Elsevier Editorial System(tm) for Biological Conservation

Manuscript Draft

Manuscript Number: BIOC-D-15-01082R2

Title: Habitat fragmentation and genetic diversity in natural populations of the Bornean elephant: implications for conservation

Article Type: Full length article

Keywords: Bornean elephant; SNPs; microsatellites; population genetics

Corresponding Author: Dr Benoit Goossens, PhD

Corresponding Author's Institution: Cardiff University

First Author: Benoit Goossens, PhD

Order of Authors: Benoit Goossens, PhD; Reeta Sharma; Nurzhafarina Othman; Célia Kun-Rodrigues; Rosdi Sakong; Marc Ancrenaz; Laurentius N Ambu; Nathaniel K Jue; Rachel J O'Neill; Michael W Bruford; Lounès Chikhi

Abstract: The Bornean elephant population in Sabah, with only 2,000 individuals, is currently mainly restricted to a limited number of forest reserves. The main threats to the species' survival are population fragmentation and isolation of the existing herds. To support and help monitor future conservation and management measures, we assessed the genetic diversity and population structure of Bornean elephants using mitochondrial DNA, microsatellites and single nucleotide polymorphisms. Our results confirmed a previously reported lack of mitochondrial control region diversity, characterized by a single widespread haplotype. However, we found low but significant degree of genetic differentiation among populations and marked variation in genetic diversity with the other two types of markers among Bornean elephants. Microsatellite data showed that Bornean elephants from the Lower Kinabatangan and North Kinabatangan ranges are differentiated and perhaps isolated from the main elephant populations located in the Central Forest and Tabin Wildlife Reserve. The pairwise FST values between these sites ranged from 0.08 to 0.14 ($p < 0.001$). Data from these markers also indicate that the Bornean elephant populations from Lower Kinabatangan Wildlife Sanctuary and North Kinabatangan (Deramakot Forest Reserve) possess higher levels of genetic variation compared to the elephant populations from other areas. Our results suggest that (i) Bornean elephants probably derive from a very small female population, (ii) they rarely disperse across current human-dominated landscapes that separate forest fragments, and (iii) forest fragments are predominantly comprised of populations that are already undergoing genetic drift. To maintain the current levels of genetic diversity in fragmented habitats, conservation of the Bornean elephants should aim at securing connectivity between spatially distinct populations.

Letter to the Editor

Date: 27 January 2016

Dr. Lian Pin Koh, Ph.D.
Indo-Pacific, Australasia Editor
Biological Conservation

Manuscript Ref. No.: BIOC-D-15-01082R1
Manuscript title: Habitat fragmentation and genetic diversity in natural populations of the Bornean elephant: implications for conservation
Authors: Benoit Goossens, Reeta Sharma, Nurzhafarina Othman, Célia Kun-Rodrigues, Rosdi Sakong, Marc Ancrenaz, Laurentius N. Ambu, Nathaniel K. Jue, Rachel J. O'Neill, Michael W. Bruford, and Lounès Chikhi

Dear Editor,

Please find our revised manuscript following the very minor suggestions made by the referees. We thank you and the referees for your time and for the positive comments.

We hope that you'll agree that our paper has greatly improved and will be approved by your journal. We look forward to your reply.

Yours sincerely,

Benoit Goossens on behalf of the authors

Reviewers' comments:

Reviewer #2:

Overview

The authors have made many improvements to the manuscript to provide additional clarity, analysis, and remove hidden objectives. I continue to disagree with the authors over the interpretation of their results (lines 507-509 and 609-610) specifically that they observed two population clusters instead of one; but we can disagree respectfully as the presentation of the data allows readers enough information for their own assessment.
I provide a few minor comments below for additional clarity.

REPLY: The authors would like to thank the reviewer for considering our manuscript interesting and accepting it for publication after minor revision. We have incorporated the following reviewer's specific comments in preparation of revised version of manuscript. Below are our responses to the comments.

1. **Abstract**
The abstract could be improved by providing details of the study. For example, instead of saying that significant structure was found, highlight the specific results of two populations and the Fst. Or provide genetic diversity estimates as that is a main result.

REPLY: We have now modified the abstract as suggested. Lines 39-45.

2. Introduction
In lines 81-82, please clarify if the density is for humans or elephants.

REPLY: The given density is for elephant. We have clarified it in the text. Line 85.

3. Methods
Lines 151-154- some of the concerns of both reviewers were not addressed here regarding first the decision making process of selecting 273 samples (only because they were < 2days old?), and second how were the matriarchs identified (using genetic data or observations).

REPLY: The 273 dung samples were only selected according to their age (less than 2 days) as stated in the text. The matriarchs were identified visually. That information is included Line 158.

4. Results
4.1 In lines 403-405, I assume the 28-plex had a lower genotype error rate than the 48-plex; however, it would be helpful to state the error rates between the two multiplex sizes as I think other researchers will be interested in that data.

REPLY: The reviewer is correct that the 28-plex had a lower genotype error rate than the

42-plex. We have now included the error rates of the two multiplex sizes in the text. Lines 413-415.

4.2 Figure S1 would be easier to read in color. Alternatively, if the legend were organized in the same vertical order as the bars, the greyscale may work.

REPLY: Now we have changed Figure S1 in color.

4.3 Table 2 is still cut off, thus unavailable for evaluation. Tables 3 and S3 also appear to be cut off.

REPLY: We apologize for this. It might have happened during the manuscript submission. It should not be a problem for the editorial office since tables are submitted separately.

5. Discussion

5.1 I disagree with the statement in lines 716-718 and feel as though you should provide references of studies that answered their research questions with small numbers of SNPs. Many SNP studies for non-model organisms (including for population structure inference) have 100s to 1000s of loci. I think this is important to the readers of Biological Conservation, as some of them may not fully understand the complexity of the issue of marker type and number specifically as it relates to question type (as you mention) and sample size (which could be emphasized in this section for a fuller discussion).

REPLY: We agree and now we have rephrased this sentence and also added references in support of our text. Lines 727-741.

5.2 Line 723- Only Haasl and Payseur said that SNP panels 15-fold larger than microsatellite panels were needed for recent divergence, Morin et al 2009 promoted increasing sample sizes for better population structure inference.

REPLY: Agreed, we have now removed the reference. Lines 736-737.

Minor

Line 528, "both" is in the wrong place in the sentence.

REPLY: It is now corrected.

Figure 2 is much improved since the last draft. However, in panel b, the blue and grey should be flipped on one of the panels (you can see this well in the 4th Central Forest sampling area).

REPLY: It is now corrected.

Habitat fragmentation and genetic diversity in natural populations of the Bornean elephant: implications for conservation

Benoit Goossens[1,2,3,4*], Reeta Sharma[5*], Nurzhafarina Othman[1,2], Célia Kun-Rodrigues[5], Rosdi Sakong[1], Marc Ancrenaz[6], Laurentius N. Ambu[3], Nathaniel K. Jue[7], Rachel J. O'Neill[7], Michael W. Bruford[1,4], Lounès Chikhi[5,8,9]

[1]Organisms and Environment Division, School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff CF10 3AX, UK

[2]Danau Girang Field Centre, c/o Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia

[3]Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia

[4]Sustainable Places Research Institute, Cardiff University, 33 Park Place, Cardiff CF10 3BA, UK

[5]Population and Conservation Genetics, Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, P-2780-156, Oeiras, Portugal

[6]HUTAN/Kinabatangan Orangutan Conservation Programme, PO Box 17793, 88874 Kota Kinabalu, Sabah, Malaysia

[7]Institute for Systems Genomics and Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, United States of America

[8]CNRS, Université Paul Sabatier, ENFA, UMR 5174 EDB (Laboratoire Evolution & Diversité Biologique), 118 route de Narbonne, F-31062 Toulouse, France

[9]Université Paul Sabatier, UMR 5174 EDB, F-31062 Toulouse, France

*These two authors contributed equally to the work.

Corresponding authors

Email addresses:

goossensbr@cardiff.ac.uk

reeta2000@gmail.com, reetasharma.igc@gmail.com

Keywords: Bornean elephant, SNPs, microsatellites, population genetics

Abstract

The Bornean elephant population in Sabah, with only 2,000 individuals, is currently mainly restricted to a limited number of forest reserves. The main threats to the species' survival are population fragmentation and isolation of the existing herds. To support and help monitor future conservation and management measures, we assessed the genetic diversity and population structure of Bornean elephants using mitochondrial DNA, microsatellites and single nucleotide polymorphisms. Our results confirmed a previously reported lack of mitochondrial control region diversity, characterized by a single widespread haplotype. However, we found low but significant degree of genetic differentiation among populations and marked variation in genetic diversity with the other two types of markers among Bornean elephants. Microsatellite data showed that Bornean elephants from the Lower Kinabatangan and North Kinabatangan ranges are differentiated and perhaps isolated from the main elephant populations located in the Central Forest and Tabin Wildlife Reserve. The pairwise F_{ST} values between these sites ranged from 0.08 to 0.14 ($p < 0.001$). Data from these markers also indicate that the

Bornean elephant populations from Lower Kinabatangan Wildlife Sanctuary and North Kinabatangan (Deramakot Forest Reserve) possess higher levels of genetic variation compared to the elephant populations from other areas. Our results suggest that (i) Bornean elephants probably derive from a very small female population, (ii) they rarely disperse across current human-dominated landscapes that separate forest fragments, and (iii) forest fragments are predominantly comprised of populations that are already undergoing genetic drift. To maintain the current levels of genetic diversity in fragmented habitats, conservation of the Bornean elephants should aim at securing connectivity between spatially distinct populations.

1. Introduction

The Asian elephant (*Elephas maximus*) once ranged from Mesopotamia in the west across the Indian subcontinent to South-east Asia (including the islands of Sumatra, Java and Borneo) and China and as far north as the Yangtze River (Olivier, 1978). As a consequence of habitat loss and fragmentation, it has been extirpated from approximately 85% of its historical range, and only exists in a number of fragmented and isolated populations in South and South-east Asia (Sukumar, 1989; Fernando et al., 2000). As a consequence it is presently classified as 'Endangered' by the IUCN (International Union for Conservation of Nature). Elephants in Borneo are morphologically, and behaviourally distinct from the elephants of mainland Asia (Cranbrook et al., 2008). Also, the genetic distinctiveness of the Bornean elephant from other mainland Asian elephant subspecies makes it one of the highest priority populations for Asian elephant conservation (Fernando et al., 2003). They are considered as an evolutionary significant unit, requiring specific conservation measures,

however, their recognition as formal subspecies (*Elephas maximus borneensis*) status still awaits more detailed study.

Bornean elephants have a very limited distribution, restricted to approximately 5% of the island, at the extreme northeast, mostly in the Malaysian state of Sabah. They are usually found in family groups (from 5 to 20 individuals) that sometimes merge together to form larger herds of up to 200 individuals (Othman, unpublished data). They are found in five major ranges, with a total of 2,000 individuals (95% CI: 1,184 - 3,652, Alfred et al., 2010; Elephant Action Plan, Sabah Wildlife Department, 2012-2016, see Figure 1). The main threat facing the Bornean elephant, as with all mainland Asian elephants, is habitat loss and fragmentation which occurs directly through conversion of existing forests to commercial plantations, such as palm oil, or for permanent human settlement (Elephant Action Plan, Sabah Wildlife Department, 2012-2016). For instance, in the last 50 years, 80% of the Lower Kinabatangan floodplain forest has been converted to agricultural land or used for human settlement (Estes et al., 2012; Abram et al., 2014). The remaining forest in this region is now highly fragmented; yet, it has the highest elephant density of the five ranges with 2.15 individuals per km^2 (Alfred et al., 2010; Estes et al., 2012). With increasing elephant density, the risk for human-elephant conflict and associated human and elephant mortality also rises (Santiapillai and Ramono, 1993; Williams et al., 2001; Alfred et al., 2011). Connectivity between ranges (i.e. between Lower Kinabatangan and North Kinabatangan, Lower Kinabatangan and Tabin Wildlife Reserve, Tabin Wildlife Reserve and Central Sabah) is now lacking, although elephants are increasingly travelling through oil palm plantations (Goossens, unpublished data), which could act as corridors, but also increase human-elephant conflicts.

Habitat fragmentation and loss can also affect the genetic structure of populations both directly and indirectly by restricting gene flow or increasing the levels of genetic drift and inbreeding (e.g. Reed and Frankham, 2003). Changes in genetic diversity associated with habitat fragmentation have been found in Bornean orang-utans that share the same habitat with the elephants (Goossens et al., 2005). Orang-utan populations in the Lower Kinabatangan region have experienced a dramatic demographic decline and are undergoing rapid genetic differentiation induced by genetic drift as a consequence of anthropogenic isolation (Goossens et al., 2006; Jalil et al., 2008; Sharma et al., 2012a). Due to severe habitat loss and range contraction, the Bornean elephant may therefore be at risk of deleterious population effects such as loss of genetic diversity, inbreeding depression and ultimately extinction (e.g. Templeton et al., 1990; Saccheri et al., 1998). However, in long-lived species with overlapping generations, such as elephants, signatures of genetic loss may be masked for decades or even centuries. Retaining of genetic diversity for longer time periods due to the low reproductive rate and long generation time means that deleterious effects of habitat fragmentation will take longer to manifest themselves. So documenting such impacts can only be done after a long time period has elapsed (Armbruster et al., 1999; Ewers and Didham, 2006). In order to make biologically sound conservation plans, an understanding of the current amount of genetic diversity remaining in natural populations of Bornean elephant and its distribution among populations is essential. This is also important for improving and informing their future management.

Previous studies have found low levels of genetic diversity in Borneo elephant. Fernando et al. (2003) compared mitochondrial DNA (mtDNA) and microsatellite diversity for Asian elephant populations and found a single mitochondrial haplotype in Bornean elephant samples analysed from Sabah. They concluded that the Bornean

elephant split from other *Elephas* subspecies around 300,000 years ago and was therefore indigenous to Borneo following a Pleistocene colonization. Sharma et al. (2012b) used previously identified set of SNPs and confirmed that Bornean elephants exhibited low genetic diversity. In both studies, however, the sample sizes analysed were very small, with a total of 20 and seven individuals, respectively.

Therefore, the main aim of the present study was to (i) assess the level of genetic diversity of Bornean elephant and its distribution across the entire range; and (ii) detect possible genetic differentiation between the different local populations in Sabah using a large sample size from across much of the elephant range. It was addressed by using mtDNA and nuclear genetic markers (i.e. microsatellites and SNPs). This will yield insights into the genetic connectivity of remnant populations and indicate which populations are the most genetically isolated and most in need of restoration management (Frankham et al., 2002).

2. Material and methods

2.1. Study sites and sampling

Elephants in Sabah are distributed in five main ranges: i) Lower Kinabatangan, ii) North Kinabatangan (Deramakot, Tangkulap and Segaliud Forest Reserves), iii) Central Forest (Ulu Segama, Malua, Kuamut, Gunung Rara, and Kalabakan Forest Reserves, Danum Valley and Maliau Basin Conservation Areas), iv) Tabin Wildlife Reserve, and v) Ulu Kalumpang (Alfred et al., 2010; Elephant Action Plan, Sabah Wildlife Department, 2012-2016). All ranges (but Ulu Kalumpang, due to its inaccessibility) were covered and systematically searched for elephant feces between October 2005 and November 2007 (Figure 1). Samples were mainly collected along logging roads where elephants consume grass and travel. The same procedure was conducted for three ranges (Central Sabah, Tabin and North Kinabatangan). For the Lower Kinabatangan population, samples were collected along the main river, in riparian feeding areas where individuals were encountered, allowing collection of fresh samples. Samples were collected during field expeditions of 5-7 days, giving a short time period during which samples were collected from every location. Fresh elephant dungs (less than 1-2 days old) were sampled by collecting approximately 5-10 grams of dung from the outermost layer of intact dung boli into a 50 mL Falcon tube filled with 70% ethanol. Dungs of up to 7-10 days old were also sampled in the absence of fresh dung piles. We collected as many samples as we could to ensure that as many different individuals from social groups could be analysed. The use of microsatellite data would then allow us to discard samples from same individuals. In total, 779 fecal samples were collected across Sabah. GPS coordinates were taken for each sample. Of these 779 samples, 273 were chosen for further analysis. These were fresh feces (between a few hours and two days old), sampled from free-ranging elephants. Out of these 273 elephant feces, 170 were from

14 distinct family groups for which each matriarch was visually identified. The samples used for DNA extraction from each of the forest range were as follows: i) Lower Kinabatangan (LK) (*n*=46), ii) North Kinabatangan (Deramakot (DER); *n*=33), iii) multiple sites in the Central Forest (CF) (Ulu Segama-Malua (USM); *n*=78, Gunung Rara (GR); *n*=32, Kalabakan (KAL); *n*=46, Kuamut (KU); *n*=6, Maliau Basin (MB) Conservation Area; *n*=9), and iv) Tabin Wildlife Reserve (TWR) (*n*=23).

Fresh whole blood samples were also collected from 20 Bornean elephants originating from different locations in Sabah between 2009 and 2011. Seven of these samples were collected from elephants in the Lok Kawi Wildlife Park in Sabah and the original source population is known for these samples. Individuals that were sampled in Sandakan and Lahad Datu areas, originated from the main elephant populations in the LK and CF ranges (Figure 1).

2.2. Molecular Analysis

DNA from elephant fecal samples was extracted using the QIAamp® DNA Stool Mini Kit (QIAGEN Ltd., West Sussex, United Kingdom) with modification. Samples were lysed in ASL buffer and an InhibitEX pill was added. The DNA product was dissolved in 200µL of buffer AE and stored at -20ºC. Two extractions per fecal samples were performed. Genomic DNA was extracted from blood samples using the protocol as described in Sharma et al. (2012b). The extracted DNA from each sample was analysed using three different markers: a) Microsatellites, b) Single nucleotide polymorphisms, and c) Mitochondrial DNA. The total number of samples analyzed for each marker type is not the same (see Table 1).

a) *Microsatellites*

Eighteen dinucleotide microsatellite loci isolated by Kongrit et al. (2007) for the Asian elephant were used to genotype 273 DNA samples and 224 unique elephant individuals were identified. All forward primers were fluorescently labeled and combined in multiplexed polymerase chain reactions (PCR) (see Table S1) and amplified in 10µL volumes using Multiplex PCR Kit containing 5µL of Multiplex Mix, 1µL of 10X primer mix, 1µL of Q buffer, 1µL of water, and 2µL of DNA template (QIAGEN). Amplifications for each sample were repeated three times as follows: initial denaturation at 95°C for 10 minutes; denaturation at 95°C for 45 seconds, annealing at 58°C for 45 seconds, extension at 72°C for one minute and final extension for 10 minutes. Amplification was completed after 45 cycles and included negative controls to check for contamination. A positive control was used to standardize our allele scoring. PCR products were sent to Macrogen Inc, South Korea for genotyping. If analysis of three positive PCRs per locus was not conclusive, we performed the analysis of four additional positive PCRs, as in Taberlet et al. (1996). For peak identification and fragment sizing, Peak Scanner version 1.0 (www.appliedbiosystems.com/peakscanner) was used.

b) *Single Nucleotide Polymorphism (SNPs)*

SNP genotypes were obtained for 70 unique individuals chosen from all four main ranges that were also included in the set of elephants genotyped for microsatellite markers and from 20 blood samples (Figure 1, Table 1). SNP genotyping was performed using the MassARRAY platform from Sequenom (San Diego, CA, USA). Five assays targeting 194 SNPs were selected (Gabriel et al., 2009; Sharma et al., 2012b). Each assay allowed us to co-amplify between 28 and 42 SNP loci and are referred to as “plexes” in the following.

206 Before genotyping the elephant DNA from fecal samples for all plexes, we
207 established the efficiency of SNP genotyping from fecal samples using the Sequenom
208 iPLEX protocol, and optimized the reactions by genotyping a representative sample of
209 48 individuals for two plexes to evaluate DNA quality in the fecal samples that were
210 collected and extracted in 2005-2007. The quality of the DNA was also screened
211 through the amplification of one microsatellite locus (amplicon size < 100 bp), retaining
212 only samples showing positive amplifications. Total DNA in extractions was measured
213 using a Nanodrop ND-1000 spectrophotometer. The final concentration of the diluted
214 DNA was kept at a working concentration of 10x (~ 200 ng/µl) according to the iPLEX
215 protocol from Sequenom. PCR reactions were conducted using the extracted fecal DNA
216 concentrations at 1, 2, 5, 10, and 20x (at five different concentrations). For each sample,
217 PCRs were performed in duplicate on independent plates to avoid cross-contamination.
218 Quality control criteria for genotyping were adopted using water as negative control and
219 inter-plate duplicates. The MassARRAY Typer 4.0 software was used for data
220 acquisition and analysis. Genotypes were called after cluster analysis using the default
221 setting of Gaussian mixture model. Genotype calls were then reviewed manually to
222 undo any uncertain calls due to clustering artifact. DNA extracted from elephant whole
223 blood was included as positive controls. For every 96-well sample plate, one well was
224 used as a blank control (water) and five wells as duplicate checks (using DNA from
225 good quality blood samples). Plex with less than 99.5% concordance in duplicate
226 checks was considered failed and was repeated. Plex with more than 25% call rate in the
227 blank control was also considered failed. MassARRAY typer 4.0 uses a three parameter
228 model to calculate significance of each putative genotype. A final genotype is called and
229 assigned as ‘conservative’, ‘moderate’, ‘aggressive’, ‘low probability’, and ‘user call
230 (manual calls)’ based on degree of confidence. SNPs were classified as "failed assays"

when the majority of genotypes could not be scored due to low probability or when the samples did not cluster well according to genotype. SNPs that were not found to conform to Hardy-Weinberg Equilibrium (HWE) in one or more samples were cross-checked. Each elephant genomic DNA isolated from fecal and blood samples was amplified in 3-5 different replicates (average 4 replicates for each plex), allowing us to quantify genotyping error rates.

c) *Mitochondrial DNA (mtDNA)*

The mtDNA of forty-seven unique elephant individuals from a number of different sites were successfully amplified and sequenced using the published primers (Fernando et al., 2000; Table 1). These individuals were identified from each site on the basis of unique genotypes across multiple eighteen microsatellite loci. Fernando et al. (2003) analysed a 630 bp fragment of mtDNA comprising cytochrome b (109 bp), $tRNA_{thr}$, $tRNA_{pro}$ (135 bp), and the hypervariable left domain of non-coding control region (386 bp), and only one haplotype was detected for all 20 Bornean elephant samples. In order to check for the existence of new haplotypes, we amplified and sequenced the same mtDNA fragment. All PCR amplifications followed Fernando et al. (2003) and each amplicon was sequenced in both directions. Sequencing was performed at DNA Sequencing Core at Cardiff University in an ABI3100 automated sequencer. Sequences for each individual were aligned in SEQUENCHER 3.1.2 (Gene Code Corporation, 1998). All contigs were manually inspected and sequences were compared with published Asian elephant sequences from Fernando et al. (2003) (GenBank accession numbers AY245538, AY245802 to AY245827).

2.3. Data analyses

255 The two nuclear data sets, microsatellite and SNP loci were treated independently for
256 each analysis. Further, the analysis was repeated by combining the microsatellites and
257 SNPs in an effort to increase statistical power (Morin et al., 2009). In this case, a total of
258 60 nuclear loci (18 microsatellites and 42 SNP loci) were used in the analysis of 70
259 individual elephants from four major ranges. We used only 42 out of 194 SNP loci in
260 the data analyses because many putative SNPs were found monomorphic (refer section
261 3.1 for details).
262
263 ***2.3.1. Genotyping error***
264 The software GIMLET v.1.3.3 (Valière, 2002) was applied to the microsatellite and
265 SNP data to estimate error rates in individual genotyping: ADO (allelic drop out), and
266 successful PCRs. GIMLET also allows to construct a consensus multilocus genotype
267 (the most likely genotype based on all amplifications of a sample) from a set of PCRs
268 and to calculate genotyping error rates comparing the repeated genotypes and their
269 consensus. Genotypes were validated independently by B.G. and N.O. for microsatellite
270 data and by R.S. and C.K.R for SNP data.
271 Because we found homozygosity excess in our microsatellite data, we searched
272 for evidence of null alleles using the program MICROCHECKER (Van Oosterhout et
273 al., 2004) and to infer the most probable technical cause of HWE departures. Deviations
274 from HWE due to inbreeding or population substructure should result in heterozygote
275 deficits across most or all loci, whereas technical causes such as null alleles should
276 result in heterozygote deficits that are variable across loci and populations. The
277 frequency of null alleles was also calculated for each locus using FreeNA (Chapuis and
278 Estoup, 2007), with the maximum-likelihood estimation from Dempster et al. (1977).
279 Due to a significant proportion of null alleles found (> 10% at any locus) in FreeNA,

false homozygote frequencies were used to adjust the number of null alleles by re-naming potential nulls as 999 (Chapuis and Estoup, 2007; Chapuis et al., 2008; Oddou-Muratoio et al., 2009). Further analysis of data used both the adjusted allele frequency data and raw data to assess the effect of null alleles on our results.

2.3.2. Relatedness analyses

We used the information derived from 18 microsatellite loci in 170 individuals (sampled within the family units in each of the site) to estimate pairwise relatedness (*r*) values using the maximum likelihood method in ML-Relate (Kalinowski et al., 2006). Relatedness estimates were adjusted for the presence of null alleles.

2.3.3. Genetic diversity and Population structure

Allele frequencies, mean number of alleles across loci (MNA), observed heterozygosities (H_O), and gene diversity (*He*) (Nei, 1978) were obtained for microsatellite and SNP data using the GENETIX 4.01 (Belkhir et al., 2000) and GenAlEx (Peakall and Smouse, 2006) softwares. Allelic richness (MNA) was adjusted for discrepancies in sample size by incorporating a rarefaction method, and was estimated for each forest range using Fstat 2.9.3.2 version (Goudet, 1995).

In order to compare parameters inferred from SNPs and microsatellites, we evaluated microsatellite diversity for the same individuals for which we also had SNP genotypes (n=70). These individuals represented elephant populations from all major forest ranges, i.e. LK, DER, CF, and TWR.

Population structure across the Bornean elephant distribution range was investigated using Bayesian clustering as implemented in STRUCTURE v. 2.3.3 (without spatial information) (Pritchard et al., 2000; Falush et al., 2003), and TESS

2.3.1 (incorporating spatial information) for both SNP and microsatellite datasets (François et al., 2006; Chen et al., 2007). STRUCTURE was used under a model assuming admixture, ignoring population affiliation and allowing for correlation of allele frequencies between clusters. We conducted ten runs for each value of K = 1-7 and each run consisted of a 50,000 burn-in followed by 250,000 iterations. The most likely value of K was assessed by comparing the likelihood of the data for different values of K and by the rate of change in the log probability of the data between successive K values (Delta K; Evanno et al., 2005). Structure analyses were repeated after removing all but a few of the individuals that were sampled within the family units. After removing individuals that were related, the final data consisted of 58 individuals from different sites within each of the forest range (LK, n=10; DER, n=9; CF (USM, n=9; GR, n=6; KAL, n=8; KU, n=5; MB, n=5); and TWR, n=6). Structure Harvester was used to calculate and plot Delta K (Earl and vonHoldt, 2011). Assignment of individuals to the inferred clusters was estimated according to the highest q-values (probability of membership). STRUCTURE results were visualized with the program DISTRUCT (Rosenberg, 2004).

TESS was run using the conditional autoregressive (CAR) admixture model with spatial interaction parameter set at 0.6, as recommended by Chen et al. (2007). In the analysis, we also considered other values of the spatial dependence parameter (Ψ) 0.0 and 1.0. This parameter weights the relative importance given to the spatial connectivities (Ψ = 0 recovers the model underlying STRUCTURE, while Ψ =0.6 and 1.0 indicate moderate and strong values, respectively). One hundred replicate runs of 100,000 sweeps (disregarding the first 30,000) were performed for K values 2 to 7. The preferred K was selected by comparing the individual assignment results and the deviance information criterion (DIC) for each K (Durand et al., 2009). DIC values

averaged over 100 independent iterations were plotted against *K*, and the most likely value of *K* was selected by visually assessing the point at which DIC first reached a plateau and the number of clusters to which individuals were proportionally assigned. The 10 runs with the lowest DIC values for the selected *K*-value were retained and their admixture estimates were averaged using CLUMPP version 1.1.2 (Jakobsson and Rosenberg, 2007), applying the greedy algorithm with random input order and 1000 permutations to align the runs and calculate G' statistics.

We also applied principal component analysis (PCA) to the microsatellite and SNP genotypes since PCA is not dependent on any model assumptions and can thus provide a useful validation of Bayesian clustering output (Patterson et al., 2006; Francois and Durand, 2010). We used the R package Adegenet v1.3.4 (Jombart et al., 2008) to carry out standard PCA analysis.

We calculated Wright's *F*-statistics for both nuclear datasets (SNP and microsatellites) according to the method of Weir and Cockerham (1984) and their significance was tested with 10,000 permutations using the GENETIX 4.01 (Belkhir et al., 2000). We also used Arlequin 3.1 (Excoffier et al., 2005) to perform a hierarchical analysis of molecular variance (AMOVA, Excoffier et al., 1992) to determine significance of genetic variation between the forest ranges and when grouped by geographic location.

In order to test for the impact of null alleles in our microsatellite data set, we also calculated global F_{ST} values with FreeNA software. These values were computed, as described in Chapuis and Estoup (2007), with 10,000 bootstrap iterations, alternatively using and not using the excluding null alleles (ENA) method.

We investigated fine-scale spatial genetic structure by analyzing isolation-by-distance (IBD) using the microsatellite genotype dataset. We performed a Mantel test to

355 investigate the correlation between the Queller and Goodnight relatedness estimator
356 (1989) and geographical distance. The significance of results was assessed by 9999
357 permutations. This was done using the software GenAlEx (Peakall and Smouse, 2006).
358
359 ***2.3.4. Departures from mutation drift equilibrium***
360 We tested for a departure from mutation drift equilibrium in the microsatellite data
361 using BOTTLENECK version 1.2.0.2 (Cornuet and Luikart, 1996; Piry et al., 1999).
362 Significant departures can be due to changes in population sizes such as expansions and
363 bottleneck under the assumption that samples are obtained from a random mating and
364 isolated population. Under that model bottlenecked populations will show an excess of
365 heterozygotes relative to that expected at equilibrium from observed allelic diversity.
366 BOTTLENECK was run under three mutation models: the infinite alleles (IAM), two-
367 phased (TPM), and stepwise mutation (SMM). The TPM was set at 95% stepwise
368 mutation model and 5% multi-step mutations, as recommended by Piry et al. (1999).
369 Wilcoxon signed-rank tests were used to identify heterozygosity excess (Piry et al.,
370 1999).
371

3. Results

3.1. *Genotyping error*

In the case of microsatellites markers, out of the 273 samples genotyped, two samples did not yield reliable results and 224 unique genotypes were identified (representing approximately 11% of the estimated elephant population of Sabah). Comparison of the observed genotypes with the distribution of randomized genotypes generated with the program MICROCHECKER (van Oosterhout et al., 2004) suggested that heterozygote deficiency is at least partly due to the presence of null alleles. However, there was no evidence for scoring errors due to stuttering or large allele dropout, as confirmed by a GIMLET analysis. Sixteen loci in CF, while four loci each in LK, TWR and nine loci in DER were not in HWE. Altogether null alleles were present at intermediate frequency for 13 out of 18 microsatellite loci and their frequency estimates ranged from 3% and 20%. The loci that displayed deviations from HWE are listed in Table S2 with, F_{IS} values being significantly different from zero and positive in all four ranges varying from 0.14 in LK to 0.38 in DER (Table 2).

SNPs were successfully genotyped for all blood samples (n=20). We generated 10,397 (97%) genotype calls (obtained out of 10,716 expected calls) from all the blood samples across all the plexes and replicates. We observed 54, 17, 8, and 6% Sequenom calls that were at “conservative”, “moderate”, “aggressive”, and “low probability” levels of confidence (listed in Table S3), respectively. Of the 194 SNP loci, corresponding to 10,716 genotypes, 319 (3%) did not successfully call any bases within our confidence limits and hence were assigned as “no alleles”. In addition, 12% of the genotype calls were noted as “user calls” because these calls were assigned manually by us. Within the SNP data set consisting of 194 SNPs and 20 blood samples, amplification success rate ranged between 93-100% per locus and 156 (80%) of the 194 SNPs were

found monomorphic and were excluded from all analyses. Three of the remaining 38 loci showed significant heterozygotes excess, thus 35 loci were used in the data analyses. Individual multilocus genotypes in this SNP data set were on average 97% complete.

For the fecal samples (*n*=70), the tests performed by using different concentrations of extracted DNA showed that the 10 and 20x concentrations could be scored unambiguously without any quality difference in genotype calls. We therefore performed further genotyping of all fecal-extracted DNA at 10x concentration. We obtained 43,354 (71%) genotype calls out of 60,284 expected calls. In total across all plexes, we observed 33, 11, 6, and 13% Sequenom calls were at "conservative", "moderate", "aggressive", and "low probability" levels of confidence, respectively. A high number of genotypes (29%) were assigned as "no alleles" due to bad spectrum (Figure S1 and Table S3). We only made "user calls" in 9% of the cases. SNP genotyping showed a low percentage of positive PCRs, and ranged from 20 to 92% (average across each sample) and 41-78% (average across each plex). Allelic dropout (ADO) ranged between 0 and 37% in 25 samples with highest ADO rate observed in 6 fecal samples (>20%). We also compared genotype error rate of SNPs across different multiplexing levels (28-plex to 42-plex) and found a significant correlation between genotype error rate of SNPs and multiplexing levels in the fecal samples. The genotyping error rate of SNPs across 28-plex and 42-plex observed was 9 and 25%, respectively.

Our results showed that five of the 194 SNP loci generated no nucleotide signal, 130 yielded monomorphic profiles and of the remaining 59 polymorphic loci, 31 were identified as polymorphic in both fecal and blood samples. We therefore found that 7 and 28 loci were only polymorphic in the blood and fecal samples, respectively. Sixteen

of the 59 loci showed a significant departure from HWE ($p < 0.001$) with five loci having positive (heterozygote deficit) and the remainder having negative F_{IS} values (heterozygote excess) and one additional locus had many genotypes (> 95%) missing. Hence, these 17 loci were excluded from further analyses, leaving a complete SNP data set of 42 reliable polymorphic loci. Individual multilocus genotypes were on average 61% complete in this data set.

3.2. Relatedness analyses

Analysis of pairwise relatedness (r) of all individual genotypes derived from 18 microsatellite loci revealed an average relatedness of -0.005 ± 0.476. However, average pairwise relatedness values between individuals examined within each site were significantly positive in four (USM, GR, KAL, TWR) out of seven populations and r ranged from 0.264 (KAL) to 0.413 (USM). Furthermore, relatedness values between individuals that were sampled within family units from USM, GR, KAL, and TWR also showed very high values (r = 0.065 to 0.783). These results are given in Table S4.

3.3. Genetic diversity

3.3.1. Mitochondrial DNA

All 47 fecal samples were successfully amplified and sequenced, and only one haplotype was detected. This haplotype corresponds to the unique β-haplotype BD, as previously reported by Fernando et al. (2003) in the elephant samples collected from three different locations. Thus sequencing additional individuals did not allow us to identify new haplotypes beyond this.

3.3.2. Microsatellites and SNPs

All 18 microsatellites were polymorphic with between two and eight alleles per locus. The mean number of alleles (MNA) per site ranged from 1.7 to 3.6 and a positive correlation between this parameter and sample size (r = 0.80) was observed. Observed heterozygosity (average *H*o across loci) was lower than expected for all sites and ranged from 0.14 in TWR to 0.41 in LK, whereas expected heterozygosity (average *H*e across loci) ranged from 0.21 in TWR to 0.47 in LK (Table 2). After correcting the microsatellite data set for null alleles using EM algorithm (as implemented in FreeNA), both the average observed and expected heterozygosity values were higher than the raw data ranging from 0.23 in CF to 0.46 in LK and 0.26 in CF to 0.49 in LK, respectively. Despite this correction, F_{IS} values were still positive and significant in two sites (DER and CF) (Table S5).

Within the SNP data set the percentage of polymorphic loci ranged from 52 to 90%. It was highest in the samples from LK in agreement with the results obtained using microsatellites, where all loci were polymorphic. Most loci were in Hardy-Weinberg equilibrium and while F_{IS} values differed markedly between datasets, they were generally lower (and non-significant) for SNPs than for microsatellites. The average *H*o and *H*e for the SNP data ranged from 0.22 in TWR to 0.31 in DER, and from 0.23 in TWR and CF to 0.32 in DER, respectively (Table 3). While microsatellites provided higher estimates of *H*o and *H*e for elephants in the LK region, average *H*o and *H*e for SNPs were consistently higher in the elephants from DER than those in the LK (Table 3). Interestingly, we found 94% of SNP loci to be polymorphic in the blood samples from LK and only 65% from CF range. However, the two types of marker do not appear to exhibit very different patterns, as Table 2 suggests for the 70 individuals for which we had genotypes derived from both SNPs and microsatellites. This supports the main result of higher genetic diversity in the elephants from the LK region. This

again is confirmed by the combined nuclear data set (microsatellite and SNP), which produced similar levels and patterns of genetic diversity in all populations (Table 2). Hence, taken together, our results agreed in suggesting that elephants in the Kinabatangan region (including LK and DER populations) are genetically more diverse than the others.

3.4. Assessment of population structure

Bayesian cluster analysis of microsatellite genotypes in STRUCTURE supported the existence of two genetic clusters. Examination of Ln P (X|*K*) and Delta *K* values also suggested a level of subdivision at $K = 2$. Cluster 1 consisted mostly of individuals from the geographically isolated LK and DER populations, with a few individuals sampled elsewhere also being assigned to this cluster (Figure 2a). Cluster 2 included individuals from all forest sites located in CF (i.e., USM, GR, KU, MB, KAL), and TWR. Across all individuals, 35% were assigned to cluster 1 with a membership proportion of $q > 0.70$. Individuals within CF and TWR (representing the remaining 65% individuals) were more strongly assigned to the alternative cluster with q varying from 0.75 to 0.95. No further sub-structuring was found within the CF. STRUCTURE analyses performed on the microsatellite data representing only unrelated individuals also produced results similar to $K = 2$ (one cluster comprising LK and DER and the other cluster composed of the other sites).

TESS gave results similar to STRUCTURE, grouping the majority of the individuals from LK and DER together (Figures 2a and 3), whereas all individuals within the CF and TWR were assigned to another cluster (cluster 1 in Figure 3). The DIC plot of the TESS runs did not show a well-defined plateau as the DIC values continuously decreased at higher *K*max values (Figure S2). Across the 10 TESS runs, *K*

497 = 2 showed the most consistent groupings of elephant populations. Higher values
498 displayed less stable clustering and did not recognize additional distinct population
499 clusters aside from the groups inferred at $K = 2$. Also, analyses performed on the data
500 eliminating the related individuals produced results similar to $K = 2$.

501 The results from the principal component analysis (PCA) of the microsatellite
502 data corroborated the aforementioned analyses. Individual genotypes from the two
503 clusters identified by STRUCTURE and TESS were separated along the first and second
504 components of the PCA, which may provide a slightly better discrimination than the
505 other two analyses. As shown in Figure 4, the first component which accounted for
506 16.43% of variation resulted in the separation of LK and DER from the other
507 populations. Simultaneously, the sites located in CF were clustered together with TWR.
508 Furthermore, the second component (accounting for 8.48%) could be interpreted as
509 separating LK and DER sites. The overlap between all these clusters further illustrates
510 their weak genetic differentiation. The following principal coordinate axes displayed
511 uninformative clustering patterns.

512 We also tested the efficiency of our methods for detecting population structure
513 using SNP data. Contrary to the microsatellite results, STRUCTURE detected only a
514 single genetic cluster (i.e., K=1, Figure 2b). All assignment percentages for the SNP
515 dataset were lower than for the microsatellite dataset. TESS detected the presence of
516 two very weak spatial genetic clusters (Figure 2b). The PCA approach based on allele
517 frequencies also showed results similar to STRUCTURE and failed to reveal the
518 presence of the two clusters (Figure 4).

519 Combining SNPs and microsatellites for 70 individual samples, which would be
520 expected to increase statistical power over either single data set, did not result in the
521 detection of a clear population structure (Liu et al., 2005).

522 The consistency of the clusters identified through the Bayesian clustering
523 approach for both microsatellite and SNP data was tested by pairwise F_{ST} analysis.
524 Using microsatellites, we found a considerable level of genetic differentiation between
525 all demographically isolated populations in different ranges (average F_{ST} = 0.10, p <
526 0.001) suggesting limited gene flow between adjacent populations of elephants. The
527 pairwise F_{ST} values between the sampling sites ranged from 0.03 to 0.14 (p < 0.001)
528 (Table 4). We observed that LK, DER and TWR were the most differentiated (F_{ST}
529 values between 0.08 - 0.14) forest ranges. Furthermore, in a separate analysis, the sites
530 in the CF range, such as GR, KAL, USM appeared to have low (but significant) genetic
531 differentiation (F_{ST} 0.04 - 0.07). Global F_{ST} and the F_{ST} values for each locus were
532 similar when calculated with (ENA) and without estimating a null allele correction (F_{ST}
533 =0.10, F_{ST}ENA= 0.11). This suggests that the presence of null allele has limited effect,
534 if any, on our calculations of genetic differentiation. AMOVA analyses revealed that
535 variation within populations, between populations, and among regions accounted for
536 88%, 7%, and 5% of the total variation, respectively (p < 0.001).
537 SNP data produced low pairwise F_{ST} estimates indicative of little to moderate
538 differentiation (pairwise F_{ST} values between 0.006 and 0.136, p < 0.05) but were not
539 significant except in the comparisons involving DER and TWR for which F_{ST} values
540 were above 0.10 and 0.13, respectively. Combined data from both SNP and
541 microsatellite marker types produced low F_{ST} estimates indicative of limited to medium
542 differentiation between subpopulations (F_{ST} values between 0.02 and 0.11).
543 Results from Mantel test on microsatellite data suggest a significant negative
544 correlation between pairwise relatedness and geographic distances (r = -0.142, p =
545 0.000) among elephant populations, implying that relatedness decreased with
546 geographic distance.

3.5. Departures from mutation drift equilibrium

For microsatellites, the results of the BOTTLENECK analysis showed that there was no consistent or strong signal for a departure from mutation drift equilibrium. Only for the DER sample we found a significant departure under three mutational models (Wilcoxon test; TPM: P = 0.00008; IAM: P = 0.00002, SMM: P = 0.00134). The LK population showed significant excess of heterozygotes only under the IAM and TPM (IAM: P = 0.00032; TPM: P = 0.01184). None of the other population showed evidence of heterozygosity excess.

4. Discussion

4.1. Genetic diversity and assessment of population structure

Our study revealed relatively low levels of genetic variation in Bornean elephants, using both mitochondrial sequence and nuclear genetic markers. The lack of mtDNA variability in Bornean elephant was not surprising as the only previous study by Fernando et al. (2003) found no variation in this region of mtDNA across 20 individuals. MtDNA studies in other Asian elephants have shown that adult females of a family or a social group share the same haplotype and are closely related to one another due to the matriarchal social system of elephants (Fernando and Lande, 2000; Vidya et al., 2005a; de Silva et al., 2011). For instance, study by Vidya et al. (2005a) found only one haplotype in the world's largest Asian elephant population (of over 9,000 elephants) in the Nilgiris, southern India. Therefore, the fixation of a single mtDNA haplotype among Bornean elephants is probably more a consequence of a recent or an ancient population bottleneck (Fernando et al., 2003). The current study, however, significantly expands the geographic and demographic importance of these findings, suggesting that an exceptionally high proportion of Bornean elephants in Sabah have a single mtDNA lineage regardless of their present geographic location.

Our microsatellite and SNP data based on larger geographic coverage across Sabah and increased sample sizes confirm earlier findings of low levels of genetic variation in Bornean elephant. Fernando et al. (2003) compared all Asian elephant populations (e.g. Borneo, Cambodia, Bhutan, Bangladesh, Thailand, Sumatra, Sri Lanka, and South India; *H*e = range 0.014 - 0.63 for five microsatellite loci) and showed that Bornean elephant have very low genetic diversity. Our results also suggest low diversity in Bornean elephants but substantially higher estimates than that were

previously shown (*He* = 0.041, *Ho* = 0.014, Fernando et al., 2003). The average expected heterozygosity (*He*) found in our study was 0.27 (SNPs) and 0.34 (microsatellite) across all populations. However, the microsatellite loci used in our study differ from those used by Fernando et al. (2003) and were polymorphic in the populations on which they were tested. Recent studies point to the fact that cross-specific amplification using highly polymorphic markers provides a biased picture of the genetic diversity when compared with randomly specific markers (Chikhi, 2008; Queirós et al., 2014). Hence, the genetic diversity values obtained in our study are likely overestimated. Such overestimated values are, as expected, in agreement with a previous study by Ahlering et al. (2011), who analyzed elephant populations from Laos using nine microsatellites (same as those used here) and found estimates of genetic diversity (*He* = 0.75) that are much higher than those previously reported (*He* = 0.48) in Fernando et al.'s 2003 study. Even though these values might be overestimated, levels of genetic variation in Bornean elephants are still similar to or lower than that observed in other rare and endangered species with low levels of genetic variation, e.g. Iberian lynx *He* = 0.31-0.46 (Casas-Marce et al., 2013), European bison *He* = 0.28 (Tokarska et al., 2009), and Ngorongoro Crater lion *He* = 0.58 (Antunes et al., 2008).

Although the Bornean elephants are characterized by low genetic variation at microsatellite loci, we found signature of a recent demographic bottleneck only in the Deramakot and Lower Kinabatangan (but not under SMM) elephant population. If all three mutation models had given the same significant result in each of the population, this would have suggested that the signal was strong enough to be detected using a summary statistics approach.

Interestingly, the pattern of genetic variation inferred by SNP genotypes obtained using elephant fecal and blood samples were also similar. While this would be

606 consistent with a higher diversity in LK in agreement with microsatellite data, the fact
607 that sample sizes differ suggests that further sampling and validation is required.

608 We found overall concordance among markers (SNPs and microsatellites) in the
609 amount of genetic variation within elephant populations. For instance, the elephants in
610 the Kinabatangan region (i.e. LK and Deramakot) exhibit relatively high levels of
611 genetic variability despite the fragmentation of their habitat. Some contemporary gene
612 flow between elephant populations within the Kinabatangan region may contribute to
613 the maintenance of these high heterozygosity levels. However, the more likely
614 explanations are that fragmentation is recent and/or that effective population size was
615 greater in the recent past. In any case, erosion of heterozygosity should occur relatively
616 quickly through genetic drift and inbreeding if populations are maintained without gene
617 flow (Hartl and Clark, 1989). Hence, continued maintenance of habitat inter-
618 connectivity in the whole Kinabatangan region is very important to retain the maximal
619 amounts of the high genetic diversity seen.

620 Overall, we found evidence of weak population structure across the distribution
621 range of the Bornean elephant. SNP markers were not powerful enough for assessment
622 of population structure. On the contrary, microsatellite data indicated that at least two
623 elephant sub-populations exist within Sabah. Further, STRUCTURE and TESS analyses
624 performed on the microsatellite data representing only unrelated individuals, produced
625 results similar to $K = 2$. This also indicates that the population structure detected in our
626 samples is not due to the occurrence of related individuals (Anderson and Dunham,
627 2008).

628 Our microsatellite data showed that Bornean elephants from the Lower
629 Kinabatangan and North Kinabatangan (Deramakot) ranges are somewhat isolated from

630 the main elephant populations located in the Central Forest and Tabin Wildlife Reserve,
631 as illustrated in the Bayesian clustering methods. Clearly, genetic grouping of these two
632 populations and their higher levels of genetic diversity than other populations is likely
633 reflective of historical connectivity between the two currently disjunct, but
634 geographically proximate locations. Further, the genetic distinctiveness of these
635 populations from the rest of the elephant populations (for both genotype-based and
636 allele frequency based analyses) is most likely attributed to the recent/historical
637 population bottleneck and associated local effects of genetic drift (e.g. fixation and loss
638 of alleles).

639 These results also revealed that Bornean elephants show genetic differentiation
640 over small geographic distances. Given the absence of obvious barriers to gene flow and
641 high dispersal abilities, Bornean elephant exhibited unexpected population
642 differentiation over relatively small distances. For instance, the observed genetic
643 differentiation among elephant subpopulations in the Kinabatangan region (i.e. LK and
644 DER) is unexpected (F_{ST} ~ 0.08), given the geographical proximity of the sub-
645 populations and the dispersal capabilities of elephants. These sub-populations are
646 separated by approx. 35 km, a distance within the range of elephant movement (Estes et
647 al., 2012). Similarly, our data suggest reduced gene flow between the elephant
648 populations in North Kinabatangan (Deramakot) and Central Forest ranges, which is the
649 largest elephant population in Sabah (Elephant Action Plan, Sabah Wildlife
650 Department, 2012-2016; Estes et al., 2012). These elephant populations are separated by
651 the Kinabatangan River but topographic features such as rivers do not appear to limit
652 elephant movement, however, elephants do not tolerate land-use activities that alter
653 habitats permanently, such as agriculture and human settlement (Zhang et al., 2015,
654 Elephant Action Plan, Sabah Wildlife Department, 2012-2016). Neither of the Bayesian

clustering methods recognized the distinctiveness of the forest sites located in Central Forest range, despite significant differentiation by pairwise F_{ST} analysis (range 0.04 - 0.07). This result is difficult to interpret based on their close geographical proximity (roughly within 50-100 km) and the presence of the largest continuous forest habitat for elephants in this region (Alfred et al., 2010). These levels of genetic differentiation in Bornean elephant populations imply a strong impact of local genetic drift, indicating that the effective population sizes are very small in each forest fragment and that current gene flow among them is likely limited. Habitat fragmentation in Sabah is very recent and cannot fully explain the overall pattern of observed genetic differentiation. Very large parts of the forest in the Kinabatangan region have been logged and converted into oil palm plantations in the last 50 years. Similarly, Tabin forest was connected to Central Sabah before the establishment of the Lahad Datu-Sandakan road that has affected the connectivity between elephant populations in the entire region and identified as significant barrier to elephant dispersal and gene flow (Elephant Action Plan, Sabah Wildlife Department, 2012-2016). However, considering the time lag between the recent landscape changes in Sabah and the beginning of genetic sampling, the Lahad Datu-Sandakan road was probably too recent for a significant barrier. It is known that significant time lags can exist for genetic discontinuities to develop after barrier formation (Landguth et al., 2010) and these time lags make it especially difficult to detect changes in gene flow for species with long generation times such as elephant. As such, some of the observed patterns might be reflective of past rather than present levels of connectivity and fragmentation.

It is also possible that part of the observed pattern of genetic differentiation is influenced by the inclusion of a relatively high percentage of closely related individuals in our sampling scheme. The results from Mantel test demonstrate high genetic

680 relatedness at small spatial scales and suggest that individuals were less genetically
681 related as the geographic distance that separated the two individuals increased.
682 Interestingly, the pairwise relatedness values observed are higher than those reported
683 within family units (r = 0.37) for an elephant population in southern India (Vidya and
684 Sukumar, 2005). Our findings concur with the few other studies, indicating how
685 isolation by distance within elephant populations may occur at fine spatiotemporal
686 scales and in the absence of obvious landscape barriers, and may be driven by
687 mechanisms, such as dispersal behaviour and species' mating system (e.g. Vidya et al.,
688 2005b). Nonetheless, understanding and modelling the social organisation of Bornean
689 elephants should provide a clearer picture of the current distribution of genetic diversity
690 and population structure (e.g. Parreira et al., 2015).
691
692 ***4.2. Null alleles and genotyping error***
693 Using microsatellite markers, we observed a statistically positive F_{IS} for all Bornean
694 elephant populations studied herein, indicating a deviation from HWE. The most likely
695 explanations include the presence of null alleles and non-random mating. Additionally,
696 a Wahlund effect (due to sampling several genetically differentiated sub-populations)
697 should not be ruled out as contributors to the deviation from HWE. Many individuals
698 also exhibited high relatedness values with members of the family units (r > 0.34). This
699 may explain the deviation of HWE for the two populations (DER and CF) after the
700 genotypes were corrected for null alleles.
701 The lower rate of SNP genotyping success observed in this study is most likely
702 due to the low DNA concentration of fecal samples and high multiplexing levels. One
703 limitation of MALDI-TOF mass spectrometer analysis, as implemented in Sequenom, is
704 the purity of the sample required by the assay (Sobrino et al., 2005). Fecal DNA

normally includes other sources of DNA from diet and is always more degraded than blood or other tissue types (Taberlet et al., 1996). Indeed, an error frequently identified in our data set was the inaccurate calling of individual genotypes with individual genotypes fell between the three main genotype clusters. In this case, the alleles were treated as missing data. Further, MassARRAY assays in Sequenom employ a classical PCR technology that requires multiple sets of primers to amplify multiple specific regions. Such specificity requirement becomes even more critical with increasing level of multiplexing.

The positive correlation between SNP genotype error rate and multiplexing level suggests that genotyping performance rate (accuracy) of SNPs from fecal samples decreased with increasing multiplexing level. Multiplex PCR is a sensitive technique and the methods for multiplexing markers, such as microsatellites have considerably improved over the last years due to the use of specialized PCR protocols, such as optimized PCR buffers, the Qiagen multiplex PCR kit, and Touch-down PCR (Guichoux et al., 2011). However, for highly multiplexed sets of SNPs (> 28), more advanced strategies might still be necessary and additional efforts to improve SNP multiplex genotyping and scoring remain critical. Unfortunately, genotyping error rates are not available for Sequenom assays in other non-model species. However, comparable values are available from a recent study that has tested the efficiency of SNP genotyping from fecal DNA of Italian wolf using a different genotyping assay (TaqMan) (Fabbri et al., 2012). This study found a high percentage of positive PCRs (86-92%) and low ADO rate (0-18%).

Our study suggests that the polymorphic SNPs were useful for estimating general level of genetic diversity in Bornean elephants but closely related populations were generally better distinguished with microsatellites than SNPs. Panels of SNPs are

730 rapidly becoming the population genomic markers in ecological and conservation
731 genetics studies (Morin et al., 2004, Kraus et al., 2015), however, given concerns
732 regarding error rates and monomorphic SNP calls, these panels have only limited value,
733 especially when working with genetically depauperate populations, as seen in this study.
734 Simulated (Morin et al., 2009) and empirical data (Tokarska et al., 2009) have shown
735 that significant power (> 0.95) to detect population structure at very recent divergence
736 times required as many as 15 times more SNPs than microsatellites (Haasl and Payseur,
737 2011). The present study includes many fewer SNP loci and we will still need to apply
738 more loci for a finer resolution of population and conservation genetics of Bornean
739 elephants in future studies. Also the choice between the genetic markers will be best
740 determined by the questions being tested, as well as laboratory specific costs and
741 technical capabilities (McMahon et al., 2014).
742
743 ***4.3. Conservation implications***
744 This study gives the first description of the genetic diversity and structure of Bornean
745 elephant populations across Sabah (i.e. most of the species range) and as such should be
746 used for their conservation management. The failure to identify variation in the mtDNA
747 control region of Bornean elephants, in combination with the microsatellite and SNP
748 data, lends support to the conclusion that Bornean elephants exhibit overall low genetic
749 variability. The detection of reduced gene flow levels among elephant populations in
750 Sabah adds to a growing body of literature that documents an increasingly fragmented
751 landscape for large mammals in Borneo. One of the most important management-related
752 results from this study is that significant genetic differentiation exists between extant
753 elephant populations. Reinforcing gene flow by re-establishing habitat connectivity
754 between populations, especially between ranges such as the Kinabatangan, Tabin and

Central Forest, may therefore be the priority given that the Kinabatangan elephant population appears to be the most genetically diverse. Our study also reveals that genetic diversity is unequally distributed between the elephant populations in Sabah that are often small and isolated. If populations decrease too much in size, they will become more sensitive to stochastic events. Inbreeding and loss of genetic variation are inevitable consequences of small population sizes (Saccheri et al. 1998; Frankham 2010). It is not known how long the elephant populations have been small but it could be that in populations that experience small sizes may be less susceptible to future inbreeding depression because they have been purged of deleterious recessive alleles (Hedrick, 1994). Also, genetic diversity can be restored with introduction of individuals from other mainland elephant population, however, it also carries the risk of outbreeding depression and needs careful evaluation (Frankham et al., 2011). Nevertheless, the elephant populations in Borneo may require more intensive genetic management and the possible expression of inbreeding depression should be carefully followed.

Acknowledgements

This work was supported by Portuguese Science Foundation ("Fundação para a Ciência e a Tecnologia" (FCT)) individual fellowship (ref. SFRH/BPD/64837/2009) to RS and FCT project (ref. PTDC/BIA-BDE/71299/2006) to LC. LC was also supported by the "Laboratoire d'Excellence (LABEX)" entitled TULIP (ANR -10-LABX-41). BG and MWB were supported by a grant from the Darwin Initiative for the Survival of Species (Grant no. 09/016, DEFRA, UK). BG was supported by grants from US Fish and Wildlife Service Asian Elephant Conservation Fund, Elephant Family, Houston Zoo and Columbus Zoo. CKR was funded by an Optimus Alive Oeiras-IGC fellowship to LC. NJ and RO were supported by the National Science Foundation. BG thanks the Economic Planning Unit (Malaysia) for permission to conduct research in Sabah. We are thankful to the staff of Lok Kawi Wildlife Park, for their help and cooperation in collecting elephant blood samples, to Isabel Marques, João Costa, Susana Ladeiro, and Isa Pais at IGC for help in the genotyping work, T. Minhós for helpful discussions regarding relatedness analyses. Blood samples were imported to Portugal under import permit from Portugal (CITES certificate no. PT/LI-0187/2010) and export permit from Federation of Malaysia (CITES certificate no. 4874). Thanks to Craig Obergfell for initial sequencing assistance.

References

Abram, N.K., Xofis, P., Tzanopoulos, J., MacMillan, D.C., Ancrenaz, M., Chung, R., Peter, L., Ong, R., Lackman, I., Goossens, B., Ambu, L., Knoght, A.T., 2014. Synergies for improving oil palm production and forest conservation in floodplain landscapes. PLoS ONE 9, e95388.

Ahlering, M.A., Hedges, S., Johnson, A., Tyson, M., Schuttler, S., Eggert, L.S., 2011. Genetic diversity, social structure, and conservation value of the elephants of the Nakai Plateau, Lao PDR, based on non-invasive sampling. Conservation Genetics 12, 413–422.

Alfred, R., Ahmad, A.H., Payne J., Williams, C., Ambu, L., 2010. Density and population estimation of the Bornean elephants (*Elephas maximus borneensis*) in Sabah. Online Journal of Biological Sciences 10, 92–102.

Alfred, R., Ambu, L., Nathan, Senthilvel, S.K.S.S., Goossens, B., 2011. Current status of Asian elephants in Borneo. Gajah 35, 29–35.

Armbruster P, Fernando P, Lande R., 1999. Time frames for population viability analysis of species with long generations: an example with Asian elephants. Animal Conservation 2:69-73.

Anderson, E.C., Dunham, K.K., 2008. The influence of family groups on inferences made with the program Structure. Molecular Ecology Resources 8, 1219–1229.

Antunes, A., Troyer, J.L., Roelke, M.E., Pecon-Slattery, J., Packer, C., Winterbach, C., Winterbach, H., Hemson, G., Frank, L., Stander, P., Siefert, L., Driciru, M., Funston, P.J., Alexander, K.A., Prager, K.C., Mills, G., Wildt, D., Bush, M., O'Brien, S.J., Johnson, W.E., 2008. The evolutionary dynamics of the lion *Panthera leo* revealed by host and viral population genomics. PLoS Genetics 4, e1000251.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F., 2000. GENETIX 4.01, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, Université de Montpellier II, Montpellier. (http://www.univ-montp2.fr/~genetix/genetix/genetix.html).

Casas-Marce, M., Soriano, L., López-Bao, J.V., Godoy, J.A., 2013. Genetics at the verge of extinction: insights from the Iberian lynx. Molecular Ecology 22, 5503–5515.

822 Chapuis, M.P., Estoup, A., 2007. Microsatellite null alleles and estimation of population
823 differentiation. Molecular Biology and Evolution 24, 621–631.
824 Chapuis, M.P., Lecoq, M., Michalakis, Y., Loiseau, A., Sword, G.A., Piry, S., Estoup,
825 A., 2008. Do outbreaks affect genetic population structure? A worldwide survey
826 in *Locusta migratori*, a pest plagued by microsatellite null alleles. Molecular
827 Ecology 17, 3640–3653.
828 Chen, C., Durand, E., Forbes, F., François, O., 2007. Bayesian clustering algorithms
829 ascertaining spatial population structure: a new computer program and a
830 comparison study. Molecular Ecology Notes 7, 747–756.
831 Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B., Beaumont, M.A., 2010. The
832 confounding effects of population structure, genetic diversity and the sampling
833 scheme on the detection and quantification of population size changes. Genetics
834 186, 983–995.
835 Chikhi, L., 2008. Genetic markers: how accurate can genetic data be? Heredity 108,
836 471–472.
837 Cornuet, J.M., Luikart, G., 1996. Description and power analysis of two tests for
838 detecting recent population bottlenecks from allele frequency data. Genetics 144,
839 2001–2014.
840 Cranbrook, E., Payne, J., Leh, C.M.U., 2008. Origin of the elephants *Elephas maximus*
841 L. of Borneo. Sarawak Museum Journal LXIII, 84.
842 Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete
843 data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1–38.
844 de Silva, S., Ranjeewa, D.G.A., Kryazhimskiy, S., 2011. The dynamics of social
845 networks among female Asian elephants. BMC Ecology 11, 17.
846 Durand, E., Jay, F., Gaggiotti, O.E., François, O., 2009. Spatial inference of admixture
847 proportions and secondary contact zones. Molecular Biology and Evolution 26,
848 1963–1973.
849 Earl, D.A., Vonholdt, B.M., 2011. Structure harvester: a website and program for
850 visualizing structure output and implementing the Evanno method. Conservation
851 Genetics Resources 4, 359–361.
852 Elephant Action Plan (2012-2016), Sabah Wildlife Department (SWD), Ministry of
853 Tourism, Culture and Environment, Sabah. ISBN: 978-983-40057-6-4.

854 Estes, J.G., Othman, N., Ismail, S., Ancrenaz, M., Goossens, B., Ambu, L.N., Estes,
855 A.B., Palmiotto, P.A., 2012. Quantity and configuration of available elephant
856 habitat and related conservation concerns in the Lower Kinabatangan Floodplain
857 of Sabah, Malaysia. PLoS ONE 7, e44601.
858 Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance
859 inferred from metric distances among DNA haplotypes: application to human
860 mitochondrial DNA restriction data. Genetics 131, 479–491.
861 Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: An integrated software
862 package for population genetics data analysis. Evolutionary Bioinformatics
863 Online 1, 47–50.
864 Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of
865 individuals using the software STRUCTURE: a simulation study. Molecular
866 Ecology 14, 2611–2620.
867 Ewers RM, Didham RK, 2006, Confounding factors in the detection of species
868 responses to habitat fragmentation, Biological Reviews, 81, 117-142.
869 Fabbri, E., Caniglia, R., Mucci, N., Thomsen, H.P., Krag, K., Pertoldi, C., 2012.
870 Comparison of single nucleotide polymorphisms and microsatellites in non-
871 invasive genetic monitoring of a wolf population. Archives of Biological
872 Sciences 64, 321–335.
873 Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using
874 multilocus genotype data: linked loci and correlated allele frequencies. Genetics
875 164, 1567–1587.
876 Fernando, P., Lande, R. 2000. Molecular genetic and behavioral analyses of social
877 organization in the Asian elephant. Behavioral Ecology and Sociobiology 48,
878 84–91.
879 Fernando, P., Vidya, T.N.C., Payne, J., Stuewe, M., Davison, G., Alfred, J.R., Andau,
880 M., Bosi, E., Kilbourn, A., Melnick, D.J., 2003. DNA analysis indicates that
881 Asian elephants are native to Borneo and are therefore a high priority for
882 conservation. PLoS Biology 1, 110–115.
883 Frankham, R., Ballou, J.D., Briscoe, D.A., 2002. Introduction to conservation genetics.
884 Cambridge University Press, Cambridge, pp 617.
885 Frankham, R. 2010. Challenges and opportunities of genetic approaches to biological

886 conservation. Biol. Conserv.143: 1919 -1927.

887 Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R.
888 and Fenster, C. B., 2011. Predicting the Probability of Outbreeding
889 Depression. Conservation Biology, 25: 465–475.

890 François, O., Ancelet, S., Guillot, G., 2006. Bayesian clustering using hidden Markov
891 random fields in spatial population genetics. Genetics 174, 805–816.
892 François, O., Duran, E., 2010. The state of the field: Spatially explicit Bayesian
893 clustering models in population genetics. Molecular Ecology Resources 10:
894 773–784.
895 Gabriel, S., Ziaugra, L., Tabbaa, D., 2009. SNP genotyping using the Sequenom
896 MassARRAY iPLEX platform. Current Protocols in Human Genetics Chapter 2,
897 Unit 2.12.
898 Girod, C., Vitalis, R., Lebois, R., Freville, H., 2011. Inferring population decline and
899 expansion from microsatellite data: a simulation-based evaluation of the MSvar
900 methods. Genetics 188, 165–179.

901 Goossens, B., Chikhi, L., Jalil, M.F., Ancrenaz, M., Lackman-Ancrenaz, I., Mohamed,
902 M., Andau, P., Bruford, M.W., 2005. Patterns of genetic diversity and migration
903 in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*)
904 populations from Sabah, Malaysia. Molecular Ecology 14, 441–456.
905 Goossens, B., Chikhi, L., Ancrenaz, M., Lackman-Ancrenaz, I., Andau, P., Bruford,
906 M.W., 2006. Genetic signature of anthropogenic population collapse in orang-
907 utans. PLoS Biology 4, e25.
908 Goudet, J., 1995. F STAT (Version 1.2): A computer program to calculate F-statistics.
909 Journal of Heredity 86, 485–486.
910 Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., 2011.
911 Current trends in microsatellite genotyping. Molecular Ecology Resources 11,
912 591–611.
913 Haasl, R. J., and B. A. Payseur (2011). Multi-locus inference of population structure: A
914 comparison between single nucleotide polymorphisms and microsatellites.
915 Heredity 106(1): 158-71.
916 Hartl, D.L., Clark, A.G., 1989. Principles of population genetics, 2nd edn.

Sunderland, MA: Sinauer Associates Inc.

Haynes, G.D., Latch, E.K. 2012. Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus* spp.) using the BovineSNP50 Beadchip. PLoS ONE 7, e36536.

Jalil, M.F., Cable, J., Sinyor, J., Lackman-Ancrenaz, I., Ancrenaz, M., Bruford, M.W., Goossens, B., 2008. Riverine effects on mitochondrial structure of Bornean orang-utans (*Pongo pygmaeus*) at two spatial scales. Molecular Ecology 17, 2898–2909.

Jombart, T., Devillard, S., Dufour, A.B., Pontier, D., 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101, 92–103.

Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806.

Kalinowski, S.T., Wagner, A.P., Taper, M.L., 2006. ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes 6, 576–579.

Kongrit, C., Siripunkaw, C., Brockelman, W.Y., Akkarapatumwong, Wright, T.F., Eggert, L.S., 2007. Isolation and characterization of dinucleotide microsatellite loci in the Asian elephant (*Elephas maximus*). Molecular Ecology Notes 8: 175–177.

Kim, S., Mishra, A., 2007. SNP Genotyping: Technologies and Biomedical Applications. Annual Review of Biomedical Engineering 9, 289–320.

Kraus, R. H. S., vonHoldt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., Förster, D. W., Fickel, J., Roos, C. and Nowak, C. (2015), A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. Molecular Ecology Resources, 15: 295–305. doi: 10.1111/1755-0998.12307.

Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M.A., Luikart, G., 2010. Quantifying the lag time to detect barriers in landscape genetics. Molecular Ecology 19, 4179–4191.

948 Liu, N., Chen, L., Wang, S., Oh, C., Zhao, H., 2005. Comparison of single nucleotide
949 polymorphisms and microsatellites in inference of population structure. BMC
950 Genetics 6, 1471–2156.
951 McMahon, B. J., E. C. Teeling, J. Höglund. 2014. How and why should we
952 implement genomics into conservation? Evolutionary Applications 7:999-1007.
953 Morin PA, Luikart G, Wayne RK, the SNP workshop group: SNPs in ecology,
954 evolution and conservation. Trends Ecol Evol 2004, 19:208-216.
955 Morin, P.A., Mccarthy, M., 2007. Highly accurate SNP genotyping from historical and
956 low-quality samples. Molecular Ecology Notes 7, 937–946.
957 Morin, P.A., Martien, K.K., Taylor, B.L., 2009. Assessing statistical power of SNPs for
958 population structure and conservation studies. Molecular Ecology Resources 9,
959 66–73.
960 Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small
961 number of individuals. Genetics 89, 583–590.
962 Olivier, R., 1978. Distribution and status of the Asian elephant. Oryx 14, 379–424.
963 Oddou-Muratorio, S., Vendramin, G.G., Buiteveld, J., Fady, B., 2009. Population
964 estimators or progeny tests: what is the best method to assess null allele
965 frequencies at SSR loci? Conservation Genetics 10, 1343–1347.
966 Parreira, B., Chikhi, L., 2015. On some genetic consequences of social structure, mating
967 systems, dispersal and sampling. Proceedings of the National Academy of
968 Sciences USA 112, 3318–3326.
969 Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis.
970 PLoS Genetics 2, 2074–2093.
971 Piry, S., Luikart, G., Cornuet, J.M., 1999. BOTTLENECK: a computer program for
972 detecting recent reductions in the effective size using allele frequency
973 data. Journal of Heredity 90, 502–503.
974 Peakall, R., Smouse, P.E., 2006. GENALEX 6: genetic analysis in Excel. Population
975 genetic software for teaching and research. Molecular Ecology Notes 6, 288–
976 295.
977 Pritchard, J.K., Stephens, M., Donnelly, P.J., 2000. Inference of population structure
978 using multilocus genotype data. Genetics 155, 945–959.
979 Queller, D.C., Goodnight, K.F., 1989. Estimating relatedness using genetic markers.
980 Evolution 43, 258–275.

Queirós, J., Godinho, R., Lopes, S., Gortazar, C., de la Fuente, J., Alves, P.C., 2014. Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross specific and species-specific amplifications. Molecular Ecology Resources 15, 747–760.

Reed, D.H., Frankham, R., 2003. Correlation between fitness and genetic diversity. Conservation Biology 17, 230–237.

Rosenberg, N.A., 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4, 137–138.

Russello, M.A., Waterhouse, M.D., Etter, P.D., Johnson, E.A., 2015. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3, e1106.

Saccheri, I.J., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., Hanski, I., 1998. Inbreeding and extinction in a butterfly metapopulation. Nature 392, 401–494.

Santiapillai, C., Ramono, W.S., 1993. Why do elephants raid crops in Sumatra? Gajah 11, 55–58.

Shim, P.S., 2003. Another look at the Borneo elephant. Sabah Society Journal 20, 7–14.

Sharma, R., Arora, N., Goossens, B., Nater, A., Morf, N., Salmona, J., Bruford, M.W., van Schaik, C.P., Krützen, M., Chikhi, L., 2012a. Effective population size dynamics and the demographic collapse of Bornean orang-utans. PLoS ONE 7, e49429.

Sharma, R., Goossens, B., Kun-Rodrigues, C., Teixeira, T., Othman, N., Boone, J.Q., Jue, N.K., Obergfell, C., O'Neill, R.J., Chikhi, L., 2012b. Two different high throughput sequencing approaches identify thousands of *de novo* genomic markers for the genetically depleted Bornean elephant. PLoS ONE 7, e49533.

Silva, S.G.D., 1968. Elephants of Sabah. Sabah Society Journal 3, 169–181.

Sobrino, B., Brión, M., Carracedo, A., 2005. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International 154, 181–194.

Sukumar, R., 1989. The Asian elephant: ecology and management. Cambridge University Press, Cambridge.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P., Bouvet, J., 1996. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Research 24, 3189–3194.

1013 Templeton, A.R., Shaw, K., Routman, E., Davis, S.K., 1990. The genetic consequences
1014 of habitat fragmentation. Annals of the Missouri Botanical Garden 77, 13–27.
1015 Tokarska, M., Kawalko, A., Wojcik, J.M., Pertoldi, C., 2009. Genetic variability in the
1016 European bison (*Bison bonascus*) population from Bialowieza forest over 50
1017 years. Biological Journal of the Linnean Society 97, 801–809.
1018 Valière, N., 2002. GIMLET: a computer program for analysing genetic individual
1019 identification data. Molecular Ecology Notes 2, 377–379.
1020 Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. MICRO-
1021 CHECKER: software for identifying and correcting genotyping errors in
1022 microsatellite data. Molecular Ecology Notes 4, 535–538.
1023 Vidya, T.N.C., Sukumar, R., 2005. Social organization of the Asian elephants (*Elephas*
1024 *maximus*) in southern India inferred from microsatellite data. Journal of
1025 Ethology 23, 205–210.
1026 Vidya, T.N.C., Fernando, P., Melnick, D.J., Sukumar, R., 2005a Population
1027 differentiation within and among Asian elephant (*Elephas maximus*) populations
1028 in southern India. Heredity 94, 71–80.
1029 Vidya, T.N.C., Fernando, P., Melnick, D.J., Sukumar, R., 2005b. Population genetic
1030 structure and conservation of Asian elephants (*Elephas maximus*) across India.
1031 Animal Conservation 8, 377–388.
1032 Weir, B.S., Cockerham, C.C., 1984. Estimating *F*-statistics for the analysis of
1033 population structure. Evolution 38, 1358–1370.
1034 Williams, A.C., Johnsingh, A.J.T., Krausman, P.R., 2001. Elephant-human conflict in
1035 Rajaji National Park, northwest India. Wildlife Society Bulletin 29, 1097–1104.
1036 Wright, S., 1978. Variability within and among natural populations. University of
1037 Chicago Press, Chicago. Evolution and the Genetics of Populations, Vol. 4.
1038 Zhang L, Dong L, Lin L, Feng L, Yan F, Wang L, et al. 2015. Asian Elephants in
1039 China: Estimating Population Size and Evaluating Habitat Suitability. PLoS
1040 ONE 10(5): e0124834. doi:10.1371/journal.pone.0124834

Figure legends

Figure 1: Map showing locations where Bornean elephant fecal and blood samples were collected. Black squares in the map show the sites where elephant blood samples were obtained (Sandakan and Lahad Datu). The fecal samples were collected from elephants in four main ranges (as described in methods): i) Lower Kinabatangan, ii) North Kinabatangan (Deramakot), iii) Central Forest (Ulu Segama Malua, Kuamut, Gunung Rara, and Kalabakan Forest Reserves, Maliau Basin Conservation Areas), and iv) Tabin Wildlife Reserve.

Figure 2: Individual assignment probabilities of Bornean elephants to genetic clusters using the model-based programs STRUCTURE (above) and TESS (below) run of K = 2. Results from a) Microsatellites (for 224 individuals), b) SNPs (for 70 individuals). Each column represents q values, the proportions in which a given genotype belongs to a cluster of the given colour. Vertical black bars separate sampling sites. Geographic sampling locations are indicated below the figure. Multiple sites were sampled in the Central Forest (Ulu Segama Malua, Kuamut, Gunung Rara, Kalabakan Forest Reserves, and Maliau Basin Conservation Areas).

Figure 3: Spatially explicit predictive map of admixture coefficients as determined by TESS for 2 clusters, K_{max}=2 for microsatellite data. The color scale represents the posterior probability of individuals having membership to a single genetic cluster. Numbered ellipses indicate the populations; 1= Lower Kinabatangan, 2 = Deramakot (North Kinabatangan), multiple sites in Central Forest: (3 = Ulu Segama Malua (USM), 4 = Gunung Rara (GR), 5 = Malua Basin (MB), 6 = Kuamut (KU), and 7 = Kalabakan (KAL)), 8 = Tabin Wildlife Reserve (TWR).

Figure 4: Principal component analysis (PCA) of the Bornean elephant fecal samples based, a) microsatellite (Msat), b) SNP. Dots represent different individuals; bottom right inset shows eigenvalues of principle components in relative magnitude.

Supplementary Material

Figure S1: Sequenom SNP genotype calls observed in elephant blood and fecal samples across different plexes.

Figure S2: Determination of K, the number of genetic clusters of Bornean elephant from: a) microsatellite, b) SNP. The number of populations (K) vs. the second order derivative of structure's natural log-likelihood output (Delta K; as described in Evanno et al., 2005). The highest Delta K indicates the most likely number of populations. STRUCTURE (left) and Tess analysis (right).

Table S1: Multiplexes and fluorescent labels for 18 microsatellite loci.

Table S2: Estimates of null allele frequencies across all elephant populations by loci.

Table S3: Details of Sequenom SNP genotype call for Bornean elephant samples: a.) Blood, and b) Feces.

Table S4: Analysis of pairwise relatedness using 18 microsatellite loci: a) Summary of

Mean within population pairwise relatedness values, and b) Summary of Mean within "elephant family units" pairwise relatedness values (Queller and Goodnight estimator, 1989).

Table S5: Genetic diversity measures using corrected data across microsatellites (mean across 18 loci).

1081

1082

Figure 1

Figure 2

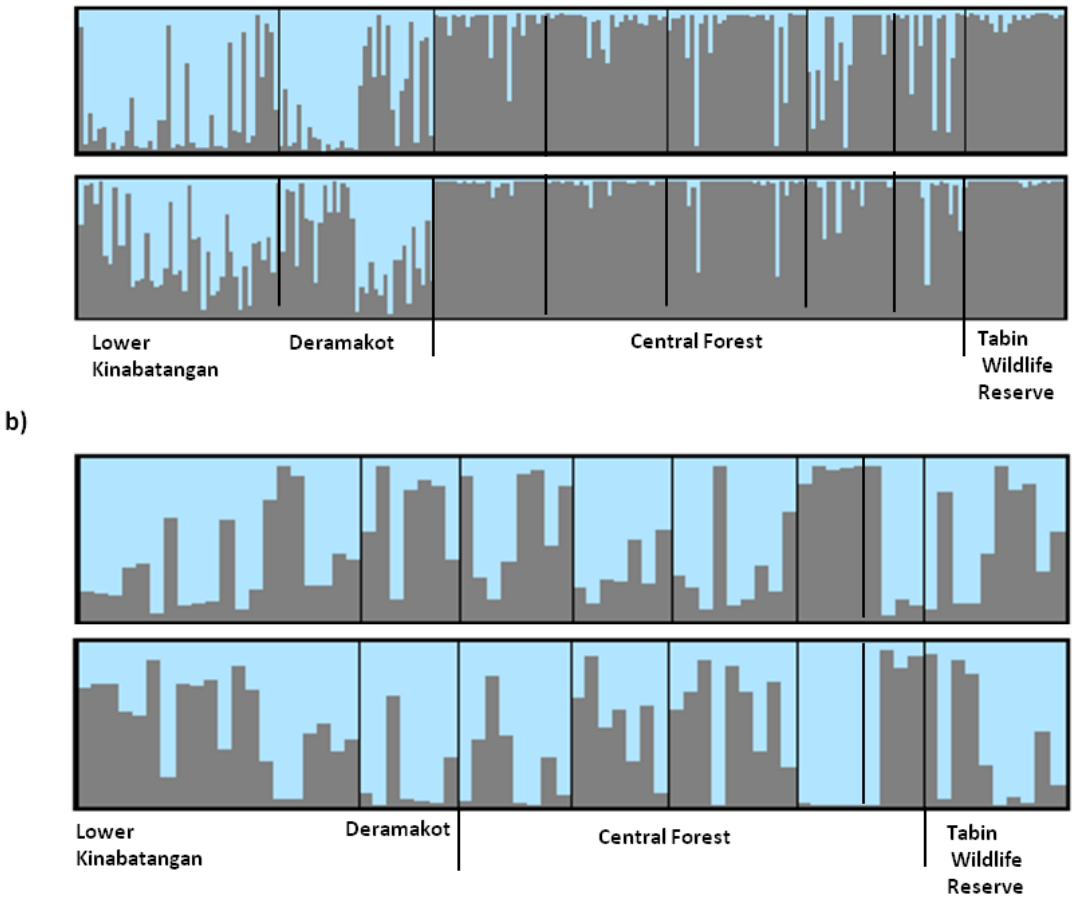
a)
Lower Kinabatangan
Deramakot
Central Forest
Tabin Wildlife Reserve
b)
Lower Kinabatangan
Deramakot
Central Forest
Tabin Wildlife Reserve

Figure 3

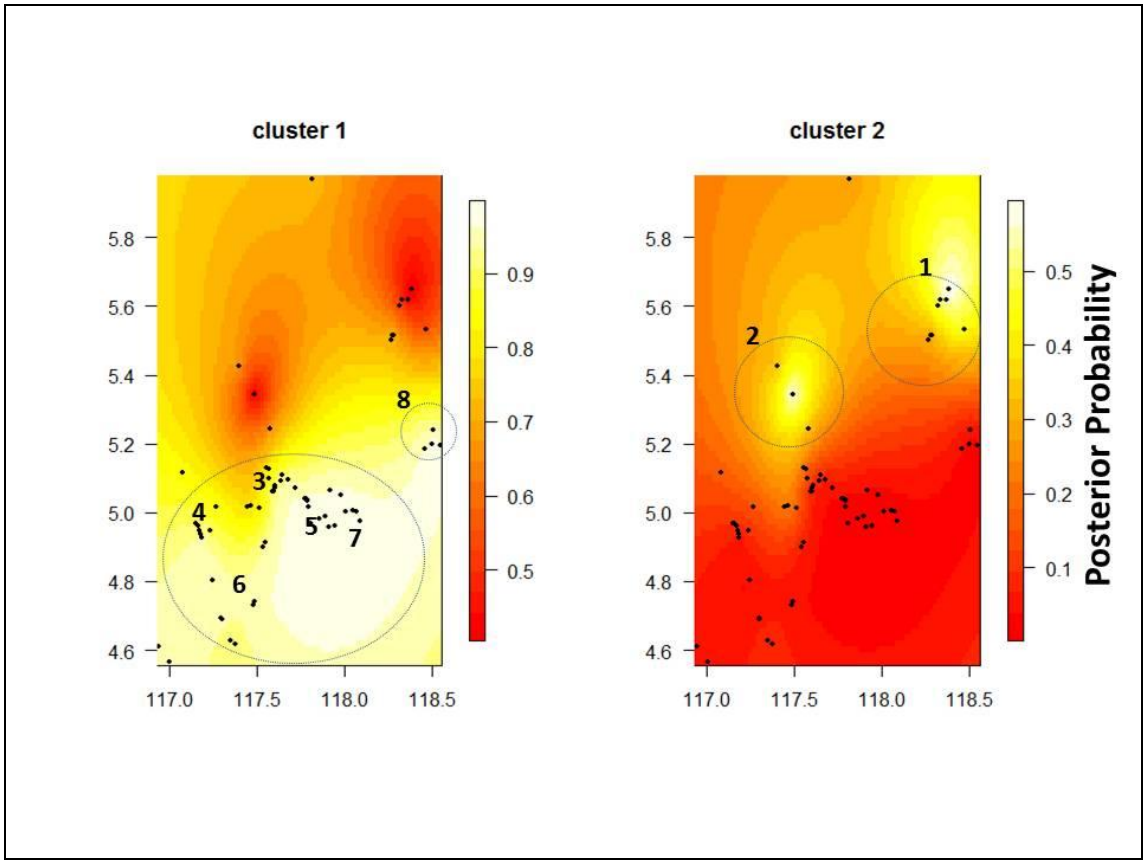
cluster 1
cluster 2
Posterior Probability
1
2
3
4
5
6
7
8
5.8
5.6
5.4
5.2
5.0
4.8
4.6
117.0
117.5
118.0
118.5
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

Figure 4

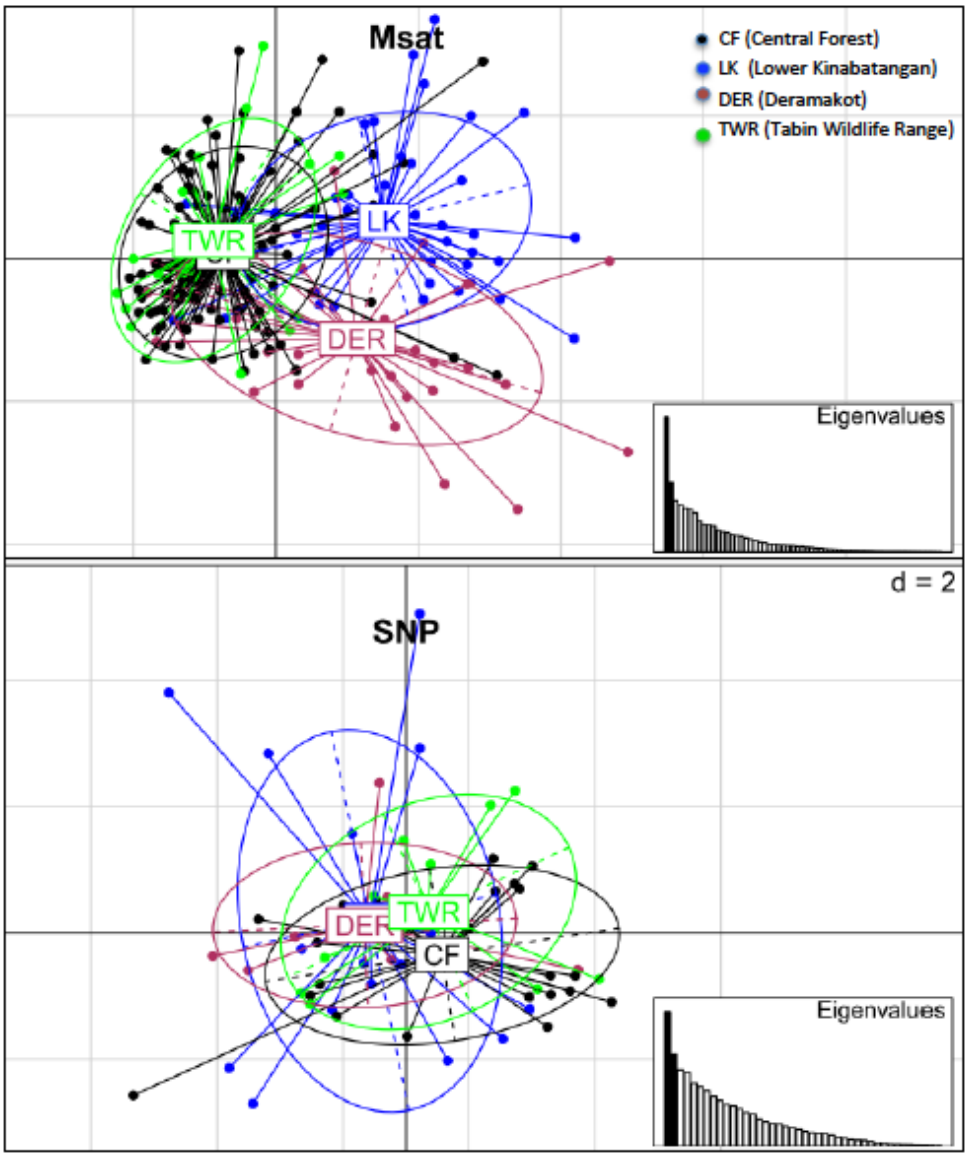
Msat
CF (Central Forest)
LK (Lower Kinabatangan)
DER (Deramakot)
TWR (Tabin Wildlife Range)
TWR
LK
DER
Eigenvalues
d = 2
SNP
DER
TWR
CF
Eigenvalues

Figure S1

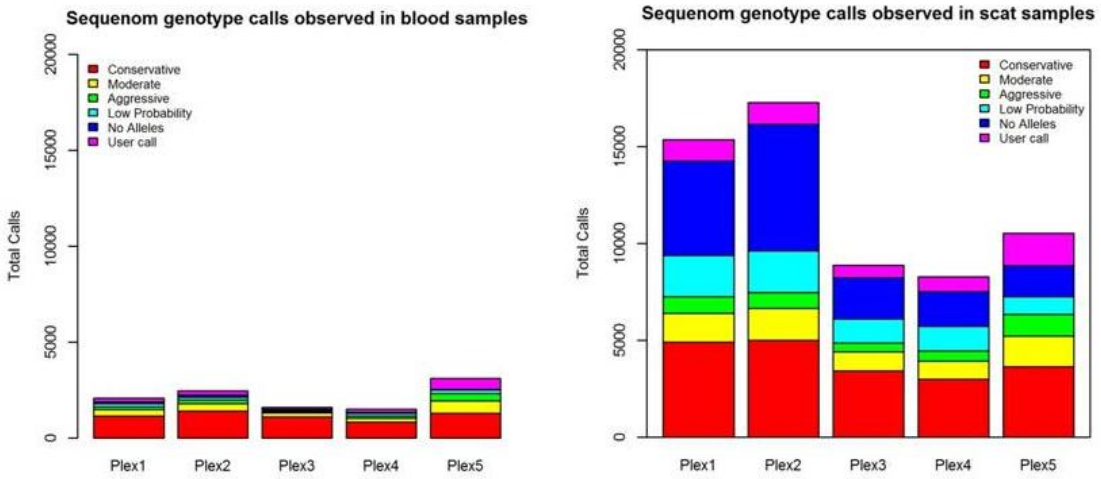
Sequenom genotype calls observed in blood samples
Total Calls
0
5000
10000
15000
20000
Conservative
Moderate
Aggressive
Low Probability
No Alleles
User call
Plex1
Plex2
Plex3
Plex4
Plex5
Sequenom genotype calls observed in scat samples
Total Calls
0
5000
10000
15000
20000
Conservative
Moderate
Aggressive
Low Probability
No Alleles
User call
Plex1
Plex2
Plex3
Plex4
Plex5

Figure S2

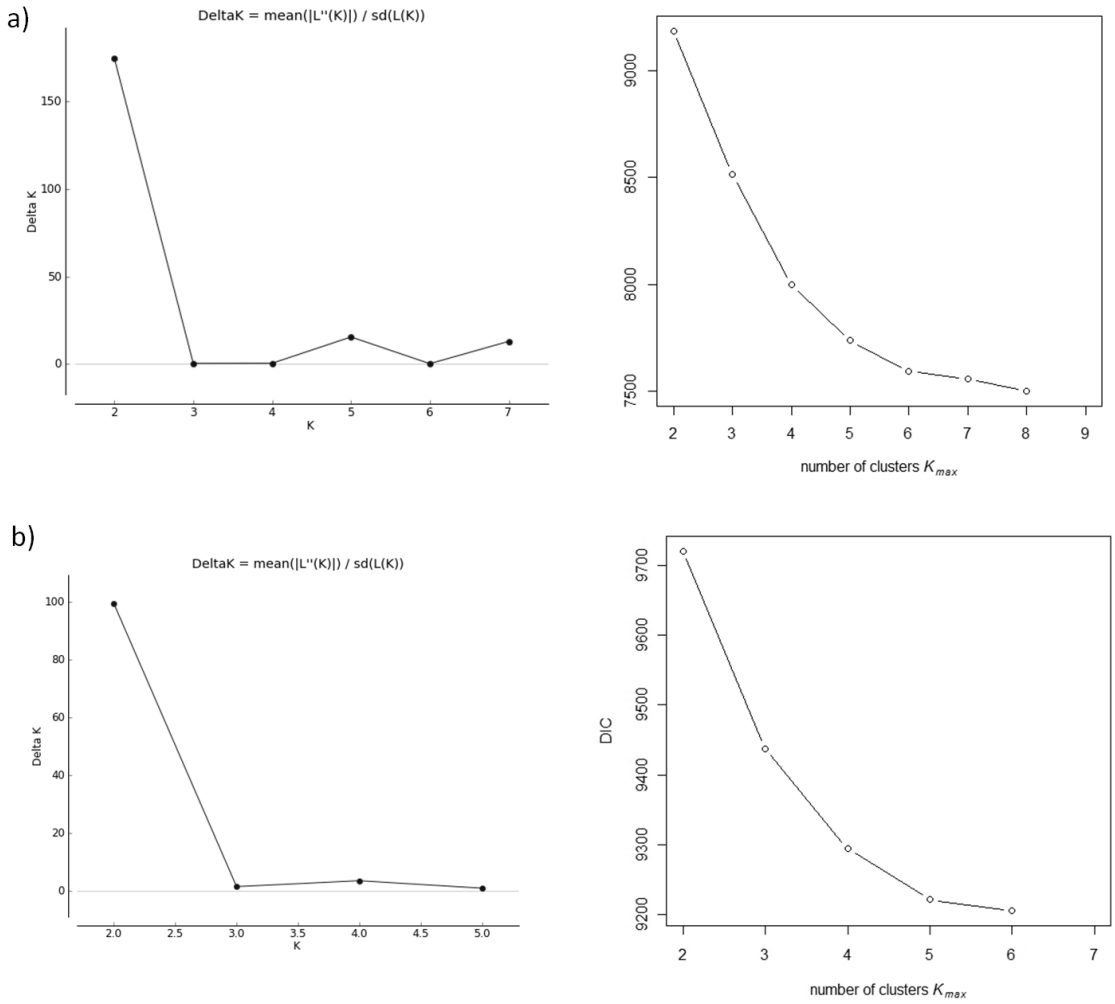
a)
DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
K
number of clusters K_{max}
b)
DeltaK = mean(|L''(K)|) / sd(L(K))
Delta K
K
DIC
number of clusters K_{max}

Figure 1

Click here to download high resolution image

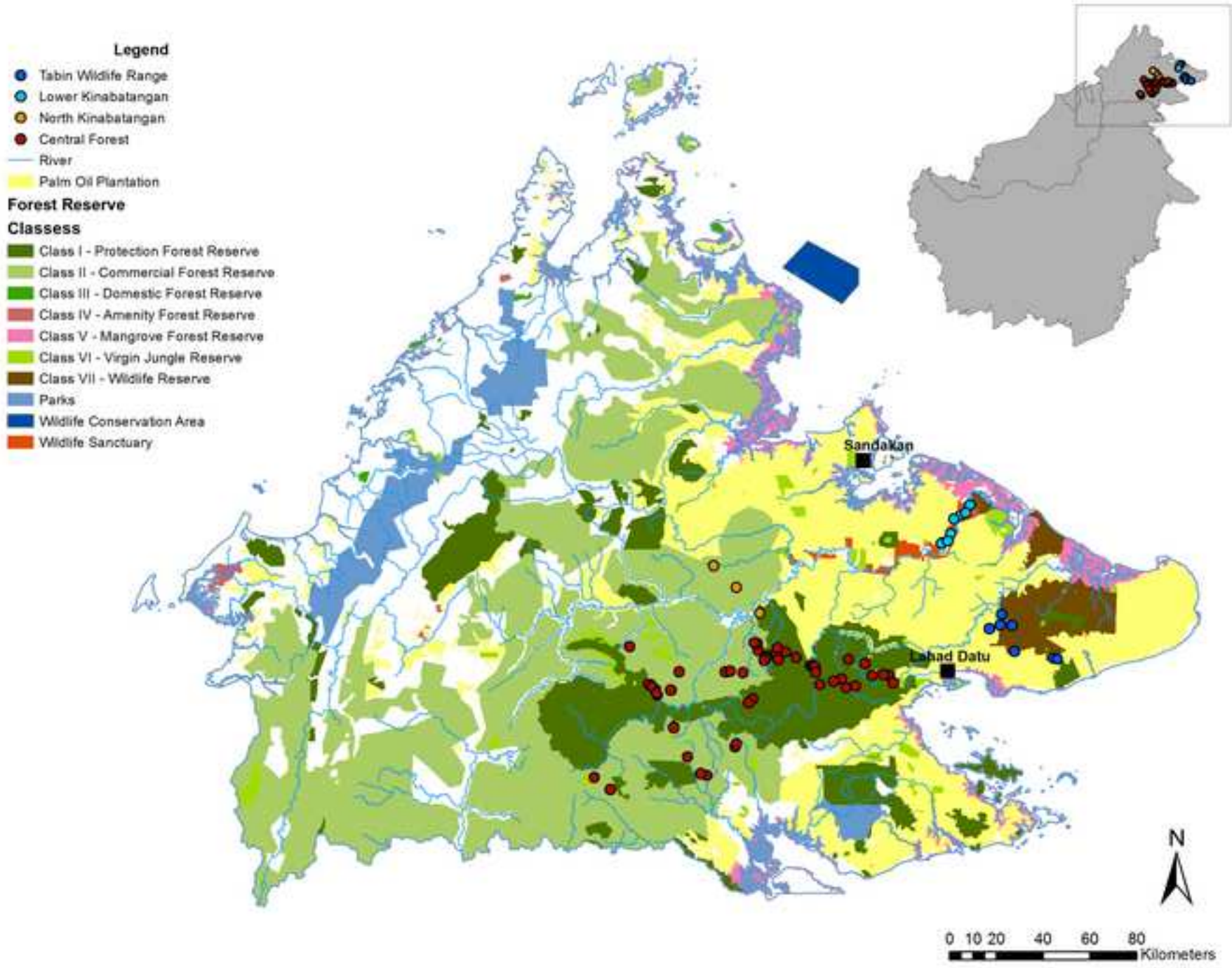

Table 1

Table 1. Details of elephant samples used for each type of marker

Major Bornean elephant range	Sampled site	Analyses for each type of marker			
		Mitochondrial DNA (mtDNA)	Microsatellites (Msat)	Single nucleotide polymorphisms (SNPs)	
		Feces	Feces	Blood	Feces
Lower Kinabatanagan	Lower Kinabatanagan (LK)	5	43	10	20
	Sandakan	-	-	6	-
North Kinabatangan	Deramakot (DER)	-	33	-	9
Central Forest (CF)	Ulu Segama Malua (USM)	25	50	-	7
	Gunung Rara (GR)	8	30	-	8
	Malua Basin (MB)	-	7	-	7
	Kuamut (KU)	-	6	-	-
	Kalabakan (KAL)	-	34	-	9
	Lahad Datu	-	-	4	-
Tabin Wildlife Reserve	Tabin Wildlife Reserve (TWR)	5	21	-	10
-	Lok Kawi Wildlife Park (captive elephants)	4	-	-	-
Total number of samples analysed for each marker type (*n*)		47	224	20	70

Table 2

Table 2. Genetic diversity measures in Bornean elephant using: 18 microsatellite loci (a), and co
Mean number of alleles (MNA), Allelic richness (AR), and departures from Hardy-Weinb

a.

Population	Sample size (*n*)	Expected Heterozygosity
Lower Kinabatanagan (LK)	43	0.47
Deramakot (DER)	33	0.42
Central Forest (CF)	127	0.25
Tabin Wildlife Reserve (TWR)	21	0.21
All	224	0.34

b.

Population	Sample size (*n*)	Expected Heterozygosity
Lower Kinabatanagan (LK)	20	0.33
Deramakot (DER)	9	0.30
Central Forest (CF)	31	0.23
Tabin Wildlife Reserve (TWR)	10	0.22
All	70	0.27

*: significant

Table S1

Table S1. Multiplexes and fluorescent labels for 18 microsatellite loci.

Multiplexes	Loci	Fluorescent labels
1	EMU-01	6-FAM
	EMU-02	VIC
	EMU-03	NED
2	EMU-04	6-FAM
	EMU-05	VIC
	EMU-14	6-FAM
3	EMU-06	NED
	EMU-17	VIC
	EMU-19	6-FAM
4	EMU-07	6-FAM
	EMU-09	NED
	EMU-12	VIC
5	EMU-08	VIC
	EMU-10	NED
	EMU-18	6-FAM
6	EMU-11	6-FAM
	EMU-13	NED
	EMU-15	VIC

Table 3

Table 3: Comparison of genetic diversity indices across SNP and microsatellite data for 70 individua

Elephant population	*n*	Expected Heterozygosity		Observed Heterozygosity	
		Microsatellites	SNPs	Microsatellites	SNPs
Lower Kinabatanagan (LK)	20	0.45	0.29	0.39	0.26
Deramakot (DER)	9	0.24	0.32	0.17	0.31
Central Forest (CF)	31	0.23	0.23	0.16	0.23
Tabin Wildlife Reserve (TWR)	10	0.21	0.23	0.15	0.22
All	70	0.28	0.27	0.22	0.25

*significant

NS:not significant

Table 4

Table 4: Pairwise F_{ST} values between the sampling sites for microsatellite data. The pairwise F_{ST} value between the two major genetic clusters identified using STRUCTURE was 0.087.

Total number of samples analysed (*n*)	Four major Bornean elephant range	Lower Kinabatangan	North Kinabatangan	Central Forest	Tabin Wildlife Reserve
43	Lower Kinabatangan (LK)	0.00			
33	North Kinabatangan (DER)	**0.08**	0.00		
127	Central Forest (CF)	**0.11**	**0.12**	0.00	
21	Tabin Wildlife Reserve (TWR)	**0.11**	**0.14**	**0.03**	0.000

bold indicates significant values of *p* value <0.001.

Table S2

Table S2. Estimates of null allele frequencies across all elephant populations by loci

Bornean elephant Population in four major range	Microsatellite loci																	
	EMU01	EMU02	EMU03	EMU04	EMU05	EMU14	EMU06	EMU17	EMU19	EMU07	EMU09	EMU12	EMU08	EMU10	EMU18	EMU11	EMU13	EMU15
Lower Kinabatanagan (LK)	0.05	0.00	0.00	0.00	0.01	0.03	**0.12**	0.10	**0.11**	0.00	**0.13**	0.07	0.07	0.00	0.04	0.00	0.00	0.00
Deramakot (DER)	0.09	**0.11**	**0.12**	0.00	**0.10**	0.10	**0.11**	**0.15**	0.00	**0.14**	**0.16**	0.00	**0.14**	0.09	**0.16**	0.09	**0.11**	**0.10**
Central Forest (CF)	0.03	0.03	0.09	0.07	0.03	0.07	0.05	0.06	0.00	0.07	0.06	0.03	0.04	0.03	0.00	0.07	0.07	0.03
Tabin Wildlife Reserve (TWR)	0.00	0.00	**0.17**	**0.11**	0.00	0.00	0.00	**0.10**	0.00	0.00	**0.11**	0.06	0.01	0.08	0.06	0.00	**0.20**	0.00

frequencies of greater than 10 percent null alleles indicated in bold.

Table S3

Table S3. Details of Sequenom SNP genotype call for Bornean elephant samples: Blood (a) and Feces (b)

a.

Plex	Sequenom call type	After user call	
		Blood (*n* =20)	%
Plex 1	Conservative	1126	54.13
	Moderate	353	16.97
	Aggressive	135	6.49
	Low Probability	162	7.79
	No Alleles	98	4.71
	User Call	206	9.90
Plex 2	Conservative	1400	57.19
	Moderate	392	16.01
	Aggressive	179	7.31
	Low Probability	151	6.17
	No Alleles	105	4.29
	User Call	221	9.03
Plex 3	Conservative	1098	68.84
	Moderate	197	12.35
	Aggressive	85	5.33
	Low Probability	60	3.76
	No Alleles	45	2.82
	User Call	110	6.90
Plex 4	Conservative	824	55.49
	Moderate	200	13.47
	Aggressive	110	7.41
	Low Probability	158	10.64
	No Alleles	42	2.83
	User Call	151	10.17
Plex 5	Conservative	1291	41.54
	Moderate	627	20.17
	Aggressive	393	12.64
	Low Probability	183	5.89
	No Alleles	29	0.93
	User Call	585	18.82

Total number of expected calls= 10,716
Total number of obtained calls= 10,397

b.

Plex	Sequenom call type	After us
		Feces (*n* =70)
Plex 1	Conservative	4904
	Moderate	1475
	Aggressive	861
	Low Probability	2138
	No Alleles	4878
	User Call	1104
Plex 2	Conservative	4989
	Moderate	1659
	Aggressive	806
	Low Probability	2161
	No Alleles	6535
	User Call	1130
Plex 3	Conservative	3411
	Moderate	972
	Aggressive	457
	Low Probability	1244
	No Alleles	2129
	User Call	661
Plex 4	Conservative	2986
	Moderate	926
	Aggressive	532
	Low Probability	1276
	No Alleles	1791
	User Call	751
Plex 5	Conservative	3622
	Moderate	1591
	Aggressive	1125
	Low Probability	899
	No Alleles	1597
	User Call	1674

Total number of expected calls= 60,284
Total number of obtained calls= 43,354

Table S4

Table S4. Analysis of pairwise relatedness based on data from 18 microsatellite loci:

a. Summary of Mean within population pairwise relatedness values (Queller and Goodnight estimator, 1989).

b. Summary of Mean within "elephant family units" pairwise relatedness values (Queller and Goodnight estimator, 1989).

a.

Elephant population	Mean	U	L	*P* (mean-rand >= mean-data)	*P* (mean-rand <= mean-data)
Lower Kinabatanagan (LK)	-0.260	0.156	-0.144	1.000	0.001
Ulu Segama Malua (USM)	0.413	0.120	-0.131	0.001	1.000
Gunung Rara (GR)	0.348	0.189	-0.167	0.001	1.000
Deramakot (DER)	-0.194	0.158	-0.158	0.994	0.007
Kalabakan (KAL)	0.264	0.166	-0.165	0.002	0.999
Tabin Wildlife Reserve (TWR)	0.316	0.243	-0.206	0.008	0.993

b.

Family Units	Mean	U	L	*P* (mean-rand >= mean-data)	*P* (mean-rand <= mean-data)
kinabG1 (LK)	-0.303	0.231	-0.229	0.998	0.003
kinabG2 (LK)	-0.249	0.442	-0.344	0.895	0.106
kinabG3 (LK)	-0.258	0.479	-0.398	0.839	0.162
kinabG7 (LK)	0.155	0.410	-0.341	0.199	0.802
UsmG12 (USM)	0.710	0.658	-0.593	0.022	0.979
UsmG15 (USM)	0.608	0.726	-0.567	0.057	0.944
UsmG29 (USM)	0.783	0.530	-0.362	0.001	1.000
UsmG32 (USM)	0.380	0.480	-0.393	0.067	0.934
UsmG33 (USM)	0.450	0.450	-0.336	0.026	0.975
GRG40 (GR)	0.341	0.210	-0.186	0.001	1.000
DermkG49 (DER)	-0.166	0.247	-0.224	0.903	0.098
DermkG50 (DER)	0.065	0.262	-0.221	0.270	0.731
kalabakG52 (KAL)	0.195	0.453	-0.351	0.169	0.832
kalabakG54 (KAL)	0.403	0.214	-0.186	0.002	0.999
tabinwrG55 (TWR)	0.246	0.351	-0.268	0.075	0.926
tabinwrG56 (TWR)	0.583	0.386	-0.328	0.002	0.999

Upper (U) and lower (L) confidence limits bound the 95% confidence interval about the null hypothesis of 'No Difference' across the populations as determined by permutation.

Table S5

Table S5: Genetic diversity measures using corrected data across microsatellites (mean across 18 loci)

Bornean elephant populations	*n*	Hexp (unb)	Hexp (unb) corrected	Hobs	Hobs corrected	F_{IS}	$F_{IS\ Corrected}$
Lower Kinabatanagan (LK)	43	0.47	0.49	0.41	0.46	0,14*	0.08^{NS}
Deramakot (DER)	33	0.42	0.48	0.26	0.40	0,38*	0,18*
Central Forest (CF)	127	0.25	0.26	0.18	0.23	0,27*	0,16*
Tabin Wildlife Reserve (TWR)	21	0.21	0.28	0.14	0.29	0,36*	-0.012
All	224	0.34	0.38	0.25	0.35	0,30*	-0.01

*: significant

NS:not significant