Identification of Diagnostic Mitochondrial DNA Single Nucleotide Polymorphisms Specific to Sumatran Orangutan (Pongo abelii) Populations

Puji Rianti,1,2* Dyah Perwitasari-Farajallah,1,2 Dondin Sajuthi,1,3 Joko Pamungkas,1,3 Alexander Nater,4,5 Michael Krützen1,5

1 Primate Research Center, Bogor Agricultural University, Bogor 16151, Indonesia.
2 Department of Biology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Darmaga Campus, Bogor 16680, Indonesia.
3 Faculty of Veterinary, Bogor Agricultural University, Darmaga Campus, Bogor 16680, Indonesia.
4 Evolutionary Biology, Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden.
5 Anthropological Institute and Museum, University of Zürich, CH-8057 Zürich, Switzerland.

1. Introduction

Mitochondrial DNA is transmitted exclusively from females to their offspring and can thus be used to trace matrilineal ancestry (Brown et al. 1982). The hypervariable region I (HVR-I) is located within the control region of mitochondrial DNA and evolves rapidly (Lau et al. 1998). Thus, the HVR-I is ideally suited to detect genetic variability at the intraspecific level (Tamura and Nei 1993; Brumfield et al. 2003). Furthermore, focusing on among-individual variation may improve the success of conservation programs aiming to revitalize declining populations and species (Forsman 2014).

Most mutational events in HVR-I are commonly known as single nucleotide polymorphisms (SNPs) (Fumagalli et al. 1989; Morin et al. 2004). SNPs displaying unique nucleotide substitutions within a particular population, but not in others, are diagnostic in that they can be used to assign individuals unequivocally to their population of origin (Yang et al. 2007; McTavish and Hillis 2015). The use of diagnostic SNPs in mitochondrial DNA has led to a detailed analysis of matrilineal ancestry in humans (der Sarkissian et al. 2014; Xavier et al. 2015), great apes and monkeys (Sharma et al. 2012; Prado-Martinez et al. 2013; Baden et al. 2014; Kopp et al. 2015) and numerous other organisms (Dudgeon et al. 2012; Matte et al. 2013; Hassanin et al. 2014; Shamblin et al. 2014). Hereinafter, HVR-I can be helpful in distinguishing populations within a species (Burckhardt et al. 1999; Arora et al. 2010), as an individual DNA fingerprinting (DeSalle and Amato 2004), especially in cases of strong female philopatry, such as orangutans (van Noordwijk et al. 2012; Kopp et al. 2014).

* Corresponding author.
E-mail address: pujirianti@ipb.ac.id (P. Rianti).
Peer review under responsibility of Institut Pertanian Bogor.

http://dx.doi.org/10.1016/j.hjb.2015.09.002
1978-3019/© 2015 Institut Pertanian Bogor. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
In orangutans (genus: *Pongo*), mitochondrial DNA haplotypes have been used to distinguish between both currently recognized orangutan species (*Groves 2001; Brandon-Jones et al. 2004*), as well as between populations within each island (Borneo: *P. pygmaeus*; Sumatra: *P. abelii*) (*Steiper 2006; Nater et al. 2011*). This is possible due to the remarkably strong natal female philopatry (*Goossens et al. 2006; Arora et al. 2012; van Noordwijk et al. 2012*) with strongly male biased dispersal (*Nietlisbach et al. 2012*), which has been documented mainly for Bornean orangutans. Although social structure differs between Sumatran and Bornean orangutans (*van Schaik et al. 2009*), general dispersal patterns are the same (*Nietlisbach et al. 2012*). To date, however, it still remains unclear whether the HVR-I alone will be sufficient to assign Sumatran orangutans to their population of origin.

Sumatran orangutans are currently listed as critically endangered (*Singleton et al. 2008*). They occur in the north of the island of Sumatra (tropical forests in the Aceh and North Sumatra provinces) and approximately 6600 individuals are left in the wild (*Figure 1; Wich et al. 2008*). The generally small population sizes of Sumatran orangutans combined with the high habitat fragmentation caused their conservation status to be declared as “critically endangered” (*Singleton et al. 2008*). Here, provenance information based on mitochondrial DNA haplotype information would allow confiscated orangutans to be traced back to their population of origin and would thus significantly enhance conservation efforts. To do so, however, detailed information about the extent and distribution of genetic diversity of the HVR-I in Sumatran orangutan is needed, which we provide in this article.

2. Materials and Methods

2.1. Samples

We used fecal and hair samples of wild Sumatran orangutan from nine sampling locations in Aceh and North Sumatra during 2005–2012 (*Figure 2; Table 1*). We also used blood from wild-born orangutan held in the Sumatran Orangutan Conservation Program “SOCP” Batu Mbeling rehabilitation center, North Sumatra (*Nater et al. 2013*). All orangutan samples (blood, hair and fecal) were collected under the research permit of the Indonesian Ministry of Forestry, and were preserved in EDTA, ethanol 90%, RNA later or silica gel, respectively (*Nsubuga et al. 2004; Nater et al. 2011*). Samples were transported to Bogor Agricultural University and from there to the Anthropological Institute and Museum, University of Zurich, Switzerland, using the Convention on International Trade in Endangered Species permit number 00670/IV/SATS-LN/2013; 09717/IV/SATS-
LN/2010; 07279/IV/SATS-LN/2009; 00961/IV/SATS-LN/2007 and 06968/IV/SATS-LN/2005. In total, we (re)analyzed 54 samples to determine population clades and diagnostic SNPs specific to Sumatran orangutan populations.

2.2. DNA sequencing

Laboratory work was carried out largely as previously described (Arora et al. 2010). We amplified the HVR-I of 54 Sumatran orangutans using primers DLF (5’CCTGCCCCTGTAGTACAAATAAGTA3’) and

Figure 2. Nine sampling locations of Sumatran orangutan (Pongo abelii) populations. North Aceh (north of Tamiang River); Blangkejeren (north of Alas River); Langkat (east of Alas River, south of Tamiang River); BatuArdan (east of Alas River, northwest of Toba Lake); Batang Toru (south of Lake Toba); Pakpak Bharat (southeast of Alas River, southwest of Toba Lake); Suaq Balimbing (Kluet swamp, northwest of Alas River); Ketambe (north of Alas River); and Tripa (north and west of Tripa River).

Table 1. List of Sumatran orangutan samples used in this study

<table>
<thead>
<tr>
<th>Collector</th>
<th>No. samples</th>
<th>Type</th>
<th>Location</th>
<th>Sample origin (information)</th>
<th>Coll. year</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>6</td>
<td>Fecal</td>
<td>SQ</td>
<td>Suaq Balimbing</td>
<td>2006–2008</td>
</tr>
<tr>
<td>AN</td>
<td>6</td>
<td>Fecal</td>
<td>KE</td>
<td>Desa Tanjung Muda, Desa Rambah Sayang, South East Aceh</td>
<td>2005; 2008</td>
</tr>
<tr>
<td>AN</td>
<td>5</td>
<td>Hair</td>
<td>BA</td>
<td>Sidikalang, Perolihen, Sidiangkat</td>
<td>2009</td>
</tr>
<tr>
<td>AN</td>
<td>1</td>
<td>Blood</td>
<td>BA</td>
<td>Quarantine (BatuArdan)</td>
<td>2008</td>
</tr>
<tr>
<td>AN</td>
<td>5</td>
<td>Blood</td>
<td>BK</td>
<td>Quarantine (Sayo Luwes, Blangkejeren)</td>
<td>2004–2005</td>
</tr>
<tr>
<td>AN</td>
<td>1</td>
<td>Hair</td>
<td>BK</td>
<td>Blangkejeren</td>
<td>2005–2009</td>
</tr>
<tr>
<td>AN</td>
<td>5</td>
<td>Blood</td>
<td>TR</td>
<td>Quarantine (Tripa Swamp)</td>
<td>2009</td>
</tr>
<tr>
<td>AN</td>
<td>1</td>
<td>Hair</td>
<td>TR</td>
<td>Tripa Swamp</td>
<td>2007</td>
</tr>
<tr>
<td>AN</td>
<td>4</td>
<td>Blood</td>
<td>NA</td>
<td>Quarantine (Takengon, North Aceh)</td>
<td>2007; 2009</td>
</tr>
<tr>
<td>AN</td>
<td>2</td>
<td>Hair</td>
<td>NA</td>
<td>Takengon, North Aceh</td>
<td>2007–2009</td>
</tr>
<tr>
<td>AN</td>
<td>5</td>
<td>Fecal</td>
<td>BK</td>
<td>Tampkahan, Sampan Getek, aras Napal, Aceh Tamiang</td>
<td>2008</td>
</tr>
<tr>
<td>AN</td>
<td>1</td>
<td>Hair</td>
<td>BK</td>
<td>Langkat</td>
<td>2010–2011</td>
</tr>
<tr>
<td>PR</td>
<td>5</td>
<td>Hair</td>
<td>PB</td>
<td>Buluh Didi, Loe Meang</td>
<td>2011–2012</td>
</tr>
<tr>
<td>PR</td>
<td>1</td>
<td>Fecal</td>
<td>PB</td>
<td>Singkil</td>
<td>2011–2012</td>
</tr>
<tr>
<td>PR</td>
<td>6</td>
<td>Fecal</td>
<td>BT</td>
<td>West Batang Toru, East Tapanuli</td>
<td>2011–2012</td>
</tr>
</tbody>
</table>

AN (Nater et al. 2013); PR (this study).

SQ = Suaq Balimbing; KE = Ketambe; BA = BatuArdan; PB = Pakpak Bharat; BK = Blangkejeren; TR = Tripa; NA = North Aceh; LK = Langkat; BT = Batang Toru.
D5 (5’TGTCCGGATATTGGATTACG3’) (Warren et al. 2001; Arora et al. 2010; Nater et al. 2013). We modified the polymerase chain reaction step by using 45 cycles of 30 seconds at 58°C for annealing. Electropherograms were analyzed using Sequencing Analysis, v5.2, and edited manually. All sequences were assembled using SeqMan (Lasergene v8; DNASTAR).

2.3. Genetic variation and statistical analysis

We aligned all resulting sequences using clustal W algorithm (Thompson et al. 1994; Larkin et al. 2007) as implemented in BioEdit, v7.2.5 (Hall et al. 2011). This was followed by DNAsp, v5.10.01 (Librado and Rozas 2009) to detect unique haplotypes within sampling locations. To identify diagnostic SNPs and population clades, we carried out a phylogenetic tree reconstruction based on maximum likelihood (ML with 1000 bootstrap replicates) in MEGA6 software (Tamura 2012), applying the Tamura–Nei model (Tamura and Nei 1993). Based on the resulting mtDNA clades, we estimated nucleotide substitution rates, polymorphic sites, and nucleotide and haplotype diversity for each clade. We also carried out an analysis of molecular variance and calculated the average population pairwise differences using Arlequin, v3.5.1.2 (Excoffier and Lischer 2010).

3. Results

We sequenced 422 base pairs (bp) of HVR-I mitochondrial DNA sequences from nine Sumatran orangutan sampling locations (54 samples). In total, we found 52 diagnostic SNPs with substitutions specific to single population clades (Table 2). Overall, we observed 20 haplotypes. One haplotype was extremely common and occurred 17 times in our data set in five sampling locations (Figure 3). Most of the other haplotypes were only observed once or twice. Population clades based on the maximum likelihood phylogenetic tree showed five major HVR-I mitochondrial DNA lineages among the nine sampling locations. Two population clades comprised a single sampling location each: Batang Toru (C-BT) and Tripa (C-TR). The other three population clades were heterogeneous with respect to sampling locations. The largest population clade combined HVR-I mitochondrial DNA haplotypes from Pakpak Bharat, Batu Ardan, Suau Balimbing, Ketame, and Blangkejeren (C-AR). Besides its own unique haplotypes, the North Aceh (C-NA) and Langkat (C-LK) population clades contained haplotypes from Blangkejeren (Figure 3), an area centrally located between C-NA, C-LK and C-AR populations (Figure 2). We also observed an insertion—deletion specific to C-BT.

Overall, there were 91 variable sites within a total of 95 mutations in the aligned sequences (Table 3), indicating high genetic variation. Haplotype diversity (standard deviation) \( h = 0.880 \) (0.035) and nucleotide diversity per site (standard deviation) \( \pi = 0.088 \) (0.043). The analysis of molecular variance with fixation index showed high variance among five population clades at 96.43% (FST = 0.964; \( p < 0.05 \)). Our results point to strong inter-population differentiation for HVR-I mitochondrial DNA, demonstrated by the high and significant \( F_{ST} \) values for all 10 population pairs (\( p < 0.05 \); Table 4).

4. Discussion

Our analyses revealed five major genetic matrilineal clades in Sumatran orangutan populations, which can be differentiated by 52 diagnostic SNPs. Given the pronounced tendency of female philopatry in orangutans, as at least documented in Borneo (Arora et al. 2012), this information provides a basis for linking Sumatran orangutans currently held in rehabilitation centers to their putative population of origin. Identifying source populations especially in Sumatra is extremely crucial, because the genetic divergence among populations is much deeper compared to Borneo (Warren et al. 2001; Arora et al. 2010). Moreover, Sumatran orangutans exhibit a large genetic differentiation of populations between north and south area of Lake Toba (Nater et al. 2011). However, this is not reflected in the current Sumatran orangutan taxonomy, which still considers all Sumatran orangutans to be member of a single species without any division into subspecies. Because of this, caution should be exerted when releasing Sumatran orangutans of unknown origin in the same area, due to potential mixing of individuals from different gene pools, which might lead to outbreeding depression (Moritz 1999; Frankham 2010). It can also present problems in terms of social interactions, for instance when unrelated females...
Figure 3. Hypervariable region I (HVR-I) mitochondrial DNA haplotypes. Maximum likelihood tree of partial HVR-I mitochondrial DNA sequences, with bootstrap support values in percent (Tamura-Nei model; bootstrap ¼ 1000). C-AR: Alas River population clade (Pakpak Bharat, BatuArdan, Ketambe, Suaq Balimbing and part of Blangkejeren sampling locations); C-TR: Tripa population clade; C-NA: North Aceh population clade; C-LK: Langkat population clade; C-BT: Batang Toru population clade.

Diagnostic SNPs to Sumatran orangutan populations 153
are forced into the home ranges of a group of related females they may be chased away.

One of the ultimate goals of genetic conservation is to ensure the long-term persistence of species, mainly through the maintenance of intra-population genetic diversity (Kahlaien et al. 2014), which still appears to be at appreciable levels in Sumatran orangutans. In contrast to previous studies where four geographically distinct haplogroups were reported (Nater et al. 2011; Nater et al. 2013), our results indicate an additional fifth mitochondrial DNA clade (Tripa location), which is different from the remaining Alas River clade (called “West Alas” in the Nater et al. 2013). Tripa is a peat-swamp forest on the west coast of Aceh, with an estimated census size of approximately 280 individuals in a 140 km² habitat (Wich et al. 2008). This location has encountered a massive oil palm conversion, with only 23.53% of its original size remaining relatively undisturbed (van Schaik et al. 2001; Gaveau et al. 2009), given the extensive habitat loss and anthropogenic pressure this location is currently facing (van Schaik et al. 2001; Wich et al. 2008; Nater et al. 2013). Therefore, further investigation is needed to address whether more detailed work is required to assess the conservation status of orangutans from this area.

We found one sampling location (Blangkejeren, a dry highland area at an altitude up to 1000 m, which is located in Central Aceh), sharing haplotypes with three other populations (North Aceh, Langkat and Alas River). Blangkejeren is located at the northern headwaters of the Alas River (Barber et al. 2005), which does not act as an effective barrier to gene flow in the area. It is also centrally positioned among the populations of North Aceh, Langkat and Alas River (Ketambe), to all of which it is roughly equidistant. Further sampling, ideally involving samples from wild orangutans or samples from rehabilitation centers with good provenance, is required to investigate the status of the Blangkejeren highland area due to its central location and unique ecology.

An earlier study based on mitochondrial genes suggested that the Batang Toru lineage diverged from all other orangutan lineages in both Sumatra and Borneo around 3.5 million years ago (Nater et al. 2011). Based on this and more recent autosomal evidence (Nater et al. 2013), the Batang Toru population is genetically the most isolated in Sumatra. Therefore, it needs further investigation whether this population might comprise a new subspecies, although currently no obvious morphological differences have been documented. Nonetheless, the Batang Toru orangutan population requires a dedicated conservation effort, due to its isolation (small census size with approximately 550 individuals in an area of ca. 975 km² at about 200–1500 m above sea level) (Wich et al. 2008). This density is much higher than that thought to be sustainable for semi-solitary orangutans (Singleton and van Schaik 2001). An adult male of Sumatran orangutan require ca. 4.50 km² of home range with ca. 1 km² daily travel distance for both sexes (Campbell-Smith et al. 2011). Sumatran orangutans need large habitats close to streams, rivers and swamps with abundant availability of soft-pulp fruit as main food sources (Delgado and van Schaik 2000; Wich et al. 2011a). We are questioning the possibility of the habitat supporting such high densities of orangutans, within this geographically isolated area, without natural chances of dispersal in any way.

From a genetic diversity perspective, orangutans have the highest genetic diversity among all great apes (Steiper and Young 2006; Prado-Martinez et al. 2013). Orangutan populations in Sumatra are thought to be very old, based on the deep phylogenetic splits which date back hundreds of thousands to millions of years (Nater et al. 2011). This is in stark contrast to Borneo, where for mitochondrial DNA all populations appear to consolidate around 176 thousands years ago (Arora et al. 2010). This remarkable stability of Sumatran populations is reflected in our data set in the higher diversity estimates in Sumatra compared to Borneo (Arora et al. 2010). However, these idiosyncrasies are not yet taken into account by the current orangutan taxonomy, which suggests three subspecies for Borneo, but none for Sumatra (Groves 2001). Thus, there is an urgent need to revise the taxonomy of Sumatran orangutans by taking genetic information into account.

In conservation genetics, methods like genetically informed demography-based approaches, cladistics diversity measures, nested clad analysis and diagnostic SNPs are frequently used to assist in comprehensive decision-making (DeSalle and Amato 2004; Morin et al. 2004). The haplotypes described in this study provide a first step at documenting the genetic diversity of all extant Sumatran orangutan populations. However, our data are not sufficient yet to advocate taxonomic revisions. Broader genetic population analyses and different marker systems will be needed to do so. A pilot survey to the southern area of Batang Toru and historic DNA analysis using museum samples will sharpen the status recommendation. Based on this present study, restriction fragment length polymorphism assays can be developed to carry out genetic assignments using basic laboratory equipment. Our study also addresses the requirements by the “Strategi dan Rencana Aksi Konservasi Orangutan Indonesia 2007–2017” (Indonesian National Orangutan Conservation Strategy and Action Plan 2007–2017; Soeharto et al. 2007). This Indonesian regulation on orangutan conservation already requires to take action for stabilizing the minimum population in the wild by educating the local communities as well as public and strengthening the national conservation law to limit hunting, illegal trading and primary land forest conversion (Indonesia 1990; Meijsnaard et al. 2011; Wich et al. 2012). The regulation also specifies the reintroduction of all individuals in captivity, back to their natural habitat by 2015. Our study supports the latter task by giving compiled genetic information, allowing the

<table>
<thead>
<tr>
<th>Nucleotide substitution rates</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
<th>Polymorphic sites</th>
<th>Singleton variable sites</th>
<th>Parsimony informative sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenine (A)</td>
<td>–</td>
<td>1.88</td>
<td>3.50</td>
<td>7.40</td>
<td>Two variants</td>
<td>4</td>
<td>83</td>
</tr>
<tr>
<td>Thymine (T)</td>
<td>2.88</td>
<td>33.27</td>
<td>6.92</td>
<td>–</td>
<td>Three variants</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Cytosine (C)</td>
<td>2.88</td>
<td>17.91</td>
<td>6.92</td>
<td>–</td>
<td>Four variants</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Guanine (G)</td>
<td>23.05</td>
<td>1.88</td>
<td>3.50</td>
<td>–</td>
<td>Total</td>
<td>4</td>
<td>87</td>
</tr>
</tbody>
</table>

Each entry on nucleotide substitution shows the rates of substitution (r) from one base (row) to another base (column); Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics; sum of r values = 100.
identification of orangutan management units, particularly for the conservation efforts of Sumatran orangutans in its natural habitat.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The authors would like to thank all field assistants and local societies for their help in sample collections. A tremendous appreciation for the field and collection support of the following institutions: the Indonesian Ministry of Forestry, Direktorat Jendral Perlindungan Hutan dan Sumber Daya Alam (PHKA), Indonesian Science Institute (LIPI), Balai Besar Konservasi Sumber Daya Alam Sumatera Utara (BBKSDA SUMUT), Badan Pengelola Kawasan Ekosistem Leuser (BPKEL), Taman Nasional Gunung Leuser (TNGL), Sumatran Orangutan Center Program—Yayasan Eko Lestari (SOCYEL), and Primate Research Center Bogor (PSS-IPB). The authors thank Serge A. Wich and Ian Singleton [Yayasan Eko Lestari—Foundation for a Sustainable Ecosystem (YEL), Medan, Indonesia] for the valuable sampling opportunities in Batang Toru and a deep discussion about this area. This study was financially supported by Swiss National Foundation (SNF) grant to Michael Krützen and Carel van Schaik, period 2010–2014.

References
