Journal of Heredity Advance Access published October 16, 2012

Journal Of Heredity
doi:10.1093/jhered/ess065

Marked Population Structure and Recent Migration in the Critically Endangered Sumatran Orangutan (*Pongo abelii*)

Alexander Nater, Natasha Arora, Maja P. Greminger, Carel P. van Schaik, Ian Singleton, Serge A. Wich, Gabriella Fredriksson, Dyah Perwitasari-Farajallah, Joko Pamungkas, and Michael Krützen

From the Anthropological Institute & Museum, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland (Nater, Arora, Greminger, van Schaik, and Krützen); Foundation for a Sustainable Ecosystem (YEL), Medan, Indonesia (Singleton and Fredriksson); Sumatran Orangutan Conservation Programme (PanEco-YEL), Medan, Indonesia (Singleton, Wich, and Fredriksson); Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, Liverpool, United Kingdom (Wich); Primate Research Center, Bogor Agricultural University, Bogor, Indonesia (Perwitasari-Farajallah and Pamungkas); Department of Biology, Bogor Agricultural University, Bogor, Indonesia (Perwitasari-Farajallah); Department of Animal Infectious Diseases and Veterinary Public Health, Bogor Agricultural University, Bogor, Indonesia (Pamungkas).

Address correspondence to Alexander Nater at the address above, or e-mail: a.nater@aim.uzh.ch.

Abstract

A multitude of factors influence how natural populations are genetically structured, including dispersal barriers, inhomogeneous habitats, and social organization. Such population subdivision is of special concern in endangered species, as it may lead to reduced adaptive potential and inbreeding in local subpopulations, thus increasing the risk of future extinctions. With only 6600 animals left in the wild, Sumatran orangutans (*Pongo abelii*) are among the most endangered, but also most enigmatic, great ape species. In order to infer the fine-scale population structure and connectivity of Sumatran orangutans, we analyzed the most comprehensive set of samples to date, including mitochondrial hyper-variable region I haplotypes for 123 individuals and genotypes of 27 autosomal microsatellite markers for 109 individuals. For both mitochondrial and autosomal markers, we found a pronounced population structure, caused by major rivers, mountain ridges, and the Toba caldera. We found that genetic diversity and corresponding long-term effective population size estimates vary strongly among sampling regions for mitochondrial DNA, but show remarkable similarity for autosomal markers, hinting at male-driven long-distance gene flow. In support of this, we identified several individuals that were most likely sired by males originating from other genetic clusters. Our results highlight the effect of natural barriers in shaping the genetic structure of great ape populations, but also point toward important dispersal corridors on northern Sumatra that allow for genetic exchange.

Key words: *conservation, gene flow, Great apes, microsatellites, Sundaland*

Most natural populations do not behave like single units, in which random mating occurs over the entire distribution (Kimura and Weiss 1964). Rather, most populations are genetically structured, the extent of which is determined by several factors. Geographical factors include both isolation by distance (Wright 1943) and physical barriers impeding gene flow across them, such as mountain ridges, rivers, and deserts. Ecological factors concern the distribution of resources and predators, which may lead to an aggregation of individuals within high-quality habitat patches (Slatkin 1987). A third category includes social, mating, and dispersal behaviors. Gregarious species, where individuals live in social groups, often show a marked population structure even in the complete absence of obvious geographical or ecological factors (Storz 1999; Ross 2001). Yet, strong genetic structuring imposed by limited dispersal has also been found in non-gregarious species. This is because in both gregarious and non-gregarious species it is potentially advantageous for individuals to show some degree of philopatry, as in the natal area food resources are familiar and kin is available

for social support (Johnson and Gaines 1990; Handley and Perrin 2007). Moreover, dispersal is usually heavily biased toward one sex, because one major benefit of dispersal is the avoidance of inbreeding (Bengtsson 1978; Pusey 1987). As a consequence, the extent of observed genetic structure may vary greatly depending on the inheritance mode of the genetic marker system used to investigate such patterns.

The underlying genetic structure of populations is especially important from a conservation perspective. Genetic structure may lead to local isolation of gene pools, resulting in effective subpopulation sizes that are only a fraction of the effective population size in a population without substructure (Charlesworth 2009). This has three important evolutionary consequences. First, lower effective sizes of subpopulations lead to stronger genetic drift effects and a reduced number of mutation events in each subpopulation. As a consequence, genetic diversity within each subpopulation will be lower compared with that of an unstructured population. Moreover, deleterious mutations that would be eliminated by background selection in unstructured populations might become fixed in small subpopulations, thus reducing the average population fitness (Hedrick and Kalinowski 2000; Reed and Frankham 2003). Second, population structuring increases the chance of mating among relatives, therefore causing potential loss of fitness due to inbreeding depression (Hedrick and Kalinowski 2000). Third, local separation of genetic variants will allow different selection pressures to act on specific subpopulations, thus allowing for adaptations to specific local environmental conditions (Williams 1966; Kawecki and Ebert 2004). While local adaptations raise the average fitness of subpopulations in a constant environment, the loss of genetic diversity reduces the potential of the subpopulations to adapt to changing environmental conditions and therefore carries greater risks of future extinctions (Reed and Frankham 2003). All these negative effects, however, can be counterbalanced by gene flow among subpopulations (Slatkin 1987). Therefore, knowledge about the extent to which genetic diversity is structured and exchanged across the range of a species is crucial to predict the long-term survival of populations and to implement effective conservation measures.

Population subdivision is a major concern in large-bodied animals with small population sizes, slow life histories, and low rates of reproduction, as such taxa are especially vulnerable to the aforementioned negative effects of population fragmentation (Hedrick and Kalinowski 2000). Great apes are of special interest in investigating the causes and consequences of population subdivision, not only because studying their population histories can reveal valuable insights into the evolution of modern humans, but also because all extant species are listed as endangered or even critically endangered (IUCN 2011). Furthermore, great apes show variation in dispersal patterns, which affects the genetic structuring of populations. For instance, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) show female-biased dispersal (Tautz et al. 1999; Mitani et al. 2002), whereas males are the dispersing sex in orangutans (*Pongo* spp.) (Singleton and van Schaik 2002; Morrogh-Bernard et al. 2011; van Noordwijk et al. 2012; Arora et al. 2012), as is the case in most mammals (Dobson 1982). In contrast, in gorillas (*Gorilla* spp.), both sexes disperse, even though mean dispersal distance is different between males and females (Douadi et al. 2007).

In the past, a substantial body of work has investigated population structure in great apes, such as in chimpanzees (Becquet et al. 2007; Gonder et al. 2011), bonobos (Eriksson et al. 2004; Eriksson et al. 2006), gorillas (Bergl and Vigilant 2007; Guschanski et al. 2008), and Bornean orangutans (*Pongo pygmaeus*) (Warren et al. 2001; Goossens et al. 2005; Jalil et al. 2008; Arora et al. 2010). Yet, a detailed population genetic analysis of Sumatran orangutans (*Pongo abelii*) is still lacking, even though Sumatran orangutans are critically endangered (IUCN 2011). As of today, only an estimated 6600 individuals remain in the wild, when compared with about 54 000 Bornean orangutans (Wich et al. 2008). In contrast to the Bornean species, where three subspecies have been defined based on morphological characters (Groves 2001), no subspecies have been proposed for Sumatran orangutans.

Historically, Sumatran orangutans populated most of the Indonesian island of Sumatra, as evidenced by fossil finds and historical records (Rijksen and Meijaard 1999; Delgado and Van Schaik 2000). The current distribution is, however, restricted to small forest patches on the northern tip of Sumatra (Wich et al. 2008). Ecological and anthropogenic factors, such as prehistoric hunting and recent deforestation, have been suggested as explanations for the drastic range collapse of orangutans (Delgado and Van Schaik 2000). The comparatively limited range of Sumatran orangutans that remains today is subdivided by major rivers and mountain ridges. Moreover, the massive forest exploitation that started in the last century (Rijksen and Meijaard 1999) has caused severe habitat fragmentation, leaving habitat blocks of continuous forest that often harbor only a few hundred individuals (Wich et al. 2008). This habitat fragmentation in combination with the potentially very strong reproductive skew in Sumatran orangutan males (Setia and van Schaik 2007; Utami Atmoko et al. 2009) might have drastically reduced the effective sizes of local subpopulations, thus minimizing genetic diversity and posing a severe threat of future extinctions.

Sumatran orangutans show the strictest arboreality among all great apes (Delgado and Van Schaik 2000) and occur in two different rain-forest habitat types. Low-altitude peat-swamp forests offer high and constant food supplies and support the highest population densities (Husson et al. 2009). At lower densities, permanent populations of Sumatran orangutans can be found in dry-land forests up to an altitude of 1500 m above sea level or more (Wich et al. 2004; Husson et al. 2009). However, in non-riverine dry-land forests, the mast fruiting phenomenon causes extreme temporal fluctuations in food availability (Knott 1998; Husson et al. 2009), which may act as a strong selective pressure for adaptive traits related to prolonged food scarcity. Unfortunately, due to the absence of long-term field studies covering the entire extant range of Sumatran orangutans, little is known about variation in behavior, physiology, and morphology within this species that could hint at the presence of habitat specific adaptations.

The current lack of knowledge about the genetic structure of Sumatran orangutans is mainly caused by difficulties in obtaining samples with reliable provenance throughout the entire species' range. This factor prevented most previous genetic studies from interpreting the extraordinary high diversity on the mitochondrial DNA (mtDNA) level they found in Sumatran orangutans when compared with their Bornean sister species (Muir et al. 2000; Kanthaswamy et al. 2006; Steiper 2006). However, using samples with a well-defined geographic origin, Nater et al. (2011) showed that mitochondrial variation is strongly geographically structured on Sumatra. This study identified four distinct mitochondrial clusters in Sumatran orangutans, with divergence times of up to 3.5 million years. Similar, albeit less-pronounced patterns of geographical structuring of mtDNA was found in Bornean orangutans (Warren et al. 2001; Arora et al. 2010). However, mtDNA is not a good indicator of population structure and gene flow in species that show a strong male-bias in dispersal, like orangutans (Galdikas 1995; Singleton and van Schaik 2002; Morrogh-Bernard et al. 2011). In fact, using Y-chromosomal markers, Nater et al. (2011) showed that the deep divergence and strong geographic clustering observed with mtDNA is not present in the male population history, indicating long-distance migration by males across Sumatra. The amount of gene flow and the resulting extent of homogenization of autosomal gene pools among local subpopulations is, however, impossible to measure using only sex-linked marker systems.

In this study, we aimed to unravel patterns of genetic diversity and differentiation in Sumatran orangutans, using a combination of mitochondrial and autosomal genetic markers. We investigated the role of geographical, ecological, and behavioral factors underlying the fine-scale population structure and tested for connectivity among subpopulations. To achieve this, we analyzed the most comprehensive and largest set of orangutan samples from Sumatra to date, using samples from wild individuals originating from the entire species' range.

Materials and Methods

Sample Collection

Three different kinds of orangutan samples were analyzed for this study: First, fecal samples were collected non-invasively at long-term study sites. Second, in areas where animals were not habituated, we collected hair samples from deserted nests. Third, we obtained blood and hair samples of confiscated wild-born orangutans from the quarantine station of the Sumatran Orangutan Conservation Program (SOCP) in Medan, North Sumatra.

We obtained orangutan samples from seven different sampling regions (Figure 1A): Tripa (TR), North Aceh (NA, north of Tamiang River), West Leuser (WL), Central Leuser (CL, west side of Alas River), Langkat (LK, east of Alas River, south of Tamiang River), Batu Ardan (BA, east of Alas River, west of Lake Toba), and Batang Toru (BT, south of Lake Toba) (see Supplementary Table S1 online). Fecal and hair samples were collected and stored following the genetic sampling protocol of the orangutan network (http://www.aim.uzh.ch/orangutannetwork, last accessed August 24, 2012). All blood samples were taken during routine veterinary examination in the SOCP quarantine station. Blood samples were collected in standard EDTA blood collection tubes and stored at −20 °C.

The amount and reliability of information about the wild origin of rehabilitant orangutans varied considerably. We classified the provenance of these individuals as reliable if the location of confiscation was known in detail and if this location was near an extant wild orangutan population. The samples from rehabilitant orangutans that did not meet these criteria were classified as having unknown provenance and excluded from certain analyses (see below).

The collection and transport of samples was carried out in compliance with Indonesian and international regulations. Samples were exported from Indonesia to Zurich under the Convention on International Trade in Endangered Species (CITES permits 09717/IV/SATS-LN/2010, 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007, 06968/IV/SATS-LN/2005).

Laboratory Procedures

DNA from fecal, hair, and blood samples was extracted and processed following the procedures described in Nater et al. (2011). We used a set of 12 human-derived (Goossens et al. 2005) and 15 species-specific microsatellite markers (Nietlisbach et al. 2010) to genotype the orangutan samples. In order to minimize genotyping errors due to allelic dropout, we followed the real-time PCR approach from Morin et al. (2001), performing between two and seven independent PCR repetitions per sample. PCR conditions and fragment length analysis are described in Arora et al. (2010) and Nietlisbach et al. (2010). We were able to genotype 112 out of 162 samples for at least 24 microsatellite loci. The identity check revealed three and two samples that were present as a triplicate and a duplicate, respectively, resulting in 109 unique genotypes.

For the sequencing of the hyper-variable region I (HVRI) of the mtDNA d-loop, we used the same primers, PCR conditions, and sequencing chemistry as Arora et al. (2010), resulting in a final alignment of 457 base pairs. Some sequences were from samples with insufficient DNA quantity for successful microsatellite genotyping. To avoid duplicates in the HVRI dataset, we only included sequences from individuals that had either a distinct genotype or were sampled more than 50 km apart from other samples in the dataset, resulting in 123 HVRI sequences. The sequences are deposited on GenBank under the accession numbers JQ962945–JQ962972.

HVRI Median-Joining Network

A median-joining network (Bandelt et al. 1999) using all HVRI sequences was drawn using NETWORK v4.6.0.0 and NETWORK PUBLISHER v1.3.0.0 (http://www.fluxus-engineering.com, last accessed August 24, 2012). An epsilon value of zero and equal weighting of all nucleotide positions

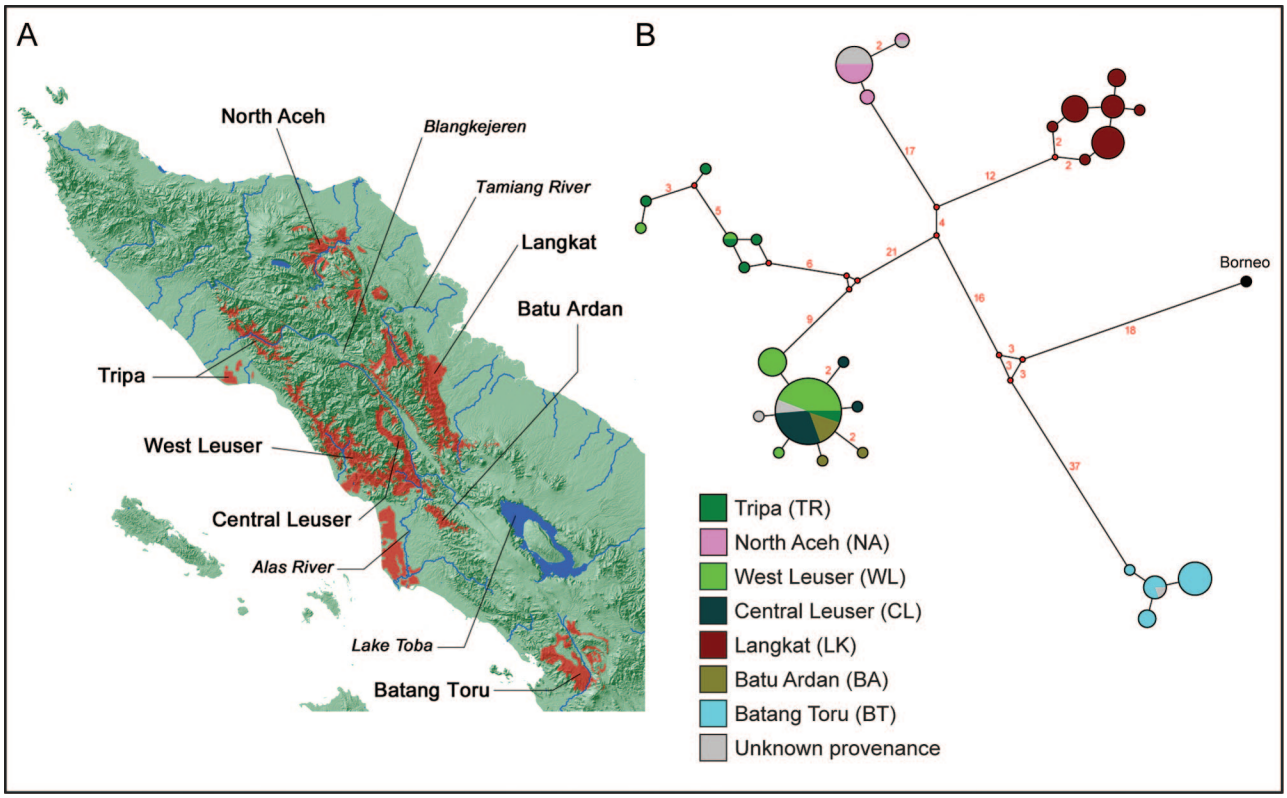


Figure 1. (**A**) Map of sampling regions in northern Sumatra. Labels in italics denote important geographic features. The red shading represents the current distribution of Sumatran orangutans. (**B**) Median-joining network of mitochondrial HVRI haplotypes. The red numbers in between the nodes indicate the number of mutational steps in between haplotypes (one step if not indicated otherwise). The size of each node is proportional to the number of individuals with the same haplotype.

was used for the network presented here. Using higher epsilon values or differently weighted transitions/transversions did not change the basic structure of the network.

Summary Statistics

We computed summary statistics and genetic differentiation measures for HVRI sequences and autosomal microsatellites using ARLEQUIN v3.5.1.2 (Excoffier and Lischer 2010). For both mitochondrial and autosomal datasets, we incorporated only samples with reliable provenance information. Based on this information, we divided the sample set *a priori* into seven sampling regions (Table 1).

To assess pairwise population differentiation, we calculated the differentiation measures Φ_{ST} (HVRI, Excoffier et al. 1992) and R_{ST} (microsatellites, Slatkin 1995). We used the Tamura and Nei distance correction (Tamura and Nei 1993) with a gamma value of 0.219 for the calculation of the genetic distance matrix for Φ_{ST}, as determined by the model selection test with jMODELTEST v0.1.1 (Posada 2008).

To infer the long-term effective population size N_e of the seven sampling regions, we calculated the estimators θ_π (based on the mean pairwise genetic distance between sequences; Tajima 1983) and θ_H (based on the heterozygosity of microsatellites; Ohta and Kimura 1973). Additionally, we used a likelihood-based estimator of θ (referred to as θ_L) using the software LAMARC v2.1.6 (Kuhner 2006). We applied the GTR+I nucleotide substitution model (Lanave et al. 1984) for the HVRI sequence data, which is the best-fitting of the supported models inferred by jMODELTEST, and the stepwise mutation model for the microsatellite data. The analysis was performed for each sampling region separately, and we used the Bayesian sampler with two chains of 1 000 000 steps each, sampling every 20th step and discarding the first 5000 samples as burn-in. The prior distribution of θ ranged from 10^{-5} to 10 (uniform on a natural logarithmic scale) and the starting value of θ was set to 0.01.

The different estimators of θ were used to calculate N_e, with θ equaling $N_e\mu$ for mitochondrial and $4N_e\mu$ for autosomal markers. Thus, these estimators allow inferring long-term N_e from a single population sample if the mutation rate is known. We used a mutation rate of 4.108×10^{-6} per site per generation for HVRI (Soares et al. 2009), assuming a generation time of 25 years (Wich et al. 2009), or 1×10^{-4} per locus per generation for the autosomal microsatellites (Schlötterer 2000).

Autosomal Genetic Structure

To assess genetic structure based on autosomal microsatellites, we first performed a principal component analysis

Table 1 Summary statistics for all examined orangutan sampling regions

		HVRI				Autosomal microsatellites				
Sampling region	Habitat[a]	$N_{Samples}$	θ_π[b]	HD[c]	N_e[d]	$N_{Samples}$	H_E[e]	θ_H[f]	N_e[g]	Census[h]
Tripa (TR)[i]	PSF	7	12.78	0.95	6808	9	0.64	1.68	4197	~380
North Aceh (NA)	DF	10	0.79	0.51	389	10	0.61	1.60	4009	~350
West Leuser (WL)	PSF	28	3.78	0.54	2013	21	0.61	1.61	4023	~3000
Central Leuser (CL)	DF	14	0.44	0.27	237	15	0.59	1.56	3901	~1100
Langkat (LK)	DF	26	1.40	0.80	747	24	0.64	1.66	4162	~1050
Batu Ardan (BA)	DF	8	0.78	0.46	417	9	0.59	1.57	3929	~300
Batang Toru (BT)	DF	18	0.96	0.65	503	8	0.63	1.63	4087	~550

[a]Prevailing habitat type; PSF, peat-swamp forest; DF, dry-land forest (Husson et al. 2009).
[b]Estimate of $\theta = N_e\mu$ based on the mean pairwise corrected nucleotide distance.
[c]Haplotypic diversity (Nei 1987).
[d]Effective population size, based on a mutation rate of 1.643×10^{-7} per site per year and a generation time of 25 years.
[e]Mean expected heterozygosity.
[f]Estimate of $\theta = 4N_e\mu$ based on the mean expected heterozygosity.
[g]Effective population size, based on a mutation rate of 10^{-4} per locus per generation.
[h]Estimated census size (Wich et al. 2008).
[i]The sampling region of Tripa includes coastal and highland areas.

(PCA) using the covariance-standardized method as implemented in the software GENALEX v6.41. Next, we used the Bayesian clustering algorithm implemented in the software STRUCTURE v2.3.3 (Pritchard et al. 2000) to identify distinct genetic clusters in the dataset. Because both methods do not require making *a priori* assumptions about genetic structure, we were able to include samples with unknown provenance. For the STRUCTURE analysis, we used the admixture model with correlated allele frequencies, a burn-in length of 3×10^5 steps followed by 3×10^6 MCMC steps. We ran the analysis with K values ranging from 1 to 10. For each K we performed 10 independent runs and averaged the ln Pr(Data | K) statistic over all iterations. Since the Pr(Data | K) estimator has been shown to overestimate K, as it frequently plateaus at higher values than the true number of K (Evanno et al. 2005), we also calculated the delta K statistic (Evanno et al. 2005), which gives a conservative estimate of K.

Migrant Detection

To assess the level of subpopulation connectivity, we identified individuals in the dataset that were either direct migrants or first generation offspring of direct migrants and local individuals. To achieve this, we used two different methods. First, given the strong geographic clustering of mtDNA haplotypes (Nater et al. 2011), we checked the median-joining network for individuals with reliable provenance that clustered with samples from another geographic region in order to detect direct migrants. Second, we used a Bayesian approach to assign individual genotypes to different subpopulations as either local individuals, direct migrants or F_1 admixed individuals, as implemented in the software BAYESASS 1.3 (Wilson and Rannala 2003). For this, we pre-assigned the individuals to the three different clusters identified in the previous STRUCTURE analysis and ran the MCMC analysis two times independently for 2.4×10^7 steps each, including a burn-in of 4×10^6 steps, with sampling every 2000 steps.

Table 2 Pairwise population differentiation values for HVRI (Φ_{ST}, above diagonal) and autosomal microsatellites (R_{ST}, below diagonal)

Φ_{ST}/R_{ST}	TR	NA	WL	CL	LK	BA	BT
TR	-	0.89***	0.58***	0.70***	0.95***	0.61**	0.97***
NA	0.05*	-	0.94***	0.99***	0.98***	0.98***	0.99***
WL	0.02	0.06**	-	0.04	0.96***	0.01	0.98***
CL	0.04*	0.11***	0.02	-	0.99***	0.02	1.00***
LK	0.02	0.02	0.05***	0.05***	-	0.98***	0.99***
BA	0.05	0.07*	0.07***	0.08**	0.00	-	0.99***
BT	0.12**	0.17***	0.14***	0.10***	0.08***	0.12**	–

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Both runs combined resulted in a total of 20 000 assignments for each individual.

Results

HVRI Median-Joining Network

The median-joining network (Figure 1B) showed a strong structuring of mtDNA haplotypes into four geographically distinct clusters: (1) Batang Toru, (2) Langkat, (3) Tripa, West Leuser, Central Leuser and Batu Ardan (referred to as West Alas cluster), and (4) North Aceh. We did not observe any haplotype sharing among these four clusters in our dataset of individuals with reliable provenance information.

Summary Statistics

The division of mitochondrial haplotypes into four distinct clusters as apparent in the mtDNA network correlated well with the Φ_{ST} statistic of genetic differentiation, as all comparisons between different clusters were highly significant (Table 2, above diagonal). However, within the West Alas

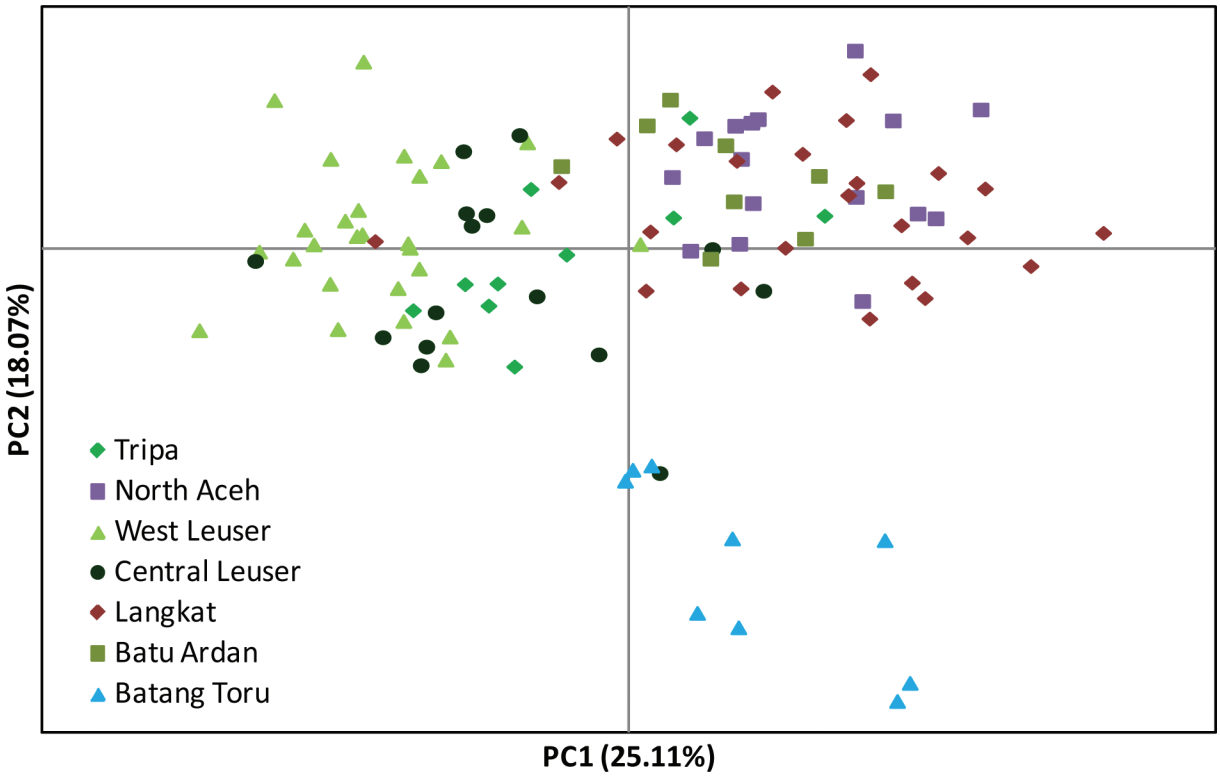


Figure 2. Principal component analysis of the autosomal microsatellite markers for all seven sampling regions.

cluster, the sampling region of Tripa was also significantly differentiated from all other regions in the same cluster. This differentiation points to highly different haplotype frequencies between Tripa and the other regions within this cluster, as these all share haplotypes among each other.

The R_{ST} measures for the microsatellites revealed additional information about the population structure beyond female philopatric patterns (Table 2, below diagonal). Three main patterns emerged. First, Batang Toru, the only sampling region south of Lake Toba, was highly differentiated from all other regions. Second, in contrast to high mtDNA differentiation, Tripa showed low R_{ST}-values to most other sampling regions, except Batang Toru. Third, the region of Langkat showed low differentiation to North Aceh, Tripa, and Batu Ardan.

The different estimators of θ revealed consistent patterns among the seven sampling regions, but estimates of θ for the microsatellite loci were consistently higher for θ_L when compared with θ_H (see Supplementary Table S2 online). We found that the genetic diversity estimates based on mtDNA and the corresponding N_e varied extensively across the different sampling regions (Table 1), as expected from the large differences in density estimates and habitat areas (Wich et al. 2008; Husson et al. 2009). In general, the estimated effective population sizes were similar to the census size estimates for most sampling regions (Wich et al. 2008). There was one striking exception. Tripa on the northwest coast exhibited the highest sequence diversity among the seven sampling regions and a N_e of nearly 7000 individuals, but contains among the smallest number of orangutans, with an estimated census size of less than 400 individuals. The Tripa region also showed a positive Tajima's *D* statistic and a multimodal pairwise mismatch distribution of HVRI sequences, indicating a recent population decline, while most other regions exhibited negative values of *D* and unimodal mismatch distributions, indicating recent expansions (Figure S3; see Supplementary Table S2 online). In contrast to the large regional variability for mtDNA, autosomal estimates of genetic diversity and N_e were remarkably similar among sampling regions (Table 1).

Autosomal Genetic Structure

The PCA revealed a geographically defined structure in the autosomal microsatellite data (Figure 2). The first principal component (PC) explained 25.11% of the total variance and distinguished between the sampling regions west and east of the Alas River. The region south of Lake Toba, Batang Toru, clusters with the regions east of the Alas River and cannot be distinguished with the first PC only. The second PC, explaining a further 18.07% of the variance, separated Batang Toru from all sampling regions north of Lake Toba. Therefore, by combining both PCs (explaining 43.18% of the total variance), there appears to be three clusters of sampling regions, separated from each other by the Alas River and Lake Toba. The separation was, however, not perfect, as the regions of WL, TR, BA, and CL showed outliers within the variation of other regions. The additional PCs did not seem to contain

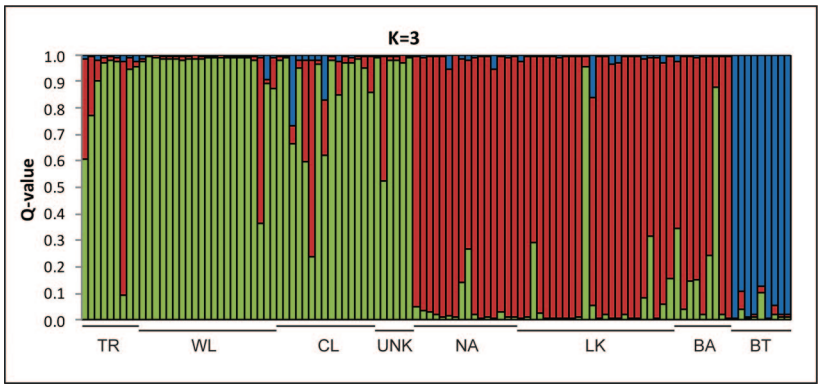


Figure 3. Results of the STRUCTURE analysis of the autosomal microsatellite markers for the most probable number of clusters ($K = 3$ according to delta K statistic). The membership coefficients Q shown are for the iteration with the highest likelihood. Samples are grouped by sampling region. The assignment is based on provenance record and mtDNA haplotype. UNK refers to samples with unknown provenance and ambiguous mtDNA assignment (belonging to the West Alas cluster).

any further information about geographic structuring of genotypes (Supplementary Figure S4).

The STRUCTURE analysis resulted in a clear signal for substructure in the Sumatran autosomal microsatellite dataset. Highest delta K was achieved for three clusters, while Pr(Data|K) peaked at five clusters (Supplementary Figure S5). At $K = 3$, the clusters corresponded largely to the mtDNA haplotype clusters described above, with some exceptions (Figure 3). First, the North Aceh and Langkat regions grouped together. Second, the region of Batu Ardan, which in the HVRI network assigned to the West Alas cluster, showed for autosomal markers a clear affinity to the Langkat and North Aceh regions. Third, the separation between the two genetic clusters north of Lake Toba (West Alas and Langkat/North Aceh) was not as sharp as for the mtDNA, as regions close to the geographic boundaries of the two clusters revealed a number of individuals with admixed genotypes. In contrast, samples from south of Lake Toba (Batang Toru) showed much less signals of admixture. Patterns of genetic admixture were also evident when the membership coefficients Q for each cluster were plotted in ranked order for all individuals for each cluster (Supplementary Figure S6). While all three curves showed two asymptotes at $Q = 0$ and $Q = 1$, multiple samples had Q-values between 0.2 and 0.8 (13 for West Alas, 13 for Langkat/North Aceh, and one for Batang Toru), indicating admixed ancestry.

A higher number of K did not result in a better resolution of sampling regions (Supplementary Figure S7). Since STRUCTURE often only identifies the uppermost level of hierarchical genetic structure (Evanno et al. 2005), we repeated the analysis for each of the three geographically defined clusters separately, using only samples that showed a membership coefficient of higher than 0.6 for a certain cluster in the first STRUCTURE analysis. None of the three clusters showed any sign of further substructure, as $K = 1$ returned the highest Pr(Data|K) values for all three clusters.

To test if part of the partitioning of the mitochondrial or autosomal genetic diversity can be explained by habitat type, we performed an AMOVA analysis with ARLEQUIN, where we divided the dataset into two groups corresponding to habitat type (peat-swamp forest versus dry-land forest, see Table 1). We included only samples from the West Alas cluster, as this is the only autosomal cluster that contains both habitat types. For autosomal microsatellites, habitat differences explain only 0.22% of the total variance, while over 97% is found within sampling regions (Table 3). For the mtDNA diversity, the variance component between habitat types is negative, indicating complete absence of any partitioning of genetic variance between habitat types.

Migrant Identification

All individuals showed congruence between their provenance record and their assigned mtDNA cluster. We did, however, identify three females and two males with high Q-values (>0.6) for a cluster that did not match their mtDNA haplotypes and provenance ($K = 3$, Figure 3). These individuals are unlikely to be direct migrants from the autosomal cluster they were assigned to in the STRUCTURE analysis. Rather, their natal range is indicated by their mtDNA haplotype, given that female orangutans have been shown to exhibit strong philopatric tendencies.

The BAYESASS analysis assigned migrant status to three of the five individuals previously identified in the STRUCTURE analysis as admixed or assigned to a cluster that did not match their mtDNA haplotype. In total, we found five individuals which have a less than 50% probability of being local in the cluster defined by their mtDNA haplotypes (Table 4). Only in one case, however, could we identify an admixed individual with significant statistical support ($P < 0.05$ of being local). This individual was a female with reliable provenance information, originating from the upper Alas valley in the Langkat region and carrying an mtDNA haplotype from the Langkat cluster. Her genotype, however,

Table 3 AMOVA of mitochondrial and autosomal microsatellite data between peat-swamp and dry-land forests within the West Alas cluster

	mtDNA		Autosomal microsatellites	
	Variance[a]	% Variance	Variance[a]	% Variance
Between habitat types	−0.74	−24.55	0.19	0.22
Among sampling regions, within habitat types	1.94*	64.71	1.87	2.17
Within sampling regions	1.80*	59.83	84.34*	97.61

*$P < 0.05$

[a]negative variance components indicate lack of genetic structure.

Table 4 List of individuals that show a probability of less than 0.5 to originate from the sampling cluster

					BAYESASS[d]		
Sample number	Sampling region[a]	Sex	mtDNA[b]	Q-value[c]	Local	Direct migrant	Admixed
BA2	BA (LK+NA)	Female	WA	0.876 (WA)	0.088 (LK+NA)	0.359 (WA)	0.553 (WA)
LK3	LK (LK+NA)	Female	LK	0.702 (LK+NA)	0.494 (LK+NA)	0.010 (WA)	0.496 (WA)
LK27	LK (LK+NA)	Female	LK	0.955 (WA)	0.004 (LK+NA)	0.365 (WA)	0.632 (WA)
LK7	LK (LK+NA)	Male	LK	0.673 (LK+NA)	0.409 (LK+NA)	0.002 (WA)	0.589 (WA)
TR4	TR (WA)	Male	WA	0.884 (LK+NA)	0.443 (WA)	0.228 (LK+NA)	0.329 (LK+NA)

[a]The autosomal genetic cluster to which most of the samples from the listed sampling regions assign is written in parentheses: WA, West Alas cluster, LK+NA, Langkat/North Aceh cluster, BT, Batang Toru cluster.
[b]mtDNA cluster assignment.
[c]Highest Q-value in the STRUCTURE analysis with $K = 3$.
[d]Posterior probabilities of the three classes in the BAYESASS analysis.

had a high membership coefficient to the West Alas cluster ($Q = 0.955$).

Discussion

Our study is the first to precisely locate and describe the geographic structuring of genetic diversity on mitochondrial and autosomal levels across the whole range of Sumatran orangutans. We were able to quantify the genetic diversity present within each of the seven sampling regions by analyzing the highly polymorphic HVRI region of the mtDNA and used that information to infer long-term effective population sizes of each sampling region. These estimates correlate strongly with recent census size estimates for most regions (Wich et al. 2008). Not surprisingly, the highest effective population sizes were observed for peat-swamp forests on the west coast of northern Sumatra, which also have the highest population density estimates (Husson et al. 2009). In one region, however, N_e and census size were in stark contrast to each other: the area of Tripa showed extraordinary high mitochondrial HVRI diversity and corresponding N_e in a comparatively small geographic region, which contains only an estimated 380 individuals. This signal points to a massive recent decline in the subpopulation size, which might have been caused by the dramatic and on-going habitat degradation in this area (van Schaik et al. 2001; Gaveau et al. 2009). It is plausible to assume that the lowland area along the northwest coast of Aceh was once completely covered with continuous peat-swamp forest and harbored thousands of orangutans (Gaveau et al. 2009). After decades of deforestation, current estimates indicate that all forests in the Tripa region will be irrecoverably lost by 2015–16 if forest destruction/conversion will continue at its current rate (Tata et al. 2010; Wich et al. 2011). There are other prominent examples in the literature highlighting discrepancies between large long-term N_e and small census sizes, which are linked to anthropogenic pressures. For example, heavy exploitation of gray (*Eschrichtius robustus*) and humpback whale (*Megaptera novaeangliae*) stocks due to whaling has led to dramatic population declines not reflected by long-term N_e (Roman and Palumbi 2003; Alter et al. 2007).

In contrast to the varying HVRI diversity found within different regions across the island, we obtained very homogenous genetic diversity estimates among sampling regions for autosomal microsatellite markers, resulting in N_e estimates of around 4000 or 10 000 individuals for each of the seven regions, depending on the estimator of θ. This striking discrepancy when compared with the HVRI estimates is most likely caused by pronounced male-biased dispersal and strong female philopatric tendencies in orangutans (Galdikas 1995; Singleton and van Schaik 2002; Morrogh-Bernard et al. 2011; Nietlisbach et al. 2012; van Noordwijk et al. 2012; Arora et al. 2012). Field studies have shown that female orangutans preferentially establish their home range overlapping with the home ranges of their maternal kin (Singleton and van Schaik 2002; van Noordwijk et al. 2012). Thus, mitochondrial DNA does get hardly, if at all, exchanged among neighboring geographic regions, and mtDNA diversity well reflects the number of orangutans in

the different local subpopulations. Males, in contrast, leave their natal area, a pattern linked to inbreeding avoidance (Pusey and Wolf 1996). Intense male-male competition (Utami Atmoko et al. 2009) may force young males to cover large distances before being able to settle down (Nietlisbach et al. 2012). Such widely dispersing males might distribute newly arisen alleles in the whole meta-population and recover alleles that have been lost locally due to genetic drift, thereby homogenizing the allele frequencies of autosomal markers among sampling regions. Thus, the highly similar levels of autosomal diversity in contrast to the large differences in mtDNA diversity across the island are a clear indicator of considerable male-mediated gene flow among these regions. The panmictic distribution of Y-haplotypes on Sumatra (Nater et al. 2011) provides further evidence for this male-driven homogenization of the gene pool.

Due to the use of multiple independent autosomal markers, we were able to investigate male-mediated gene flow in more detail. The cluster analysis with STRUCTURE showed that the strength of male-driven gene flow is not sufficient to completely homogenize allele frequencies among sampling regions, thus resulting in a clear pattern of geographically structured autosomal variation. The three clusters identified in the autosomal dataset were defined by geographical features. It appears that eruptions of the Toba volcano (Chesner et al. 1991) isolated the orangutans from Batang Toru, the region south of it, from the rest of the species occurring north of it. The high pairwise R_{ST}-values across Lake Toba provide further evidence of strong separating effects of the Toba eruptions, which have also led to a deep divergence of mtDNA haplotypes north and south of the caldera (Nater et al. 2011). The forests between these two areas might have been connected between major eruptions, but the combination of periodic separation and strong female philopatry has served to keep the populations from homogenizing. North of Lake Toba, the Alas River, part of the Barisan graben running the length of Sumatra (Verstappen 1973), divides the remaining regions into two distinct genetic clusters. The Alas valley was likely repeatedly blocked by volcanic material from the nearby Toba eruptions, turning the upper Alas river into a large lake for prolonged periods (van Schaik and Mirmanto 1985). This damming of the Alas River might have promoted the structuring of the gene pool north of Lake Toba. Interestingly, the habitat type does not seem to play a significant role in the structuring of autosomal diversity in Sumatran orangutans, indicating that dispersing males do not prefer to migrate to areas that ecologically resemble their natal habitat, and thus prevent more fine-tuned adaptation of orangutans to local habitat types.

Even though the STRUCTURE analysis revealed strong geographical structuring of the autosomal gene pool, we nevertheless found clear signals for recent gene flow across the island. First, the two sample regions of Langkat and North Aceh cannot be distinguished in the STRUCTURE analysis, even though these regions show a mitochondrial divergence of 0.85 Ma (Nater et al. 2011). Therefore, the observed low autosomal differentiation (R_{ST} = 0.02) points towards considerable levels of male-mediated gene flow after the two subpopulations were separated from each other. If this migratory contact with the Langkat region can be maintained, it will greatly help reducing inbreeding pressure on the small North Aceh subpopulation. As a second signal of gene flow, we found many admixed individuals in the STRUCTURE plot (Figure 3). Interestingly, these individuals were mostly sampled in regions close to the boundary of autosomal clusters, like Tripa, Central Leuser, and Langkat, supporting the idea of recent gene flow. Third, we were able to identify multiple individuals with substantial likelihoods of having paternal ancestry from another cluster. While only one individual shows good statistical support for being admixed ($P < 0.05$), it should be kept in mind that we sampled only an estimated 0.7–4.6% of all individuals per sampling region. Moreover, we only investigated migration among major autosomal clusters and not individual sampling regions, due to the impossibility to reliably discriminate them genetically.

Further investigation of the provenance of admixed individuals hinted toward an important corridor for gene flow between genetic clusters. Three of the five individuals identified as having admixed ancestry originate from the upper Alas valley near Blangkejeren, while a fourth admixed individual has been confiscated in the highlands of the Tripa area. These locations are all close to the area where the supposed boundaries of the West Alas, North Aceh, and Langkat clusters meet, and this highland area contains orangutan habitat with resident subpopulations. The presence of clear migration signals in this area underlines its critical importance as a connection among major subpopulations of Sumatran orangutans and therefore deserves special habitat conservation efforts.

Special consideration also needs to be given to the region of Batu Ardan, where there is a clear discrepancy between autosomal data and mtDNA structure, possibly due to male-mediated migration. This region, located between the Alas River and Lake Toba, shows a strong affinity of mtDNA haplotypes to the West Alas cluster, even though it is located on the opposite (eastern) side of the major Alas River. In fact, Batu Ardan shares a common haplotype with all regions on the western side, but also has two derived haplotypes that do not occur elsewhere. This supports the notion that the small Batu Ardan subpopulation could be the result of a recent colonization event from the western side of the Alas, probably due to a loop cut-off of the meandering river (Nater et al. 2011). However, for autosomal markers, we found that Batu Ardan reveals a high affinity to the adjacent Langkat/North Aceh cluster, from which it is separated by a deep river valley. This river might be passable by orangutans near its headwaters, allowing males to bring in autosomal alleles from the Langkat region. The notion that the recolonization from the west side of the Alas and subsequent influx of males from Langkat occurred after the forests recovered from the devastating Toba super-eruption around 73 kya (Chesner et al. 1991) is tempting but cannot yet be proven with the data at hand.

Sumatran orangutans are genetically deeply structured into at least three autosomally distinct clusters, despite regular male-mediated gene flow between the West Alas and the Langkat/North Aceh clusters, which occurred at least up to very recently and is probably still on-going. However, continuing habitat degradation is threatening the existence of orangutans on Sumatra in two ways. First, due to the shrinkage of suitable habitat area, the local subpopulation census sizes will be further reduced. Already today, only one of the three autosomal clusters, West Alas, harbors well over 1000 individuals. Second, through the destruction of important corridors for migration, genetic exchange with neighboring subpopulations will be disrupted. Both effects combined will inevitably lead to a substantial loss of genetic diversity with all its negative consequences (Reed and Frankham 2003). Especially the only remaining subpopulation south of Lake Toba, Batang Toru, is highly threatened in this regard. Given the genetic uniqueness of the orangutans in this area on both the mitochondrial and autosomal level and the fact that most of the forest in this area has no protected status (Wich et al. 2011), urgent measures are needed to preserve this indispensable reservoir of genetic diversity of Sumatran orangutans.

Orangutans are the least gregarious and the most arboreal of all great apes (Delgado and Van Schaik 2000). As such, comparing the observed patterns in Sumatran orangutans with those of other great ape species will aid the inference of factors underlying the observed population structure in these taxa. Previous genetic studies on great apes showed that rivers are one of the most important factors in shaping population structure and subspecies boundaries (e.g., gorillas: Anthony et al. 2007; Bornean orangutans: Goossens et al. 2005, Arora et al. 2010; chimpanzees: Becquet et al. 2007; bonobos: Eriksson et al. 2004). Our study supports these findings by identifying the Alas River as a major division line of genetic diversity within the range of Sumatran orangutans. Moreover, volcanic activities of the Toba region during the last 1.2 million years (Chesner et al. 1991) played another major role in the structuring of genetic diversity in Sumatran orangutans. Such a pattern of long-lasting isolation caused by volcanic activities has so far not been documented for great apes.

Given that Sumatran orangutans are critically endangered, knowledge of the extent to which human-induced habitat degradation is affecting the population structure is of critical importance for conservation efforts. Bergl and Vigilant (2007) revealed a pronounced substructure in the small Cross River gorilla population (*Gorilla gorilla diehli*) largely following the patterns of forest connectivity. Likewise, Goossens et al. (2005) showed that in Bornean orangutans, subpopulations in many of the isolated forest lots on the same side of the Kinabatangan River in Sabah, Malaysia, are significantly differentiated from each other, despite their close geographic proximity. Both studies highlight the adverse effects of anthropogenic forest degradation on the dispersal abilities of forest dwelling primates. Interestingly, we did not observe similar signals in Sumatran orangutans, despite their strict arboreality and the heavy forest exploitation within their range (Rijksen and Meijaard 1999). The Sumatran subpopulations appear to be more effectively connected through male dispersal for two reasons. First, the uninhabited mountain regions connecting subpopulations are forested, and thus dispersing males, who have been sighted at altitudes of up to 2000 m above sea level (Rijksen 1978), can move through them. Second, Sumatran forests provide suitable habitat to higher altitudes than Bornean ones due to the *Massenerhebung* effect (van Schaik et al. 1995), and this makes it easier for migrating males to cross rivers at their headwaters.

The example of the Sumatran orangutan demonstrates that even species with a geographically very limited range can show strong underlying genetic structure, caused by geographical barriers, habitat discontinuities, limited dispersal, and long population persistence. Correspondingly, genetic diversity might be mainly found among local subpopulations rather than within, and local extinctions carry a serious risk of losing a substantial part of a species' total genetic diversity. Our study highlights the need to assess the genetic make-up of endangered species in detail, identify local subpopulation boundaries, and focus conservation efforts on maintaining dispersal corridors among genetic clusters.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

Swiss National Science Foundation (3100A-116848 to M.K. and C.vS.); Messerli Foundation; A.H.-Schultz Foundation; Claraz Schenkung.

Acknowledgments

We are very grateful to Ellen Meulman, Andrea Permana, Izumi, Zulfikar, Tony Weingrill, Helga Peters, Gregoire Bertagnolio, Gail Campbell-Smith, Yenny Saraswati and Rachmad Wahyudi for collecting samples. Corinne Ackermann and Kai Ansmann performed valuable laboratory work for this study. We thank Erik Willems for contributing the map of northern Sumatra. Furthermore, we highly appreciate the support of the following institutions: Indonesian State Ministry of Research and Technology (RISTEK), Indonesian Institute of Sciences (LIPI), Sabah Wildlife Department, Taman National Gunung Leuser (TNGL), Borneo Orangutan Survival Foundation (BOSF), Leuser International Foundation (LIF), and Badan Pengelola Kawasan Ekosistem Leuser (BPKEL).

References

Alter SE, Rynes E, Palumbi SR. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. Proceedings of the National Academy of Sciences of the United States of America 104:15162–15167.

Anthony NM, Johnson-Bawe M, Jeffery K, et al. 2007. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. Proceedings of the National Academy of Sciences of the United States of America 104:20432–20436.

Arora N, Nater A, van Schaik CP, et al. 2010. Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). Proceedings of the National Academy of Sciences 107:21376–21381.

Arora N, van Noordwijk MA, van Schaik CP, et al. 2012. Parentage-based pedigree reconstruction reveals female matrilineal clusters and male-biased dispersal in the non-gregarious Asian great apes, the Bornean orangutans *(Pongo pygmaeus)* Mol Ecol. 21:3352–3362.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol and Evol. 16:37–48.

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D. 2007. Genetic structure of chimpanzee populations. PLoS Genet. 3:e66. doi:10.1371/journal.pgen.0030066.

Bengtsson BO. 1978. Avoiding inbreeding—at what cost? Journal of Theor Biol. 73:439–444.

Bergl RA, Vigilant L. 2007. Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). Mol Ecol. 16:501–516.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10:195–205.

Chesner CA, Rose WI, Deino A, Drake R, Westgate JA. 1991. Eruptive history of Earth's largest Quaternary caldera (Toba, Indonesia) clarified. Geology. 19:200–203.

Delgado RA, Van Schaik CP. 2000. The behavioral ecology and conservation of the orangutan (*Pongo pygmaeus*): a tale of two islands. Evol Anthrop. 9:201–218.

Dobson FS. 1982. Competition for mates and predominant juvenile male dispersal in mammals. An Behav. 30:1183–1192.

Douadi MI, Gatti S, Levrero F, et al. 2007. Sex-biased dispersal in western lowland gorillas (Gorilla gorilla gorilla). Mol Ecol. 16:2247–2259.

Eriksson J, Hohmann G, Boesch C, Vigilant L. 2004. Rivers influence the population genetic structure of bonobos (*Pan* paniscus). Mol Ecol. 13:3425–3435.

Eriksson J, Siedel H, Lukas D, et al. 2006. Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). Mol Ecol. 15:939–949.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14: 2611–2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10:564–567.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes—application to human mitochondrial DNA restriction data. Genetics. 131:479–491.

Galdikas BMF. 1995. Social and reproductive behavior of wild adolescent female orangutans. In: Nadler RD, Galdikas BFM, Sheeran LK, Rosen N, editors. The neglected ape. New York: Plenum Press. p. 163–182.

Gaveau DLA, Wich S, Epting J, et al. 2009. The future of forests and orangutans (*Pongo abelii*) in Sumatra: predicting impacts of oil palm plantations, road construction, and mechanisms for reducing carbon emissions from deforestation. Environmental Research Letters 4. doi:10.1088/1748-9326/4/3/034013

Gonder MK, Locatelli S, Ghobrial L, et al. 2011. Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. Proceedings of the National Academy of Sciences of the United States of America 108:4766–4771.

Goossens B, Chikhi L, Jalil MF, et al. 2005. Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*) populations from Sabah, Malaysia. Mol Ecol. 14:441–456.

Groves CP. 2001. Primate taxonomy. Washington, DC/London: Smithsonian Institution Press.

Guschanski K, Caillaud D, Robbins MM, Vigilant L. 2008. Females shape the genetic structure of a gorilla population. Current Biol. 18:1809–1814.

Handley LJL, Perrin N. 2007. Advances in our understanding of mammalian sex-biased dispersal. Mol Ecol. 16:1559–1578.

Hedrick PW, Kalinowski ST. 2000. Inbreeding depression in conservation biology. Annu Rev Ecol Syst. 31:139–162.

Husson SJ, Wich SA, Marshall AJ, et al. 2009. Orangutan distribution, density, abundance and impacts of disturbance. In: Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP, editors. Orangutans: geographic variation in behavioral ecology and conservation. Oxford: Oxford University Press.

IUCN. 2012. IUCN Red List of Threatened Species. Version 2012.1. http://www.iucnredlist.org, last accessed August 24, 2012.

Jalil MF, Cable J, Inyor JS, et al. 2008. Riverine effects on mitochondrial structure of Bornean orang-utans (*Pongo pygmaeus*) at two spatial scales. Mol Ecol. 17:2898–2909.

Johnson ML, Gaines MS. 1990. Evolution of dispersal—theoretical models and empirical tests using birds and mammals. Annu Rev Ecol Syst. 21:449–480.

Kanthaswamy S, Kurushima JD, Smith DG. 2006. Inferring *Pongo* conservation units: a perspective based on microsatellite and mitochondrial DNA analyses. Primates. 47:310–321.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecology Letters 7:1225–1241.

Kimura M, Weiss GH. 1964. Stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576.

Knott CD. 1998. Changes in orangutan caloric intake, energy balance, and ketones in response to fluctuating fruit availability. Intl J Primatol. 19:1061–1079.

Kuhner MK. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics. 22:768–770.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. J Mol Evol. 20:86–93.

Mitani JC, Watts DP, Muller MN. 2002. Recent developments in the study of wild chimpanzee behavior. Evol Anthrop. 11:9–25.

Morin PA, Chambers KE, Boesch C, Vigilant L. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Mol Ecol. 10:1835–1844.

Morrogh-Bernard HC, Morf NV, Chivers DJ, Krutzen M. 2011. Dispersal patterns of orang-utans (*Pongo* spp.) in a Bornean peat-swamp forest. Intl J Primatol. 32:362–376.

Muir CC, Galdikas BMF, Beckenbach AT. 2000. mtDNA sequence diversity of orangutans from the islands of Borneo and Sumatra. Journal of Mol Evol. 51:471–480.

Nater A, Nietlisbach P, Arora N, et al. 2011. Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: *Pongo*). Mol Biol Evol. 28:2275–2288.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nietlisbach P, Arora N, Nater A, et al. 2012. Heavily male-biased long-distance dispersal of orang-utans (*genus: Pongo*), as revealed by Y-chromosomal and mitochondrial genetic markers. Mol Ecol. 21:3173–3186.

Nietlisbach P, Nater A, Greminger MP, Arora N, Krützen M. 2010. A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orang-utans (*genus: Pongo*). Conserv Genet Resour. 2:153–158.

Ohta T, Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res. 22:201–204.

Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol and Evol. 25:1253–1256.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Pusey A, Wolf M. 1996. Inbreeding avoidance in animals. Trends Ecol Evol. 11:201–206.

Pusey AE. 1987. Sex-biased dispersal and inbreeding avoidance in birds and mammals. Trends Ecol Evol. 2:295–299.

Reed DH, Frankham R. 2003. Correlation between fitness and genetic diversity. Conser Biol. 17:230–237.

Rijksen HD. 1978. A field study on Sumatran orangutans (*Pongo pygmaeus abelii* Lesson 1827): ecology, behavior and conservation. Wageningen (The Netherlands): H. Veenman and Zonen.

Rijksen HD, Meijaard E. 1999. Our vanishing relative: the status of wild orang-utans at the close of the twentieth century. Dordrecht (The Netherlands): Kluwer Academic Publishers.

Roman J, Palumbi SR. 2003. Whales before whaling in the North Atlantic. Science 301:508–510.

Ross KG. 2001. Molecular ecology of social behaviour: analyses of breeding systems and genetic structure. Mol Ecol. 10:265–284.

Schlötterer C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma. 109:365–371.

Setia TM, van Schaik CP. 2007. The response of adult orang-utans to flanged male long calls: inferences about their function. Folia Primatologica. 78:215–226.

Singleton I, van Schaik CP. 2002. The social organisation of a population of Sumatran orang-utans. Folia Primatologica. 73:1–20.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. Science. 236:787–792.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 139:457–462.

Soares P, Ermini L, Thomson N, et al. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Human Genet. 84:740–759.

Steiper ME. 2006. Population history, biogeography, and taxonomy of orangutans (Genus: *Pongo*) based on a population genetic meta-analysis of multiple loci. J Human Evol. 50:509–522.

Storz JF. 1999. Genetic consequences of mammalian social structure. J Mammal. 80:553–569.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512–526.

Tata HL, van Noordwijk M, Mulyoutami E, et al. 2010. Human livelihoods, ecosystem services and the habitat of the Sumatran orangutan: rapid assessment in Batang Toru and Tripa. Bogor (Indonesia): World Agroforestry Centre (ICRAF) Southeast Asia Regional Office.

Tautz D, Gerloff U, Hartung B, Fruth B, Hohmann G. 1999. Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (*Pan paniscus*) determined from DNA analysis of faecal samples. Proceedings of the Royal Society of London Series B—Biological Sciences. 266:1189–1195.

Utami Atmoko SS, Singleton I, van Noordwijk MA, van Schaik C, Setia TM. 2009. Male-male relationships in orangutans. In: Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP, editors. Orangutans: geographic variation in behavioral ecology and conservation. Oxford: Oxford University Press.

van Noordwijk MA, Arora N, Willems EP, et al. 2012. Female philopatry and its social benefits among Bornean orangutans. Behav Ecol and Sociobiol. 66:823–834.

van Schaik CP, Mirmanto E. 1985. Spatial variation in the structure and litterfall of a Sumatran rain forest. Biotropica. 17:196–205.

van Schaik CP, Monk KA, Robertson JMY. 2001. Dramatic decline in orangutan numbers in the Leuser ecosystem, Northern Sumatra. Oryx. 35:14–25.

van Schaik CP, Priatna A, Priatna D. 1995. Population estimates and habitat preferences of orangutans based on line transects of nests. In: Nadler RD, Galdikas BFM, Sheeran LK, Rosen N, editors. The Neglected Ape. New York: Plenum Press. p. 129–147.

Verstappen HT. 1973. A geomorphological reconnaissance of Sumatra and adjacent islands (Indonesia). Groningen (The Netherlands): Wolters-Noordhoff B.V.

Warren KS, Verschoor EJ, Langenhuijzen S, et al. 2001. Speciation and intrasubspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. Mol Biol and Evol. 18:472–480.

Wich S, Riswan, Jenson J, Refish J, Nelleman C. 2011. Orangutans and the economics of sustainable forest management in Sumatra. UNEP/GRASP/PanEco/YEL/ICRAF/GRID-Arendal. http://www.unep.org/pdf/orangutan_report_scr.pdf, last accessed August 24, 2012.

Wich SA, Buij R, van Schaik C. 2004. Determinants of orangutan density in the dryland forests of the Leuser ecosystem. Primates. 45:177–182.

Wich SA, de Vries H, Ancrenaz M, et al. 2009. Orangutan life history variation. In: Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP, editors. Orangutans: geographic variation in behavioral ecology and conservation. Oxford: Oxford University Press.

Wich SA, Meijaard E, Marshall AJ, et al. 2008. Distribution and conservation status of the orang-utan (*Pongo* spp.) on Borneo and Sumatra: how many remain? Oryx. 42:329–339.

Williams GC. 1966. Adaptation and natural selection: a critique of some current evolutionary thought. Princeton (NJ): Princeton University Press.

Wilson GA, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 163:1177–1191.

Wright S. 1943. Isolation by distance. Genetics. 28:114–138.

Received February 3, 2012; Revised July 9, 2012; Accepted July 10, 2012

Corresponding Editor: Adalgisa Caccone

Nater_SupMat_S1-S5 - Population structure in Sumatran orangutans - JHered.

S1. List of samples used for analyses

Sample ID	Type	Status[a]	Relia-ble[b]	Region[c]	Origin	Sex	HVRI[d]	STR[e]
BA1	Blood	R	yes	BA	Sidikalang, Dairi, North Sumatra	female		X
BA2	Blood	R	yes	BA	Sidikalang, Dairi, North Sumatra	female	wa1	X
BA3	Blood	R	yes	BA	Sidikalang, Dairi, North Sumatra	female		X
BA4	Blood	R	yes	BA	Perolihen, Pakpak Bharat, North Sumatra	male	wa1	X
BA5	Faeces	W		BA	Perolihen, Pakpak Bharat, North Sumatra	n/a	wa1	X
BA6	Hair	W		BA	Perolihen, Pakpak Bharat, North Sumatra	male		X
BA7	Hair	R	yes	BA	Salak, Pakpak Bharat, North Sumatra	female	wa1	X
BA8	Hair	R	yes	BA	Perolihen, Pakpak Bharat, North Sumatra	male	wa1	X
BA9	Hair	R	yes	BA	Singkohor, Aceh Singkil, Aceh	female		X
BA10	Hair	W		BA	Puncak Sidiangkat, Pakpak Bharat, North Sumatra	n/a	wa2	
BA11	Hair	W		BA	Siranggas, Pakpak Bharat, North Sumatra	n/a	wa1	
BA12	Hair	W		BA	Salak, Pakpak Bharat, North Sumatra	n/a	wa3	
BT1	Blood	R	yes	BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	male	bt2	X
BT2	Blood	R	yes	BT	Tukka, Tapanuli Tengah, North Sumatra	female	bt4	X
BT3	Blood	R	yes	BT	Pinang Sori, Tapanuli Tengah, North Sumatra	male	bt4	X
BT4	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	male	bt1	X
BT5	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	n/a	bt3	
BT6	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	bt3	X
BT7	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	male	bt4	
BT8	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	bt3	X
BT9	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	bt4	X
BT10	Hair	R	yes	BT	Marancar, Tapanuli Selatan, North Sumatra	male	bt4	
BT11	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT12	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT13	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT14	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT15	Hair	W		BT	Rambasiashur, Tapanuli Selatan, North Sumatra	n/a	bt1	
BT16	Hair	W		BT	Rambasiashur, Tapanuli Selatan, North Sumatra	n/a	bt1	
BT17	Hair	W		BT	Rambasiashur, Tapanuli Selatan, North Sumatra	n/a	bt1	X
BT18	Hair	W		BT	Sitandiang, Tapanuli Selatan, North Sumatra	n/a	bt4	

CL1	Blood	R	yes	CL	Ladang Lawe Bengkari, Aceh Tenggara, Aceh	male	wa1	X
CL2	Blood	R	yes	CL	Lawe Sikap, Aceh Tenggara, Aceh	male	wa1	X
CL3	Blood	R	yes	CL	Lawe Sikap, Aceh Tenggara, Aceh	male	wa1	X
CL4	Blood	R	yes	CL	Lawe Sikap, Aceh Tenggara, Aceh	male	wa1	X
CL5	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	
CL6	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	wa1	
CL7	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	wa5	
CL8	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male		X
CL9	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female		X
CL10	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	
CL11	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	
CL12	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female		X
CL13	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	X
CL14	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa4	X
CL15	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	X
CL16	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female		X
CL17	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	X
CL18	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male		X
CL19	Hair	R	yes	CL	Lawe Pupus, Aceh Tenggara, Aceh	male		X
CL20	Hair	R	yes	CL	Pindie, Gayo Lues, Aceh	female	wa1	X
LK1	Blood	R	yes	LK	Tangkahan, Langkat, North Sumatra	male	lk1	X
LK2	Blood	R	yes	LK	Langkat, North Sumatra	male	lk2	X
LK3	Blood	R	yes	LK	Pinding, Gayo Lues, Aceh	female	lk2	X
LK4	Blood	R	yes	LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK5	Blood	R	yes	LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK6	Blood	R	yes	LK	Blangkejeren, Gayo Lues, Aceh	female	lk2	X
LK7	Blood	R	yes	LK	Rikit Gaib, Gayo Lues, Aceh	male	lk2	X
LK8	Blood	R	yes	LK	Tenggulun, Aceh Tamiang, Aceh	male	lk7	X
LK9	Blood	R	yes	LK	Tamiang Hulu, Aceh Tamiang, Aceh	male	lk7	X
LK10	Blood	R	yes	LK	Kejuruan Muda, Aceh Tamiang, Aceh	female	lk7	X
LK11	Faeces	W		LK	Ketambe, East Side of Alas River, Aceh Tenggara, Aceh	female	lk2	X
LK12	Faeces	W		LK	Ketambe, East Side of Alas River, Aceh Tenggara, Aceh	female	lk2	X
LK13	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk3	
LK14	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	male	lk1	X
LK15	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK16	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	male	lk1	X
LK17	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	

LK18	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	
LK19	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK20	Faeces	W		LK	Sikundur, Langkat, North Sumatra	female	lk4	X
LK21	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female		X
LK22	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	male		X
LK23	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	n/a	lk7	
LK24	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	n/a	lk4	
LK25	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	n/a	lk4	X
LK26	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	female	lk7	X
LK27	Hair	R	yes	LK	Kota Panjang, Gayo Lues, Aceh	female	lk5	X
LK28	Hair	R	yes	LK	Pinding, Gayo Lues, Aceh	male		X
LK29	Hair	R	yes	LK	Barak Gajah, Langkat, North Sumatra	female	lk6	X
NA1	Blood	R	yes	NA	Langsa, Aceh	male	na2	
NA2	Blood	R	yes	NA	Pondok Baru, Aceh Tengah, Aceh	male	na1	X
NA3	Blood	R	yes	NA	Lhokseumawe, Aceh	male	na1	X
NA4	Blood	R	yes	NA	Gayo Lues, Aceh	male	na1	X
NA5	Blood	R	yes	NA	Langsa, Aceh	male	na1	X
NA6	Blood	R	yes	NA	Aceh Besar, Aceh	male		X
NA7	Blood	R	yes	NA	Blang Jerango, Agara, Gayo Lues, Aceh	male	na1	X
NA8	Blood	R	yes	NA	Indra Makmu, Aceh Timur, Aceh	male	na3	X
NA9	Blood	R	yes	NA	Indra Makmu, Aceh Timur, Aceh	male	na3	X
NA10	Hair	R	yes	NA	Takengon, Aceh Tengah, Aceh	female	na1	X
NA11	Hair	R	yes	NA	Takengon, Aceh Tengah, Aceh	male	na1	X
WL1	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	male	wa6	
WL2	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female	wa1	
WL3	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female	wa7	
WL4	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	male	wa8	
WL5	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female	wa6	
WL6	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female		X
WL7	Blood	R	yes	WL	Singkil, Aceh Singkil, Aceh	female	wa1	X
WL8	Blood	R	yes	WL	Gunung Merutung, Aceh Selatan, Aceh	male	wa1	X
WL9	Blood	R	yes	WL	Cot Siaumantouk, Aceh Selatan, Aceh	female	wa9	X
WL10	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL11	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL12	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL13	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL14	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa1	X

WL15	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL16	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	X
WL17	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL18	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL19	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa1	X
WL20	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL21	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa6	X
WL22	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL23	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa6	
WL24	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL25	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL26	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa6	X
WL27	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL28	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL29	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa6	
WL30	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL31	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL32	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL33	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL34	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL35	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL36	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL37	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL38	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL39	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa6	
WL40	Hair	R	yes	WL	Singkil, Aceh Singkil, Aceh	male	wa1	
WL41	Hair	R	yes	WL	Singkil, Aceh Singkil, Aceh	male	wa1	
TR1	Blood	R	yes	TR	Trangon, Gayo Lues, Aceh	female		X
TR2	Blood	R	yes	TR	Aceh Barat Daya, Aceh	male	wa10	X
TR3	Blood	R	yes	TR	Aceh Barat, Aceh	male	wa11	X
TR4	Blood	R	yes	TR	Aluebillie, Aceh Nagan Raya, Aceh	male	wa1	X
TR5	Blood	R	yes	TR	Aceh Barat Daya, Aceh	female	wa12	X
TR6	Blood	R	yes	TR	Trangon, Gayo Lues, Aceh	male	wa13	X
TR7	Hair	R	yes	TR	Meulaboh, Aceh Barat, Aceh	female		X
TR8	Hair	R	yes	TR	Meulaboh, Aceh Barat, Aceh	female	wa9	X
TR9	Hair	R	yes	TR	Trangon, Gayo Lues, Aceh	male	wa1	X
UNK1	Blood	R	no	UNK	Sumatra	male	bt1	X

UNK2	Blood	R	no	UNK	Gayo Lues, Aceh	male	na1	X
UNK3	Blood	R	no	UNK	Northern Aceh	male	na1	X
UNK4	Blood	R	no	UNK	Sumatra	male	na1	X
UNK5	Blood	R	no	UNK	Sumatra	female	na2	X
UNK6	Blood	R	no	UNK	Sumatra	female	na1	X
UNK7	Blood	R	no	UNK	Northern Aceh	male	na1	X
UNK8	Blood	R	no	UNK	Aceh	female	wa14	X
UNK9	Blood	R	no	UNK	Aceh	male	wa1	X
UNK10	Blood	R	no	UNK	Aceh	female		X
UNK11	Blood	R	no	UNK	Aceh	male	wa1	X
UNK12	Blood	R	no	UNK	Aceh	female	wa1	X
UNK13	Blood	R	no	UNK	Aceh	female		X

[a] sample status, W=sampled in the wild, R=rehabilitant individual; [b] provenance reliability of samples from rehabilitant individuals; [c] sampling region used for analyses (see Materials & Methods); [d] mitochondrial HVRI haplotype; [e] autosomal microsatellite data included in analyses.

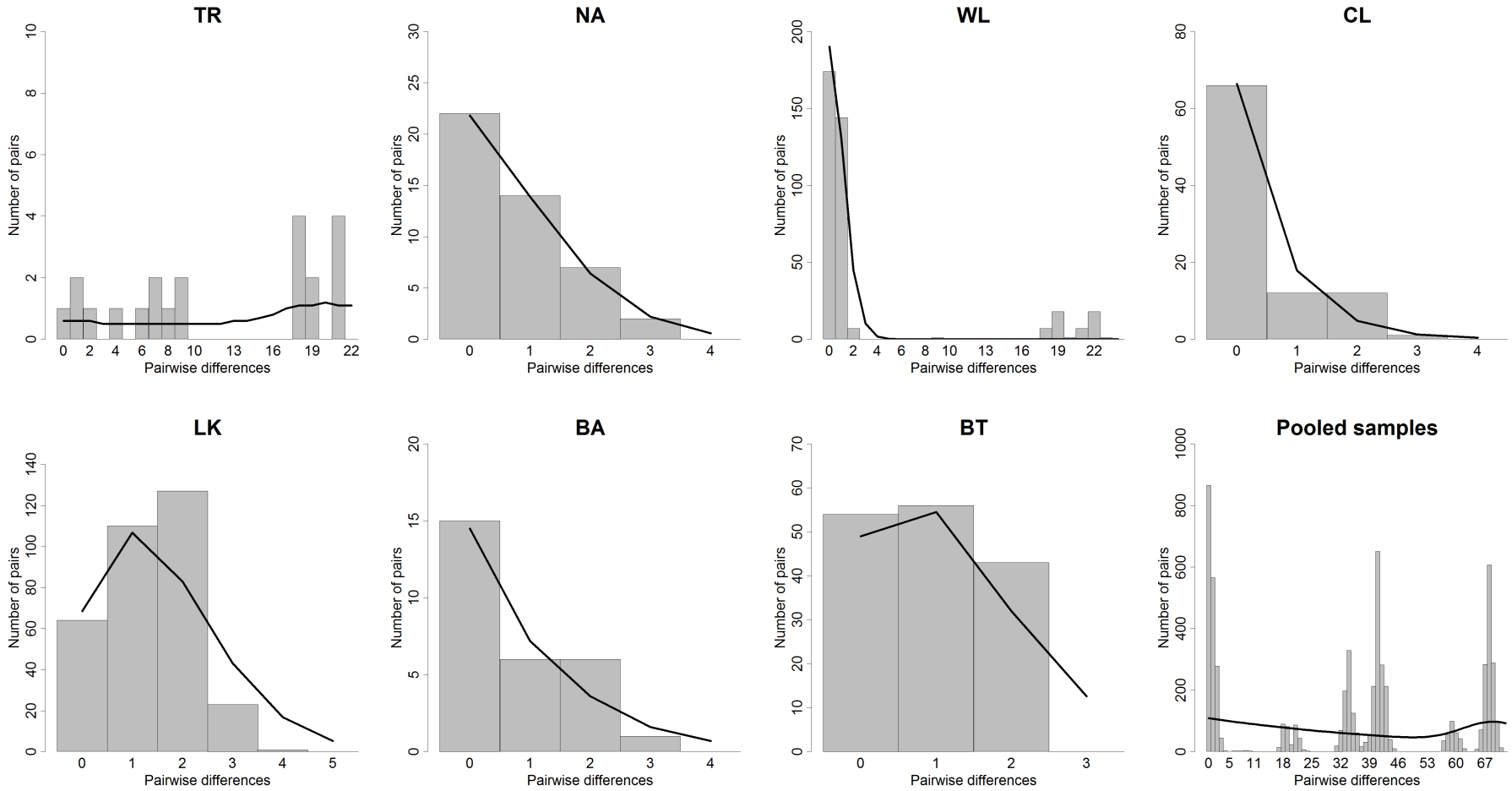
S2. Summary statistics for the seven sampling regions and pooled samples

	TR	NA	WL	CL	LK	BA	BT	Pooled
HVRI								
θ_π	12.78	0.79	3.78	0.44	1.40	0.78	0.96	68.40
θ_S	8.98	1.06	5.91	0.94	1.57	1.16	0.87	18.74
θ_L mode	10.09	1.05	3.88	1.12	1.91	1.16	1.00	19.03
θ_L 95%-HPD[a]	4.33–30.64	0.20–4.13	1.92–7.89	0.19–3.53	0.62–4.46	0.23–4.91	0.17–2.91	14.32–26.35
$N_e(\theta_\pi)$	6,808	419	2,012	236	747	417	512	36,432
$N_e(\theta_S)$	4,783	565	3,148	502	838	616	465	9,983
$N_e(\theta_L)$	5,448	566	2,095	604	1,032	627	541	10,271
Tajima's D[b]	0.76	-1.03	-1.79*	-1.67*	-0.43	-1.45	0.18	2.28
Fu's F_S[c]	0.44	-0.05	2.97	-0.76	-1.96	-0.30	-0.29	15.30
SSD[d]	0.07	0.00	0.02	0.01	0.02	0.01	0.01	0.05*
RI[e]	0.16	0.07	0.14	0.37	0.13	0.14	0.09	0.03***
Microsatellites								
θ_H	1.68	1.60	1.61	1.56	1.66	1.57	1.63	1.70
θ_L mode	3.95	4.04	4.11	4.47	4.28	4.06	3.40	4.27
θ_L 95%-HPD[a]	2.94–4.77	3.97–4.67	4.00–4.53	4.30–4.73	4.22–4.69	3.93–5.14	2.63–4.30	4.05–4.55
$N_e(\theta_H)$	4,197	4,009	4,023	3,901	4,162	3,929	4,087	4,257
$N_e(\theta_L)$	9,870	10,098	10,286	11,174	10,696	10,153	8,502	10,685

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

[a] 95%-highest posterior density interval; [b] Tajima's D test (Tajima 1989), significantly positive value indicate population contraction, negative values population expansion; [c] Fu's F_S test (Fu 1997), significantly negative values indicate population expansion; [d] sum of squared deviations from the expectations under a model of population expansion (Schneider & Excoffier 1999), significance indicates stationarity; [e] raggedness index (Harpending et al. 1998), significance indicates stationarity.

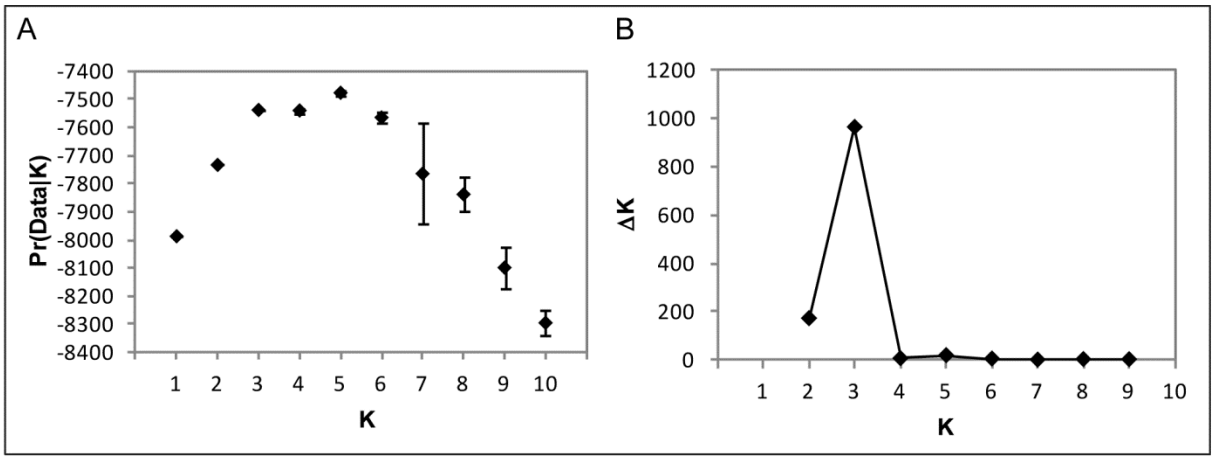
S3. Pairwise mismatch distributions for the HVRI haplotypes of seven sampling regions and all pooled samples



The gray bars indicate the observed number of sequence pairs for each category of pairwise sequence differences. The black line indicates the expectation under a model of spatial expansion.

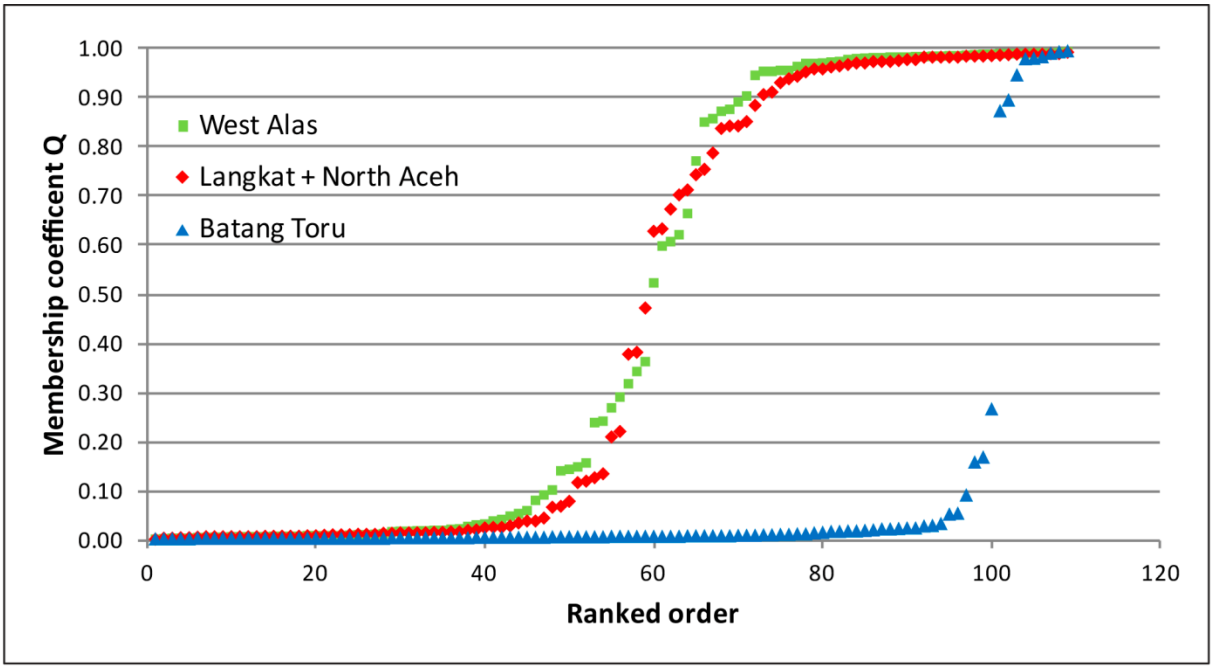
S4. Additional principal components for the microsatellite genotypes of the seven sampling regions

S5. Inference of the number of cluster K in the STRUCTURE runs

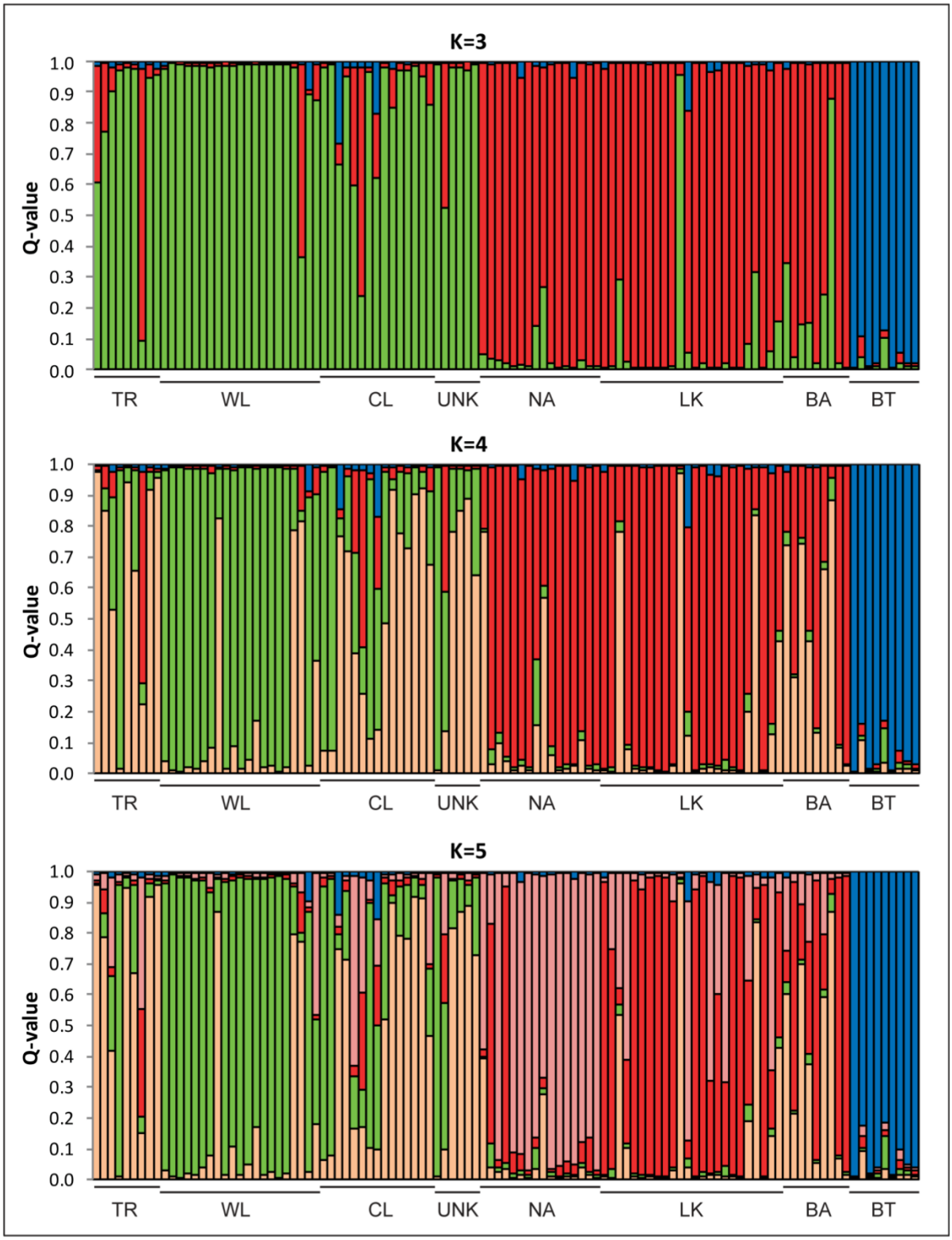


For each number of clusters K, we ran 10 iterations with the same settings. (A) Mean probability of the data given K clusters. The error bars represent ±1 standard deviation. The highest probability is achieved with K=5. (B) Delta K statistic, showing a clear mode at K=3.

S6. Membership coefficients Q plotted in ranked order for each of the three clusters inferred in the STRUCTURE analysis with K=3



S7. Results of the STRUCTURE analysis for K=3 to K=5



The membership coefficients Q shown are for the run with the highest likelihood for each K. The most probable number of cluster according to the delta K statistic is K=3.

References

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.

Harpending HC, Batzer MA, Gurven M, et al. (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1961-1967.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**, 1079-1089.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.