. W. (2005). Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*) populations from Sabah, Malaysia. *Molecular Ecology*, *14*(2), 441–456.

Goossens, B., Chikhi, L., Utami, S. S., de Ruiter, J., & Bruford, M. W. (2000). A multi-samples , multi-extracts approach for microsatellite analysis of faecal samples in an arboreal ape. *Conservation Genetics*, *1*, 157–162. https://doi.org/10.1023/a:1026535006318

Goossens, B., Funk, S. M., Vidal, C., Latour, S., Jamart, A., Ancrenaz, M., … Bruford, M. W. (2002). Measuring genetic diversity in translocation programmes: Principles and application to a chimpanzee release project. *Animal Conservation*, *5*(3), 225–236. https://doi.org/10.1017/S1367943002002275

Goossens, B., Setchell, J. M., James, S. S., Funk, S. M., Chikhi, L., Abulani, A., … Bruford, M. W. (2006). Philopatry and reproductive success in Bornean orang-utans (*Pongo pygmaeus*). *Molecular Ecology*, *15*(9), 2577–2588.

https://doi.org/10.1111/j.1365-294X.2006.02952.x

Hacia, J. G., Fan, J. B., Ryder, O., Jin, L., Edgemon, K., Ghandour, G., … Collins, F. S. (1999). Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nature Genetics*, *22*(2), 164–167. https://doi.org/10.1038/9674

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*, 618–620.

Harrison, M. E., & Chivers, D. J. (2007). The orang-utan mating system and the unflanged male: A product of increased food stress during the late Miocene and Pliocene? *Journal of Human Evolution*, *52*(3), 275–293. https://doi.org/10.1016/j.jhevol.2006.09.005

Hoffman, J. I., Thorne, M. A. S., McEwing, R., Forcada, J., & Ogden, R. (2013). Cross-amplification and validation of SNPs conserved over 44 million years between seals and dogs. *PLoS ONE*, *8*(7), 1–10. https://doi.org/10.1371/journal.pone.0068365

Inoue, E., Inoue-Murayama, M., Takenaka, O., & Nishida, T. (2007). Wild chimpanzee infant urine and saliva sampled noninvasively usable for DNA analyses. *Primates*, *48*(2), 156–159. https://doi.org/10.1007/s10329-006-0017-y

Jones, O. R., & Wang, J. (2010). COLONY: A program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, *10*(3), 551–555. https://doi.org/10.1111/j.1755-0998.2009.02787.x

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*(5), 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Krawczak, M. (1999). Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis*, *20*(8), 1676–1681.

Leigh, S. R., & Shea, B. T. (1995). Ontogeny and the evolution of adult body size dimorphism in apes. *American Journal of Primatology*, *36*, 37–60.

Lynch, M., & Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers. *Tbe Genetics Society of America*, *152*, 1753–1766.

Marshall TC , Slate, J., Slate, J, Kruuk, L. E. B. and P. J. (1997). Statistical confidence for likelihood-based paternity. *Molecular Ecology*, (1998), 639–655.

Miller, J. M., Poissant, J., Kijas, J. W., & Coltman, D. W. (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources*, *11*(2), 314–322.

Morin, P. A., Chambers, K. E., Boesch, C., & Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, *10*(7), 1835–1844.

Morrogh-Bernard, H. C., Morf, N. V., Chivers, D. J., & Krützen, M. (2011). Dispersal patterns of orang-utans (*Pongo* spp.) in a Bornean peat-swamp forest. *International*

Journal of Primatology, *32*(2), 362–376.

Nietlisbach, P., Arora, N., Nater, A., Goossens, B., Van Schaik, C. P., & Krützen, M. (2012). Heavily male-biased long-distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y-chromosomal and mitochondrial genetic markers. *Molecular Ecology*, *21*(13), 3173–3186.

Nietlisbach, P., Nater, A., Greminger, M. P., Arora, N., & Krützen, M. (2010). A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orang-utans (Genus: *Pongo*). *Conservation Genetics Resources*, *2*, 153–158.

Nsubuga, A. M., Robbins, M. M., Roeder, A. D., Morin, P. A., Boesch, C., & Vigilant, L. (2004). Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology*, *13*(7), 2089–2094. https://doi.org/10.1111/j.1365-294X.2004.02207.x

Ogden, R., Baird, J., Senn, H., & McEwing, R. (2012). The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: A case study from Arabian and scimitar-horned oryx. *Conservation Genetics Resources*, *4*(2), 471–473.

Okitsu, C. Y., Berg, D. J. Van Den, Lieber, M. R., & Hsieh, C. (2013). Reproducibility and reliability of SNP analysis using human cellular DNA at or near nanogram levels. *BMC Research Notes*, *6*(515), 1–6.

Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology*, *19*(24), 5332–5344.

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, *43*(2), 258–275.

R Core Team. (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna, Austria URL Http://Www.R-Project.Org/*.

Rodman, P. S. (1979). Individual activity patterns and the solitary nature of orangutans. In D. A. Hamburg & E. R. McCown (Eds.), *The great apes : Perspectives on human evolution Volume V* (pp. 235–255). Menlo Park, Calif.: Benjamin/Cummings Pub. Co.

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

Ross, C. T., Weise, J. A., Bonnar, S., Nolin, D., Satkoski Trask, J., Smith, D. G., … Kanthaswamy, S. (2014). An empirical comparison of short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) for relatedness estimation in Chinese rhesus macaques (*Macaca mulatta*). *American Journal of Primatology*, *76*(4), 313–324.

Rousset, F. (2008). Genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Russon, A. E., & Galdikas, B. M. F. (1993). Imitation in free-ranging rehabilitant

orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *107*(2), 147–161.

Rutledge, L. Y., Holloway, J. J., Patterson, B. R., & White, B. N. (2009). An improved field method to obtain DNA for individual identification from wolf scat. *Journal of Wildlife Management*, *73*(8), 1430–1435.

Shapiro, G. L. (1982). Sign acquisition in a home-reared/free-ranging orangutan: Comparisons with other signing apes. *American Journal of Primatology*, *3*(1–4), 121–129.

Singleton, I., Knott, C. D., Morrogh-bernard, H. C., Wich, S. A., & Schaik, C. P. Van. (2009). Ranging behavior of orangutan females and social organization. In S. A. Wich, S. S. Utami-Atmoko, T. M. Setia, & C. P. van Schaik (Eds.), *Orangutans: geographic variation in behavioral ecology and conservation* (pp. 205–214). Oxford Universtiy Press USA - OSO.

Stopher, K. V., Nussey, D. H., Clutton-Brock, T. H., Guinness, F., Morris, A., & Pemberton, J. M. (2012). Re-mating across years and intralineage polygyny are associated with greater than expected levels of inbreeding in wild red deer. *Journal of Evolutionary Biology*, *25*(12), 2457–2469.

Stumpf, R. M., Emery Thompson, M., & Knott, C. D. (2008). A comparison of female mating strategies in *Pan troglodytes* and *Pongo* spp. *International Journal of Primatology*, *29*(4), 865–884.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., …

Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, *24*(16), 3189–3194.

Tajima, T., Malim, T. P., & Inoue, E. (2018). Reproductive success of two male morphs in a free-ranging population of Bornean orangutans. *Primates*, *59*(2), 127–133.

Utami-Atmoko, S. S. (2000). Bimaturism in orangutan males. *Utrecht University, Utrecht.*

Utami-Atmoko, S. S., Traylor-Holzer, K., Rifqi, M. A., Siregar, P. G., Achmad, B., Priadjati, A., … Lees, C. M. (eds. . (2017). Orangutan population and habitat viability assessment: final report. *IUCN/SSC Conservation …*, 1–130. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Orangutan+population+and+habitat+viability+assessment:+Final+Report.#0

Utami, S. S., Goossens, B., Bruford, M. W., De Ruiter, J. R., & van Hooff, J. a. R. a. M. (2002). Male bimaturism success and reproductive success in Sumatran orang-utans. *Behavioral Ecology*, *13*(5), 643–652.

Vallender, E. J. (2011). Expanding whole exome resequencing into non-human primates. *Genome Biology*, *12*(9), R87. https://doi.org/10.1186/gb-2011-12-9-r87

van Adrichem, G. G. J., Utami, S. S., Wich, S. A., van Hooff, J. A. R. A. M., & Sterck, E. H. M. (2006). The development of wild immature sumatran orangutans (*Pongo abelii*) at Ketambe. *Primates*, *47*(4), 300–309. https://doi.org/10.1007/s10329-006-0193-9

van Noordwijk, M. A., Arora, N., Willems, E. P., Dunkel, L. P., Amda, R. N., Mardianah, N., … van Schaik, C. P. (2012). Female philopatry and its social benefits among Bornean orangutans. *Behavioral Ecology and Sociobiology*, *66*(6), 823–834. https://doi.org/10.1007/s00265-012-1330-7

van Schaik, C. P., Damerius, L., & Isler, K. (2013). Wild orangutan males plan and communicate their travel direction one day in advance. *PLoS ONE*, *8*(9), e74896. https://doi.org/10.1371/journal.pone.0074896

Vigilant, L., & Guschanski, K. (2009). Using genetics to understand the dynamics of wild primate populations. *Primates*, *50*(2), 105–120. https://doi.org/10.1007/s10329-008-0124-z

Vigilant, L., Hofreiter, M., Siedel, H., & Boesch, C. (2002). Paternity and relatedness in wild chimpanzee communities. *Proceedings of the National Academy of Sciences*, *98*(23), 12890–12895. https://doi.org/10.1073/pnas.231320498

Voigt, M., Wich, S. A., Ancrenaz, M., Meijaard, E., Abram, N., Banes, G. L., … Kühl, H. S. (2018). Global demand for natural resources eliminated more than 100,000 Bornean orangutans. *Current Biology*, *28*(5), 761-769.e5. https://doi.org/10.1016/j.cub.2018.01.053

Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, *160*(3), 1203–1215.

Wang, J. (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research*, *89*(3), 135–153.

https://doi.org/10.1017/S0016672307008798

Wang, J. (2011). Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, *11*(1), 141–145. https://doi.org/10.1111/j.1755-0998.2010.02885.x

Wang, J. (2014). Marker-based estimates of relatedness and inbreeding coefficients: An assessment of current methods. *Journal of Evolutionary Biology*, *27*(3), 518–530. https://doi.org/10.1111/jeb.12315

Wang, J. (2017). Estimating pairwise relatedness in a small sample of individuals. *Nature Publishing Group*, *119*(5), 302–313. https://doi.org/10.1038/hdy.2017.52

Wang, J., & Santure, A. W. (2009). Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*, *181*(4), 1579–1594. https://doi.org/10.1534/genetics.108.100214

Weir, B. S., & Cockerham, C. C. (1984). Estimating F -statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, *16*, 97–159.

APPENDIXES

APPENDIX A – SNP LOCI

Di-allelic human SNP loci, human chromosome number and MapInfo coordinate for SNP, number of individuals genotyped, observed heterozygosity H_O and expected heterozygosity H_E, polymorphic information content PIC, p-value of probability test for deviation from Hardy-Weinberg equilibrium, and average F_{IS} (Weir & Cockerham, 1984) * indicates significant deviation from HW equilibrium

SNP Name	Human Chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} W&C
1:25617206-CT	1	25290715	25	0.16	0.15	0.136	1.000	0.000	-0.067
1:25729163-GA	1	25402672	23	0.04	0.04	0.042	No information		
19:49206962-GA	19	48703705	24	0.13	0.19	0.169	0.207	0.002	0.349
19:579157-TC	19	579157	20	0.45	0.45	0.342	1.000	0.000	0.000
6:31239829-CT	6	31272052	20	0.35	0.30	0.247	1.000	0.000	-0.188
6:32610134-GA	6	32642357	20	0.55	0.45	0.342	0.607	0.002	-0.229
9:136131022-C-T	9	133255635	17	0.88	0.52	0.375	0.004*	0.000	-0.752
9:136131415-CT	9	133256028	20	0.55	0.48	0.359	0.646	0.002	-0.148
9:136137555-G-A	9	133262152	22	0.09	0.24	0.208	0.025*	0.001	0.628
9:136146449-TAAGAC-T	9	133271018	10	0.70	0.48	0.351	0.221	0.002	-0.500
9:139925644-GA	9	137031192	20	0.30	0.26	0.222	1.000	0.000	-0.152
9:139925843-CA	9	137031391	19	0.37	0.31	0.255	1.000	0.000	-0.200
exm2229707	19	48596811	18	0.44	0.51	0.375	0.655	0.002	0.139
exm224876	2	126696000	19	0.47	0.46	0.349	1.000	0.000	-0.025
exm2260060	1	240579605	20	0.55	0.41	0.319	0.256	0.002	-0.357
exm2260204	13	41549067	19	0.16	0.15	0.135	1.000	0.000	-0.059
exm2260552	16	50669787	21	0.14	0.14	0.124	1.000	0.000	-0.053
exm2261221	2	236300554	14	0.43	0.35	0.280	1.000	0.000	-0.238
exm2261348	3	10664912	17	0.35	0.30	0.248	1.000	0.000	-0.185
exm2262610	9	137345126	18	0.17	0.25	0.211	0.274	0.002	0.329
exm2264375	9	138066115	21	0.24	0.22	0.188	1.000	0.000	-0.111
exm2265018	1	74396083	20	0.25	0.51	0.374	0.030	0.001	0.518
exm2265648	3	188976960	16	0.44	0.42	0.323	1.000	0.000	-0.050
exm2266502	7	5793154	13	0.39	0.32	0.262	1.000	0.000	-0.200

SNP Name	Human Chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} W&C
exm2266554	7	101448189	25	0.32	0.27	0.233	1.000	0.000	-0.171
exm2267112	10	81963701	6	0.33	0.55	0.375	0.476	0.002	0.412
exm2267114	10	84338568	16	0.50	0.48	0.359	1.000	0.000	-0.035
exm2268218	19	12581187	23	0.61	0.46	0.351	0.180	0.002	-0.322
exm2269623	3	51378937	21	0.33	0.29	0.239	1.000	0.000	-0.177
exm2270539	6	161727303	21	0.62	0.44	0.336	0.116	0.002	-0.429
exm2271402	10	55891194	19	0.16	0.15	0.135	1.000	0.000	-0.059
exm2271881	12	129933255	15	0.47	0.37	0.294	0.530	0.002	-0.273
exm2272151	14	20349972	21	0.24	0.22	0.188	1.000	0.000	-0.111
exm2272325	15	90960641	15	0.20	0.19	0.164	1.000	0.000	-0.077
exm2272572	17	8124275	13	0.31	0.27	0.226	1.000	0.000	-0.143
exm51163	1	42830512	21	0.29	0.48	0.360	0.077	0.002	0.415
exm518984	6	18143724	11	0.36	0.52	0.373	0.540	0.002	0.310
exm526563	6	29828746	19	0.26	0.24	0.202	1.000	0.000	-0.125
exm537081	6	32938875	19	0.21	0.19	0.171	1.000	0.000	-0.091
exm537383	6	33068728	21	0.24	0.29	0.239	0.451	0.002	0.167
exm537454	6	33069863	21	0.62	0.47	0.354	0.188	0.002	-0.327
exm537513	6	33080851	18	0.33	0.36	0.286	1.000	0.000	0.064
exm612728	7	30922175	20	0.15	0.14	0.129	1.000	0.000	-0.056
exm-rs3117034	6	33119581	19	0.21	0.27	0.231	0.371	0.002	0.234
exm-rs8176746	9	133255935	22	0.68	0.50	0.370	0.186	0.002	-0.370
JHU_1.3691239	1	3774676	20	0.20	0.26	0.222	0.352	0.002	0.240
JHU_11.35177589	11	35156043	19	0.16	0.15	0.135	1.000	0.000	-0.059
JHU_11.35216457	11	35194911	11	0.27	0.52	0.375	0.220	0.002	0.492
JHU_17.42329003	17	44251636	24	0.29	0.31	0.258	1.000	0.000	0.064
JHU_17.42330696	17	44253329	20	0.15	0.14	0.129	1.000	0.000	-0.056
JHU_2.127436468	2	126678893	11	0.18	0.17	0.152	1.000	0.000	-0.053
JHU_22.43100132	22	42704127	23	0.22	0.20	0.175	1.000	0.000	-0.100
JHU_6.10535520	6	10535288	19	0.21	0.27	0.231	0.374	0.002	0.234
JHU_6.10535603	6	10535371	20	0.10	0.10	0.090	1.000	0.000	-0.027
JHU_6.32607324	6	32639548	18	0.44	0.36	0.286	0.526	0.002	-0.259
JHU_6.32607610	6	32639834	19	0.63	0.48	0.357	0.315	0.003	-0.333
JHU_6.32608034	6	32640258	16	0.81	0.50	0.366	0.014*	0.001	-0.667

SNP Name	Human Chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} W&C
JHU_6.32608356	6	32640580	16	0.31	0.27	0.229	1.000	0.000	-0.154
JHU_6.32610682	6	32642906	17	0.35	0.30	0.248	1.000	0.000	-0.185
JHU_6.32628305	6	32660529	15	0.07	0.43	0.332	0.002*	0.000	0.851
JHU_6.32629270	6	32661494	17	0.65	0.51	0.372	0.347	0.002	-0.285
JHU_6.32629370	6	32661594	17	0.35	0.50	0.367	0.324	0.002	0.299
JHU_6.32629548	6	32661772	21	0.67	0.51	0.374	0.205	0.003	-0.315
JHU_6.32629602	6	32661826	20	0.70	0.51	0.372	0.172	0.002	-0.393
JHU_6.32629617	6	32661841	14	0.21	0.50	0.363	0.085	0.002	0.576
JHU_6.32629679	6	32661903	17	0.82	0.50	0.367	0.010*	0.001	-0.684
JHU_6.32630966	6	32663190	19	0.47	0.46	0.349	1.000	0.000	-0.025
JHU_6.32633225	6	32665449	16	0.56	0.50	0.366	1.000	0.000	-0.135
JHU_6.33032864	6	33065088	14	0.64	0.45	0.341	0.220	0.002	-0.444
JHU_6.33045658	6	33077882	21	0.24	0.22	0.188	1.000	0.000	-0.111
JHU_6.33053788	6	33086012	15	0.47	0.43	0.332	1.000	0.000	-0.077
JHU_6.33089374	6	33121598	19	0.21	0.27	0.231	0.369	0.002	0.234
kgp13606542	7	95295992	16	0.38	0.39	0.305	1.000	0.000	0.032
kgp15099441	22	42129132	13	0.15	0.27	0.226	0.235	0.002	0.442
kgp3038063	16	89553920	19	0.21	0.19	0.171	1.000	0.000	-0.091
kgp451798	10	72013241	12	0.67	0.52	0.375	0.563	0.002	-0.294
kgp9521982	8	69832577	18	0.39	0.32	0.264	1.000	0.000	-0.214
rs1000709	9	114475474	20	0.20	0.19	0.164	1.000	0.000	-0.086
rs1034063	20	3051186	23	0.44	0.43	0.334	1.000	0.000	-0.005
rs1042544	6	33086680	12	0.42	0.43	0.328	1.000	0.000	0.035
rs1055055	20	5546645	10	0.30	0.27	0.222	1.000	0.000	-0.125
rs1058433	1	93154836	17	0.12	0.30	0.248	0.042*	0.001	0.615
rs1060622	9	37974746	17	0.71	0.47	0.352	0.049*	0.001	-0.524
rs1138374	20	16260771	25	0.24	0.49	0.365	0.014*	0.001	0.515
rs12480506	3	160086741	11	0.55	0.52	0.373	1.000	0.000	-0.053
rs12634498	1	151874041	22	0.00	0.09	0.083	0.024*	0.001	1.000
rs13320	5	102335711	19	0.63	0.50	0.369	0.358	0.003	-0.271
rs1584717	13	28734932	8	0.63	0.53	0.371	1.000	0.000	-0.207

SNP Name	Human Chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} W&C
rs1617234	6	29747985	16	0.38	0.52	0.375	0.341	0.002	0.280
rs1633086	11	66504671	21	0.00	0.09	0.087	0.024*	0.001	1.000
rs1671063	7	87504154	17	0.18	0.17	0.148	1.000	0.000	-0.067
rs17064	7	130331348	19	0.21	0.19	0.171	1.000	0.000	-0.091
rs1760921	17	12747460	21	0.29	0.25	0.215	1.000	0.000	-0.143
rs1800462	22	29232752	12	0.50	0.51	0.368	1.000	0.000	0.015
rs1809627	12	129872147	19	0.37	0.37	0.296	1.000	0.000	0.008
rs1980889	2	240419949	17	0.18	0.17	0.148	1.000	0.000	-0.067
rs1984661	3	193682987	21	0.05	0.05	0.045	No information		
rs1997719	14	106852518	17	0.59	0.50	0.367	0.624	0.002	-0.185
rs2047709	6	33007734	19	0.53	0.44	0.339	0.610	0.002	-0.192
rs2078402	17	50683744	18	0.61	0.48	0.355	0.321	0.002	-0.299
rs2088335	16	2895089	16	0.56	0.42	0.323	0.257	0.002	-0.364
rs2105992	15	25197251	20	0.05	0.05	0.048	No information		
rs2267647	19	15929482	19	0.21	0.34	0.277	0.143	0.002	0.390
rs2277624	6	6168985	25	0.24	0.27	0.233	0.485	0.002	0.127
rs2301763	5	601532	22	0.23	0.21	0.181	1.000	0.000	-0.105
rs2739765	10	5899990	22	0.27	0.30	0.253	0.539	0.002	0.106
rs3740066	6	32938451	12	0.42	0.43	0.328	1.000	0.000	0.035
rs3765070	11	69261743	18	0.22	0.20	0.178	1.000	0.000	-0.097
rs3823193	4	21094522	18	0.11	0.11	0.099	1.000	0.000	-0.030
rs3828570	19	17366509	21	0.48	0.42	0.325	0.631	0.002	-0.143
rs591510	17	61596942	18	0.61	0.44	0.334	0.120	0.002	-0.417
rs6934645	1	5335808	16	0.25	0.23	0.195	1.000	0.000	-0.111
rs7122786	22	42716955	20	0.40	0.47	0.351	0.633	0.002	0.146
rs720853	10	124648781	20	0.60	0.49	0.365	0.385	0.003	-0.226
rs7248564	10	133239619	17	0.29	0.26	0.219	1.000	0.000	-0.143
rs725900	6	33082268	13	0.46	0.49	0.361	1.000	0.000	0.065
rs729206	6	33087470	13	0.54	0.41	0.316	0.500	0.002	-0.333
rs738527	6	33088972	23	0.26	0.23	0.201	1.000	0.000	-0.128
rs876352	21	43283415	21	0.14	0.22	0.188	0.232	0.002	0.341
rs880340	1	42414845	17	0.29	0.40	0.314	0.527	0.002	0.273
rs907100	14	20349972	15	0.27	0.24	0.204	1.000	0.000	-0.120
rs9277361	6	18143724	13	0.54	0.41	0.316	0.502	0.002	-0.333
rs9277542	9	89284330	17	0.59	0.47	0.352	0.589	0.002	-0.260
rs9277561	10	99844450	24	0.38	0.40	0.317	1.000	0.000	0.072
rs976531	2	238654938	18	0.39	0.32	0.264	1.000	0.000	-0.214
		Average	18	0.36	0.34	0.268	0.658	0.001	-0.037
		SD		0.19	0.14	0.092			0.308

APPENDIX B – IDB COEFFICIENTS

Table of common IDB coefficients and relatedness.

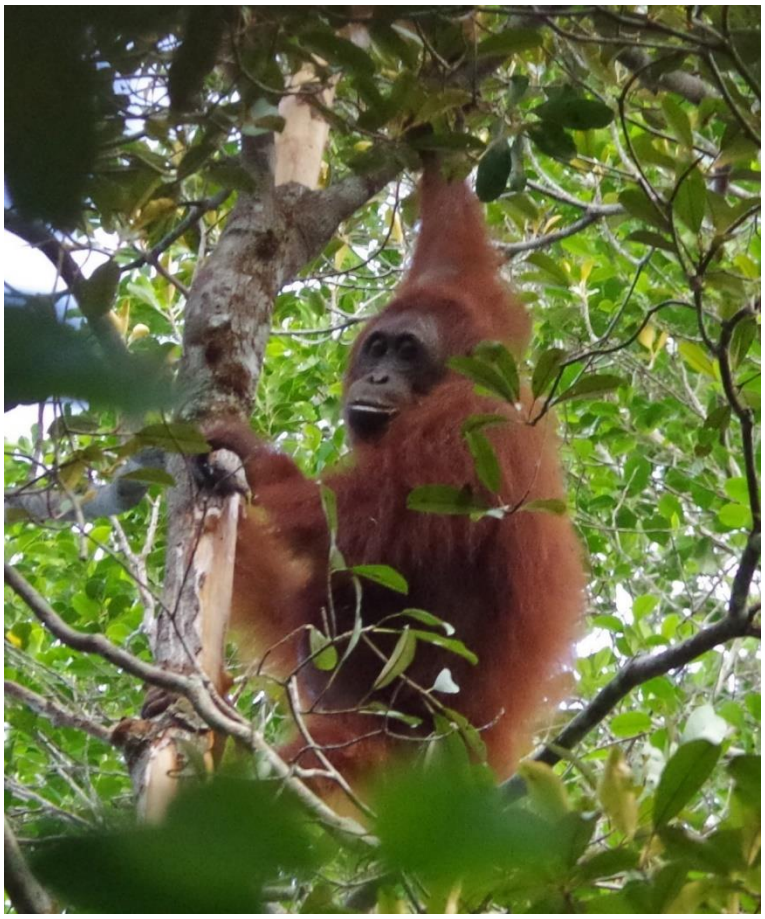
Relationship	Δ_7	Δ_8	Δ_9	r_{XY}
Monozygotic twins	1	0	0	1
Parent-Offspring	0	1	0	1/2
Fullsibs	¼	1/2	1/4	1/2
Halfsibs, Avuncular, Grandparent-Grandchild	0	1/2	1/2	1/4
Double first cousins	1/16	6/16	9/16	1/4
First cousins	0	1/4	3/4	1/8
Second cousins	0	1/16	15/16	1/32
Unrelated	0	0	1	0

Source. COANCESTRY User Guide (Wang, 2011).

APPENDIX C – CAMP LEAKEY ORANGUTANS AND RELATEDNESS

Study Subjects and pairwise *TrioML* relatedness in parentheses

1. Adult Female 01 (BD01) STR and SNP Data
    - No high relatedness other than ABD01(0.5673)
    - Potentially first cousin to BD08 (0.1496) and JDP1 (0.1624)
    - Potentially half sibling or aunt to JR02 (0.2371)
    - Potentially first cousin or half aunt to ABD07 (0.1307) JR09 (0.0974) and BDM3 (0.088)



2. Offspring of Adult Female 01 (ABD01) STR and SNP Data (did not get ID pic)
    - Confirmed genetically female offspring of BD01 (0.5673)
    - Potentially first cousin or half niece to BD03 (0.1191)
    - Half sibs with ABD06 (0.2461)
    - Potentially half great niece to BDN1 (0.0942)
    - Potentially double first cousins with BD02 (0.1864) and BR01(0.1185)

APPENDIX C – (Continued)

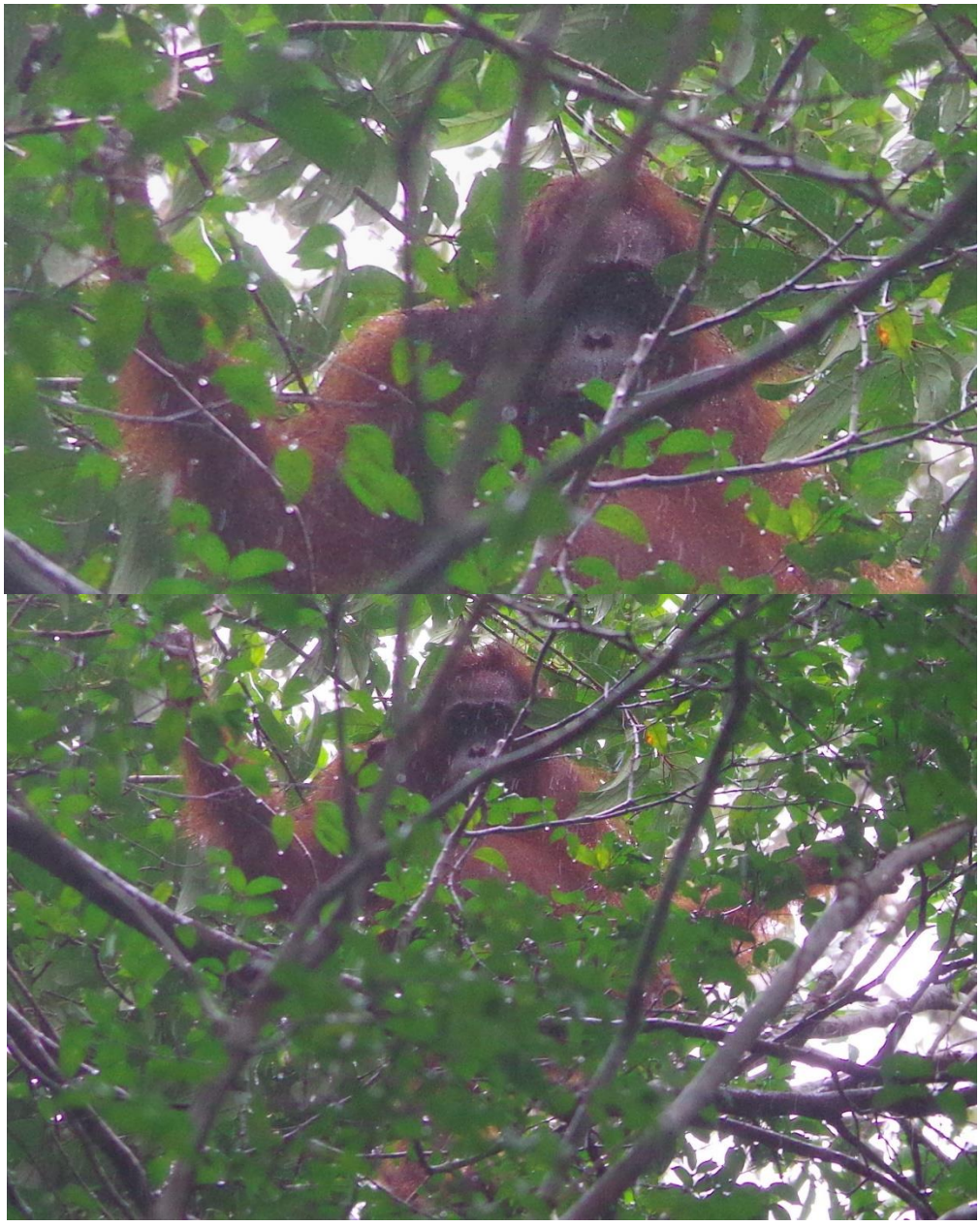
3. Adult Female 02 (BD02) also sampled as BD13 (confirmed same identity genetically) STR and SNP data
    - Half sibling or Aunt to AB10 (0.2005)
    - Possibly siblings with parent siblings to ABD05(0.4406)
    - Fist cousins or half aunt to ABD01 (0.1864)
    - First cousins with BDM1(0.1315) or half niece

Offspring of Adult Female 02 (ABDO2) (Data not analysed)

APPENDIX C – (Continued)

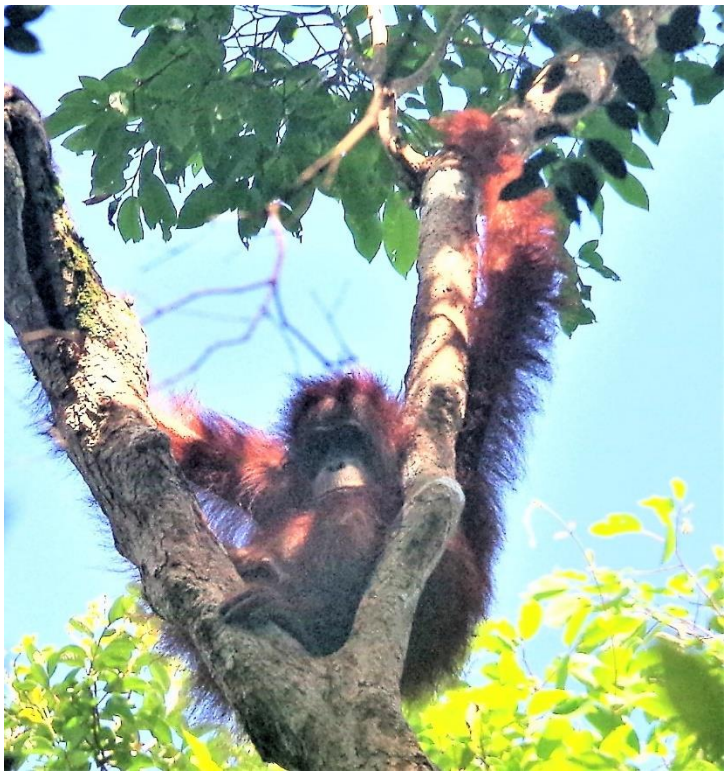
4. Adult Female 02Q (BD02Q) STR and SNP data
    - Half aunt to BR01 (0.1311)
    - Half sib or half aunt to BDR1(0.3291) and JR01 (0.2962)
    - First Cousin or Half aunt to JR08 (0.1519), JR02 (0.0975), and ABDM21 (0.0968), ABD06(0.158), and ABDR1(0.1089)

Offspring of Adult Female 02Q (ABD02Q Data not processed)



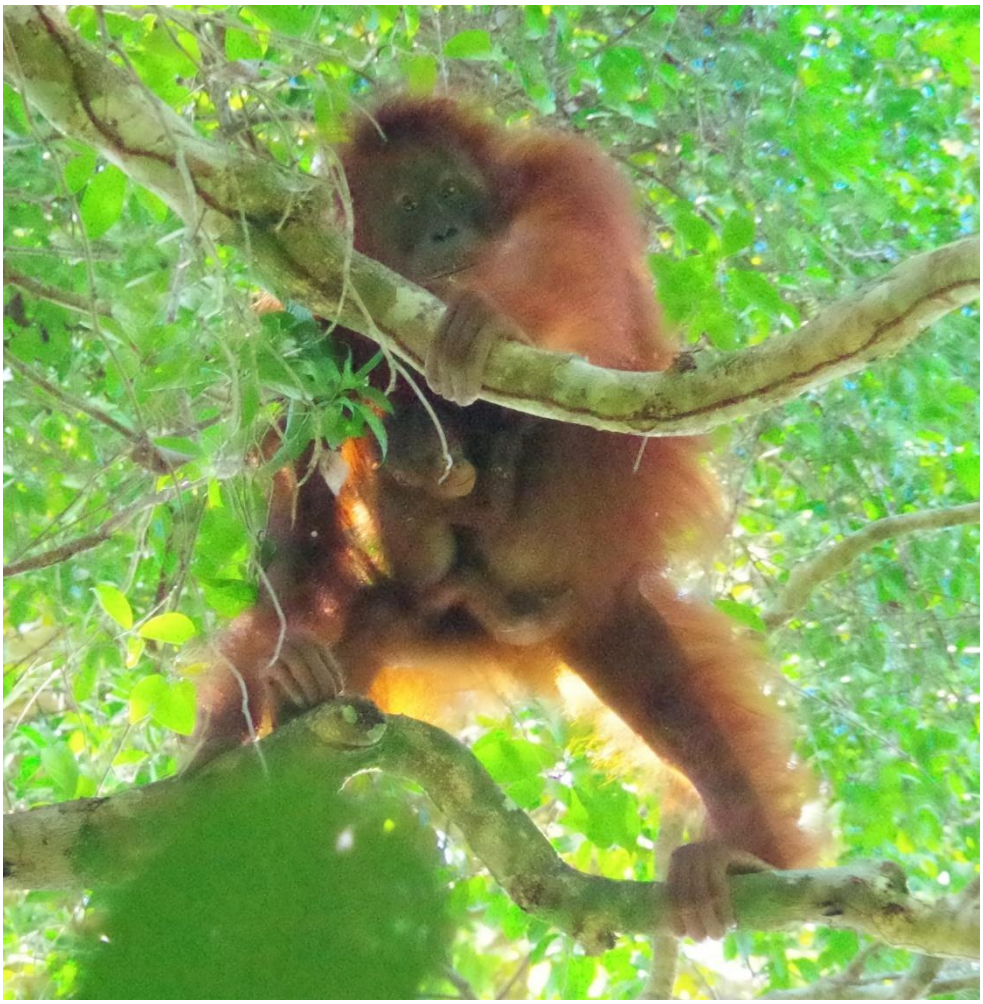
APPENDIX C – (Continued)

5. Adult Female 03 (BD03) with small infant (not sampled) only STR data*
    - Sibling to BD04 (0.6474) and maybe also AB10 (0.4863)
    - Half sibling or aunt/niece to BDR1 (0.1554) and ABD05 (0.2434)
    - First cousins or half aunt to JR02 (0.0994) and ABDR1(0.1149) and ABD01(0.1191)



APPENDIX C – (Continued)

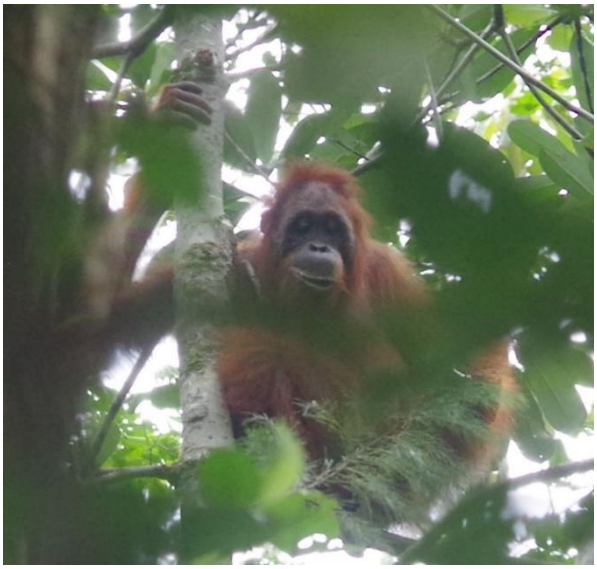
6. Adult Female 04 (BD04) with small infant (not sampled) *STR and SNP Data
    - Full Sibling to BD03 (0.6474) and maybe also AB10 (0.4474)
    - First cousins or half aunt to BR01 (0.152)

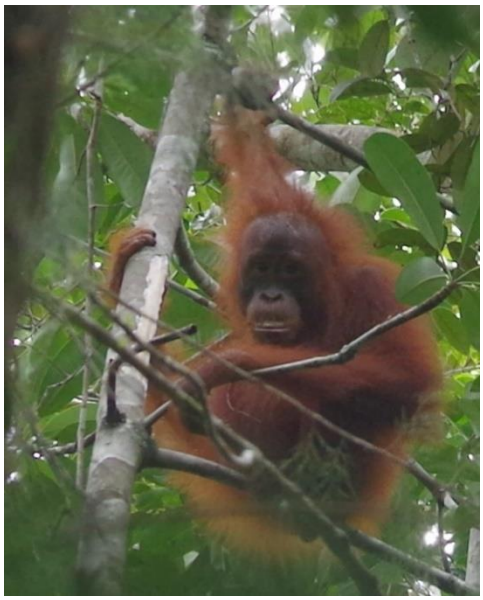


APPENDIX C – (Continued)

Adult Female 05 (BD05) Sample not analysed

7. Offspring of Adult Female ABD05 *Both STR and SNP data
    - Confirmed genetically male
    - Full sibling to BD02 (0.4406) and AB10 (0.716)
    - Half sibling to ABDM12 (0.2226)





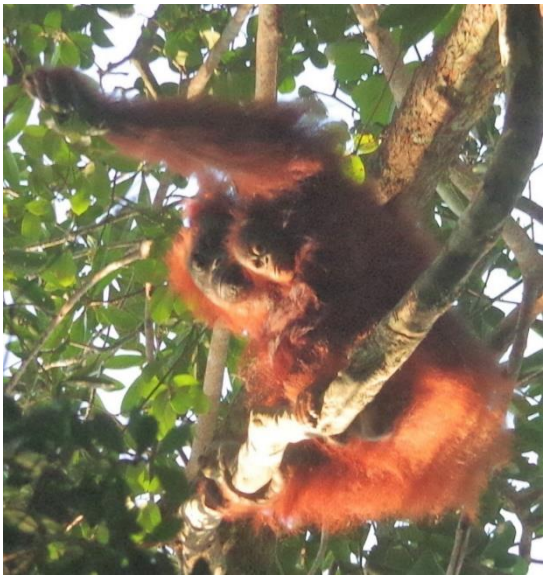
APPENDIX C – (Continued)

8. Adult Female 06 (BD06) Only STR data*
    - Confirmed mother to ABD06 (0.5825)
    - Also mother to BD07 (0.4828) and grand mother to ABD07 (0.312)
    - Mother to JR08 (0.472) and JR02 (0.5)
    - First cousin or half aunt to BDM1 (0.1662)

9. Offspring of Adult Female 06 (ABD06) Only STR data* confirmed to be female offspring of BD06 (0.5825)
    - Half sibling to ABD01 (0.2461) and JR08 (0.2746) and JR09 (0.1009)
    - Half sibling or niece to JD01(0.207)
    - First cousin or half niece to BD02Q (0.158)

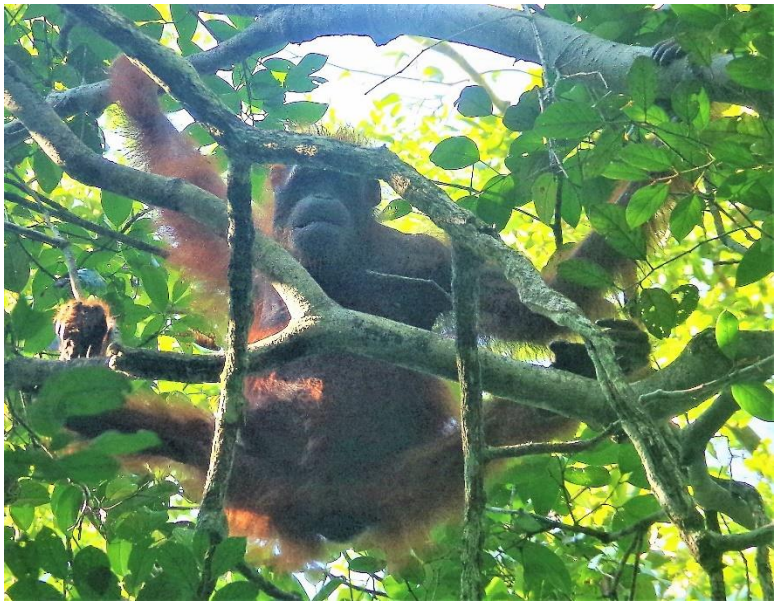
APPENDIX C – (Continued)

10. Adult Female BD07 *STR Data only
    - Daughter of BD06 (0.4828)
    - Half sibling or Niece to BDM1 (0.3471)
    - Half sibling to BDP1 (0.3104) and Half aunt to BR01(0.1768)
11. Offspring of Adult Female ABD07 *STR and SNP Data
    - Daughter to BD07 (0.5)
    - Granddaughter of BD06(0.312)
    - Possibly half niece and granddaughter to BDM1 (0.6453) and BDP1 (0.6099)
    - Possibly half cousin and niece to ABDM12 (0.6531) and BR01(0.4096)
    - First cousin or half niece to BD01(0.1247)
    - Half aunt to ABD06 (0.1204)



APPENDIX C – (Continued)

12. Adult Female BDO8 *STR Data Only
    - Mother or Sister to BDN1 (0.5) and BDP1 (0.5) and BR01 (0.5)
    - Aunt or Grandmother to BDM3 (0.3439) and JMM11 (0.4045)
    - First cousin or half aunt to BD01 (0.1483)



APPENDIX C – (Continued)

13. Adult Female 09 (BD09) *STR Data only No Photo
    - First cousin or half aunt to BD07 (0.0931) and ABDR1(0.0854) and BD06(0.1413)

APPENDIX C – (Continued)

14. Adult Female BDM1*Both STR and SNP data – Known Mother to ABDM12 and JJM11 and known maternal sibling with BDM3 and half aunt to ABDM21
    - Confirmed genetically Mother of ABDM12 (0.5892) but not genetically confirmed mother to JJM11(0.3078)
    - Half Sibling or Aunt to BD07 (0.3471) and JR09 (0.2485)
    - Aunt or great aunt and Grandmother to ABD07 (0.6453)
    - Possibly half sister cousins to BDP1 (0.5544) and half Aunt to BR01 (0.2037)
    - First cousin or half aunt to BD02 (0.1315)
    - Low relatedness to maternal sibling BDM3 (0.1027) and half niece ABDM21(0.0249)
    - Half sibling or Niece to BD06 (0.1662)
    - Maybe grandmother also to ABD01(0.1793)

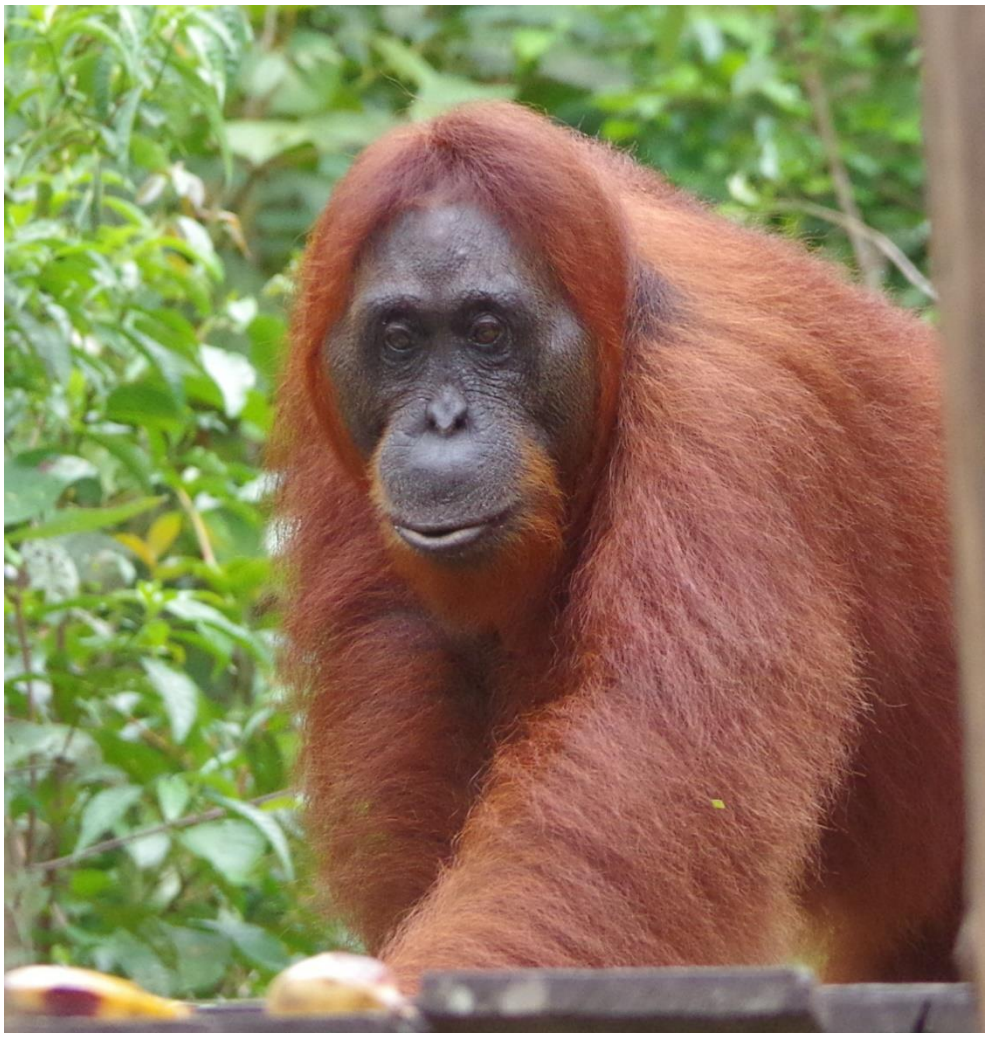
15. Offspring of BDM1 - ABDM12 is genetically confirmed female offspring of BDM1 (0.5892)
    - Sister cousin to ABD07 (0.6531)
    - Half sibling to ABD05 (0.2226) and ABD01 (0.2553)
    - Second cousin and half sibling to BD02 (0.3089)
    - Possibly first cousin and half niece to BR01 (0.2949) and AFP1 (0.2424)
    - Confirmed half niece to BDM3 (0.0848)

APPENDIX C – (Continued)

16. Juvenile Male JMM11 - (known male offspring of BDM1 before ABDM12) Did not have good photo. *STR and SNP Data
    - Nephew or grandson to BD08 (0.4045)
    - First cousin to ABDM21 (0.1274)

APPENDIX C – (Continued)

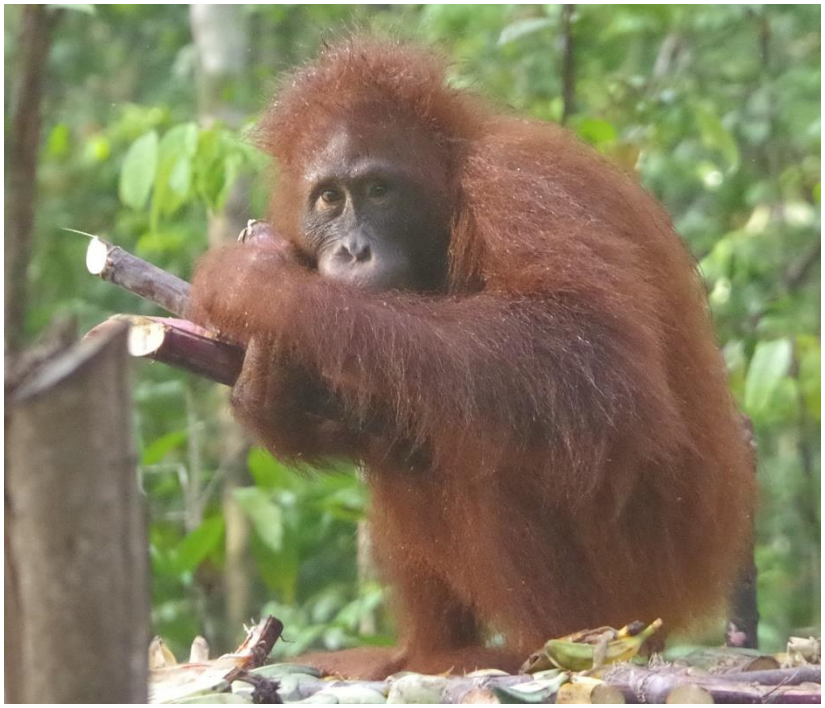
17. Adult Female BDN1–STR and SNP data* - Known mother to BDN11 (0.2813)
    - Possibly daughter to BD08 (0.5)
    - First cousin half aunt to JDP1 (0.1189)
    - Aunt or grandmother to ABD06 (0.2635) and ABD05 (0.2319)



18. Adult Female BDN11 – Adult Daughter of BDN1 (0.2813) *SNP and STR No Photo
    - Half sibs or niece to JD01 (0.3786), JR01 (0.3174), and JD03 (0.2929)
    - Possibly mother of AB10 (0.5)

APPENDIX C – (Continued)

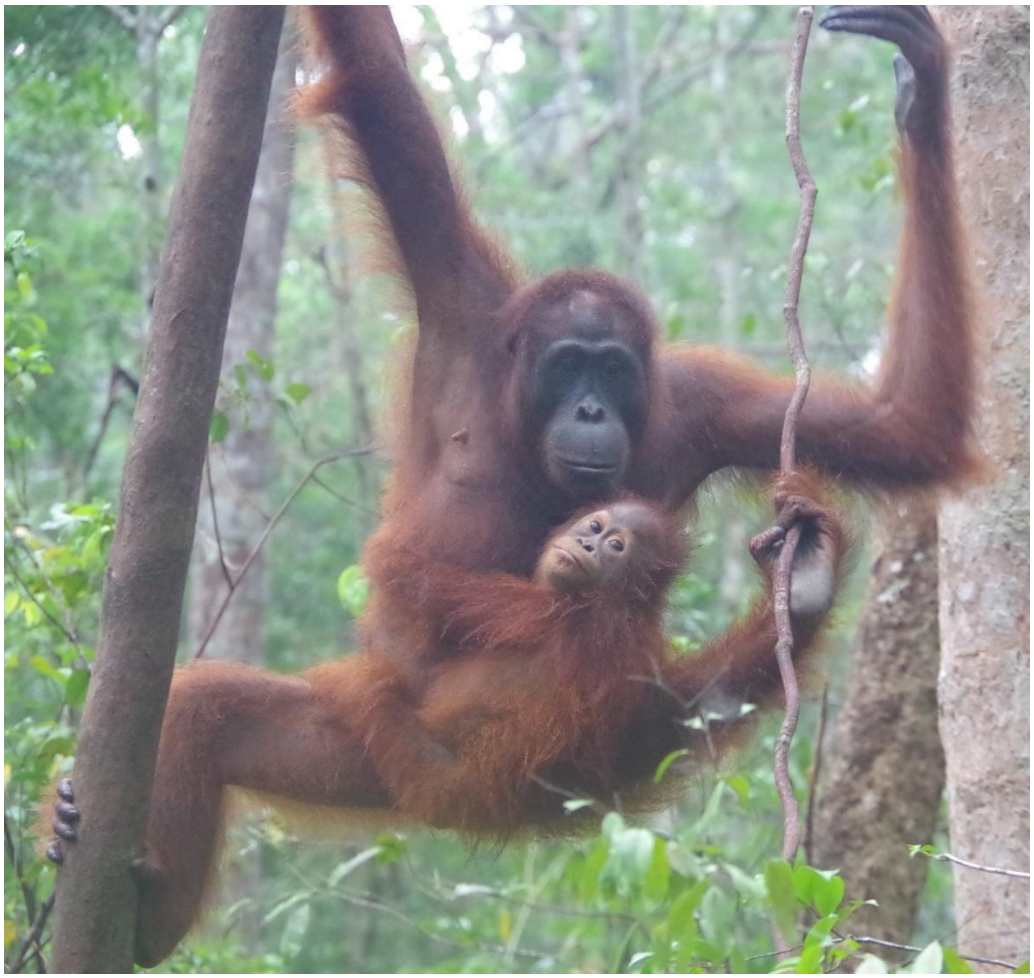
19. BDM3 – Adolescent Female – Younger maternal sibling to BDM1 (0.4209) * STR and SNP Data
    - Half sibling to AB10 (0.5) and BR01 (0.2744)
    - First cousins half niece to JDP1 (0.1144)
    - Niece or granddaughter to BD02Q (0.2432) and BD08 (0.3439)



APPENDIX C – (Continued)

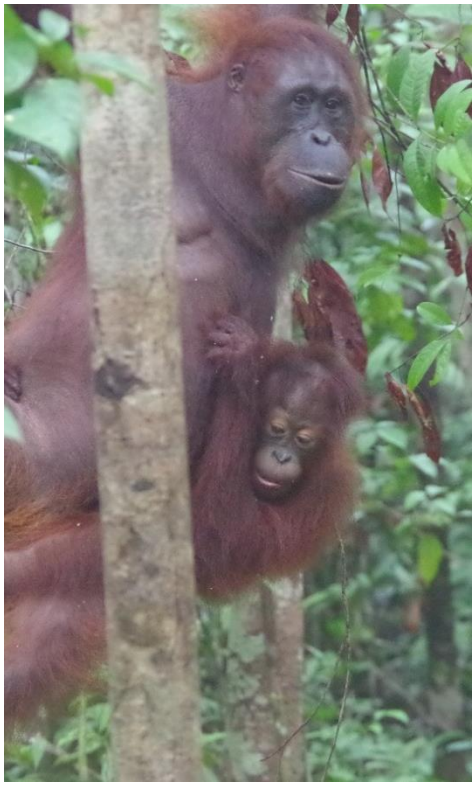
Adult Female BDM2 *Sample not analyzed

20. Female Offspring of BDM2 – ABDM21 SNP data only. BDM2 is younger maternal sibling to BDM1 and Older maternal sibling to BDM3
    - Cousin to JMM12 (0.1274)
    - High relatedness to ABDR1 (0.7967) possibly half sister cousins
    - Niece to BDR1 (0.3194)
    - Half great niece to BD02Q (0.0968)



APPENDIX C – (Continued)

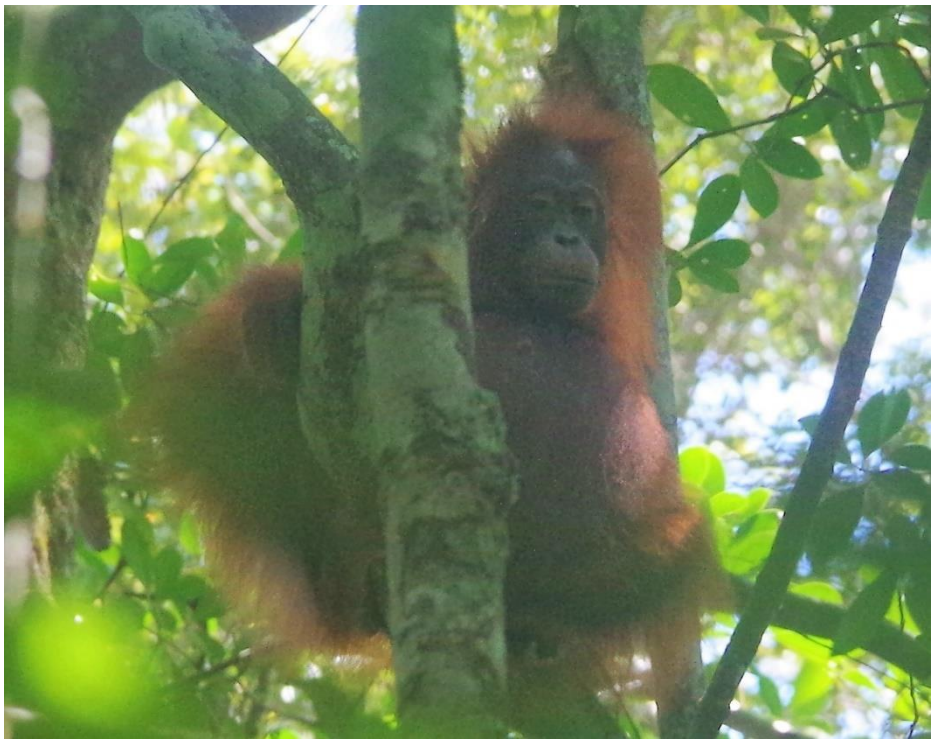
21. Adult female BDR1 – SNP and STR
    - Sibling to BD03 (0.4032)
    - Aunt to ABDM21 (0.3194)
    - Half sibling to JR01 (0.4623), JR02 (0.2184), and JR08 (0.2867)
    - Niece or half sib to BD02Q (0.3291)
22. Female offspring of BDR1 – ABDR1 – SNP and STR
    - Has unexpectedly low relatedness to Known Mother BDR1 (0.2481)
    - Half niece to BDP1(0.1487)
    - High relatedness to ABDM21 (0.7967) possibly half sister cousins
    - Half siblings with AB10 (0.4501)



APPENDIX C – (Continued)

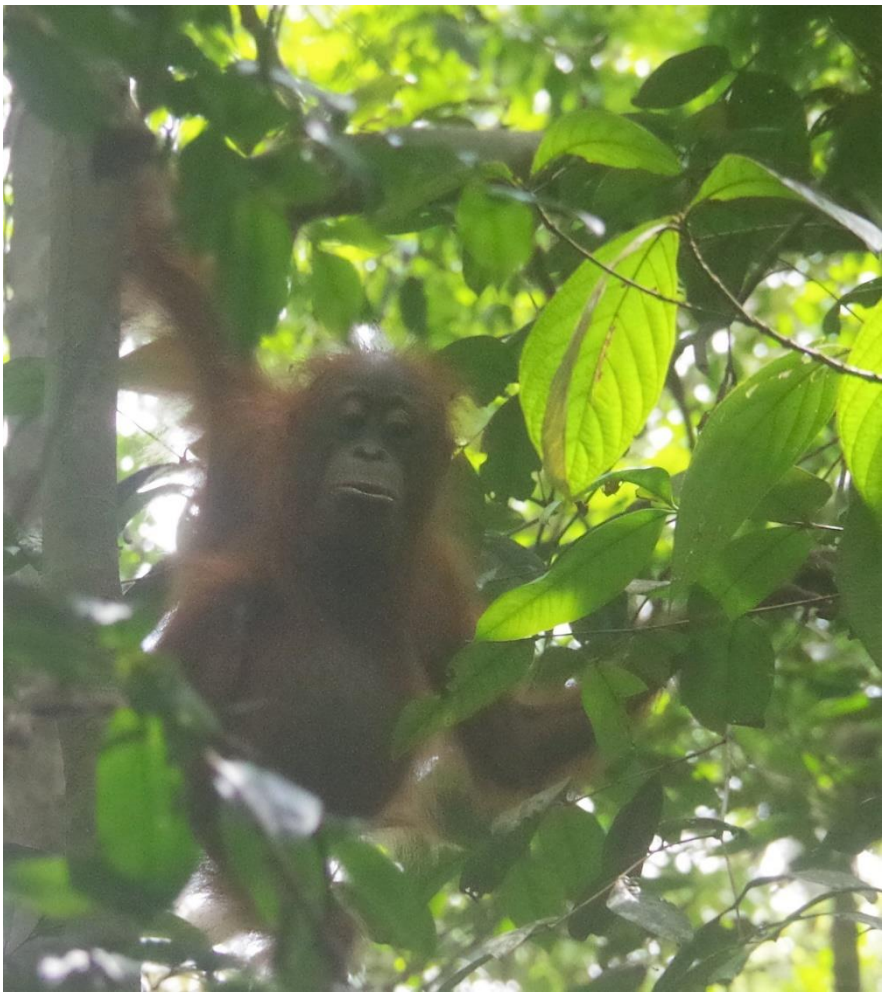
23. Adult Female BDP1 No Photo *STR and SNP data
    - Possible daughter to BD08 (0.5)
    - Mother to BR01 (0.5225)
    - Has unusually high relatedness to ABD07 (0.6099), BDM1 (0.5544), ABDM12(0.5276)
    - Aunt or grandmother to ABD01 (0.3039), and BD07 (0.3104)
    - First cousin or half aunt to ABD05 (0.1545), JR09 (0.1318), and BDM3 (0.1008).

24. Juvenile Female 01 and 02- BR01-BR02 *STR and SNP data
    - Daughter to BDP1 (0.5225)
    - Possibly granddaughter of BD08 (0.3757)
    - Half sister to BDM3 (0.2744), and ABD07 (0.4096)
    - First cousin to JR02 (0.1466) and BD04 (0.152), and BD03 (0.1542), BDM1 (0.2037), and ABDM12 (0.2949)
    - Half niece to BD02Q (0.1311)



APPENDIX C – (Continued)

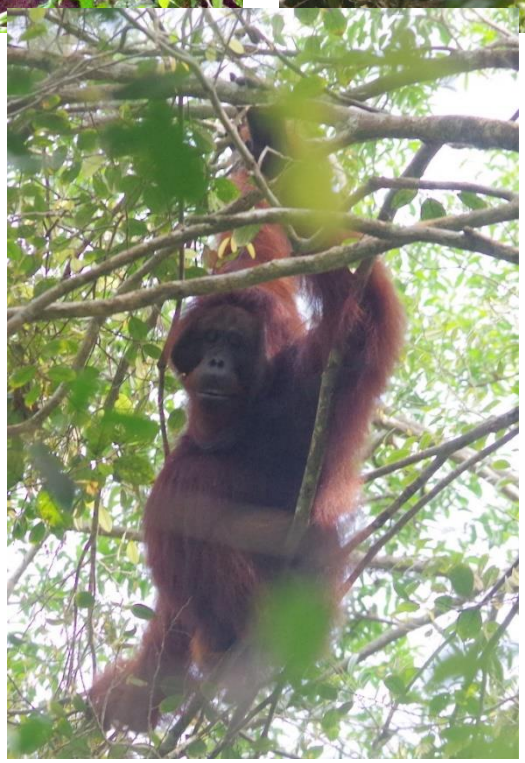
25. Large Infant (Small Juvenile) AB10 *STR data only
    - Confirmed genetically female
    - Sibling to ABD05 (0.716)
    - Niece or half sib to BD04 (0.4863), and BD03 (0.4863)
    - Maybe Daughter of BDN11 (0.5)
    - Half sibling to ABDR1 (0.4501), and BD02 (0.2005), and BDM3 (0.215)



APPENDIX C – (Continued)

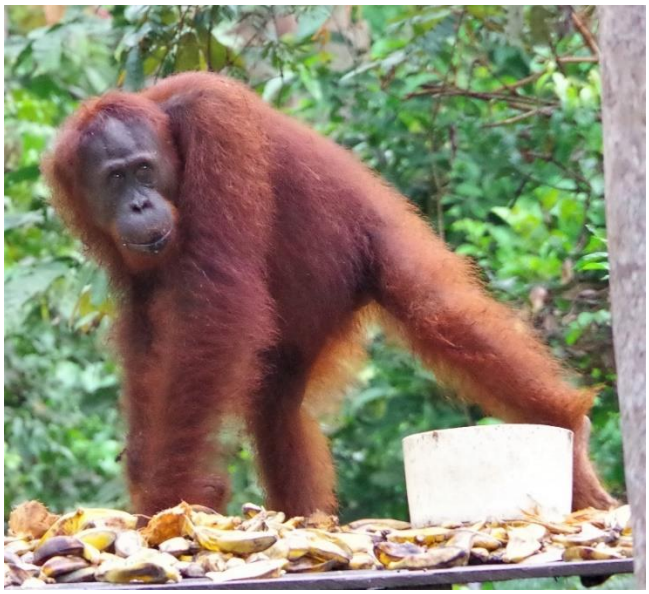
26. Adult Male JDP1 -Flanged *STR and SNP data

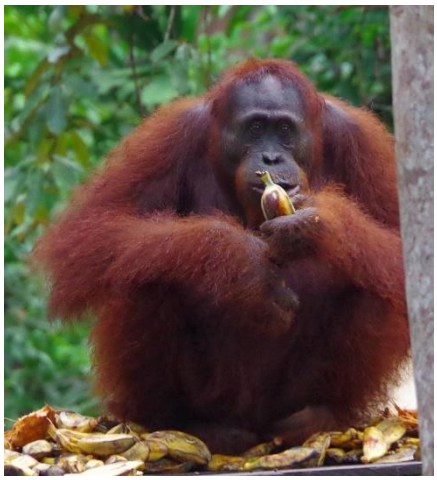
- First cousins with BDN1 (0.1189), BD02 (0.1226), JR09 (0.1751), BDM3 (0.1144), BD04 (0.103), and ABD05 (0.1137)
- Full sib or nephew to BD08 (0.2532), and BD01 (0.1624)



APPENDIX C – (Continued)

27. Juvenile Male JR01 – SNP Data only
    - Half sibling to BDR1(0.4623), BDN11(0.3174), JD03 (0.443) and JR08 (0.461),
    - First cousin or nephew to JD01 (0.2812), and BD02Q (0.2962), and JR02 (0.1401)



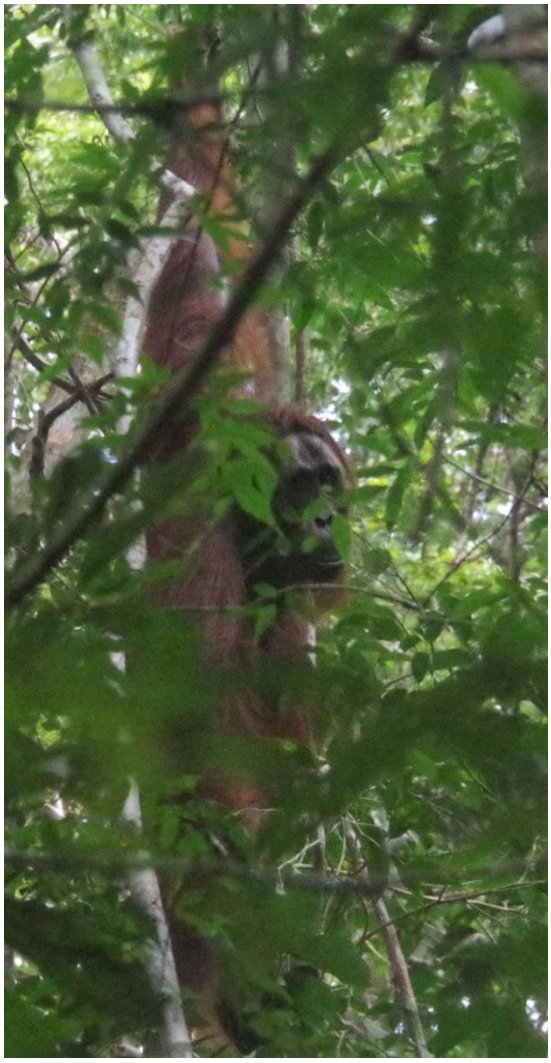


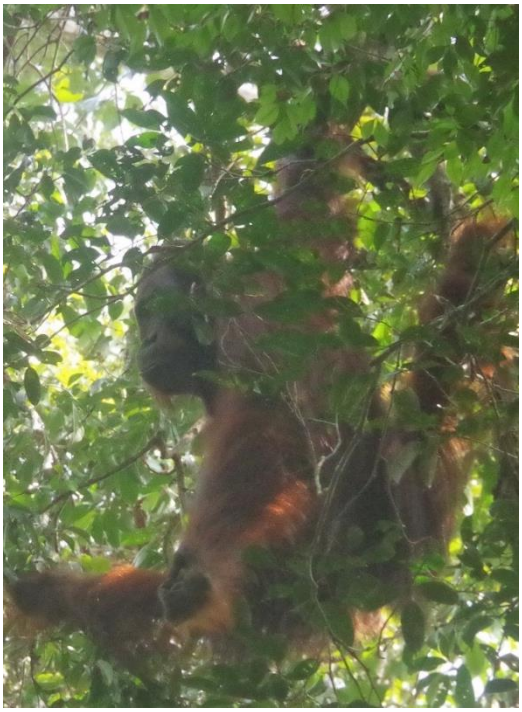
APPENDIX C – (Continued)

28. Juvenile Male JR02 *SNP and STR Data
    - Half sibling or nephew to JD03 (0.2392)
    - Cousin to JR01 (0.1401)
    - Possibly son of BD06 (0.2737)
    - Half sib or nephew to BD01 (0.2416), BDR1(0.2184), JD01 (0.1834), JR08 (0.2692), and BR01 (0.1466)

APPENDIX C – (Continued)

29. Juvenile-Sub Adult Male – JR08* STR and SNP Data
    - Son of BD06 (0.472)
    - Half brother or nephew/uncle to JR02 (0.2692), JR01 (0.461), ABD06 (0.2746), BDR1 (0.2867), JD01 (0.227),
    - Unusually high relatedness to JD03 (0.6354) possibly brother cousins
    - First cousin or half nephew to BD02Q (0.1519)





APPENDIX C – (Continued)

30. Juvenile male - JR09 – *STR and SNP data No Photo
    - First cousin or half nephew to BD06 (0.1407), ABD07 (0.1724), and ABD06 (0.1009), JDP1 (0.1751)
    - Half sibling or nephew to BDM1 (0.1751) or uncle to ABDM12 (0.2954)

APPENDIX C – (Continued)

31. Adult Male – JD01 – flanged* STR and SNP Data
    - First cousin or half nephew/uncle to BD07 (0.1028), BDR1 (0.126), JR08 (0.227), JR02 (0.1834), JR01 (0.2812), and JD03 (0.3862),
    - Half sibling or uncle to ABD06 (0.207), and BDN11 (0.3786)

APPENDIX C – (Continued)

32. Sub Adult Male – JD03 * SNP data only
    - Sibling with BDN11 (0.2929)
    - Half sib with BDR1 (0.2708), JD01 (0.3862), JR02 (0.2392), JR01 (0.443)
    - Very high relatedness with JR08(0.6354)