1-2022

Cross Species use of Human Microarray Genotyping Technology for Bornean Orangutan (Pongo pygmaeus) SNP Discovery

Ruth Ella Linsky

R. Steven Wagner

Reniastoetie Djojoasmoro

Joseph G. Lorenz

Biruté M. F. Galdikas

HAYATI
Journal of Biosciences

Vol. 29 No. 1, January 2022 62-75
DOI:10.4308/hjb.29.1.62-75
ISSN: 1978-3019
EISSN: 2086-4094

Cross Species use of Human Microarray Genotyping Technology for Bornean Orangutan (*Pongo pygmaeus*) SNP Discovery

Ruth Ella Linsky[1,2*], R. Steven Wagner[1,3], Reniastoetie Djojoasmoro[4], Joseph Lorenz[1,5], Biruté M. F. Galdikas[2,4]

[1]*Primate Behavior Program, Central Washington University, Ellensburg, USA*
[2]*Simon Fraser University, Burnaby, Canada*
[3]*Biotecon Diagnostics, Hermannswerder 17, Potsdam, Germany*
[4]*Orangutan Foundation International, Jakarta, Indonesia*
[5]*Eichendorffstrasse 32 68167 Mannheim, Germany*

ARTICLE INFO

Article history:
Received November 18, 2019
Received in revised form September 30, 2021
Accepted October 19, 2021

KEYWORDS:
Bornean orangutan,
Pongo pygmaeus,
SNP,
microarray,
Tanjung Puting

ABSTRACT

Previous genetic studies of orangutans (*Pongo* spp.) have relied mainly upon mitochondrial DNA or microsatellite short tandem repeats (STR) for genomic genotyping analysis. Scientists have yet to take advantage of the genetic closeness of the great apes to humans for genomic analysis by using advanced techniques available for human genotyping. To genotype orangutans at Tanjung Puting National Park, we developed a novel combination of a methyl-based magnetic enrichment capture of genomic fecal DNA with genotyping on a human targeted single nucleotide polymorphism (SNP) microarray, and compared this to additional microsatellite (STR) micro-capillary genotyping. We successfully isolated 125 known human genomic SNP loci (0.08% of those targeted) which hybridized orangutan DNA on the human targeted Illumina Infinium QC array. We estimated genetic diversity and relatedness (*r*) using three estimators for a total of 32 (21 female and 9 male) wild orangutans at the Camp Leakey study site. Average *Trio*ML relatedness within the sample, estimated from our combo SNP/STR dataset, was at a range consistent with half and first cousins (*r* = .082). All sampled males and females had relatives within the study site indicating we have verified a local, closely related community of wild orangutans at Camp Leakey.

1. Introduction

In accordance with regulations surrounding endangered species to avoid invasive sampling practices, geneticists have moved towards non-invasive methods by collecting feces, urine, food waste, and other discarded materials (Inoue *et al.* 2007; Rutledge *et al.* 2009). It is now possible to extract viable DNA from these wildlife by-products and primate conservationists have begun to build protocols to investigate populations using these sample types (Goossens *et al.* 2000; Nsubuga *et al.* 2004; Simons *et al.* 2012).

Orangutans (genus: *Pongo*) as semi-solitary arboreal, deep forest primate group, present an extreme example of how difficult non-invasive genetic sampling can be in the wild. However, over the past two decades there have been several studies using non-invasive fecal sampling investigating relatedness and metrics of genetic diversity within populations from study sites in Borneo and Sumatra. These studies which used different methods and genetic markers produced a range of results for each site; for example higher female relatedness was found in Sebangau (*Q* and *G r* = 0.046 and *Wang r* = 0.166; Morrogh-Bernard *et al.* 2011) and in Tuanan (*Trio*ML, *Wang*, and *Q* and *G*; Arora *et al.* 2012) and more equal but low relatedness within males and females at Ketambe (*Q* and *G r* = -0.095 and *r* = -0.108, respectively; Utami *et al.* 2002) and more equal but high relatedness within males and females at the Lower Kinabatangan Wildlife Sanctuary (*Q* and *G r* = 0.142 and *r* =0.148, respectively, Goossens *et al.* 2006). Understanding how these relatedness values are influenced by dispersal and reproductive patterns have direct consequences for species viability in the wild. Thus, having high quality and efficient methods to study these patterns is important for assessing management options.

Despite much improvement in genotyping practices, microsatellite (or Short Tandem Repeat,

* Corresponding Author
E-mail Address: ruth_linsky@sfu.ca

STR) analysis used in the majority of past orangutan genotyping studies is time consuming and requires multiple sample replicates to ensure all loci are amplified during the PCR process; further it is often plagued with issues of low initial template DNA and failure of amplification of some loci due to inhibitors from co-extracted fecal matter. STR DNA fragments also require visualization through gel electrophoresis which is another labor intensive multistep process.

Recent improvements in next generation sequencing and genotyping techniques have provided a relatively swift genotyping process which is increasingly affordable for often inadequately funded primate conservation projects (Vigilant and 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 (Kanthaswamy *et al.* 2009; Norman *et al.* 2013; Rianti *et al.* 2015; Bourgeois *et al.* 2018). The use of commercially available sequencing kits and microarray genotyping chips across closely related species has helped discovery of conserved SNP loci (Miller *et al.* 2011; Ogden *et al.* 2012; Hoffman *et al.* 2013). However, researchers 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 a previously unsampled local wild Bornean orangutan (*Pongo pygmaeus*) population at the Camp Leakey research site, we designed and assessed a new protocol for microarray SNP genotyping of orangutan DNA isolated from feces. This paper details a novel process which combines fecal DNA extraction with a modified magnetic bead enrichment capture technique, FecalSeq (Chiou and Bergey 2018), followed by orangutan genomic SNP genotyping by cross species use of human targeted microarray chips. Relatedness estimates produced by genotyping with this new process are complimented and assessed by comparison with a micro-capillary STR marker genotype dataset.

2. Materials and Methods

2.1. Study Site

With an area of approximately 4,150 km^2 (1,886 km^2 of orangutan habitat), Tanjung Puting National Park (Figure 1) is one of the largest protected areas in Central Kalimantan (Utami-Atmoko *et al.*, 2017). The Camp Leakey study area was initially established within a 35 km^2 area and contains a mix of dry ground

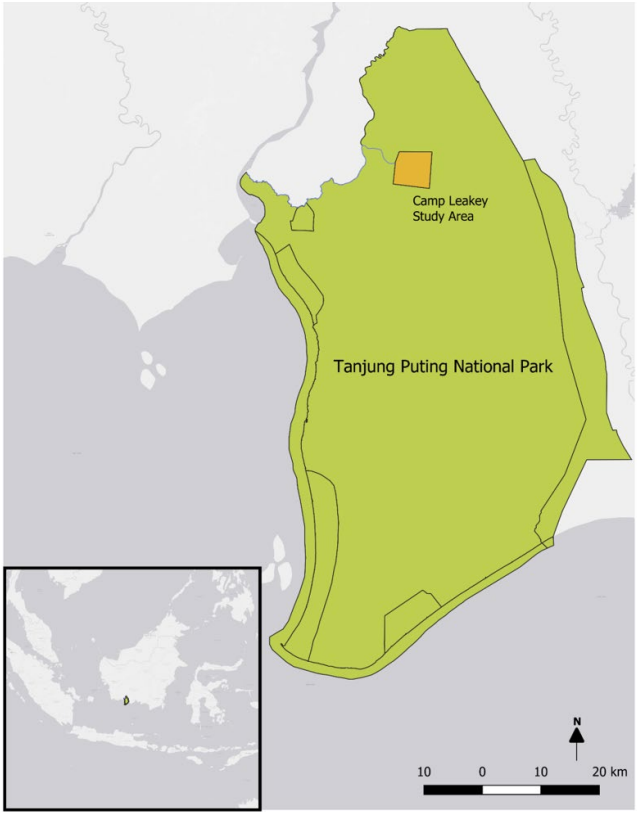

Figure 1. The Camp Leakey Study Area is located along the Sekonyer River in the northern portion of Tanjung Putting National Park in Central Kalimantan, Indonesian Borneo, Indonesia

tropical heath and dipterocarp forests with veins of permanently wet and seasonally flooded peat swamp threaded throughout (Galdikas 1979). The local wild orangutans are behaviorally observed on an ongoing basis within the study area which contains maintained trails (Galdikas 1982, 1985, 1988).

In ongoing data collection at Camp Leakey since its establishment, local wild individuals who are identifiable by local staff and researchers, as well as any unidentified individuals encountered, are behaviorally studied through follows from nest to nest each day. Data are collected by focal follows of orangutans for ten day periods. Matrilineal lines have been recorded for at least three generations on females whose home range included the Camp Leakey study area site. Thus, some relatedness values between sampled individuals are known and mother offspring pairs were identified in Table 1.

2.2. Sample Collection

Fecal sampling of wild orangutans within the Camp Leakey study area took place between January and August 2016. Once located and visually identified and

Table 1. Orangutans genotyped in this study

SNP and STR	Orangutan ID	Age class	Sex	Previously known relationships
SNP and STR	BDM1 (Mooch)	Adult	Female	BDM1 is mother to JMM12 and ABDM12 and older sister to BDM3
SNP and STR	JMM12 (Mario)	Juvenile	Male	
SNP and STR	ABDM12 (Molly)	Infant	Female	
SNP and STR	BDM3 (Macey)	Adolescent	Female	Younger sister to BDM1
SNP and STR	BDN1 (Noisy)	Adult	Female	No known
SNP and STR	BDN11	Adult	Female	No known
SNP and STR	BDR1 (Renie)	Adult	Female	BDR1 is mother to ABDR1
SNP and STR	ABDR1 (Rutha)	Infant	Female	
SNP and STR	BDP1 (Pete)	Adult	Female	BDP1 is mother to BR01
SNP** and STR	BR01 (Pamela)	Adolescent	Female	
SNP and STR	BD01 (Erma)	Adult	Female	BD01 is mother to ABD01
SNP and STR	ABD01 (E2)	Infant	Female	
SNP** and STR	BD02 (Aluh)	Adult	Female	No known
SNP and STR	BD02Q	Adult	Female	No known
SNP and STR	BD04	Adult	Female	No known
SNP and STR	ABD05	Infant	Male	No known
SNP and STR	JDP1 (Ponorogo)	Adult	Male	No known
SNP and STR	JD01	Adult	Male	No known
SNP and STR	JR02	Juvenile/adolescent	Male	No known
SNP and STR	JR08	Juvenile/adolescent	Male	No known
SNP and STR	JR09	Sub adult	Male	No known
SNP	JD03	Sub adult	Male	No known
SNP	JR01	Juvenile/adolescent	Male	No known
SNP	ABDM21 (Maureen)	Infant	Female	ABDM21 is niece to BDM1 and BDM3
STR	BD07	Adult	Female	BD07 is mother to ABD07
SNP and STR	ABD07	Infant	Female	
STR	BD06	Adult	Female	BD06 is mother to ABD06
STR	ABD06	Infant	Female	
STR	BD03	Adult	Female	No known
STR	AB10	Juvenile	Female	No known
STR	BD08 (Beth)	Adult	Female	No known
STR	BD09	Adult	Female	No known

**two sample extracts from these individual orangutans were genotyped on the SNP chip

Orangutans with names are those who were identified 100% by at least two local field staff and witnessed on multiple occasions in the study area during sample collection. The others represent either unknown individuals or those whose identity was not absolutely verified at the time of collection

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). Samples (approx. 2-10 g) were collected using sterile gloves and a sterile collection spoon to avoid contamination, and initially stored in 30 ml of 97% ethanol solution (step one). Ethanol solution was discarded 24 to 36 hours later and ~10 g of silica gel beads were placed inside sample container (step two). Samples were then transferred stored in refrigerator at -40°C until processing. Samples were collected under approved Central Washington University IACUC, with permissions from the Indonesian Institute of Sciences (LIPI), the Director general of Indonesian Directorate of Biodiversity Conservation, with prior informed consent from Tanjung Puting National Park with assistance from Orangutan Foundation International staff, and exported from Indonesia to Central Washington University under the Convention on International Trade in Endangered Species export permit 01152/IV/SATS-LN/2017.

2.3. 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™ robot 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 Camp Leakey. 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 and found traveling on their own), and two unknown subadult males. In total 15 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 during sample collection. The others represent either unknown individuals or those whose identity was not absolutely verified at the time of collection. The orangutans sampled and genotyped in this study are listed in Table 1.

To gauge initial quantity of orangutan DNA extractions, quantitative real-time PCR (qPCR) was conducted on samples using universal mammalian MYCBP primers (Higuchi *et al.* 1993) 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 three individuals were only able to be extracted once due to a low quantity sample) and initial 69 DNA samples were evaluated for quality and relative quantity by the ability to be amplified with real time PCR. Only samples with total orangutan DNA greater than 20 ng were used for the SNP microarray analysis. Extracts that were less than 20 ng were pooled for each individual for either further 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 evaluated and quantified using qPCR.

2.4. Genotyping

For microarray SNP analysis, the FecalSeq (Chiou and Bergey 2018) technique, based on the New England Bio-labs NEBnext Microbiome DNA Enrichment Kit, was used on a subset of 27 individuals to separate host orangutan DNA from co-extracted fecal 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 magnet. The resulting host enriched DNA was utilized for microarray SNP analysis. Samples post enrichment were further evaluated and quantified using the above outlined qPCR technique, with the addition of a universal bacterial 16S rRNA primer (Corless *et al.* 2000), to test for a decrease in bacterial concentration in order to assess sample enrichment success.

SNP microarray analysis was conducted on 48 samples which consisted of the following: 1 human DNA extract used as a positive control, 27 extracts from the FecalSeq enrichment process, and the remaining 20 were un-enriched extraction products. 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 causes fluorescence of the bound 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), (Table 1) were coded visually using the gel-like densitometry plot data comparison view. High quality *Pongo* DNA, and human DNA, as positive control, were 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 were done). In total 29 individuals underwent genotyping using this method (e.g., 22 of the same individuals with SNP genotypes).

2.5. Statistical Analysis

Allele frequencies and distinct individual identities were confirmed using Cervus (Kalinowski *et al.* 2007) software for both genotyping techniques (e.g., microarray and STR). Resulting genotypes from both methods were analyzed for 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 for the 22 individuals within both datasets using the triadic likelihood estimator, *Trio*ML (Wang 2007), and two moment estimators used in past studies, the coefficient of *Wang* (Wang 2002), and *Queller* and *Goodnight* pairwise relatedness estimator (Queller and Goodnight 1989), r_{xy}, using the Colony 2.0 (Jones and Wang 2010) and COANCESTRY 1.0 (Wang 2011) software. Relatedness values for the subset of 22 individuals within both datasets were calculated using allele frequencies from adults in each entire dataset. Overall group and pairwise relatedness values calculated with each estimator were tested for correlation through paired and unpaired *t*-tests and Mantel matrix correlation tests in the ade4 (Dray and Dufour 2007) package in the R statistical environment (R Core Team 2014).

3. Results

3.1. Genotyping

Average total orangutan DNA proportion per 100 µl extract was 3.31% (range <0.01–82.7%). 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 = 57% initial DNA remaining, range 12-95.8%), but was enriched compared to bacterial DNA. 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 38 of 45 samples run on the microarray (with the 7 unconfirmed due to low signal and none providing opposite sex assignment) as well as identified the unknown sex of three infants. Total DNA concentrations for enriched samples ranged from 0.85 ng/µl to 7.39 ng/µl and total DNA concentrations for un-enriched samples were 14.3ng/µl to 48 ng/µl.

Microarray data from un-enriched samples was poor, with light intensity (Norm R) and color (Norm Theta) results scattered across the spectrum (grey scattered dots in Figure 2). Clustering did not occur as should be expected in comparison to human DNA results. However, for the enriched samples visual inspection identified 125 of the SNP loci out of 15, 949 on the microarray that provided clear assignment signals and clusters (green dots in Figure 2). These 125 loci (0.78% of the total on the microarray) were further identified as presenting similar clustering patterns to those of human data (examples of this at four SNP loci discovered are shown in Figure 2) and presenting high enough minor allele frequencies (125 > 0.018, 104 > 0.1, 61 > 0.2) to be used for genotyping.

The 125 homologous bi-allelic SNP loci (Supplementary S1) were used to create genotypes for 27 individuals at a minimum of 65 (52% of the total 125) loci. Through identity analysis, two pairs of samples originally assumed to represent four unique individuals in the 27 individuals in this SNP dataset were identified as representing two individuals each sampled twice, matching at respectively all 93 and 81 loci in common and both pairs mismatching at 0 loci. These two pairs of individuals were confirmed as representing just two individuals sampled in repetition but whose identities were not 100% confirmed in the field at the time of collection. The mean proportion of the 125 SNP loci typed for the 27 individuals 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 we were unable to genotype them using this method. Thus, we 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 $6.6e^{-7}$.

3.2. Genetic Diversity

Overall average observed heterozygosity (Ho) for the 125 SNP markers was 0.36, SD = 0.19, average expected heterozygosity (He) was 0.34, SD = 0.14, with average polymorphic information content (PIC) of 0.27, SD = 0.092. Average inbreeding co-efficient (FIS) was -0.037, SD = 0.32. Of the 125 individual SNPs,

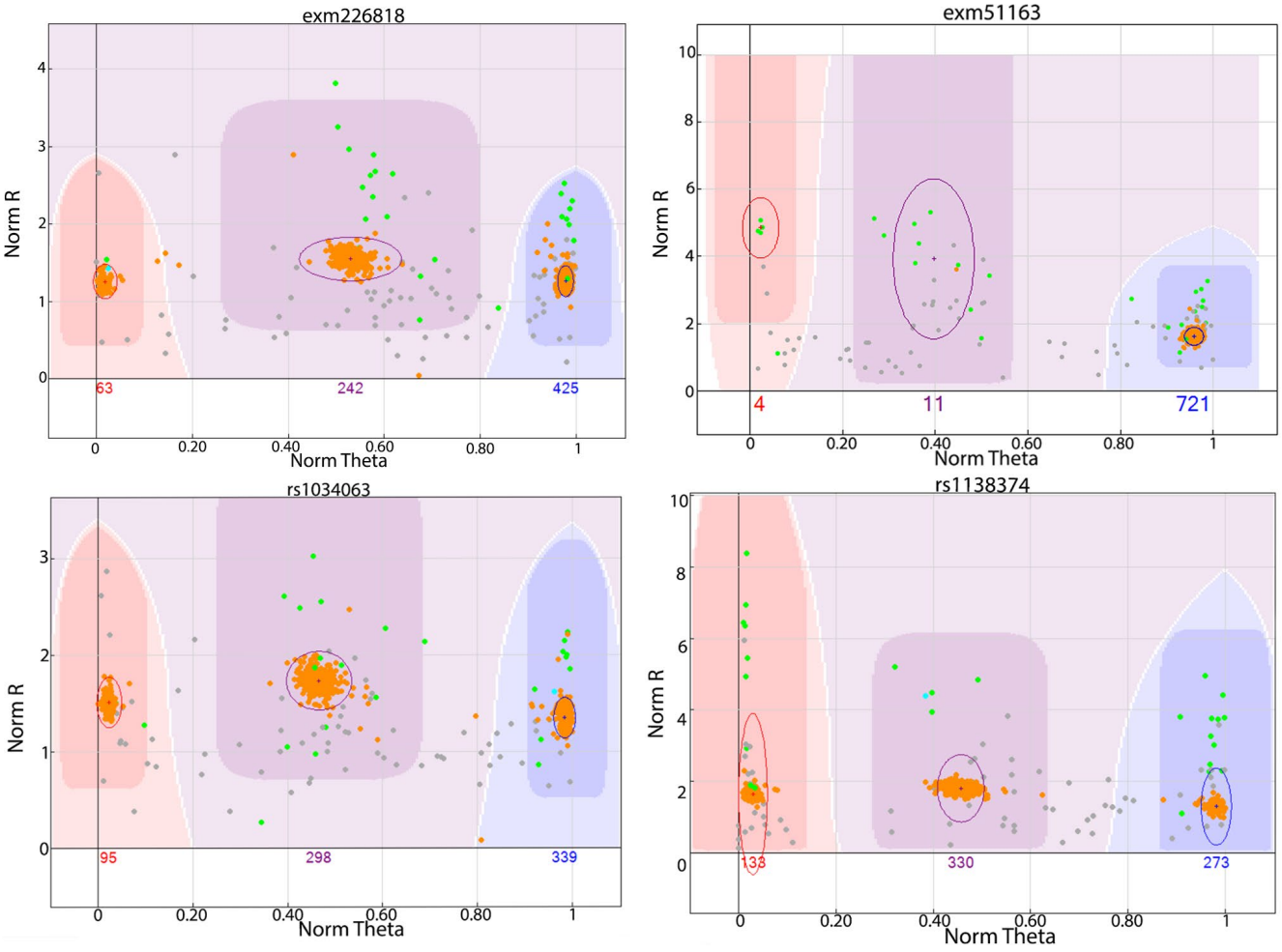

Figure 2. GenomeStudio SNP graph clustering results for four SNP loci. Orange dots are sample human data, green dots are the 27 enriched samples and dark grey dots are unenriched samples. Norm Theta represents the light color read and Norm R is the light intensity. Circles and dark colouration are areas where allele assignments or calls are made. The two exterior red and blue circles represent homozygous calls and center purple is heterozygous call

10 loci differed significantly from H-W equilibrium ($\alpha = 0.05$), with both lower and higher than expected heterozygosity and corresponding high and low FIS values. These values can be found in Supplementary S1.

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 (Ho) to be 0.65 with an expected (He) 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 FIS inbreeding coefficient for the 8 autosomal markers was 0.003 (SD = 0.14). These values can be found in Table 2.

3.3. Relatedness

Initial Cervus identity analysis of the SNP dataset also confirmed the shared identity of the two pairs of individuals each sampled twice within the SNP samples as well as 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 of the coefficient of *Wang* produced by *Coancestry* and parentage analysis using Colony confirmed four of five known mother offspring pairs in the SNP Data and the one non-confirmed pair gave a maximum *r* estimate of 0.23, and confirmed all seven within the STR data (Table 3). Average mother-offspring relatedness across all three estimators was 0.46 sd = 0.19 (SNP data) and 0.43 sd = 0.16 (STR data).

Pairwise relatedness values for the 22 individuals with both SNP and STR genotypes were compared using a Mantel correlation test and showed a positive

Table 2. Autosomal microsatellite (STR) markers

Name	Marker type	A	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} Wa and 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

*indicates a statistical departure from HW equilibrium

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

Table 3. Expected and estimated pairwise relatedness (*Wang* (*r*)) for known mother-offspring pairs

Related mother-offspring Pairs		Expected relatedness	STR relatedness	SNP relatedness
BDM1	JMM12	~0.5	0.49	0.45
BDM1	ABDM12	~0.5	0.44	0.71
BDR1	ABDR1	~0.5	0.48	0.23
BDP1	BR01	~0.5	0.67	0.61
BD01	ABD01	~0.5	0.38	0.64
BD07	ABD07	~0.5	0.57	-
BD06	ABD06	~0.5	0.61	-
		Average	0.52	0.53
		SD	0.10	0.19

correlation for all three estimators *Trio*ML (r = 0.81, p <0.001), Wang (r = 0.34, p = 0.018), and *Queller* and *Goodnight* (r = 0.42, p = 0.0032). The two datasets were then combined for all 32 individuals and a third new "Combo" dataset was produced and compared.

Overall relatedness 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 and STR) dataset using the *Trio*ML estimator was 0.082 (*var* = 0.021).

Results of a paired t-test to compare the overall *Trio*ML 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).

We determined all males were related to two or more individuals in our sample at the level of first cousins or higher. A mother was able to be assigned to two unknown males (one subadult and one juvenile) within the study area. The two adult fully flanged males were estimated to have first cousin and higher relatedness with both females and other males within the sample.

Average adult female relatedness within the sample is approximately between the levels of half cousin (or first cousin once removed) and first cousins (e.g., r values range 0.0625-0.125), with all adult and juvenile females having a close (e.g., 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.

4. Discussion

This is the first known study to combine the use of non-invasive fecal DNA sampling and extraction, methyl based enrichment FecalSeq™ (Chiou and 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 fecal extractions which is consistent with published numbers (Perry *et al.* 2010; Chiou and Bergey 2018). 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 on the microarray. By decreasing these non-specific DNA concentrations through enrichment, the orangutan DNA was able to be amplified and bind 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 *et al.* 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*, e.g., divergent at least 23 million years from each other; 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*; e.g., 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 showed 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 genomic polymorphic loci for *P. pygmaeus* of the 15,949 human loci (0.78%) probed for and 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 *et al.* 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 (Utami *et al.* 2002; Nietlisbach *et al.* 2010) 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 FIS values across all loci was negative across SNP loci and very close to 0 for STR loci. This negative FIS value in the SNP data suggests an excess of heterozygosity and the possibility of two formerly distinct groups now having admixture. However, more replicates across individuals are needed to confirm SNP assignment and reduce variance in order to help verify this phenomenon. Seven loci (i.e., 1 STR and 6 SNP) also showed a high departure from HW equilibrium could be the result of loss of heterozygosity through inbreeding within the sampled individuals.

Several comparisons of pairwise estimators of relatedness have detailed differences between various statistical methods categorized as either moment and likelihood methods. Wang (2007) compared the *Trio*ML maximum likelihood estimator to several moment estimators including his newest moment estimator (referred to as *Wang* in Table 4) as well as more commonly used moment estimator by *Queller* and *Goodnight* (*Q* and *G*) and others. This comparison showed that the *Trio*ML produces the most accurate estimates for large datasets both SNP and STR. Subsequent publications have shown

Table 4. Mean sample wide relatedness values for each dataset

Dataset	#of loci	N	*Trio*ML (*r*)	*var*	*Wang* (*r*)	*var*	*Q* and *G* (*r*)	*var*
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

Each dataset, number of loci used and how many individuals were included n, each of the three relatedness estimators TrioML, *Wang*, and *Queller* and *Goodnight Q* and *G*, followed by the variance for each estimate

that in small datasets, where allele frequencies are estimated from the sample for which one is calculating relatedness (as is the case in our dataset), that population wide average moment estimates are expected to be negative and very close to 0. However, the modified product moment estimators by Wang (2014, 2017) and Ritland (Lynch and Ritland 1999) give the least biased pairwise estimates for high and low related individuals respectively. The *Wang* pairwise relatedness estimated for the known mother-offspring pairs in our dataset indeed appeared to be closest to the expected 0.5. Through simulations with empirical data Taylor (2015) found that the *Trio*ML estimator correlated closest to "true" relatedness and was least biased for estimates of pairs in high relatedness categories. Thus, it appears that there may be combinations of dataset sizes and types, and estimator types that produce the most accurate estimates for different levels of relatedness. The three relatedness estimators clearly produce quantitatively different sample wide averages in our study across all three of the datasets, with *Trio*ML producing the highest *r* values, *Wang* producing a lower and very close to 0 overall *r*, and *Q* and *G* producing more negative relatedness values (Table 4.). These differences in our overall r estimates reflect different estimator biases. The inclusion of further genotyping from more individuals as well as more individuals with known relatedness would help to confirm if this is the case.

However, relatedness estimates calculated using the three estimators in our study were not significantly different between the STR and SNP datasets. 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. Additionally, we were able to confirm a majority of the known mother-offspring pairs with both datasets. The one mother offspring pair showing lower a relatedness of 0.23 could be due to misidentification of a sample taken with related mothers and offspring in near vicinity during sample collection. Overall, the combined results suggest sufficiently similar pairwise relatedness estimates between both of the SNP and STR genotyping methods and that this method can be used to assign relatedness in unknown individuals and is useful for understand orangutan demographics.

This study verified the existence of a closely related group of wild orangutans including adult males and females who are local to the Camp Leakey Study area. All sampled individuals appear to be from a dispersed but related community. More work will be needed to explicate the relationships in detail. 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. 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 reduce inhibitors before conducting SNP analysis. The fact we produced similar relatedness and diversity estimates 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. New genotyping technologies based on the genetic closeness between humans and apes provide opportunities for understanding wild ape populations and providing critical genetic data to support conservation efforts. 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.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This study would not have been possible without the support of The Indonesian Ministry of Research, Technology and Higher Education (RISTEKDIKTI), the Indonesian Institute of Sciences (LIPI), The Directorate General of Nature Resources and Ecosystem Conservation, Indonesian Ministry of Environment and Forestry, Tanjung Putting National Park (TNTP), The University of Palangka Raya, Central Kalimantan, and Center for International Cooperation in Sustainable Management of Tropical Peatland (CIMTROP). The authors wish to thank Orangutan Foundation International (OFI) and all the local field assistants associated with Camp Leakey and OFI who helped in sample collection, transport, and processing. This study was funded in part by CWU Graduate and Research office, Graduate Fellowship from CWU GEAR up, Pete and Sandra Barlow Award and Arlen and Debra Prentice Scholarship, as well as Orangutan Foundation International (OFI).

References

Arora, N., Nater, A., Schaik, C.P. van, Willems, E.P., van Noordwijk, M.A., Goossens, B., Morf, N., Bastian, M., Knott, C., Morrogh-Bernard, H., Kuze, N., Kanamori, T., Pamungkas, J., Dyah Perwitasari-Farajallah, E., Krützen, M., Goodman, M., 2010. Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). Proc. Natl. Acad. Sci. U. S. A. 107, 21376–21381. https://doi.org/10.1073/pnas. 1010169107

Arora, N., van Noordwijk, M.A., Ackermann, C., Willems, E.P., Nater, A., Greminger, M., Nietlisbach, P., Dunkel, L.P., Utami Atmoko, S.S., Pamungkas, J., Perwitasari-Farajallah, D., Schaik, C.P. van, 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*). Mol. Ecol. 21, 3352–3362.

Bourgeois, S., Mcewing, R., Senn, H., Kaden, J., Taggart, J.B., Ogden, R., Jeffery, K.J., Bunnefeld, N., Abernethy, K., McEwing, R., 2018. Single nucleotide polymorphism discovery and panel characterization in the African forest elephant. Ecol. Evol. 24, 2207–2217. https://doi.org/10.1002/ece3.3854

Chiou, K.L., Bergey, C.M., 2018. Methylation-based enrichment facilitates low-cost, noninvasive genomic scale sequencing of populations from feces. Sci. Rep. 8, 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. J. Clin. Microbiol. 38, 1747–1752.

Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20. https://doi.org/10.18637/jss.v022.i04

Fraige, K., Travensolo, R.F., Carrilho, E., 2013. Analysis of seven STR human loci for paternity testing by microchip electrophoresis. Brazilian Arch. Biol. Technol. 56, 213–221. https://doi.org/10.1590/S1516-89132013000200006

Galdikas, B.M.F., 1979. Orangutan adaptation at Tanjung Puting Reserve: mating and ecology, in: Hamburg, D.A., McCown, E.R. (Eds.), The Great Apes: Perspectives on Human Evolution Volume V. The Benjamin/Cummings Publishing Company, Menlo Park, Calif, pp. 195–233.

Galdikas, B.M.F., 1982. Orangutan tool-use at Tanjung Puting Reserve, central Indonesian Borneo (Kalimantan Tengah). J. Hum. Evol. 10, 19–33.

Galdikas, B.M.F., 1985. Orangutan sociality at Tanjung Puting. Am. J. Primatol. 9, 101–119.

Galdikas, B.M.F., 1988. Orangutan diet, range, and activity at Tanjung Puting, Central Borneo. Int. J. Primatol. 9, 1–35. https://doi.org/10.1007/BF02740195

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. Conserv. Genet. 1, 157–162. https://doi.org/10.1023/a:1026535006318

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

Goossens, B., Setchell, J.M., James, S.S., Funk, S.M., Chikhi, L., Abulani, A., Ancrenaz, M., Lackman-Ancrenaz, I., Bruford, M.W., 2006. Philopatry and reproductive success in Bornean orang-utans (*Pongo pygmaeus*). *Molecular Ecology*, 15, 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., Mayer, R.A., Sun, B., Hsie, L., Robbins, C.M., Brody, L.C., Wang, D., Lander, E.S., Lipshutz, R., Fodor, S.P.A., Collins, F.S., 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat. Genet. 22, 164–167. https://doi.org/10.1038/9674

Higuchi, R., Fockler, C., Dollinger, G., Watson, R., 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Bio/Technology 11, 1026-1030. https://doi.org/10.1038/nbt0993-1026

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, 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, 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. Mol. Ecol. Resour. 10, 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. Mol. Ecol. 16, 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kanthaswamy, S., Capitanio, J.P., Dubay, C.J., Ferguson, B., Folks, T., Ha, J.C., 2009. Resources for genetic management and genomics research on non-human primates at the National Primate Research Centers (NPRCs). J. Med. Primatol. 38, 17–23. https://doi.org/10.1111/j.1600-0684.2009.00371.x

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

Lynch, M., Ritland, K., 1999. Estimation of pairwise relatedness with molecular markers. Tbe Genet. Soc. Am. 152, 1753–1766.

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. Mol. Ecol. Resour. 11, 314–322.

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. Int. J. Primatol. 32, 362–376.

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*). Conserv. Genet. Resour. 2, 153–158.

Norman, A.J., Street, N.R., Spong, G., 2013. De novo SNP discovery in the scandinavian brown bear (*Ursus arctos*). PLoS One 8, 1–12. https://doi.org/10.1371/journal.pone.0081012

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. Mol. Ecol. 13, 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. Conserv. Genet. Resour. 4, 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 Res. Notes 6, 1–6.

Perry, G.H., Marioni, J.C., Melsted, P., Gilad, Y., 2010. Genomic-scale capture and sequencing of endogenous DNA from feces. Mol. Ecol. 19, 5332–5344.

Queller, D.C., Goodnight, K.F., 1989. Estimating relatedness using genetic markers. Evolution. 43, 258–275.

R Core Team, 2014. R: a language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. Available at: http//www.R-project.org/. [Date accessed: 4 December 2019]

Rianti, P., Perwitasari-Farajallah, D., Sajuthi, D., Pamungkas, J., Nater, A., Krützen, M., 2015. Identification of diagnostic mitochondrial DNA single nucleotide polymorphisms specific to Sumatran Orangutan (*Pongo abelii*) populations. HAYATI J. Biosci. 22, 149–156. https://doi.org/10.1016/j.hjb.2015.09.002

Ross, C.T., Weise, J.A., Bonnar, S., Nolin, D., Satkoski Trask, J., Smith, D.G., Ferguson, B., Ha, J., Kubisch, H.M., Vinson, A., 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*). Am. J. Primatol. 76, 313–324.

Rousset, F., 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

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. J. Wildl. Manage. 73, 1430–1435.

Simons, N.D., Lorenz, J.G., Sheeran, L.K., Li, J.H., Xia, D.P., Wagner, R.S., 2012. Noninvasive saliva collection for DNA analyses from free-ranging Tibetan Macaques *(Macaca thibetana). American Journal of Primatology*. 74, 1064–1070. https://doi.org/10.1002/ajp.22062

Taylor, H. R. 2015. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. Ecology and Evolution. 5, 3140–3150. https://doi.org/10.1002/ece3.1541

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. Behav. Ecol. 13, 643–652.

Utami-Atmoko, S.S., Traylor-Holzer, K., Rifqi, M.A., Siregar, P.G., Achmad, B., Priadjati, A., Husson, S., Wich, S., Hadisiswoyo, P., Saputra, F., Campbell-Smith, G., Kuncoro, P., Russon, A., Voigt, M., Santika, T., Nowak, M., Singleton, I., Sapari, I., Meididit, A., Chandradewi, D.S., Ripoll Capilla, B., Ermayanti, Lees, C.M., 2017. Orangutan population and habitat viability assessment: final report. IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, MN.

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

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

Wang, J., 2007. Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet. Res. 89, 135–153. https://doi.org/10.1017/S0016672307008798

Wang, J., 2011. Coancestry: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol. Ecol. Resour. 11, 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. J. Evol. Biol. 27, 518–530. https://doi.org/10.1111/jeb.12315

Wang, J., 2017. Estimating pairwise relatedness in a small sample of individuals. Nat. Publ. Gr. 119, 302–313. https://doi.org/10.1038/hdy.2017.52

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

Supplementary S1. SNP Loci
Di-allelic human SNP loci, human chromosome number and MapInfo coordinate for SNP, number of individuals genotyped, observed heterozygosity HO and expected heterozygosity HE, polymorphic information content PIC, p-value of probability test for deviation from Hardy-Weinberg equilibrium, and average FIS (Weir and Cockerham 1984)

SNP name	Human chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} Wa and 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
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

Supplementary S1. Continued

SNP name	Human chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} Wa and C
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
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
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

Supplementary S1. Continued

SNP name	Human chrom	Coordinate	*n*	H_O	H_E	PIC	P-val	S.E.	F_{IS} Wa and C
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

*indicates significant deviation from HW equilibrium