Conservation genomics reveals possible illegal trade routes and admixture across pangolin lineages in Southeast Asia

Helen C. Nash [a*], Wirdateti [b], Gabriel W. Low [a], Siew Woh Choo [c, d], Ju Lian Chong [e], Gono Semiadi [b], Ranjeev Hari [c,f], Muhammad Hafiz Sulaiman [e], Samuel T. Turvey [g], Theodore A. Evans [a, h], Frank E. Rheindt [a]

[a] Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117543, Singapore

[b] Research Center for Biology, Indonesian Institute of Sciences (LIPI), Jl Raya Jakarta-Bogor Km 45, Cibinong 16911, Indonesia

[c] Genome Informatics Research Laboratory, High Impact Research (HIR) Building, University of Malaya, 50603 Kuala Lumpur, Malaysia

[d] Department of Biological Sciences, Science Building B, Xi'an Jiaotong-Liverpool University, 111 Ren'ai Road, Suzhou Dushu Lake Science and Education Innovation District, Suzhou Industrial Park, Suzhou, P. R. China, 215123

[e] Department of Biological Sciences, Faculty of Science & Technology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

[f] Centre for Bioinformatics, School of Data Sciences, Perdana University, 43400 Serdang, Selangor, Malaysia

[g] Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

[h] School of Biological Sciences, The University of Western Australia (M092), 35 Stirling Highway, Crawley, WA 6009, Australia

*Corresponding author (helencatherinenash@yahoo.co.uk, +65 84313054)

Abstract

The use of genome-wide genetic markers is an emerging approach for informing evidence-based management decisions for highly threatened species. Pangolins are the most heavily trafficked mammals across illegal wildlife trade globally, but Critically Endangered Sunda pangolins (*Manis javanica*) have not been widely studied in insular Southeast Asia. We used > 12,000 single nucleotide polymorphic markers (SNPs) to assign pangolin seizures from illegal trade of unknown origin to possible geographic sources via genetic clustering with pangolins of known origin. Our SNPs reveal three previously unrecognized genetic lineages of Sunda pangolins, possibly from Borneo, Java and Singapore/Sumatra. The seizure assignments suggest the majority of pangolins were traded from Borneo to Java. Using mitochondrial markers did not provide the same resolution of pangolin lineages, and to explore if admixture might explain these differences, we applied sophisticated tests of introgression using > 2,000 SNPs to investigate secondary gene flow between each of the three Sunda pangolin lineages. It is possible the admixture which we discovered is due to human-mediated movements of pangolins. Our findings impact a range of conservation actions, including tracing patterns of trade, repatriation of rescue animals, and conservation breeding. In order to conserve genetic diversity, we suggest that, pending further research, each pangolin lineage should as a precaution be protected and managed as an evolutionarily distinct conservation unit.

Keywords: SNPs; mitochondrial markers; gene flow; illegal wildlife trade; population assignment; conservation breeding

1. Introduction

The importance of using evidence-based conservation to inform effective management decisions is increasingly recognized by conservation researchers and practitioners (Sutherland et al. 2004; Segan et al. 2010; Nash et al. 2016). The use of genetics to inform evidence-based management decisions for highly threatened species can improve conservation outcomes (Allendorf et al. 2010; Corlett 2016;

Pierson et al. 2016). For example genetic tools have been used to define species delimitations accurately (e.g. in passerine birds, Lohman et al. 2010; in crocodiles; Shirley et al. 2014), trace illegal wildlife trade (e.g. in sharks, Clarke et al. 2006; in elephant ivory, Wasser et al. 2008), assess population viability (e.g. for the Komodo dragon Ciofi et al. 1999; and sturgeon Schueller and Hayes 2011), and to inform conservation breeding, which includes captive breeding and genetic rescue of wild populations (e.g. Florida panther, Johnson et al. 2010; Burmese roofed turtle, Cilingir et al. 2017).

The conservation of pangolins would benefit from the application of genetic tools. There are eight pangolin species which are insectivorous, scaly mammals, native to Asia and Africa. Populations of all species are in decline due to habitat clearing and high levels of poaching driven by demand for traditional medicines and meat (Challender et al. 2014a). Consequently, pangolins are considered to be the most heavily trafficked mammals across illegal wildlife trade globally (Challender et al. 2015). The Sunda pangolin (*Manis javanica*) is distributed across several countries in Southeast Asia. 'Sunda' refers to Sundaland, which is a biogeographical region including the Malay Peninsula and Indonesian archipelago. Sunda pangolins are listed as Critically Endangered in the IUCN Red List, especially populations in Indonesia (Challender et al. 2014b). The need for increased genetic research is highlighted as a priority activity for pangolins in the global conservation action plan of the IUCN Species Survival Commission's Pangolin Specialist Group (Challender et al. 2014a).

A broad variety of genetic methods and different genetic markers can help to inform conservation actions for pangolins (Alacs et al. 2010; Hassanin et al. 2015; Tan et al. 2016). Among these methods, genome-wide markers such as single nucleotide polymorphisms (SNPs) are a powerful tool to provide detailed information about population structure, sometimes with greater resolution than other markers including microsatellites (Malenfant et al. 2014). Moreover, next-generation sequencing methods such as double digest restriction site-associated DNA sequencing (ddRADseq) have been used successfully to investigate population structure of mammals (Knowles et al. 2016).

There are existing genetic studies for the Sunda pangolin. A whole-genome of the Sunda pangolin was recently sequenced and published (Choo et al. 2016), providing the genomic infrastructure for further genetic research to inform robust conservation actions and management plans. The population genetic structure of Sunda pangolins across Indonesia has previously been investigated using the mitochondrial (mtDNA) control region (Wirdateti & Semiadi, 2017), with this study suggesting that there might be more than one pangolin lineage across Indonesia. However, mtDNA is only a single marker, prone to biases such as selective sweeps and introgression (Ballard & Whitlock, 2004). Wider-scale techniques such as ddRADseq could instead be used to look more rigorously at Sunda pangolin populations with genome-wide genetic markers.

The illegal trade of wildlife generates billions of USD per year, and a worryingly high proportion of this trade includes pangolins (UNODC, 2016). Unsustainably large seizures of pangolins have occurred across insular Southeast Asia, for example, >5 tonnes in Medan, Indonesia (WCS News Releases, 2015), and tonnes of scales and meat destined for countries such as China (Cheng et al. 2017). A paucity of wild samples of known geographic origin has hindered genetic assignments to investigate the sources of illegally traded pangolins (Zhang et al. 2015), and further widespread sampling of pangolins of known geographic origin with full chains of custody and expert identification is required. Meanwhile, the origin of illegally traded pangolins is an urgent issue which needs to be addressed.

In this study, we used ddRADseq to generate genome-wide genetic markers to assign pangolins of unknown origin from seizures of illegal trade to genetic clusters with wild samples of known origin in Sundaland. This is the first application of next-generation sequencing methods to the population assignment of pangolin seizures. We conducted analyses for > 12,000 SNPs and two mitochondrial genes. Our results are consequential to a range of conservation actions for pangolins, such as tracing illegal wildlife trade, repatriation of rescued pangolins, and conservation breeding.

2. Methods

2.1. Sample Collection

We aimed to collect wild Sunda pangolin samples of known origin from across insular Southeast Asia (Figure 1), and with the Indonesian Institute of Sciences (LIPI) we also acquired pangolin samples from seizures of illegal wildlife trade of unknown origin from across Indonesia (Figure 1). In total, we obtained 97 Sunda pangolin samples between July 2008 and January 2016. The total included 89 Indonesian tissue samples from the muscle of dead pangolins, eight of which had a known origin from wild populations in Java (Jember), Sumatra (Lampung) and Kalimantan (Pangkalanbun, Indonesian Borneo), plus seven blood samples from wild Singaporean pangolins, and one tissue sample from a dead pangolin in Sarawak (Malaysian Borneo). Pangolin specialists were present at each site for the collection of wild samples and were able to confirm via morphological features, such as hind scale counts (Gaubert 2011), that the wild samples were all Sunda pangolins. LIPI also maintain official national records and documentation of their samples. Veterinarians at Wildlife Reserves Singapore (WRS) collected the blood samples from anaesthetized Singaporean pangolins. Similarly, veterinarians at Kadoorie Farm and Botanic Garden in Hong Kong, China, collected a blood sample from one Chinese pangolin (Manis pentadactyla) in July 2014, which was used as an outgroup.

2.2. DNA extraction and ddRADseq library preparation

We extracted DNA at LIPI, the National University of Singapore (NUS), the Universiti Malaysia Terengganu, and Kadoorie Farm and Botanic Garden, using Qiagen DNeasy Blood & Tissue Kits. Chinese and Sunda pangolin DNA were exported to Singapore with appropriate Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) permits in 2014 and 2015/16 respectively. To measure DNA yields we used fluorometric quantitation of double-stranded DNA via Qubit 2.0™. For tissue samples with low yields, we re-extracted DNA using phenol chloroform.

To obtain genome-wide markers (SNPs) with next-generation sequencing, we modified a protocol for ddRADseq (Peterson et al. 2012) following Tay et al. (2016) and used the restriction enzymes EcoRI-HF and MspI because they worked well for other taxa (Garg et al. 2016; Ng et al. 2017). During optimisation of the RADseq protocol, we selected a Sera-Mag® bead ratio that produced DNA fragments within a range of 250–650 base pairs (bp) (Appendix A). We used additional control samples

with molecular grade water instead of DNA throughout all procedures to confirm there had been no contamination. Singapore Centre on Environmental Life Sciences Engineering checked the quality of each DNA library and sequenced each library in two lanes of one flowcell of an Illumina HiSeq 2500 Rapid Sequencing Run to produce 2 x 150 bp paired end reads. We spiked both lanes with 5% PhiX to increase the quantity of data obtained.

2.3. Bioinformatic Analysis

2.3.1. Identification of SNPs

To call SNPs across our reads, we first checked the quality of the 150 bp paired end reads with FastQC version 0.11.5 (Andrews 2016). We used a Phred Score of 20 as our quality threshold, which meant we had to truncate reads to 135 bp for further analysis (Appendix B). We then demultiplexed the reads using process_radtags in STACKS version 1.35 (Catchen et al. 2013), which grouped the uniquely labelled reads of each sample (Appendix C). We indexed the nuclear genome sequence of a Sunda pangolin (Choo et al. 2016) using Burrows-Wheeler Alignment Tool version 0.7.1. (Li et al. 2013), and we aligned our reads to it using bwa_memscript (Li et al. 2013) (Appendix D). We used Samtools version 1.4 (Li et al. 2009) to convert SAM to BAM files, and to sort the BAM files (Appendix E). Our ref_map.pl pipeline in STACKS included pstacks, cstacks and sstacks, with a minimum stacks depth of 5, to call SNPs in each sample, and match loci across populations according to alignment positions (Catchen et al. 2013) (Appendix F). In addition, we ran population analysis in ref_map.pl, with a minimum of 90% of individuals in a population required to process a locus for that population. For the population labels we assigned Singaporean versus non-Singaporean samples. Later, we checked if a default model of no population substructure changed the overall results, which it did not (Appendix G).

For further quality control of our SNP calling, we removed SNPs with 10% or more missing data, and individuals with more than 15% of loci missing, using PLINK version 1.9 (Purcell et al. 2007). We also tested higher missing data cut-offs to see if we could retain additional samples for analysis but it was not feasible. We pruned SNPs that were correlated (Appendix H), which applied a sliding window of 25 bp and removed correlated SNPs of R-squared ≥ 0.9 using a normal distribution curve

across each window. We also used PGDSpider version 2.1.0 (Lischer & Excoffier 2012) and BayeScan version 2.1 (Foll & Gaggiotti 2008) to further confirm that no loci were under selection (Appendix I).

2.3.2. Population genomic analysis

2.3.2.1. Principal Component Analysis and Fst estimation

As a preliminary analysis to explore how many genetic clusters there were across our samples, we applied Principal Component Analysis (PCA) to the remaining 83 Sunda pangolin samples to compare principal components of 12,150 SNPs using the SNPRelate package in R version 3.2 (R Core Team, 2016) (Figure 2a, and Appendix J). We also used SNPRelate to investigate pairwise Fst between clusters to get a sense of the extent of genetic differentiation (Appendix J), which applied the method of Weir & Cockerham (1984) to estimate Fst.

2.3.2.2. Bayesian clustering approaches and Network Analyses

Informed by the PCA results, we next applied a Bayesian clustering approach, using STRUCTURE and CLUMPP, which required an a priori understanding of the potential number of clusters. For STRUCTURE (Pritchard et al. 2000), we tested K = 1 to K = 7 to investigate whether there might be 1 to 7 genetic clusters across our 83 samples (Appendix K). We then used CLUMPP version 1.1.2 to determine the optimal alignment of clusters (Jakobsson & Rosenberg 2007) (Appendix L).

We were aware that STRUCTURE sometimes generates erroneous results due to uneven sample sizes between subpopulations (Puechmaille 2016), so we double-checked our genetic cluster results by using Network Analyses in NetView version 1.0, available in RStudio version 0.99.903 (RStudio Team 2015). The Network Analyses were based on a genetic distance matrix which we made from PLINK version 1.9 for our 83 samples using the 12,150 SNPs. We investigated three network algorithms: Fast-greedy, Infomap and Walktrap (Appendix M). Network Analyses provide a range of results which are all valid clustering arrangements (Appendix M). We visualized the clustering arrangement with the highest genetic distances as a mutual k-nearest neighbour graph in RStudio (Figure 2b).

During the course of our research, new methods to understand population genetic structure became available, so in addition we applied fineRADstructure package v0.2 (Malinksy et al. 2016) to quantify the ancestry sources in each population (Figure 2d). FineRADstructure utilizes a fineSTRUCTURE MCMC clustering algorithm (Lawson et al. 2012) to infer a co-ancestry matrix, which is a summary of nearest neighbour haplotype relationships across the dataset (Malinksy et al. 2016). We used our haplotypes.tsv file generated from the above populations analysis in ref_map.pl, and we converted it to a FineRADstructure input using Python scripting contributed by Emiliano Trucchi (Appendix S, Malinksy et al. 2016). The conversion script was run in PyCharm 2017.3 (Professional Edition), using 1 as the maximum number of SNPs per locus, and 50% as the maximum percentage of missing loci to be included in the PCA. The amount of missing data to allow was based on the missing data plot from fineRADstructure (Appendix S) in order to exclude only samples with high missing data. The clustered fineRADstructure co-ancestry matrix for 80 samples was visualized in RStudio (Figure 2c, and Appendix S).

2.3.3. Phylogenetic Analyses

2.3.3.1. Phylogenomics using SNPs

We felt the consistent emerging trend across our genetic clustering results warranted some further investigation to try and better understand the evolutionary trajectory of these lineages. In order to generate nucleotide sequences containing SNPs for the construction of maximum likelihood phylogenies in RAxML version 8.2.9 (Stamatakis, 2014), we aligned demultiplexed 135bp sequence reads in pyRAD version 3.0.64 (Eaton, 2014). This included the Chinese pangolin as an outgroup. SNPs were called using the ddrad option with the following parameters: clustering threshold of 95%, minimum cluster coverage of five, maximum of six low-quality sites per locus, minimum of 79 samples present in a final locus, and maximum of three individuals with a shared heterozygous site per locus (Appendix N). A total of 2,365 SNPs was generated and inputted in RAxML version 8.2.9 with the following parameters: GTRGAMMA option provided, 1000 rapid bootstraps inferences, and a final maximum likelihood search (Figure 3a, and Appendix O). We tested other runs with increasing and decreasing amounts of missing data, but the bootstrap support

values did not improve, so we only present the concatenation method with 2,365 SNPs (Figure 3a).

2.3.3.2. Phylogenetics using mitochondrial DNA

It was useful to compare the genetic clustering results of our genome-wide markers with mitochondrial DNA (mtDNA) markers which are commonly used to assign species in illegal wildlife trade, so we sequenced two mtDNA coding genes, cytochrome b (*Cytb*) and cytochrome oxidase c subunit 1 (*CO1*). We didn't have sufficient DNA remaining from every pangolin sample, but we were able to include the majority, 59 pangolins. We conducted Sanger sequencing at NUS and LIPI using primer sequences provided by LIPI (Appendix Q). Sequences were aligned and then *Cytb* and *CO1* sequences were manually concatenated in MEGA 7.0 (Kumar et al. 2016). The complete concatentation sequence consisted of 1575 base pairs (*Cytb* = 787 bp, *CO1* = 788 bp), and the sequences are available on GenBank (respective accession numbers MG825495-MG825551 and MG825552-MG825610). We added the mitogenome of a Chinese pangolin (GenBank KT445978.1) and one additional Sunda pangolin (NC_026781.1), then used MEGA 7.0 (Kumar et al. 2016) to construct a phylogenetic tree with maximum likelihood (ML), with a General Time Reversible Model, partial deletion and 1000 bootstraps (Figure 4a). In order to facilitate comparison of results, the first letter of the mtDNA tree labels reflect the genetic clustering result from the SNPs, e.g. J = Java, and the seizure location is also given at the end of each label. We used asterisks beside the labels to indicate when the seizure location and genetic cluster result from the SNPs were the same geographic area (Figure 4a). We also generated a haplotype network from the concatenated mtDNA sequences using the Median-Joining method in PopART (Leigh & Bryant 2015) (Figure 4b). The colour of each sample label represents the SNP cluster results to further aid comparison between the mtDNA and SNP results (Figure 4b).

2.3.4. Tests for Introgression

The contrasts between our genetic clustering results from the SNPs (Figure 2) versus the topology of the mtDNA tree/haplotype network (Figure 4) raised our suspicion that potential introgression might explain these differences. It was not possible to test every sample for introgression because the computational run time is

prohibitive. Instead we selected a few anomalous results from the mtDNA tree, and we applied ABBA BABA tests to investigate secondary gene flow in those samples from our wild samples of known origin (Zinenko et al. 2016). The ABBA BABA tests required nucleotide sequences containing SNPs so we used our PHYLIP files from pyRAD including 2,365 SNPs (Appendix R). Our first ABBA BABA test used a wild sample from Borneo, MZBR 1163, as group A; an anomalous sample, MZBR 1040, which falls within the Bornean cluster of SNPs, but has suspected introgression from Java based on the mtDNA result, as group B; and a wild sample from Java, MZBR 1184, as group C; the outgroup was our Chinese pangolin. This tested MZBR 1040 for suspected introgression from the Javan lineage. Our second and third tests used a wild sample from Java, MZBR 1184, as group A; and an anomalous sample, MZBR 0270, which falls within the Javan cluster of SNPs, but has suspected introgression from both Singapore/Sumatra and Borneo based on the mtDNA result, as group B; group C was initially a wild sample from Singapore/Sumatra, rescue 1, which we then switched to MZBR 1163 a wild sample from Borneo in another run; the outgroup was always our Chinese pangolin. This initially tested MZBR 0270 for suspected introgression from the Singapore/Sumatra lineage, and next tested the same sample for suspected introgression from the Bornean lineage. The ABBA BABA results were summarised in a simple cartoon figure (Figure 3b). These ABBA BABA tests provided examples of introgression.

2.3.5. Tracing illegal trade

The genetic clustering results from each genome-wide population genomic analysis, including PCA, STRUCTURE, NetView and fineRADstructure, all gave compatible results. Hence, any of those methods could be used to assign the seized pangolin samples of unknown origin to the wild samples of known origin (Figure 2 and Appendix L). The assignment results were the same for every method, except that fineRADstructure only used 80/83 samples due to differences in sample filtering (Figure 2c). The additional three results came from only PCA (Figure 2a), STRUCTURE (Appendix L) and NetView (Figure 2b). The phylogenetic trees (Figures 3 and 4) were not used for the illegal trade assignments due to the presence of introgression and the poor resolution of the trees.

There is perhaps some further grey support for our genetic assignments according to the locations of the illegal seizures. Therefore, we highlighted in bold cases where the sample seizure location is similar to its genetic cluster result, and the wild samples of known origin labelled as WILD are also highlighted in bold (Figure 5).

All necessary research permits and ethics approvals were granted prior to commencement of this project. In particular, NUS Institutional Animal Care and Use Committees (IACUC) approved research methods, and we obtained CITES permits for all pangolin samples.

3. Results

3.1. ddRAD sequencing and SNP discovery

Our collection of samples included 97 Sunda pangolins and 1 Chinese pangolin. Following library preparation, 2 Sunda pangolin tissue samples did not yield sufficient DNA fragments within the 250–650 bp range, so only 96 samples underwent Illumina sequencing (Appendix P). In total we obtained 49.67 GB of data from 96 samples across two lanes (reads one = 24.12 GB, reads two = 25.55 GB), and there were no lane differences in the results.

After demultiplexing our reads, 7 Sunda pangolin samples were discarded due to low sequencing coverage (Appendix P). Across the remaining 89 samples 60,197 SNPs were called via the STACKS pipeline. In PLINK 43,439 loci were removed due to having ≥10% missing data. Only 83 samples met our <15% missing data requirement; the Chinese pangolin sample was excluded at this stage. A further 4,608 loci were removed due to correlation. Consequently, for further analysis we had 83 Sunda pangolin samples with 12,150 SNPs. BayeScan did not detect selection.

3.2. Population Genomic Structure

3.2.1. Principal Component Analysis and Fst estimation

The default settings of SNPRelate removed a further 223 SNPs prior to PCA of 11,927 SNPs. The PCA results suggested that we were dealing with three distinct genetic clusters of Sunda pangolins, possibly from Sumatra/Singapore, Java, and Borneo according to the genetic cluster labels of wild pangolins of known origin

(Figure 2a). The first two eigenvectors of PCA held the largest percentage of variance among the population, principal component 1 = 4.47 % and principal component 2 = 4.38 %. The mean Fst between clusters is: Java versus Sumatra/Singapore = 0.0684, Borneo versus Java = 0.0556, Borneo versus Sumatra/Singapore = 0.0446, and across all three clusters = 0.0545. The mean Fst values increased when we removed samples showing signals of introgression (see section 3.3.2 for details of introgression), for example, the mean Fst across all three clusters increased when MZBR 0270 was removed, mean Fst = 0.0556 (Appendix J).

3.2.2. Bayesian clustering and Network Analyses

The results from STRUCTURE, Structure Harvester and CLUMPP supported the PCA results, similarly indicating that three genetic clusters across the 83 Sunda pangolins is the most likely arrangement (Appendix L). Moreover, all of the clustering approaches across PCA (Figure 2a), STRUCTURE (Appendix L), NetView (Figure 2b) and fineRADstructure (Figure 2c) produced compatible clustering results.

3.2.2.1. Additional insight to population substructure

The co-ancestry matrix from fineRADstructure used 46, 274 loci to illustrate the levels of co-ancestry across 80 Sunda pangolin samples (Figure 2c and Appendix S), with yellow representing the lowest levels of co-ancestry. The matrix confirmed that there are three main clusters, Borneo, Java and Sumatra/Singapore, and revealed additional differences in co-ancestry among clusters/samples. The black colour represents the highest level of co-ancestry and suggests that MZBR 1030 and 1031 might be highly related pangolins from Sumatra. The purple coloured MZBR 1166 and 1167 could be closely related pangolins from Borneo. The deep red colour represents samples MZBR 1184, 1185, 1190, 1063 and 1060 and suggests these might be close relatives perhaps from a similar area in Java. We were aware of the relationships between MZBR 1030 and 1031, also 1166 and 1167, from the NetView analyses (Appendix M). However, we were not aware about the groupings of pangolins labelled with deep red colours which show very high levels of co-ancestry. FineRADstructure provided more compelling detail about population substructure than our other population genomic analyses.

3.3. Phylogenetic trees

3.3.1. Phylogenomic tree using SNPs

Phylogenomic analysis with the concatenation method RAxML used 2,365 SNPs, and although the Javan samples in the RAxML tree did not form a monophyletic group (Figure 3a), the results were otherwise compatible with our population clustering results from PCA, STRUCTURE, NetView and fineRADstructure (Figures 2a-c and Appendix L). Overlaid coloured shading illustrates the population genomic clustering results on top of the RAxML tree to aid comparison (Figure 3a). The RAxML trees generally had low bootstrap support. A few samples also had high missing data, including MZBR 1189, 1183 and 1179 (Appendix O). We tested a wide variety of parameters in both pyRAD and RAxML to try to improve the bootstrap support, for example, minimum coverage for a cluster (5 to 10), maximum number of sites with quality score less than 20 (4 to 6), clustering threshold (0.85 to 0.95), and minimum samples in a final locus (70 to 81), however, the results did not improve.

3.3.2. Mitochondrial DNA tree and the issue of introgression

The results of the mtDNA tree (Figure 4a) and the mtDNA haplotype network (Figure 4b) were congruent. The mtDNA tree was not well resolved (Figure 4a) and the haplotype network clearly shows the differences to the SNP results (Figure 4b). Our ABBA BABA tests of secondary gene flow using 2,365 SNPs (Appendix R) demonstrated that there are signals of introgression across our samples, for example, MZBR 0270 showed a signal of introgression from both Singapore/Sumatra and Borneo. The test result for MZBR 1040 was not statistically significant. The statistically significant ABBA BABA test results are summarised in a simple cartoon diagram (Figure 3b).

3.4. Tracing illegal trade

Based on the congruent genetic clustering results from PCA (Figure 2a), STRUCTURE (Appendix L), NetView (Figure 2b) and fineRADstructure (Figure 2c), which all showed three key distinct clusters which were geographically labelled by the wild samples of known origin, we conclude that 20 of our samples possibly originated from Java, 21 are possibly from Sumatra/Singapore, and 42 are possibly from Borneo (Figure 5). These three key clusters are represented by black squares (Figure 5) beneath which a full list of each sample's seizure location or wild location

has been provided. The samples highlighted in bold do not appear to have been translocated overseas, their seizure location is similar to their genetic origin. The pie charts (Figure 5) show the trade route at point of seizure of the samples, they do not include the wild samples. The arrows help to clarify the pie charts by pointing in the geographic direction of the trade.

Among the 49 pangolins seized from illegal wildlife trade in Java (Figure 1), only 18 may have been sourced within Java, while 23 may instead have been imported from Borneo, and 8 from Sumatra. The 6 pangolins seized in Medan, North Sumatra (Figure 1), may have originated in Sumatra as these samples group with the Sumatra/Singapore cluster. All of the 16 pangolins seized in Borneo (Figure 1) likely originated there. Only one pangolin across our dataset seems to have been traded from Borneo into Sumatra.

4. Discussion

Our study is the first application of next-generation sequencing methods to the critically endangered Sunda pangolin. The SNP data from each analysis that we have employed suggests there are three key genetic clusters across our samples, which are likely from Borneo, Java and Singapore/Sumatra. The majority of trade across our samples seems to be from Borneo into Java. The presence of introgression across our samples likely explains the poor bootstrap support of the phylogenetic trees. Only the population genomic analyses, PCA, STRUCTURE, NetView and fineRADstructure, provided sufficient clarity to genetically match the pangolins of unknown origin from illegal trade to the wild samples of known origin.

The most compelling detail about population substructure was generated by the fineRADstructure method, and we recommend the use of this programme for further research. FineRADstructure not only confirmed our three main clusters, Borneo, Java and Singapore/Sumatra, it revealed further insights into the substructure of these clusters, such as highly related pangolins and other shared co-ancestry. There might be a geographic basis for some of the further substructures which fineRADstructure identified, such as the deep red clusters showing very high levels of shared co-ancestry (Figure 2c), but we require further wild samples of known origin to geographically label these clusters to more concise geographic areas. Currently, the deep orange coloured group within the Singapore/Sumatra cluster is

the only subgroup which we can geographically label. The wild Singaporean samples label that cluster as Singaporean.

Genetic data are essential to help trace illegal wildlife trade and create forensic genetic databases (Ogden et al. 2009; Wasser et al. 2015; Ogden & Linacre 2015). Although we understand that without the use of trained conservation dogs it is very difficult to capture rare and nocturnal pangolins, which cannot be easily baited, we urge further sampling of wild pangolins of known origin to help inform geographic population assignments. In Singapore our wild sampling was more extensive than elsewhere in Southeast Asia due to a 24hr /7 day per week rescue service for wildlife, which is managed by ACRES, as well as rehabilitation of pangolins at Singapore Zoo. We hope that other partnerships such as this between NGOs and researchers could facilitate improved pangolin sampling elsewhere. It is important that during the collection of reference samples full chains of custody are documented with expert verification.

The possible directions of illegal trade which we have revealed in this study (Figure 5) are similar to patterns of illegal trade of other wild species, for example, it has been documented that birds are also poached in large numbers across Indonesia and transported to markets in Jakarta and Java for sale or onwards distribution (Chng et al. 2015). There is some further grey support for our population assignments from our seizure locations, for example, from the local police seizures in Borneo all of those seized pangolins clustered within our probable Bornean cluster (Figure 5).

The distinct genetic clustering of populations from Borneo, Java and Singapore/Sumatra matches divergence patterns seen in other vertebrate taxa (Wilson & Reeder, 2005; Leonard et al. 2015). It is possible that our Sunda pangolin lineages might not only be genomic but also ecologically differentiated, since the habitats on Java are primarily monsoonal dryland and savannah, whereas the habitats on the other sampled areas are primarily rainforest (Sundevall, 1843; Whitten et al. 1996). Studies of ecology and morphology across the three Sunda pangolin lineages revealed in this study should be conducted to help inform their taxonomic classification (Gaubert & Antunes 2005), and phenotypic inquiry may uncover traits that support the genomic divergences shown by our data.

459 The genetic divisions were not well detected by our mtDNA data, partly due to
460 admixture and introgression. It is possible that genetic introgression explains the
461 differences between our tentative RAxML (Figure 3a) and mtDNA trees (Figure 4a).
462 Also mtDNA is a single marker that only reflects the maternal history (Ballard &
463 Whitlock, 2004). We suggest that for the population assignment of illegally traded
464 pangolins the use of genome-wide SNP markers can provide higher resolution than
465 mtDNA markers alone. We suggest that future population clustering results using
466 SNP data could be obtained quickly with just fineRADstructure (Malinsky et al. 2016)
467 and more georeferenced samples of known origin are needed.

468 Sunda pangolins are also distributed across Thailand, Myanmar, Cambodia, Laos
469 and Vietnam, and it is possible that greater substructure exists across their full
470 geographic range (Zhang et al. 2015). Protecting genetic diversity is important for the
471 resilience and survival of species (Pierson et al. 2016), and this genetic diversity
472 must therefore be considered in conservation management plans, not least to inform
473 conservation breeding (Hua et al. 2015), including captive breeding and genetic
474 rescue, and repatriation of rescued pangolins. Selection of pangolin release
475 locations should be mindful of the appropriate genetic source population (Challender
476 et al. 2014b). Currently, there is no consideration of genetic substructure when
477 pangolin seizures of unknown origin are released into the wild, and the impacts of
478 introducing and mixing pangolins from different lineages into the same area are not
479 fully understood. The genetic introgression uncovered in our study might possibly be
480 due to human influences, such as translocation of pangolins outside of their natural
481 range (Pantel & Chin, 2009). Conservation breeding records also need to document
482 and consider the highest resolution of population genetic structure to ensure that
483 genetic diversity is well-managed (Allendorf et al. 2010). We wish to emphasize the
484 importance of genetic screening of all individuals involved in releases and
485 conservation breeding.

486 As stated in the IUCN Pangolin Specialist Group action plan (Challender et al. 2014),
487 all pangolin species require further genetic investigation. In our study we chose the
488 double-digest form of RADseq instead of other forms, because this method might be
489 more reproducible for other pangolin researchers to follow, as well as less expensive
490 for researchers with smaller budgets (Andrews et al. 2016).

The findings of our research impact a range of conservation actions, including tracing illegal wildlife trade, delimitation of pangolin conservation units, repatriation of rescued pangolins to appropriate locations, and conservation breeding. To maximise the resolution of genetic tracing of pangolins, we recommend that genome-wide markers are used in combination with mtDNA genes. Our findings provide a new baseline to help begin to understand Sunda pangolin populations, and we hope these findings will inspire future research and management actions that can support effective conservation of pangolins in Asia.

Acknowledgments

All authors contributed equally to this work. All authors discussed the results and implications and commented on the manuscript at all stages. We thank the Indonesian Institute of Sciences (LIPI), Lee Kong Chian Natural History Museum (LKCNHM), Agri-Food & Veterinary Authority of Singapore (AVA), Wildlife Reserves Singapore (WRS), Department of Wildlife and National Parks Peninsular Malaysia (DWNP), Universiti Malaysia Terengganu (UMT) and Kadoorie Farm and Botanic Garden (KFBG) for assistance with sample collection and the arrangement of relevant permits and permissions. Special thanks are given to Yulianto (LIPI) and H. Zhang (KFBG) who helped to extract DNA, and S. Oh (WRS), A. Ali (WRS), S. Luz (WRS), P. Lee (WRS), C. F. Maosheng (LKCNHM), M. Chua (LKCNHM), R. Meier (LKCNHM), C.Y. Gwee (NUS), G. Ades (KFBG) and A. Grioni (KFBG). We also thank R. Asher (University of Cambridge) for his encouragement and advice, and the IUCN-SSC Pangolin Specialist Group and Singapore Pangolin Working Group who provided logistical support throughout this research. The Rheindt Lab at NUS shared a ddRADseq protocol. L. Wijedasa gave us base maps for the figures. S. Thompson helped with coding for fineRADstructure. Funding was provided by the Dennis Gould Foundation; H.C.N is supported by a SINGA PhD Research Scholarship at NUS; S.T.T is supported by a Royal Society University Research Fellowship (UF130573). We also acknowledge internal departmental funding at the Department of Biological Sciences at NUS; and the Research Center for Biology-LIPI Competitive Project 3400.001.002.021 SEAMEO BIOTROP DIPA 060.12/PSRP/SPK-PNLT/2014.

Literature Cited

Alacs EA, et al. 2010. DNA detective: A review of molecular approaches to wildlife forensics. Forensic Science, Medicine, and Pathology. **6**(3): 180–194.

Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics. **11**(10): 697–709.

Andrews KR, et al. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics. **17**: 81–92.

Andrews S. 2016. FastQC: a quality control tool for high throughput sequence data. Version 0.11.5 Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. Molecular Ecology. **13**(4): 729–744.

Catchen J, et al. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology. **22**(11): 3124–40.

Challender DWS, Waterman C, Baillie JEM. 2014a. Scaling up pangolin conservation. IUCN SSC Pangolin Specialist Group Conservation Action Plan. Zoological Society of London, London, UK.

Challender DWS, et al. 2014b. Manis javanica. The IUCN Red List of Threatened Species 2014: e.T12763A45222303. Available at: http://dx.doi.org/10.2305/IUCN.UK.2014-2.RLTS.T12763A45222303.en.

Challender DWS, Harrop SR, MacMillan DC. 2015. Understanding markets to conserve trade-threatened species in CITES. Biological Conservation. **187**: 249–259.

Cheng W, Xing S, Bonebrake TC. 2017. Recent Pangolin Seizures in China Reveal Priority Areas for Intervention. Conservation Letters. doi:10.1111/conl.12339

Chng SCL, Eaton JA, Krishnasamy K, Shepherd CR, Nijman V. 2015. In the Market for Extinction: An inventory of Jakarta's bird markets. TRAFFIC. Petaling Jaya, Selangor, Malaysia.

549 Choo SW et al. 2016. Pangolin genomes and the evolution of mammalian scales and
550 immunity. Genome Research. **26**(10): 1312-1322.

551 Çilingir FG, Rheindt FE, Garg KM, Platt K, Platt SG, Bickford DP. 2017.
552 Conservation genomics of the endangered Burmese roofed turtle. Conservation
553 Biology. **28**: in press.

554 Ciofi C, et al. 1999. Genetic divergence and units for conservation in the Komodo
555 dragon Varanus komodoensis. Proceedings of the Royal Society B. **266**: 2269-
556 2274.

557 Clarke S, et al. 2006. Identification of shark species composition and proportion in
558 the Hong Kong shark fin market based on molecular genetics and trade records.
559 Conservation Biology. **20**: 201-211.

560 Corlett RT. 2016. A Bigger Toolbox: Biotechnology in Biodiversity Conservation.
561 Trends in Biotechnology. **35**(1): 55-65.

562 Drummond AJ, et al. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7.
563 Molecular Biology and Evolution. **29**(8): 1969–1973.

564 Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: A website and program
565 for visualizing STRUCTURE output and implementing the Evanno method.
566 Conservation Genetics Resources. **4**(2): 359–361.

567 Eaton DAR. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic
568 analyses. Bioinformatics. **30**(13): 1844-1849.

569 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of
570 individuals using the software STRUCTURE: A simulation study. Molecular
571 Ecology. **14**(8): 2611–2620.

572 Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci
573 appropriate for both dominant and codominant markers: A Bayesian perspective.
574 Genetics. **180**(2): 977–993.

575 Garg KM, et al. 2016. Genome-wide data help identify an avian species-level lineage
576 that is morphologically and vocally cryptic. Molecular Phylogenetics and
577 Evolution. **102**: 97–103.

578 Gaubert P. 2011. Family Manidae. In: Wilson, D.E., Mittermeier, R.A. (Eds.),
579 Handbook of the Mammals of the World, Vol. 2: Hoofed Mammals. Lynx
580 Edicions, Barcelona, Spain.

581 Gaubert P, Antunes A. 2005. Assessing the taxonomic status of the Palawan
582 pangolin Manis Culionensis (Pholidota) using discrete morphological characters.
583 Journal of Mammalogy. **86**(6): 1068–1074.

584 Hassanin A, Hugot JP, van Vuuren BJ. 2015. Comparison of mitochondrial genome
585 sequences of pangolins (Mammalia, Pholidota). Comptes Rendus Biologies.
586 **338**(4): 260–265.

587 Hua L, et al. 2015. Captive breeding of pangolins: Current status, problems and
588 future prospects. ZooKeys. **507**: 99–114.

589 Jakobsson M, Rosenberg NA. 2007. CLUMPP: A cluster matching and permutation
590 program for dealing with label switching and multimodality in analysis of
591 population structure. Bioinformatics. **23**(14): 1801–1806.

592 Johnson WE, et al. 2010. Genetic restoration of the Florida panther. Science **329**:
593 1641-1645.

594 Knowles LL, et al. 2016. Quantifying the similarity between genes and geography
595 across Alaska's alpine small mammals. Journal of Biogeography. **43**(7): 1464–
596 1476.

597 Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics
598 Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. **33**(7):
599 1870-4.

600 Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population
601 Structure using Dense Haplotype Data. PLoS Genetics. **8**(1): e1002453.

602 Leigh JW, Bryant D. 2015. POPART: full-feature software for haplotype network
603 construction. Methods in Ecology and Evolution. **6**: 1110–1116.

604 Leonard JA, et al. 2015. Phylogeography of vertebrates on the Sunda Shelf: a multi-
605 species comparison. Journal of Biogeography. **42**: 871–879.

606 Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics,
607 **25**(16): 2078–2079.

608 Li H, et al. 2013. Burrows-Wheeler Alignment Tool v0.7.1. Available at:

609 http://bio-bwa.sourceforge.net/bwa.shtml

610 Lischer HEL, Excoffier L. 2012. PGDSpider: An automated data conversion tool for
611 connecting population genetics and genomics programs. Bioinformatics. **28**(2):
612 298–299.

613 Lohman DJ, et al. 2010. Cryptic genetic diversity in “widespread” Southeast Asian
614 bird species suggests that Philippine avian endemism is gravely underestimated.
615 Biological Conservation. **143**: 1885-1890.

616 Malenfant RM, Coltman DW, Davis CS. 2014. Design of a 9K illumina BeadChip for
617 polar bears (Ursus maritimus) from RAD and transcriptome sequencing.
618 Molecular ecology resources. **15**(3): 587-600.

619 Malinsky M, Trucchi E, Lawson D, Falush D. 2016. RADpainter and
620 fineRADstructure: population inference from RADseq data. BioRxiv 057711.
621 (pre-print). doi: https://doi.org/10.1101/057711.

622 Nash HC, Wong MHG, Turvey ST. 2016. Using local ecological knowledge to
623 determine status and threats of the critically endangered Chinese pangolin
624 (Manis pentadactyla) in Hainan, China. Biological Conservation. **196**: 189-195.

625 Ng NS, Wilton PR, Prawiradilaga DM, Tay YC, Indrawan M, Garg KM, Rheindt FE.
626 2017. The effects of Pleistocene climate change on biotic differentiation in a
627 montane songbird clade from Wallacea. Molecular Phylogenetics and Evolution.
628 **114**: 353-366.

629 Ogden R, Dawnay N, McEwing R. 2009. Wildlife DNA forensics - Bridging the gap
630 between conservation genetics and law enforcement. Endangered Species
631 Research. **9**(3): 179–195.

632 Ogden R, Linacre A. 2015. Wildlife forensic science: A review of genetic geographic
633 origin assignment. Forensic Science International: Genetics. **18**: 152–159.

634 Pantel S, Chin SY. 2009. Proceedings of the Workshop on Trade and Conservation
635 of Pangolins native to South and Southeast Asia. TRAFFIC Southeast Asia.
636 Petaling Jaya, Selangor, Malaysia.

637 Peterson BK, et al. 2012. Double digest RADseq: An inexpensive method for de
638 novo SNP discovery and genotyping in model and non-model species. PLoS
639 ONE. **7**(5).

640 Pierson JC, et al. 2016. Genetic factors in threatened species recovery plans on
641 three continents. Frontiers in Ecology and the Environment. **14**(8): 433–440.

642 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using
643 multilocus genotype data. Genetics. **155**(2): 945–959.

644 Puechmaille SJ. 2016. The program structure does not reliably recover the correct
645 population structure when sampling is uneven: Subsampling and new estimators
646 alleviate the problem. Molecular Ecology Resources. **16**(3): 608–627.

647 Purcell S, et al. 2007. PLINK: A tool set for whole-genome association and
648 population-based linkage analyses. American Journal of Human Genetics. **81**(3):
649 559–575.

650 Rambaut A, et al. 2014. Tracer v1.6. Available at: http://beast.bio.ed.ac.uk/Tracer

651 Rambaut A, Drummond AJ. 2016. TreeAnnotator v2.4.2. Institute of Evolutionary
652 Biology, University of Edinburgh.

653 R Core Team 2016. R: A Language and Environment for Statistical Computing. R
654 Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-
655 project. org/

656 RStudio Team 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston,
657 MA. Available at: http://www.rstudio.com/

658 Schueller AM, Hayes DM. 2011. Minimum viable population size for lake sturgeon
659 (Acipenser fulvescens) using an individual-based model of demographics and
660 genetics Canadian Journal of Fisheries and Aquatic Sciences. **68**: 62-73.
661

662 Segan DB, et al. 2010. Using conservation evidence to guide management.
663 Conservation Biology. **25**(1): 200-202.
664
665 Shirley MH, et al. 2014. Rigorous approaches to species delimitation have significant
666 implications for African crocodilian systematics and conservation. Proceedings
667 of the Royal Society B. **281**: 20132483.

668 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
669 analysis of large phylogenies. Bioinformatics. **30**.

670 Sundevall CJ. 1843. Wer sicht der gattung Manis. Kongl. Svenska vetenskaps-
671 akademiens handlingar. 1842: 245-283.

672 Sutherland WJ, et al. 2004. The need for evidence-based conservation. Trends in
673 Ecology and Evolution. **19**: 305-308.

674 Tan TK, et al. 2016. PGD: a pangolin genome hub for the research community.
675 Database (Oxford). Available at:
676 http://database.oxfordjournals.org/content/2016/baw063.full.

677 Tay YC, Chng MWP, Sew WWG, Rheindt FE, Tun KPP, Meier R. 2016. Beyond the
678 Coral Triangle: high genetic diversity and near panmixia in Singapore's
679 populations of the broadcast spawning sea star Protoreasternodosus. Royal
680 Society Open Science. **3**: 160253.

681 United Nations Office on Drugs and Crime (UNODC). 2016. World Wildlife Crime
682 Report: Trafficking in protected species. Available at:
683 https://www.unodc.org/documents/data-and-
684 analysis/wildlife/World_Wildlife_Crime_Report_2016_final.pdf

685 Wasser SK, et al. 2008. Combating the illegal trade in African elephant ivory with
686 DNA forensics. Conservation Biology. **22**: 1065-1071.

687 Wasser SK, et al. 2015. Genetic assignment of large seizures of elephant ivory
688 reveals Africa's major poaching hotspots. Science. **349**(6243): 84–88.

689 WCS News Releases. April 2015. https://newsroom.wcs.org/News-
690 Releases/articleType/ArticleView/articleId/6715/April-27-Indonesian-National-

691 Police-Seize-Major-Shipment-of-Pangolins-Arrest-Smuggler.aspx

692 Whitten T, Soeriaatmadja RE, Afiff SA. 1996. The Ecology of Java and Bali. The
693 Ecology of Indonesia Series, II. Periplus Editions (HK) Limited.

694 Wilson DE, Reeder DM. (editors). 2005. Mammal Species of the World. A
695 Taxonomic and Geographic Reference (3rd ed). Johns Hopkins University Press.

696 Wirdateti, Semiadi G. 2017. Genetic Variation of Confiscated Pangolins of Sumatra,
697 Java, and Kalimantan based on Control Region Mitochondrial DNA. Jurnal
698 Veteriner. **18**(2): 181-191.

699 Zhang H, et al. 2015. Molecular tracing of confiscated pangolin scales for
700 conservation and illegal trade monitoring in Southeast Asia. Global Ecology and
701 Conservation. **4**: 414–422.

702 Zinenko O, et al. 2016. Hybrid origin of European Vipers (Vipera magnifica and
703 Vipera orlovi) from the Caucasus determined using genomic scale DNA markers.
704 BMC Evolutionary Biology. **16**:76.

705

Figure Captions

Fig. 1 Locations of seizures of illegally traded pangolins and wild Sunda pangolins of known origin across Indonesia (Java, Kalimantan and Sumatra), Malaysia and Singapore that were used in this study. Numbers in brackets represent the number of pangolins sampled per location

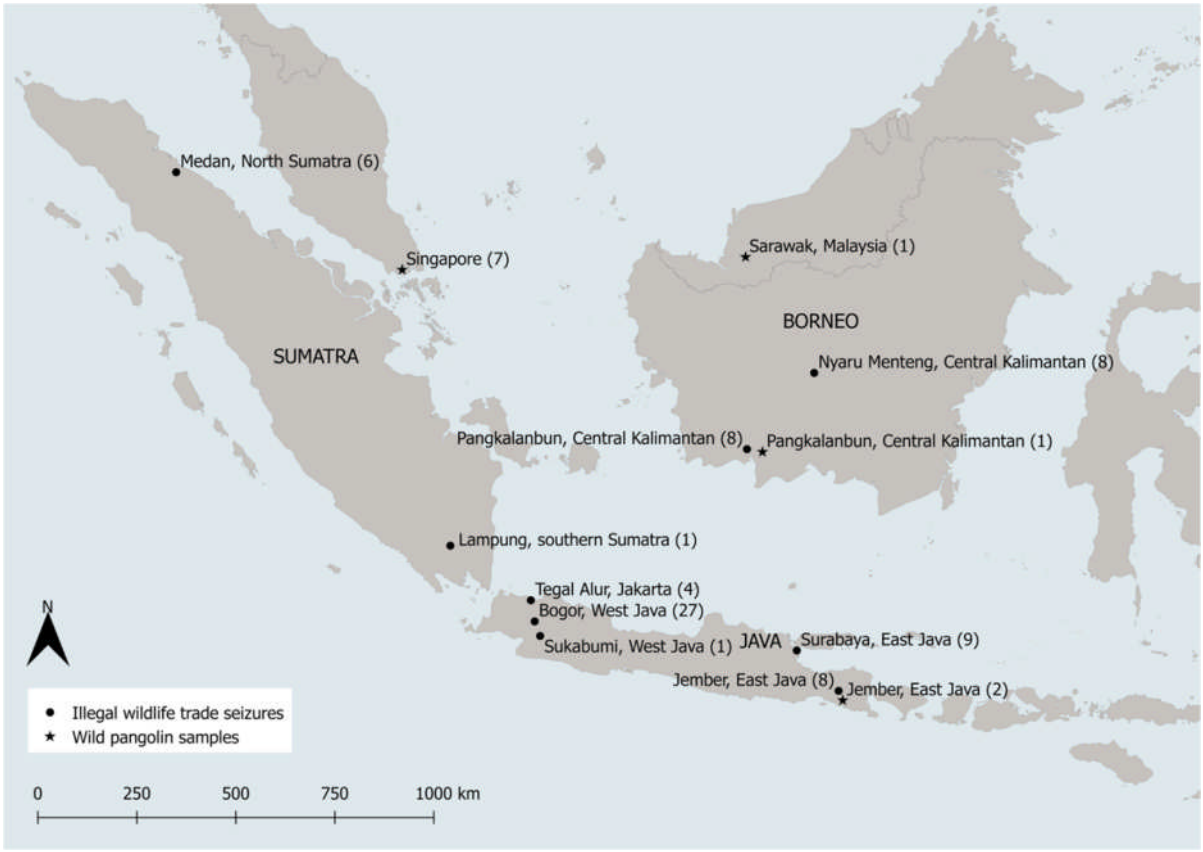
Fig. 2 Population genetic clusters. The pangolin samples within each geographically labelled cluster are similar across each figure. (a) PCA of 11,927 SNPs across 83 Sunda pangolins. Principal component 1 = 4.47 % and principal component 2 = 4.38 %. (b) Network Analysis with the highest genetic distance of clusters using 12,150 SNPs across 83 Sunda pangolins from the Walktrap model. (c) Clustered fineRADstructure coancestry matrix using 46, 274 loci across 80 Sunda pangolins

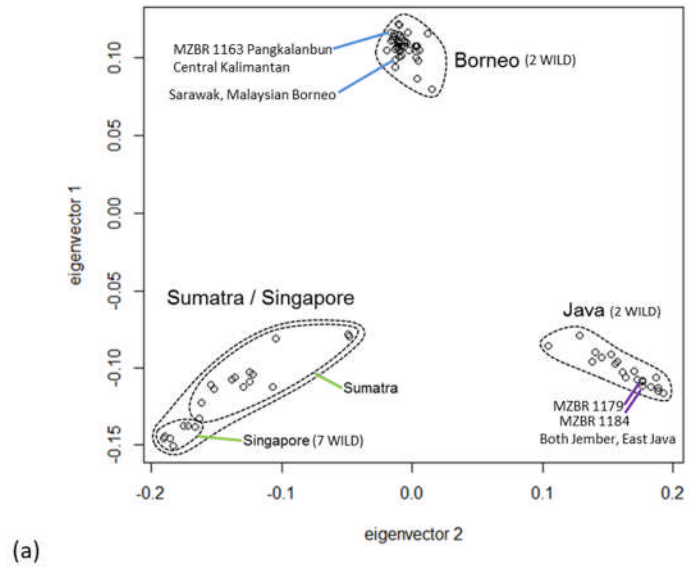
Fig. 3 (a) Maximum likelihood phylogeny in RAxML using 2,365 SNPs. The coloured overlaid shading illustrates the clustering results from other methods, PCA, Structure, NetView and fineRADstructure to facilitate comparison with those results: blue = Borneo, green = Singapore/Sumatra, no colour = Java. (b) The directions of detected introgression between pangolin lineages indicated by dashed arrows, from ABBA BABA tests using 2,365 SNPs

Fig. 4 (a) mtDNA phylogeny, with concatenated *Cytb* and *CO1* data. Sample labels begin with a key of the SNP cluster result to facilitate comparison: B = Borneo, J = Java, S = Singapore/Sumatra. Seizure locations are given following the sample name. Asterisks indicate samples where seizure location and SNP cluster result were the same geographic area. (b) mtDNA haplotype network. The colour of sample label represents the SNP cluster result to facilitate comparison: blue = Borneo, green = Singapore/Sumatra, purple = Java, black = mtDNA only. Wild samples of known origin are labelled as WILD.

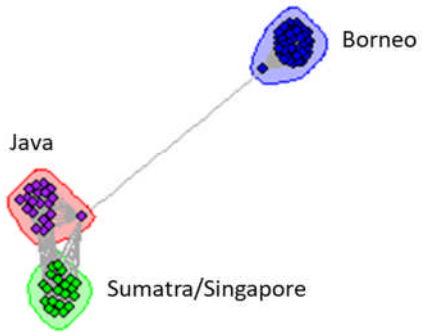
Fig. 5 Inferred directions of the illegal trade of pangolins. Varied population genomic analyses using > 12,000 SNPs all provided compatible results (PCA, Structure, NetView and fineRADstructure). A full list of each sample's seizure location or wild location is provided below each genetic cluster. The pangolins highlighted in bold were not translocated overseas, their seizure location is similar to their genetic origin. Pie charts show the trade route at point of seizure of the pangolin (they do not

737 include the wild samples). Arrows show the direction of trade, with the largest arrow
738 reflecting the highest volume of trade, and the dashed arrows reflecting less trade

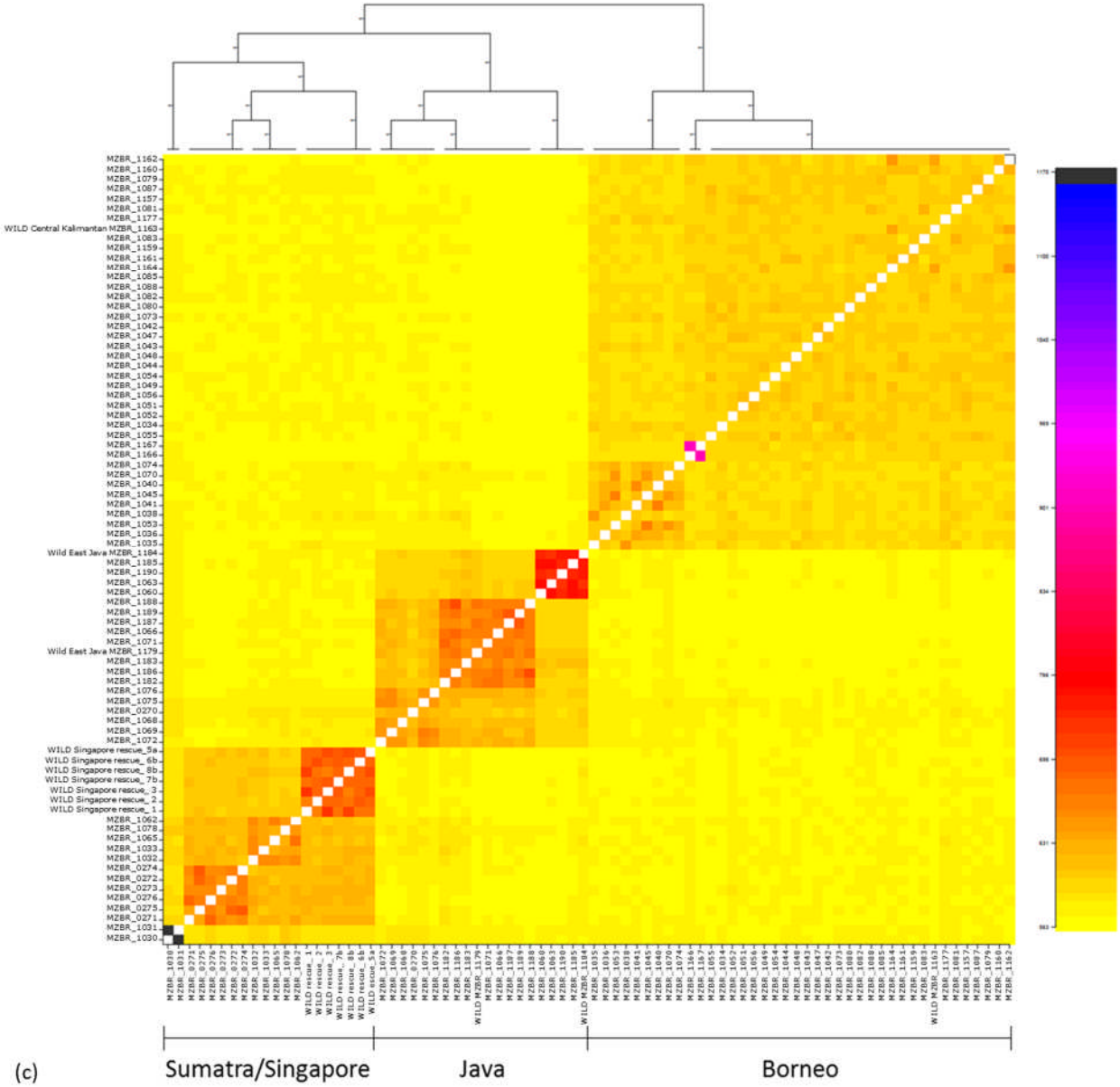

Medan, North Sumatra (6)
Singapore (7)
Sarawak, Malaysia (1)
BORNEO
SUMATRA
Nyaru Menteng, Central Kalimantan (8)
Pangkalanbun, Central Kalimantan (8)
Pangkalanbun, Central Kalimantan (1)
Lampung, southern Sumatra (1)
Tegal Alur, Jakarta (4)
Bogor, West Java (27)
Sukabumi, West Java (1)
JAVA
Surabaya, East Java (9)
Jember, East Java (8)
Jember, East Java (2)
N
Illegal wildlife trade seizures
Wild pangolin samples
0
250
500
750
1000 km


MZBR 1163 Pangkalanbun
Central Kalimantan
Sarawak, Malaysian Borneo
Borneo (2 WILD)
Sumatra / Singapore
Sumatra
Singapore (7 WILD)
Java (2 WILD)
MZBR 1179
MZBR 1184
Both Jember, East Java
eigenvector 1
eigenvector 2
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.2
-0.1
0.0
0.1
0.2

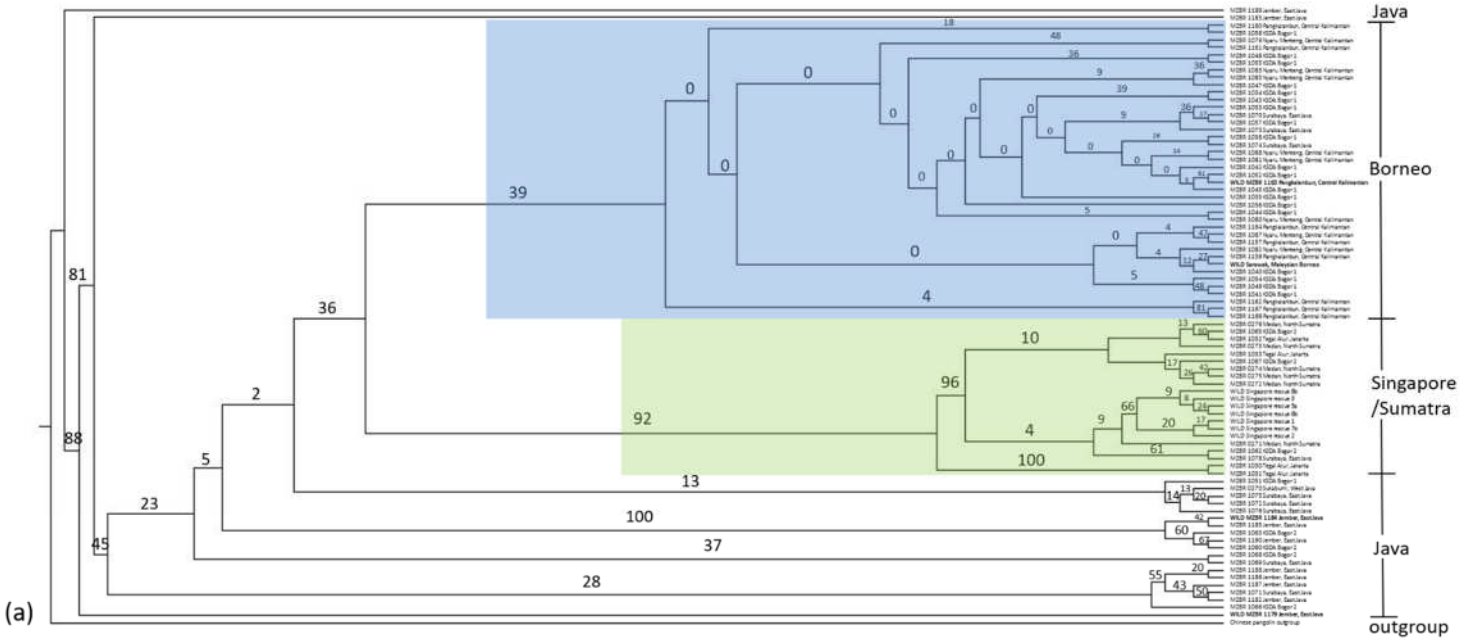
(a)

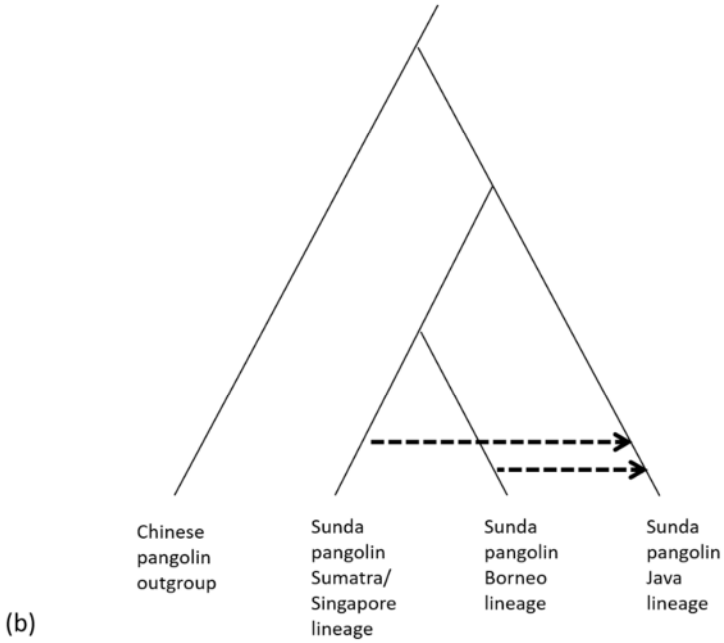

Borneo
Java
Sumatra/Singapore

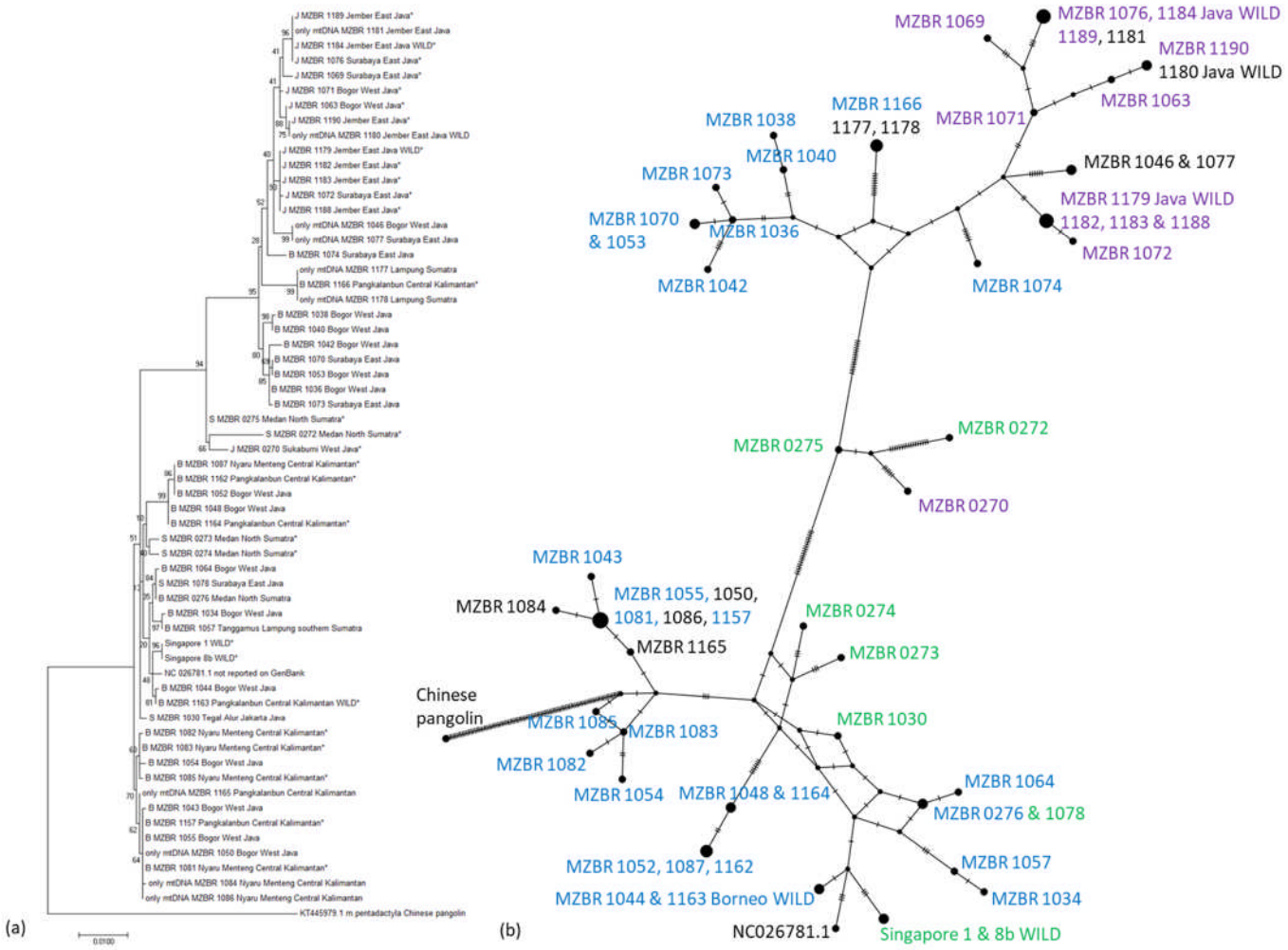
(b)

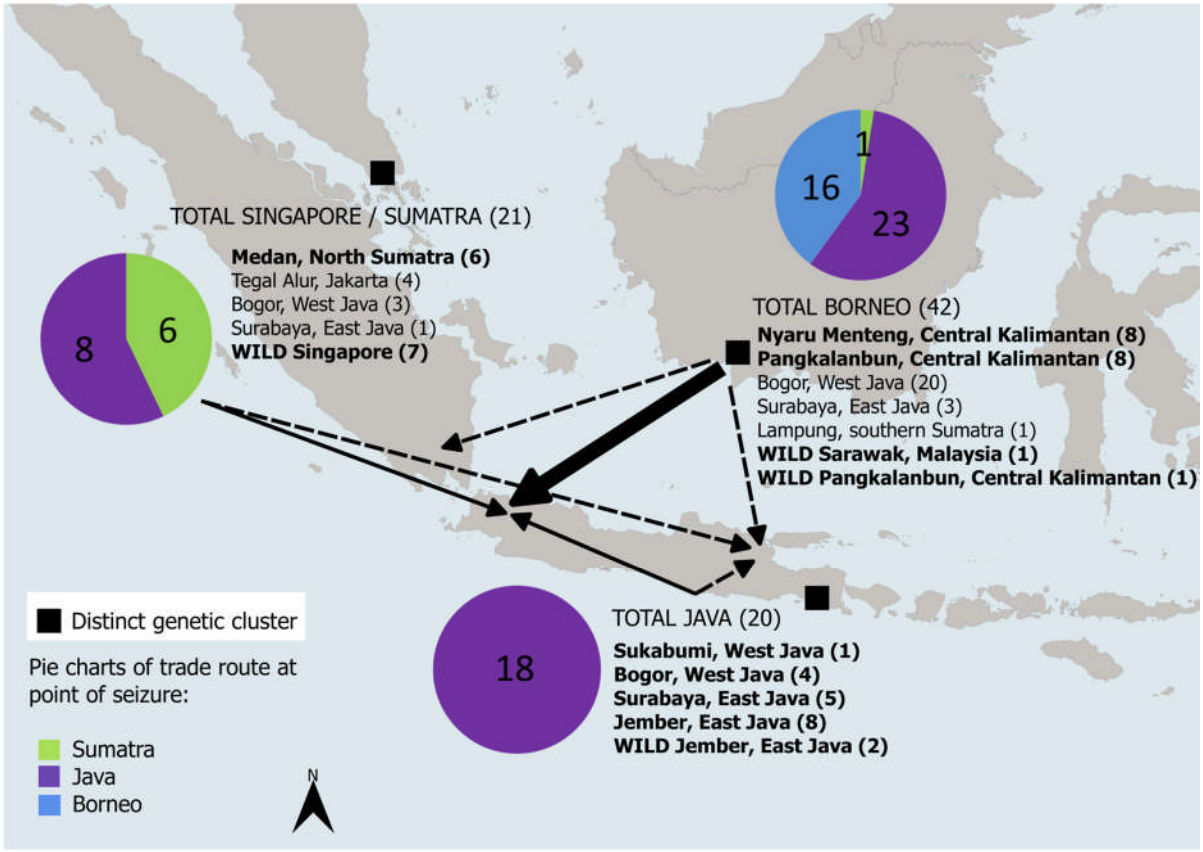

WILD Central Kalimantan MZBR_1163
Wild East Java MZBR_1184
Wild East Java MZBR_1179
WILD Singapore rescue_5a
WILD Singapore rescue_ 6b
WILD Singapore rescue_ 8b
WILD Singapore rescue_ 7b
WILD Singapore rescue_ 3
WILD Singapore rescue_ 2
WILD Singapore rescue_ 1
estimated coancestry (RAD loci)
Sumatra/Singapore
Java
Borneo

(c)


Java
Borneo
Singapore
/Sumatra
Java
outgroup
(a)


Chinese
pangolin
outgroup
Sunda
pangolin
Sumatra/
Singapore
lineage
Sunda
pangolin
Borneo
lineage
Sunda
pangolin
Java
lineage
(b)


(a)
J MZBR 1189 Jember East Java*
only mtDNA MZBR 1181 Jember East Java
J MZBR 1184 Jember East Java WILD*
J MZBR 1076 Surabaya East Java*
J MZBR 1069 Surabaya East Java*
J MZBR 1071 Bogor West Java*
J MZBR 1063 Bogor West Java*
J MZBR 1190 Jember East Java*
only mtDNA MZBR 1180 Jember East Java WILD
J MZBR 1179 Jember East Java WILD*
J MZBR 1182 Jember East Java*
J MZBR 1183 Jember East Java*
J MZBR 1072 Surabaya East Java*
J MZBR 1188 Jember East Java*
only mtDNA MZBR 1046 Bogor West Java
only mtDNA MZBR 1077 Surabaya East Java
B MZBR 1074 Surabaya East Java
only mtDNA MZBR 1177 Lampung Sumatra
B MZBR 1166 Pangkalanbun Central Kalimantan*
only mtDNA MZBR 1178 Lampung Sumatra
B MZBR 1038 Bogor West Java
B MZBR 1040 Bogor West Java
B MZBR 1042 Bogor West Java
B MZBR 1070 Surabaya East Java
B MZBR 1053 Bogor West Java
B MZBR 1036 Bogor West Java
B MZBR 1073 Surabaya East Java
S MZBR 0275 Medan North Sumatra*
S MZBR 0272 Medan North Sumatra*
J MZBR 0270 Sukabumi West Java*
B MZBR 1087 Nyaru Menteng Central Kalimantan*
B MZBR 1162 Pangkalanbun Central Kalimantan*
B MZBR 1052 Bogor West Java
B MZBR 1048 Bogor West Java
B MZBR 1164 Pangkalanbun Central Kalimantan*
S MZBR 0273 Medan North Sumatra*
S MZBR 0274 Medan North Sumatra*
B MZBR 1064 Bogor West Java
S MZBR 1078 Surabaya East Java
B MZBR 0276 Medan North Sumatra
B MZBR 1034 Bogor West Java
B MZBR 1057 Tanggamus Lampung southern Sumatra
Singapore 1 WILD*
Singapore 8b WILD*
NC 026781.1 not reported on GenBank
B MZBR 1044 Bogor West Java
B MZBR 1163 Pangkalanbun Central Kalimantan WILD*
S MZBR 1030 Tegal Alur Jakarta Java
B MZBR 1082 Nyaru Menteng Central Kalimantan*
B MZBR 1083 Nyaru Menteng Central Kalimantan*
B MZBR 1054 Bogor West Java
B MZBR 1085 Nyaru Menteng Central Kalimantan*
only mtDNA MZBR 1165 Pangkalanbun Central Kalimantan
B MZBR 1043 Bogor West Java
B MZBR 1157 Pangkalanbun Central Kalimantan*
B MZBR 1055 Bogor West Java
only mtDNA MZBR 1050 Bogor West Java
B MZBR 1081 Nyaru Menteng Central Kalimantan*
only mtDNA MZBR 1084 Nyaru Menteng Central Kalimantan
only mtDNA MZBR 1086 Nyaru Menteng Central Kalimantan
KT445979.1 m pentadactyla Chinese pangolin
0.0100
(b)
MZBR 1069
MZBR 1076, 1184 Java WILD
1189, 1181
MZBR 1190
1180 Java WILD
MZBR 1063
MZBR 1071
MZBR 1166
1177, 1178
MZBR 1038
MZBR 1040
MZBR 1073
MZBR 1070
& 1053
MZBR 1036
MZBR 1042
MZBR 1046 & 1077
MZBR 1179 Java WILD
1182, 1183 & 1188
MZBR 1072
MZBR 1074
MZBR 0275
MZBR 0272
MZBR 0270
MZBR 1043
MZBR 1084
MZBR 1055, 1050,
1081, 1086, 1157
MZBR 1165
MZBR 0274
MZBR 0273
Chinese
pangolin
MZBR 1085
MZBR 1083
MZBR 1082
MZBR 1054
MZBR 1030
MZBR 1048 & 1164
MZBR 1064
MZBR 0276 & 1078
MZBR 1052, 1087, 1162
MZBR 1044 & 1163 Borneo WILD
NC026781.1
MZBR 1057
MZBR 1034
Singapore 1 & 8b WILD


TOTAL SINGAPORE / SUMATRA (21)
Medan, North Sumatra (6)
Tegal Alur, Jakarta (4)
Bogor, West Java (3)
Surabaya, East Java (1)
WILD Singapore (7)
8
6
1
16
23
TOTAL BORNEO (42)
Nyaru Menteng, Central Kalimantan (8)
Pangkalanbun, Central Kalimantan (8)
Bogor, West Java (20)
Surabaya, East Java (3)
Lampung, southern Sumatra (1)
WILD Sarawak, Malaysia (1)
WILD Pangkalanbun, Central Kalimantan (1)
Distinct genetic cluster
Pie charts of trade route at point of seizure:
Sumatra
Java
Borneo
N
18
TOTAL JAVA (20)
Sukabumi, West Java (1)
Bogor, West Java (4)
Surabaya, East Java (5)
Jember, East Java (8)
WILD Jember, East Java (2)