

ARTICLE

Divergence and introgression in small apes, the genus *Hylobates*, revealed by reduced representation sequencingKazunari Matsudaira ^{1,2}✉ and Takafumi Ishida^{1,2}

© The Author(s), under exclusive licence to The Genetics Society 2021

Gibbons of the genus *Hylobates*, which inhabit Southeast Asia, show great diversity and comprise seven to nine species. Natural hybridisation has been observed in several species contact zones, but the history and extent of hybridisation and introgression in possibly historical and the current contact zones remain unclear. To uncover *Hylobates* species phylogeny and the extent of introgression in their evolution, genotyping by random amplicon sequencing-direct (GRAS-Di) was applied to 47 gibbons, representing seven *Hylobates* species/subspecies and two outgroup gibbon species. Over 200,000 autosomal single-nucleotide variant sites were identified. The autosomal phylogeny supported that divergence from the mainland species began ~3.5 million years ago, and subsequently occurred among the Sundaic island species. Significant introgression signals were detected between *H. lar* and *H. pileatus*, *H. lar* and *H. agilis* and *H. albibarbis* and *H. muelleri*, which all are parapatric and form ongoing hybrid zones. Furthermore, the introgression signals were detected in every analysed individual of these species, indicating a relatively long history of hybridisation, which might have affected the entire gene pool. By contrast, signals of introgression were either not detected or doubtful in other species pairs living on different islands, indicating the rarity of hybridisation and introgression, even though the Sundaic islands were connected during the Pliocene and Pleistocene glacial events.

Heredity (2021) 127:312–322; <https://doi.org/10.1038/s41437-021-00452-7>

INTRODUCTION

Primates are a highly diversified mammalian order, with ~500 (and at least of 350) extant species (Groves 2001; Rylands and Mittermeier 2014). Recently, genome-wide studies of non-human primates have become possible, revealing evolutionary processes (e.g. speciation) and demographic histories of various taxa (Prado-Martinez et al. 2013; Xue et al. 2016; Nater et al. 2017). Historical hybridisation and introgression have been described in several studies (e.g. chimpanzees and bonobos (de Manuel et al. 2016), macaques (Fan et al. 2018), baboons (Rogers et al. 2019) and Dryas and green monkeys (van der Valk et al. 2020)). Their potential roles in non-human primate evolution (e.g. adaptive introgression) have been widely discussed (Nye et al. 2018; Rogers et al. 2019; van der Valk et al. 2020). Among primates, gibbons are considered a good model for exploring the contribution of hybridisation and introgression to evolution, especially as an analogy for hominin evolution in terms of the number of taxa (up to 20 species) and the evolutionary timeframe (divergence dated in the late Miocene) (Zichello 2018). Moreover, gibbons have typically monogamous reproductive patterns (Oka and Takenaka 2001; Barelli et al. 2013) that may have resulted in similar demographic consequences of hybridisation. In addition, several cases of ongoing natural hybridisation between gibbon species have been observed (Brockelman and Gittins 1984; Marshall and Sugardjito 1986), which is a great advantage for studying hybridisation mechanisms and conditions in apes.

Gibbons that inhabit the Indomalayan realm exhibit great diversity among four extant genera, each having different chromosome numbers (*Hoolock* [$2n = 38$], *Hylobates* [$2n = 44$], *Symphalangus* [$2n = 50$] and *Nomascus* [$2n = 52$]) and comprising 15–20 species (Roos 2016; Fan et al. 2017). Among these, *Hylobates* is distributed throughout the mainland and islands of Southeast Asia. Seven allopatric species with various pelage colours and vocalisations are recognised: *Hylobates pileatus*, *H. lar*, *H. agilis*, *H. albibarbis*, *H. muelleri*, *H. moloch* and *H. klossii* (Fig. 1; Groves 2001). In addition, some studies hypothesised that the three subspecies of *H. muelleri* (i.e. *H. m. muelleri*, *H. m. abbotti* and *H. m. funereus*) are independent species, based on the estimated divergence time (~1.4–1.8 million years ago [MYA]) of their cytochrome b (cytb) sequences of mitochondrial DNA (mtDNA) (Thinh et al. 2010a; Roos 2016). A whole mitochondrial genome (mitogenome) study resolved the mtDNA phylogeny of six of the seven *Hylobates* species (except *H. albibarbis*, which was not included in the analysis), demonstrating that *H. pileatus* diverged first and *H. lar* diverged second, followed by the divergence between proto-*H. agilis*/*H. muelleri* and proto-*H. moloch*/*H. klossii*, *H. agilis* and *H. muelleri* and *H. moloch* and *H. klossii* in the Pliocene (3.5–2.2 MYA) (Chan et al. 2010 and its correction). These results supported an evolutionary scenario where the common ancestor of *Hylobates* originated on the mainland and the distribution later expanded to Southeast Asian islands (Groves 1972; Chivers 1977; Whittaker et al. 2007). Contrary to mitogenome studies, nuclear DNA markers have yielded ambiguous results on species

¹Department of Biological Sciences, School of Science, The University of Tokyo, Tokyo, Japan. ²Present address: Division of Southeast Asian Area Studies, Graduate School of Asian and African Area Studies, Kyoto University, Kyoto, Japan. Associate editor: Xiangjiang Zhan ✉email: kaz.matsudaira@gmail.com

Received: 1 June 2020 Revised: 15 June 2021 Accepted: 15 June 2021
Published online: 29 June 2021

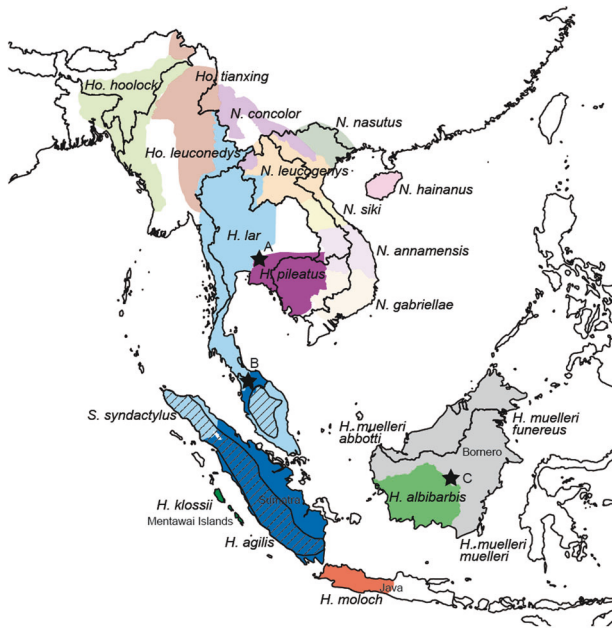


Fig. 1 Distribution of extant gibbons. Adapted from Thinh et al. (2010a, b); Fan et al. (2017). Stars indicate the hybrid/contact zones. A: Khao Yai National Park, B: Muda River and its tributary, C: headwaters of Barito and Kapuas Rivers (Brockelman and Gittins 1984).

phylogenies (Kim et al. 2011; Chan et al. 2013), likely due to the limited number of markers (~20 loci), which could not overcome the influence of incomplete lineage sorting and possible introgression. Since mtDNA phylogenies are not always consistent with species phylogenies, as has been demonstrated in, e.g. baboons (Zinner et al. 2013) and macaques (Fan et al. 2018; Osada et al. 2021), a genome-wide study that resolves nuclear genome phylogeny is warranted to understand the evolutionary history of *Hylobates*.

Ongoing natural hybridisation has been observed in several species contact zones: between *H. lar* and *H. pileatus* at Khao Yai National Park in Thailand, *H. lar* and *H. agilis* at the Muda River and its tributary near the Thai-Malaysian border and *H. albibarbis* and *H. muelleri* at the headwaters of the Barito and Kapuas Rivers in central Borneo (Fig. 1; Brockelman and Gittins 1984; Marshall and Sugardjito 1986). This observation suggests that hybridisation and consequent introgression may have affected the genetic diversity of *Hylobates* and its evolutionary history. However, the historical aspects of the frequency of contacts and the extent of introgression in the ongoing hybrid zones remain unclear. The contact zones may have existed for a long time and were probably broader than the current zones (e.g. Geissmann 1991), and introgression might have contributed to the entire gene pool of the species. Alternatively, the contact zones may have appeared only recently or were limited to small areas similar to those today. Also, introgression may have affected only a few individuals living in the contact zones or a nearby local gene pool.

Furthermore, although the current distributions of many *Hylobates* species are separated by seas, the historical distribution of these species may have been connected during the Pliocene and Pleistocene glacial periods when the sea level was low, and Sundaland emerged (Cannon et al. 2009; Woodruff 2010). Such historical distribution range connections may have resulted in hybridisation and introgression between species. Conversely, genetic studies on some forest-dependent mammals and birds in the Sunda region have uncovered deep divergence in the Miocene–Pliocene with limited gene flow/introgression during the Pleistocene glacial events (e.g. murine rodents (Gorog et al. 2004);

colugos, lesser mouse deer and pangolins (Mason et al. 2019); babblers and bulbuls (Lim et al. 2017; Cros et al. 2020)), following the presence of fragmented forests on the emerged Sundaland under arid climate conditions (Heaney 1991; Wurster et al. 2019). Since gibbons strongly depend on forests, introgression between *Hylobates* species in the Sundaic islands may have also been limited. In a previous study, historical introgression between some *Hylobates* species was detected between relatively unusual combinations (e.g. between *H. lar* in mainland Southeast Asia and northern Sumatra and *H. moloch* on Java) (Chan et al. 2013). This result might be due to the limited number of markers used; therefore, further studies have been required.

In this study, we aimed to uncover (1) the species phylogeny of *Hylobates* and (2) the whole picture of historical introgression between *Hylobates* species pairs showing ongoing hybridisation or possible past hybridisation events by applying a genome-wide approach to multiple individuals of the *Hylobates* species. For genome-wide analyses, we applied a recently developed method termed genotyping by random amplicon sequencing-direct (GRAS-Di) that uses polymerase chain reaction (PCR) and a set of primers to amplify several tens of thousands of loci, by randomly annealing the primers to genomes (Enoki and Takeuchi 2018). This method has been successfully applied to various fish species and has detected several thousands of single-nucleotide variants (SNVs) in each species (Hosoya et al. 2019). In addition, about a hundred SNVs were identified in leopard cats (Ito et al. 2020). Some other reduced representation sequencing methods, such as restriction-site-associated DNA sequencing (RAD-seq) and genotype by sequencing have been applied in other primate taxa (Bergey et al. 2013; Scally et al. 2013; Baiz et al. 2019) and were successfully obtained from ~6000 SNV sites in a paired-end RAD-seq study of marmosets (Malukiewicz et al. 2017) to ~5.8-million (M) SNV sites in a double-digest RAD-seq study of howler monkeys (Boubli et al. 2019). Thus, we have expected that GRAS-Di would also yield enough SNV markers to reveal the species phylogeny and introgression among *Hylobates* species.

MATERIALS AND METHODS

Samples

DNA was extracted from 46 Epstein–Barr virus-transformed gibbon B lymphoblastoid cells and one gibbon muscle tissue sample, using the QIAamp DNA Mini Kit (Qiagen, Germany). The DNA samples comprised *H. agilis* ($n = 2$), *H. albibarbis* ($n = 3$), *H. lar* ($n = 20$), *H. moloch* ($n = 1$), *H. m. abboti* ($n = 2$), *H. m. muelleri* ($n = 2$), *H. pileatus* ($n = 10$), *Nomascus leucogenys* ($n = 1$) and *Symphalangus syndactylus* ($n = 6$) (Table S1). The samples were derived from captive zoo gibbons, as described in previous studies (Ishida et al. 1985; Nakayama and Ishida 2006).

At the time of sample collection, the species status and key external characteristics of some individuals were not well-recorded. Hence, we confirmed the maternal species status of all 47 gibbons by comparing partial mtDNA sequences with publicly available data (Supplementary Materials, Table S3). In addition, to minimise the inclusion of unidentifiable hybrid individuals born in captivity, we evaluated multiple individuals from each species/subspecies in the GRAS-Di analysis and examined if any individual had autosomal genetic characteristics that were largely deviated from the other individuals of the same species/subspecies, when possible. No individual showed explicit genetic evidence of hybridisation in captivity. In other words, no individual showed expected genetic characteristics of filial one hybrid nor early backcross generations (see ‘Results’, viz. the phylogenetic network). Thus, we considered that the unintentional effect of hybridisation in captivity was negligible among the 47 samples.

Whole-mitogenome sequencing and analysis

We determined and analysed the whole-mitogenome sequences of 6 of the 47 gibbons (Table S1) to assess the mtDNA phylogeny and divergence time of *Hylobates*, especially regarding the phylogenetic position of *H. albibarbis*. Sanger sequencing was performed for two overlapped

fragments that were PCR-amplified using primers described by Finstermeier et al. (2013) (Supplementary Materials, Table S2). The six whole-mitogenome sequences were combined with the published whole-mitogenome sequences of 44 gibbons (which included one *S. syndactylus* in the GRAS-Di data set [Sample ID: G75]) (Arnason et al. 1996; Chan et al. 2010; Matsudaira and Ishida 2010; Finstermeier et al. 2013; Fan et al. 2017) and eight outgroup primate species: *Homo sapiens*, *Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, *Pongo pygmaeus*, *Pongo abelli*, *Macaca mulatta* and *Papio anubis* (Horai et al. 1995; Xu and Arnason 1996; Zinner et al. 2013; Liedigk et al. 2014) (Table S4). After aligning the sequences using MUSCLE (Edgar 2004), which was implemented in MEGA 6 (Tamura et al. 2013), we extracted the sequences of 13 protein-coding genes and two rRNA-coding regions, which resulted in 13,798-bp sequences. We excluded the D-loop region and 22 tRNA-coding regions from the data set as per previous studies (Chan et al. 2010; Matsudaira and Ishida 2010; Fan et al. 2017), mainly due to the difficulty obtaining good alignment across the diverged primate taxa (Pozzi et al. 2014).

The maximum likelihood (ML) tree was constructed using IQ-TREE 1.6.10 (Nguyen et al. 2015). Among the 15 identified genes, seven partitions with different substitution models were selected by ModelFinder (Kalyaanamoorthy et al. 2017) (Table S5), implemented in IQ-TREE. An ultrafast bootstrap approximation test (Minh et al. 2013) was conducted, for 1000 iterations.

Divergence time estimation was conducted using BEAST 2.5.2 (Bouckaert et al. 2014). The relaxed, lognormal clock model and birth–death model tree priors were selected. The partition scheme and substitution models used in the ML tree inference were also applied. Assuming normal distribution, three divergence time priors were used: (1) hominoids and cercopithecoids, as 30.5 MYA (95% lower and upper limit: 25.0–36.0) (Stevens et al. 2013; Bond et al. 2015); (2) *Pongo* and *Homo/Pan/Gorilla*, as 15.5 MYA (13.0–18.0) (Kelley 2002; Besenbacher et al. 2019) and (3) *Homo* and *Pan*, as 6.5 MYA (6.0–7.0) (Senut et al. 2001; Brunet et al. 2002). Four independent Markov chain Monte Carlo (MCMC) runs were performed for 25,000,000 generations, and parameters were recorded every 1000 generations. The convergence of the results was assessed by Tracer 1.7.1. (Rambaut et al. 2018) by comparing the posterior distribution of each parameter across the four independent runs. The first 25% of samples from each run were discarded as burn-in, and the remaining results were merged, using LogCombiner in BEAST, resulting in an effective sampling size >500 for each parameter. A consensus tree was constructed from 75,004 trees, using TreeAnnotator in BEAST, and visualised using FigTree v1.4.4 (<https://github.com/rambaut/figtree/releases>).

GRAS-Di

RNA in the extracted DNA samples was digested by RNase A (Nippon Gene, Japan) for 1 h. Then, ethanol precipitation was performed, and the DNA was eluted using Low-EDTA TE (Thermo Scientific, USA). Library preparation and sequencing using the GRAS-Di method were outsourced to GeneBay, Inc. (Japan). The library preparation protocols were described by Hosoya et al. (2019), except that the PCR reaction volume was 25 μ l in the first PCR run and 50 μ l in the second PCR run, instead of 10 μ l as previously described. In the first PCR run, 1 μ l (15 ng/ μ l) of DNA was used as the template, and in the second PCR run, 1 μ l of the first PCR product was used as the template (Ito et al. 2020). The sequences of the 64 primers used to obtain amplicons during the first PCR run were as per Hosoya et al. (2019). The library was sequenced on a HiSeq 4000 sequencer (Illumina, USA) with 151-bp paired-end reads.

Read mapping and genotyping

Sequence reads were de-multiplexed to each individual and provided as FASTQ files. We removed 3' adapter sequences and low-quality reads using Cutadapt 1.12 (Martin 2011), with the following command: (cutadapt -a CTGTCTTATACATCTCCGAGCCACGAGAC-A CTGTCTTATACATCTGACGCTGCCGACGA -overlap 10 -minimum-length 51 -quality-cutoff 20). Cleaned reads were mapped to the reference genome sequence of *N. leucogenys*, nomLeu3 (Carbone et al. 2014), using the Burrows–Wheeler Aligner (BWA-MEM) 0.7.17 (Li 2013). The generated SAM files were converted into BAM files and sorted using the SortSam command of the Genome Analysis Toolkit (GATK) 4.1.0.0 (McKenna et al. 2010). After adding read group information using GATK AddOrReplaceReadGroups, each BAM file was indexed by SAMtools 1.9 (Li et al. 2009). To check the effect of karyotype difference among the four genera to the mapping process, the number of mapped reads was counted using Qualimap 2.2.1 (Okonechnikov et al. 2016), and the ratio of mapped reads to all adapter-trimmed

cleaned reads was calculated. Also, the numbers of mapped reads to each nucleotide position were counted using SAMtools. To call variants in each individual, GATK HaplotypeCaller with -emit-ref-confidence and GVCF option, was used. The resulting gVCF files for the 47 gibbons were merged using GATK CombineGVCFs. GATK GenotypeGVCFs, with the -all-sites option, was used for joint genotyping variants among the 47 gibbons. The fraction of the genome and of each chromosome mapped by all the 47 samples was calculated using BCftools 1.9 (Li et al. 2009) and VCFtools 0.1.17 (Danecek et al. 2011).

SNV sites were extracted using GATK SelectVariants. Then, GATK hard filters were applied to remove ambiguous results (QD < 2.0; FS > 60.0; MQ < 40.0; MQRankSum < -12.5; ReadPosRankSum < -8.0). Subsequently, we extracted SNV sites that satisfied the following criteria: (1) genotyped in all 47 gibbons, with at least 10 \times depth of coverage for each gibbon, (2) biallelic (i.e. no more than two alleles) and (3) variable among the 47 gibbons. Then, we filtered out SNV sites that deviated from Hardy–Weinberg equilibrium using chi-squared tests with a *p* value threshold of <0.001. The tests were conducted using PLINK 1.9 (Chang et al. 2015). This filtering based on Hardy–Weinberg equilibrium was performed especially for removing sites with mapping errors at paralogous or repeated loci. We mapped the short reads of multiple species/subspecies to the *N. leucogenys* reference genome. If such loci were absent in the reference, short reads from multiple loci in the samples would be mapped to one locus, and cause an excess of heterozygosity. All detected sites showed excess of heterozygosity, which was potentially derived from such mapping errors. A deficiency of heterozygosity, probably reflecting the presence of population structure among the samples, was not detected. The chi-squared tests were conducted on *H. lar* because only for *H. lar*, sufficient samples were available for the tests (*n* = 20). Last, we only extracted SNV sites located on the 25 autosomal long scaffolds of nomLeu3. We used GATK VariantFiltration, VCFtools and vcfliib ver1.0 (<https://github.com/vcfliib/vcfliib>) to perform the filtering. The resulting data set comprised of 232,261 SNV sites.

We evaluated the genetic diversity of each species/subspecies by calculating heterozygosity across the 232,251 SNV sites for each individual using BCftools and nucleotide diversity (Nei and Li 1979) of each species/subspecies using VCFtools.

Species tree and genetic divergence

To uncover the phylogeny of the nuclear genomes, we constructed a phylogenetic tree of the autosomal SNVs in the ML framework using IQ-TREE. Since the data set only comprised SNV sites, we used an ascertainment bias correction (ASC) model (Lewis 2001) for the ML tree construction. In the ASC model, heterozygous sites (and nonvariant sites, excluded by the filter above) were ignored. Thus, 127,151 of the 232,261 SNV sites were used for the analyses. ModelFinder selected the K3Pu + F + ASC + R4 model, and an ultrafast bootstrap approximation test was performed with 1000 iterations.

We also estimated the divergence time by autosomal SNVs based on a coalescence model using SNAPP 1.4.2 (Bryant et al. 2012), which was implemented in BEAST with a model known to be functional for concatenated SNV data with a small number of individuals per taxa (Stange et al. 2018). The absence of a linkage between SNV sites, a strict molecular clock, and the same effective population size for every branch in the analysed tree were assumed in the model. Our original SNV data set was too large to compute in SNAPP, and a few thousand sites (~3000 sites) were appropriate for estimating divergence time with high precision in a similar divergence timeframe and number of species (Stange et al. 2018). Therefore, we selected a subset of the original data set. Two samples (or one sample if two were not available) per species or subspecies were extracted. Then, variable SNV sites at least 500,000 bases separated from each other were extracted, which generated 3980 SNV sites among 16 samples of nine species/subspecies. The input file for SNAPP was generated using snapp_prep.rb (https://github.com/mmatschiner/snapp_prep). In the analysis, a time constraint was set for the common ancestor of Hylobatidae estimated by the mitogenome analysis in this study (i.e. a normal distribution with mean = 8.12 MYA and sigma = 0.75). Two independent MCMC runs were performed for 1000,000 generations, and parameters were recorded once every 500 generations. The convergence of the results was assessed using Tracer. The first 25% of samples from each run were discarded as burn-in, and the remaining results were merged, using LogCombiner. A consensus tree was constructed from 3002 trees using TreeAnnotator, and was visualised using FigTree.

To evaluate the genetic divergence between species/subspecies, d_{xy} (Smith and Kronfrost 2013) was calculated using popgenWindows.py (https://github.com/simonhmartin/genomics_general). Mean d_{xy} was calculated across regions with a window size of five million bases (Mb) with at least 100 SNV sites per window.

Individuals of the same species or subspecies may have considerably different genetic characteristics, for example, due to the difference in the extent of introgression from other species. To reveal genetic relationships among individuals, we calculated the allele-sharing distances (ASD) (Bowcock et al. 1994) between each pair of individuals using asd v1.0.0 (<https://github.com/szpiech/asd>). Based on the ASD matrix, multi-dimensional scaling (MDS) plots (two-dimensional plots) were drawn, using R 3.6.2 (R Core Team 2019). Individual relationships were also assessed by a phylogenetic network of autosomal SNVs generated by the NeighborNet algorithm in SplitsTree 4.15.1 (Huson and Bryant 2006). To elucidate the results, we included only *Hylobates* gibbons in this analysis.

Detection of introgression

Several approaches were used to assess the presence of introgression among *Hylobates*. Patterson's D-statistic tests (also known as ABBA-BABA tests) (Green et al. 2010; Patterson et al. 2012) were used to detect introgression. Since phylogenetic relationships among *Hylobates* species/subspecies have been inferred from both mitogenomes and autosomal SNVs and are well-supported and consistent (except for one mitogenome from *H. agilis*; see 'Results'), we used the autosomal tree topology and calculated D-statistics for all 35 possible combinations of *Hylobates* species by setting the outgroup as either *N. leucogenys* or *S. syndactylus*. In other words, the D-statistics were calculated in the form of ((P1, P2), P3), outgroup) and three of the seven *Hylobates* species/subspecies from the present study were placed in P1, P2 and P3 following the species phylogeny. The D-statistics were calculated for the population data combinations for species/subspecies and the combinations of one individual per species/subspecies. The D-statistics were calculated using qpDstat in AdmixTools v5.1 (Patterson et al. 2012). Standard errors, Z-scores and corresponding p values of D-statistics were calculated through the weighted block jackknife approach with a block size of 5 Mb. There is no consensus in the threshold of significance for the D-statistics, and the related f_4 -statistics deviated from zero (e.g. Reich et al. 2009; Cahill et al. 2015; Chattopadhyay et al. 2019). We applied the Bonferroni correction for multiple comparisons and set the alpha level of rejecting the null hypothesis to 0.001. In AdmixTools, using its D-statistics formula, negative D-statistics significantly deviating from zero indicate the introgression between P2 and P3, and positive D-statistics significantly deviating from zero indicate the introgression between P1 and P3.

The presence of introgression was also tested by making ML graphs allowing introgression between taxa using TreeMix 1.13 (Pickrell and Pritchard 2012). To take the linkage of SNV sites into account, a block size of 1000 SNVs was used for analyses. Graphs with zero to five introgression edge(s) were analysed. We assessed the goodness of fit of the graphs by calculating the degree of explained variance with each number of introgression edges.

To infer the extent of introgression between species/subspecies, we made admixture graphs using qpGraph in AdmixTools. In qpGraph, f_2 , f_3 and f_4 statistics were calculated species/subspecies combinations based on a given graph (Patterson et al. 2012). Starting from a graph following the topology supported by the ML phylogenetic tree, the graph was revised by connecting nodes until the significant f_4 -statistics deviations from zero were resolved. Standard errors and Z-scores were calculated with the weighted jackknife with 5-Mb blocks. Here, an absolute Z-score more than 3 (corresponding to the p value of ~ 0.00135) was considered a significant deviation because it was difficult to assess the number of f_4 statistic required for the Bonferroni corrections. For this analysis, we used *N. leucogenys* as an outgroup.

When an introgression event occurs, most genomic regions from other species are expected to be removed, due to incompatibility and deleteriousness (Fu et al. 2016; but see Petr et al. 2019). However, some regions may remain as introgressed loci due to adaptive advantage or stochasticity. Potentially introgressed regions can be detected by comparing D-statistics and genetic divergence in each genome segment (Smith and Kronfrost 2013). Since introgressed regions should have shorter coalescent times than average, regions with high absolute D-statistics and low genetic divergence can be considered candidates for introgressed regions. To detect potentially introgressed loci among *Hylobates* species/subspecies, we calculated the corrected D-statistics, f_{dr} (Martin et al. 2015), and genetic divergence d_{xy} , using ABBABABAWindows.

py (https://github.com/simonhmartin/genomics_general) and popgenWindows.py, respectively. Note that by the definition and formula of f_{dr} , the sign of f_{dr} is opposite to the sign of the D-statistics calculated by AdmixTools, and only the degree of introgression between P2 and P3 (and not between P1 and P3) in the order of (((P1, P2), P3), outgroup) is testable (Martin et al. 2015). We generated SNV data sets by selecting samples of only four species or subspecies before filtering SNV sites. This regeneration increased the number of SNV sites for each species quartet (303,744 [30.8% increased] to 507,416 [118.5% increased]). However, we calculated the values for regions with a large window size of 1 Mb with at least 100 SNV sites per window due to the limited number of SNV sites in our data sets.

RESULTS

Genotyping

We obtained a mean \pm standard deviation (S.D.) of 8.6 M \pm 0.87 (min = 6.6 M; max = 10.2 M) raw-read pairs, per sample. Adapter trimming and quality filtering by Cutadapt yielded 8.5 M \pm 0.85 (min = 6.5 M; max = 10.0 M) read pairs, per sample. The ratios of mapped reads to all adapter-trimmed reads per sample were 97.6–99.2%. The ratio of one *N. leucogenys* sample was 99.1%, and within the range of 30 *Hylobates* samples (97.6–99.2%), but slightly larger than the range of six *S. syndactylus* samples (97.6–98.6%) (Table S6). The difference might be derived from the differences in the chromosome positions among the four genera. In the nomLeu3 genome, 3.6–5.2% (95,030,378–139,035,756 sites) of the 2654,007,897 autosomal sites were mapped by at least one read in each sample, and these mapped sites represented 18.9% (502,527,591 sites) of the genome. Meanwhile, 0.6% (16,871,135 sites) was mapped by at least one read in every sample. In addition, 0.7–1.1% (19,292,914–28,944,712 sites) were mapped by at least ten reads in each sample, and 0.2% (5247,911 sites) was mapped by at least ten reads in every sample (the fraction of each chromosome is shown in Table S7). The mean depth of coverage at the mapping sites in each sample was 16.3 \times \pm 1.4.

After hard filtering by GATK, a data set containing 232,261 biallelic, autosomal SNV sites was generated that successfully genotyped in all 47 gibbons. We also generated SNV sets with filters of 5 \times and 15 \times minimum depth of coverage per sample, which resulted in 356,846 and 170,329 SNV sites, respectively. The results of further analyses were substantially similar by the minimum depth of coverage, but the 15 \times set did not apply to some analyses due to the smaller number of SNV sites (data not shown).

Across the 232,261 autosomal SNV sites, *H. pileatus* showed the lowest heterozygosity (mean = 0.0186 \pm 0.0021 S.D.). The three Bornean species/subspecies (*H. albibarbis*, *H. m. muelleri* and *H. m. abbotti*) showed the highest heterozygosity (mean = 0.0536–0.0628). The other three *Hylobates* species (*H. moloch*, *H. lar* and *H. agilis*), *S. syndactylus* and *N. leucogenys* were in the middle (mean = 0.0346–0.0452) (Fig. S1). The degree of nucleotide diversity showed the same tendency as the per sample heterozygosity (Table S8).

Phylogeny and divergence

Phylogenetic relationships among *Hylobates* were well resolved, both in the mitogenome and the autosomal SNV trees (Fig. 2), and the topologies of species/subspecies relationships were consistent, except for one mitogenome sequence from *H. agilis* (Sample ID: G80). This individual clustered with the other *H. agilis* individual in the autosomal SNV tree (Fig. 2B) but clustered with *H. lar* individuals and separated from other *H. agilis* individuals in the mitogenome tree (Fig. 2A; labelled as *H. agilis* [group 2]). In the cytb tree, which comprised four *H. lar* subspecies reference sequences, this *H. agilis* individual (G80) tightly clustered with the Sumatran subspecies sequence, *H. l. vestitus* (Fig. S2).

Both the autosomal and mitogenome phylogenetic trees demonstrated that the divergence of the *Hylobates* species started in the mainland species (*H. pileatus* and *H. lar*), followed by the

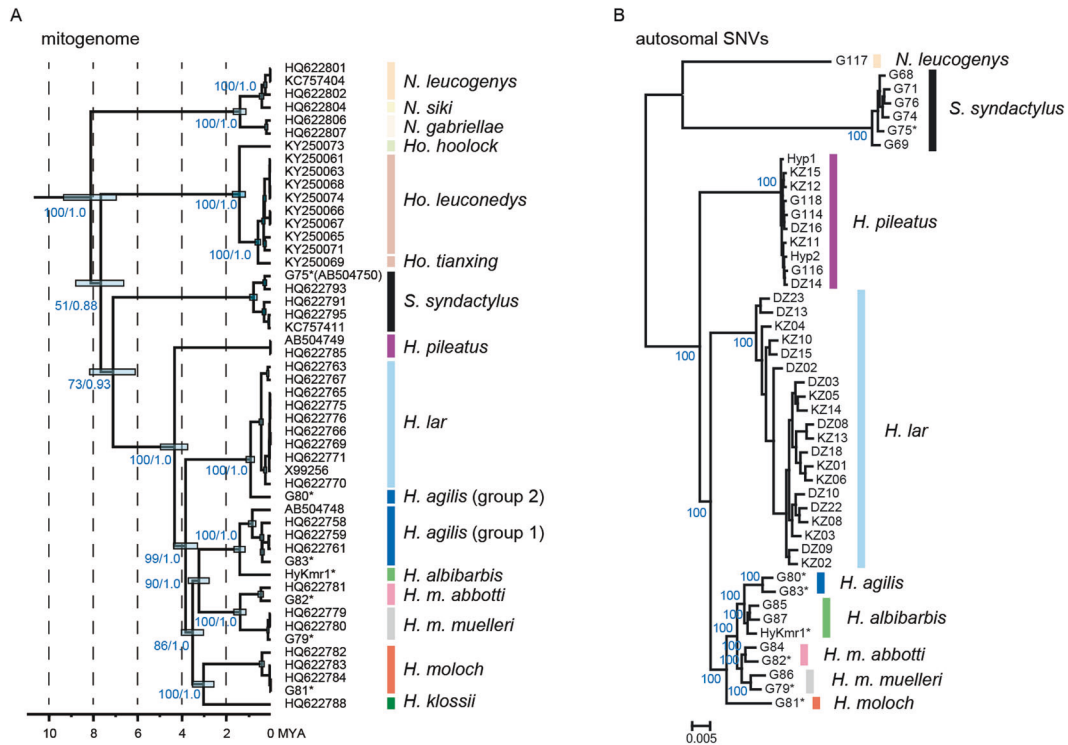


Fig. 2 Phylogenetic trees of gibbons. Values on the nodes (shown in blue) are bootstrap values for the ML analysis and the posterior probabilities for the Bayesian analysis. **A** Mitogenome phylogeny and divergence times, as estimated by Bayesian analysis, using BEAST 2. The outgroup species were omitted from this figure. The means and 95% highest posterior density credibility intervals of the divergence time are shown in Table S7. **B** Species phylogeny, based on the autosomal SNVs, produced by ML analysis using IQ-TREE. * indicates the individuals ($n = 7$) that were analysed both for mitogenomes and autosomal SNVs.

other Sundaic species/subspecies (Figs. 2, S3, and Tables S9, S10). Based on the estimated divergence time indicated by the autosomal SNVs, divergence among *Hylobates* first occurred between *H. pileatus* and the other *Hylobates* species, ~3.5 MYA (95% highest posterior density credibility interval [HPD CI]: 2.8–4.1). Next, *H. lar* diverged from the Sundaic species, ~2.6 MYA (2.1–3.2). Then, *H. moloch* diverged from the Sumatran/Bornean species, ~2.0 MYA (1.6–2.4). *H. muelleri* and *H. agilis*/*H. albibarbis* diverged ~1.3 MYA (1.0–1.6). *H. agilis* and *H. albibarbis* diverged ~1.0 MYA (0.8–1.2). *H. m. muelleri* and *H. m. abbotti* diverged ~0.9 MYA (0.7–1.1). The order of mitogenome divergence was consistent with that of autosomes; however, the estimated time was relatively older than those of the autosomes (Table S9).

Genetic divergence (d_{xy}) of the autosomes between the species/subspecies was consistent with the phylogenetic relationships and divergence time between species/subspecies shown above (Table 1). Among *Hylobates*, the highest d_{xy} was observed between *H. pileatus* and the other *Hylobates* species/subspecies (0.115–0.124) followed by between *H. lar* and the remaining *Hylobates* species/subspecies (0.104–0.108). In addition, d_{xy} between *H. m. muelleri* and *H. m. abbotti* (0.073) was comparable with that between *H. agilis* and *H. albibarbis* (0.070).

MDS plots of ASD showed genetic similarity in individuals within the same species/subspecies, and similar species/subspecies relationships with the phylogenetic tree analyses. In the MDS plot containing all *Hylobates* gibbons ($n = 40$), three clusters appeared. *H. pileatus* formed one cluster, *H. lar* formed another, and the Sundaic *Hylobates* species/subspecies formed the final cluster (Fig. 3A). In addition, among the Sundaic *Hylobates* species/subspecies, *H. moloch* separated from the others. When we plotted only the Sundaic *Hylobates* species/subspecies ($n = 10$), the species/subspecies were all separated from each other (Fig. 3B). *H. agilis*, *H. albibarbis* and *H. muelleri* (both *H. m. muelleri*

and *H. m. abbotti*) were plotted on a line, with *H. agilis* at one end and *H. muelleri* at the other end. *H. albibarbis* was located between the two species, closer to *H. agilis* than to *H. muelleri*. Furthermore, *H. m. muelleri* ($n = 2$) and *H. m. abbotti* ($n = 2$) were slightly separated from each other in the plot.

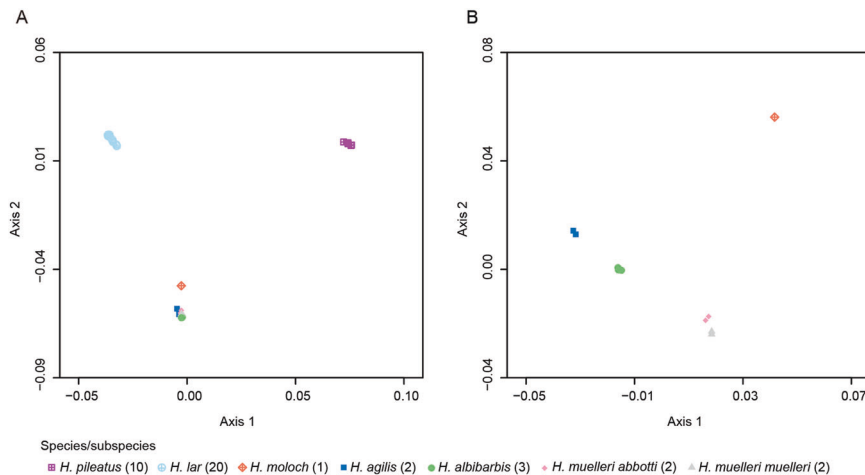
The phylogenetic network (Fig. S4) showed consistent relationships among the *Hylobates* species/subspecies, as shown in the phylogenetic trees and MDS plots of ASD. A clear reticulation, which indicates the potential of introgression, was observed in *H. albibarbis* and *H. muelleri* (both *H. m. muelleri* and *H. m. abbotti*). Contrary to the MDS plots of ASD, the two *H. agilis* individuals (Sample ID: G80 and G83) formed a reticulation and showed some genetic difference, which may be related to the difference in their mitogenomes (Fig. 2A). Furthermore, the network highlighted the population structure within *H. lar*, wherein 14 of 20 individuals were tightly clustered together. This result was comparable to the cyt b tree, wherein the same 14 individuals and one additional individual were clustered more tightly than the other five individuals (Fig. S1).

Introgression

The D-statistics calculated for the species quartets showed introgression signals between *H. albibarbis* and *H. muelleri* (both *H. m. muelleri* and *H. m. abbotti*), *H. lar* and *H. pileatus* and *H. lar* and *H. agilis* (Tables 2 and S11). These results were consistent, regardless of the combinations of quartets (i.e. regardless of P1 and the outgroup species) tested in the analysis. No other signals of introgression were detected among the other testable quartets of *Hylobates* species/subspecies. Therefore, introgression was detected among the species pairs that form ongoing natural hybrid/contact zones. The D-statistics calculated for individual quartets also showed introgression signals in the same species/subspecies pairs (Figs. 4, S5 and Table S12). In this case, however, some combinations of individuals showed D-statistics not

Table 1. d_{xy} (across variable sites) between species/subspecies.

	<i>N. leucogenys</i>	<i>S. syndactylus</i>	<i>H. pileatus</i>	<i>H. lar</i>	<i>H. moloch</i>	<i>H. agilis</i>	<i>H. albibarbis</i>	<i>H. m. muelleri</i>	<i>H. m. abbotti</i>
<i>N. leucogenys</i>									
<i>S. syndactylus</i>	0.238								
<i>H. pileatus</i>	0.230	0.243							
<i>H. lar</i>	0.226	0.239	0.115						
<i>H. moloch</i>	0.230	0.243	0.124	0.108					
<i>H. agilis</i>	0.228	0.241	0.122	0.104	0.093				
<i>H. albibarbis</i>	0.226	0.240	0.120	0.104	0.091	0.070			
<i>H. m. muelleri</i>	0.227	0.240	0.121	0.105	0.091	0.080	0.075		
<i>H. m. abbotti</i>	0.227	0.241	0.122	0.105	0.092	0.081	0.077	0.073	

**Fig. 3** MDS plot of ASD. **A** A two-dimensional plot of ASD for seven *Hylobates* species/subspecies ($n = 40$). **B** A two-dimensional plot of ASD for five Sundaic *Hylobates* species/subspecies ($n = 10$).

significantly deviated from zero. For example, the differences in the genetic makeups of the two *H. agilis* individuals (Sample ID: G80 and G83) influenced the D-statistics and Z-scores, and the signals of introgression could be higher and lower than the threshold, in some case, e.g. between *H. agilis* and *H. lar*.

The TreeMix analysis indicated that the likelihood of a graph reaching a plateau at the number of introgression events (m) was set to two, and 99.99% of the variance was explained with the graph (Fig. S6E). When $m = 2$, the edges of introgression were drawn from *H. lar* to *H. agilis* and from *H. lar* to *H. pileatus* (Fig. S6C). These two introgression events were consistent with the signal of introgression detected by D-statistics. When $m = 3$, an additional edge of introgression from *H. moloch* to *H. agilis* was detected (Fig. S6D), but this was not consistent with D-statistics results. This might be derived from the close phylogenetic relationship among the Sundaic species/subspecies.

Admixture graphs were generated, based on models that integrated the three introgression events suggested by the D-statistics and the direction inferred by TreeMix. However, it was insufficient to solve the deviations of f_4 statistics from zero. Including another introgression event from *H. albibarbis* to *H. m. muelleri* successfully solved the deviations (Fig. 5). Note that due to the genetic similarity between *H. m. muelleri* and *H. m. abbotti*, we also tested the models where only one of them was introgressed with *H. albibarbis*. However, those models were also insufficient to solve the deviations of f_4 statistics from zero (data not shown). In the admixture graph, the contributions of introgression to the genomes were estimated as follows: 6% contribution from *H. lar* (node P9) to *H. agilis* (node P11), 9% contribution from *H. lar* (node P12) to *H. pileatus* (node P14), 31% contribution from the common

ancestor of *H. m. muelleri* and *H. m. abbotti* (node P6) to *H. albibarbis* (node P8) and 8% contribution from *H. albibarbis* (node P8) to *H. m. muelleri* (node P16). When the order of the introgression from *H. lar* to *H. agilis* and from *H. lar* to *H. pileatus* was inverted, the ratio of the contribution from *H. lar* to *H. pileatus* considerably changed from 9 to 22%; however, there was no substantial change in the ratios of other introgression events (Fig. S7).

Calculations of f_{d_i} and d_{xy} were performed for the species pairs whose introgression was detected by D-statistics. The number of windows tested for each quartet was 434–757, representing 16.9–29.5% of 2570 possible windows across the autosomes. We considered autosomal regions with the top 1% f_{d_i} values and the bottom 1% d_{xy} values to be potentially introgressed loci (Table S11). Among those, two consecutive loci situated at chr7b of *N. leucogenys* (chr7b: 50,000,001–51,000,000 and chr7b: 51,000,001–52,000,000) were detected as candidates of introgressed loci between *H. lar* and *H. pileatus*, in all the tested quartets (Fig. S8 and Table S13). No candidates were detected between *H. lar* and *H. agilis*, *H. albibarbis* and *H. m. muelleri* and *H. albibarbis* and *H. m. abbotti*. Note that the genomic positions shown here are those for *N. leucogenys* ($2n = 52$) and thus differ from the exact genomic positions found in *Hylobates* ($2n = 44$) (c.f. Müller et al. 2002).

DISCUSSION

Phylogeny and taxonomy

The autosomal phylogeny identified in this study was consistent with the mitogenome phylogeny reported in the present and previous studies (Chan et al. 2010 and its correction), except that

Table 2. Patterson's D-statistics by species.

P1	P2	P3	Outgroup	D	SE	Z	p value
<i>H. agilis</i>	<i>H. albibarbis</i>	<i>H. m. abbotti</i>	<i>N. leucogenys</i>	-0.134	0.011	-12.00	<<1.0E-16*
<i>H. agilis</i>	<i>H. albibarbis</i>	<i>H. m. muelleri</i>	<i>N. leucogenys</i>	-0.138	0.012	-11.75	<<1.0E-16*
<i>H. moloch</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.129	0.016	-8.24	1.11E-16*
<i>H. m. abbotti</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.112	0.015	-7.44	5.11E-14*
<i>H. albibarbis</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.112	0.015	-7.37	8.86E-14*
<i>H. m. muelleri</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.112	0.015	-7.35	1.01E-13*
<i>H. albibarbis</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.100	0.015	-6.78	6.13E-12*
<i>H. agilis</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.100	0.016	-6.39	8.57E-11*
<i>H. m. muelleri</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.089	0.016	-5.51	1.77E-8*
<i>H. m. abbotti</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.080	0.015	-5.18	1.10E-7*
<i>H. moloch</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.087	0.017	-5.05	2.19E-7*
<i>H. agilis</i>	<i>H. albibarbis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.044	0.014	-3.05	1.16E-3
<i>H. moloch</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.047	0.017	-2.74	3.06E-3
<i>H. agilis</i>	<i>H. m. abbotti</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.041	0.016	-2.59	4.81E-3
<i>H. m. abbotti</i>	<i>H. m. muelleri</i>	<i>H. albibarbis</i>	<i>N. leucogenys</i>	-0.026	0.011	-2.45	7.16E-3
<i>H. m. abbotti</i>	<i>H. m. muelleri</i>	<i>H. agilis</i>	<i>N. leucogenys</i>	-0.023	0.012	-1.92	2.76E-2
<i>H. albibarbis</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.029	0.015	-1.90	2.89E-2
<i>H. moloch</i>	<i>H. m. muelleri</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.030	0.018	-1.69	4.53E-2
<i>H. moloch</i>	<i>H. m. abbotti</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.029	0.017	-1.66	4.83E-2
<i>H. moloch</i>	<i>H. albibarbis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.029	0.018	-1.63	5.14E-2
<i>H. agilis</i>	<i>H. m. muelleri</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.027	0.016	-1.62	5.23E-2
<i>H. moloch</i>	<i>H. m. abbotti</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.025	0.017	-1.49	6.79E-2
<i>H. m. abbotti</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.024	0.017	-1.46	7.27E-2
<i>H. moloch</i>	<i>H. albibarbis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.023	0.016	-1.39	8.21E-2
<i>H. m. muelleri</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.023	0.017	-1.34	9.01E-2
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.017	0.014	-1.22	0.112
<i>H. moloch</i>	<i>H. m. muelleri</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.018	0.017	-1.04	0.148
<i>H. m. muelleri</i>	<i>H. albibarbis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.011	0.014	-0.78	0.218
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.010	0.014	-0.72	0.236
<i>H. m. muelleri</i>	<i>H. albibarbis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.007	0.014	-0.46	0.321
<i>H. albibarbis</i>	<i>H. m. abbotti</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.005	0.013	-0.39	0.350
<i>H. albibarbis</i>	<i>H. m. abbotti</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.003	0.013	-0.24	0.406
<i>H. albibarbis</i>	<i>H. m. muelleri</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.002	0.016	-0.14	0.446
<i>H. m. abbotti</i>	<i>H. m. muelleri</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.002	0.016	-0.10	0.459
<i>H. albibarbis</i>	<i>H. m. abbotti</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.001	0.015	-0.03	0.487

N. leucogenys was used as the outgroup. Deviation of D-statistics from zero was assessed by alpha level = 0.001 after Bonferroni correction. Negative D-statistics significantly deviated from zero indicate the introgression between P2 and P3.

*Significant p values < 2.86×10^{-5} (=0.001/35).

H. klossii was not analysed using GRAS-Di. This indicates that GRAS-Di analysis successfully produced a large amount of SNV data to analyse phylogeny in rapidly diverged taxa. Our phylogenetic analysis supported the evolutionary scenario suggested by previous studies: the most recent common ancestor of *Hylobates* originated in mainland Southeast Asia, expanded its distribution southward, emerged in Sundaland during the Pliocene and speciated due to subsequent isolations (Whittaker et al. 2007; Thinh et al. 2010a; Chan et al. 2013).

With regard to the genetic characteristics of *H. albibarbis*, which was scarcely included in previous genetic studies, the results of this study confirmed the sister relationship between *H. albibarbis* and *H. agilis*. This result is consistent with a previous population-based genetic study (Hirai et al. 2009), a cytb phylogenetic analysis (Thinh et al. 2010a) and a previous taxonomic classification, which

included *H. albibarbis* as a subspecies of *H. agilis*, based on vocalisations (Marshall and Sugardjito 1986).

In addition, the estimated divergence time between *H. m. muelleri* and *H. m. abbotti* (0.92 MYA for the autosome) was very similar to that between *H. agilis* and *H. albibarbis* (0.97 MYA for the autosome) and was consistent with the results of Thinh et al. (2010a) who analysed cytb sequence of mtDNA. Following the proposal of Thinh et al. (2010a) who used the divergence time as the criteria of taxonomic classification, *H. m. abbotti* may be better elevated as species, although caution is needed about what kind of species concept should be applied in this case.

Introgression

The GRAS-Di analysis detected signals of introgression between several pairs of *Hylobates* species/subspecies. Interestingly, the

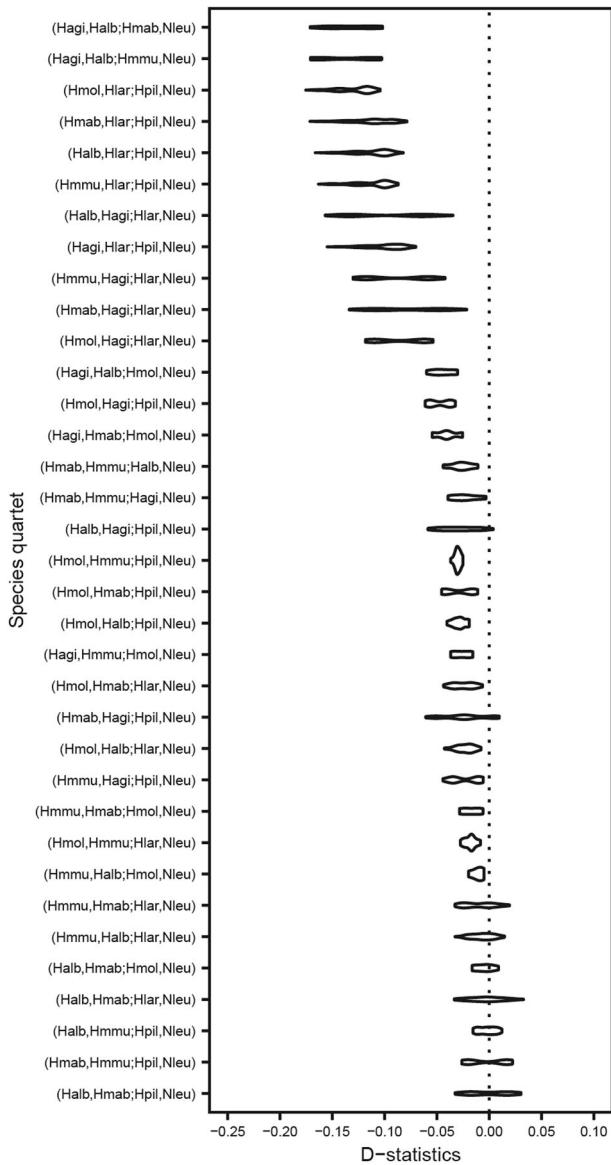


Fig. 4 Patterson's D-statistics among *Hylobates* species/subspecies. *N. leucogenys* was used as the outgroup. Species quartets are shown as (P1, P2; P3, outgroup). Negative D-statistics significantly deviated from zero indicate introgression between P2 and P3 (see Table S10). Hagi *H. agilis*, Halb *H. albibarbis*, Hlar *H. lar*, Hmab *H. m. abbotti*, Hmmu *H. m. muelleri*, Hmol *H. moloch*, Hpil *H. pileatus*, Nleu *N. leucogenys*.

introgression was detected between the species pairs, which all showed ongoing contact/hybrid zones at the species boundaries. Even when calculating the D-statistics using an individual of each species or subspecies, most of the individuals exhibited introgression signals. The exact origins of the captive gibbons analysed in this study were unknown, and therefore, it was not conclusive whether these samples represented the genetic characteristic of the entire gene pool of each species/subspecies. Nevertheless, this result supported the view that the introgression between species pairs has spread over the wide range of their distribution and is not limited to the current contact/hybrid zone. The ongoing hybridisation may not be a transient phenomenon and may have occurred for a relatively long time.

Nevertheless, the degree of admixed ancestry that can be identified in individuals from the same species/subspecies may not be uniform across their distribution. Although the signals of

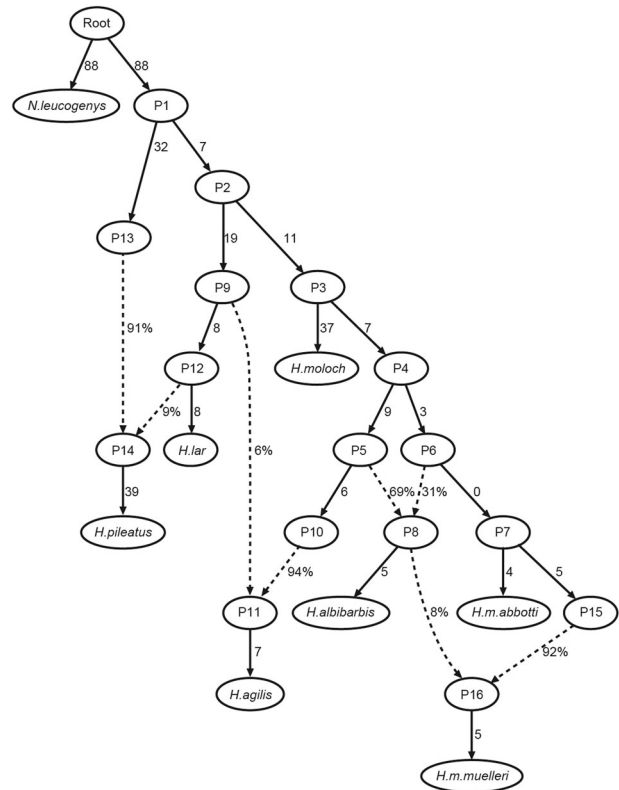


Fig. 5 Admixture graph among *Hylobates* made for a model integrating the introgression suggested by Patterson's D-statistics tests. Branch lengths are shown in *f*-statistic units.

introgression between *H. albibarbis* and *H. muelleri* and between *H. lar* and *H. pileatus* were robust in any combination of individuals, this was not the case between *H. lar* and *H. agilis*. Two *H. agilis* individuals (Sample ID: G80 and G83; both probably derived from Sumatra and not from the Malay Peninsula) showed different levels of D-statistic and Z-score values in several combinations, which likely reflects differences in the amounts of *H. lar*-derived genetic components in their genetic makeups. Individual G80, which has a mitogenome that is closely related to *H. lar* from Sumatra (*H. l. vestitus*) (Fig. S2), showed higher absolute D-statistics in tests with *H. lar*. In a previous study (Hirai et al. 2009), this type of mtDNA has been observed in several *H. agilis* individuals from Sumatra. Thus, the mtDNA of G80 likely introgressed from *H. lar* to *H. agilis* in the past, comparable to the observed mtDNA introgression from *H. pileatus* to *H. lar* in the ongoing hybrid zone (Matsudaira et al. 2013). The frequency and distribution of this mtDNA group among *H. agilis* are unknown; however, the differences in the mtDNA and D-statistics may reflect the distance from the species boundary between *H. lar* and *H. agilis*, which is located around Lake Toba on Sumatra. Since differences in the genetic makeup between G80 and G83 influenced the D-statistic values, the small sample size for some *Hylobates* species in this study should be considered cautiously. Especially, only one *H. moloch* sample was available in this study; thus, we could not confirm whether this sample accurately represents the genetic makeup of *H. moloch*. Future studies of *Hylobates* gibbons should recognise potential hidden genetic structures within species.

In addition, the degrees of introgression may have varied among hybrid zones, as suggested by previous observations. Although the degree of ongoing introgression may be limited between *H. lar* and *H. pileatus* within the hybrid zone at Khao Yai (Brockelman and Gittins 1984), a hybrid swarm (a stable hybrid/admixed population)

has formed between *H. albibarbis* and *H. muelleri* within a hybrid zone at the Barito headwaters (Mather 1992). Previous studies have suggested possible factors that may limit hybridisation and introgression in these contact/hybrid zones, such as positive assortative mating and the reduced fitness of hybrid gibbons (Brockelman and Gittins 1984; Mather 1992; Suwanvecho and Brockelman 2012; Asensio et al. 2017). Further genetic studies that focus on such hybrid zones are expected to reveal factors that influence hybridisation frequency and the degree of introgression among species.

No signals of introgression were detected between other species pairs that had potentially overlapping distribution ranges during the glacial periods. Considering that several combinations of *Hylobates* species can hybridise in captivity (Geissmann 1993), this pattern probably reflects geographic patterns rather than the phylogeny. The results were consistent with a recent finding in colugos (arboreal mammals closely related to primates), which also showed relatively deep divergence times and no evidence of recent gene flow between different Sundaic island populations (Mason et al. 2016), and in some birds (Lim et al. 2017; Cros et al. 2020). This pattern was most pronounced in species whose mobility was largely restricted by forest distribution (Mason et al. 2019; Cros et al. 2020). Since gibbons are strongly dependent on forests and rarely use the ground, the species currently living on different islands are unlikely to have interacted or hybridised with each other during glacial events.

Overall, the GRAS-Di analysis, which successfully detected hundreds of thousands of SNV sites, effectively resolved the species phylogeny of *Hylobates* gibbons (except *H. klossii*) and revealed the presence of hybridisation and introgression during *Hylobates* evolution. This study highlighted the relatively long history of ongoing hybridisation and the absence or scarcity of hybridisation/introgression between species living on the different Sundaic islands. Note that these results were influenced by how the threshold of significant D-statistics deviations from zero was set. To the best of our knowledge, there is no consensus method to set the threshold, and thus, the result should be carefully interpreted. Future studies using different analytical methodologies may detect more signals of introgression, even with the same data set.

Furthermore, although we could confirm that the effectiveness of the GRAS-Di analysis in primates was comparable to other types of reduced representation sequencing (Malukiewicz et al. 2017; Boubli et al. 2019), the numbers of SNV sites were not sufficient to fully assess potentially introgressed regions. For this purpose, more intensive sequencing efforts, such as whole-genome sequencing (WGS) should be applied. Future WGS studies of *Hylobates* gibbons that include more samples with known origins will likely reveal additional details regarding the contributions of genetic changes that have occurred during divergence, introgression to species/subspecies differences and diversity among *Hylobates*, together with detailed demographic histories such as effective population size and the amount of introgression.

DATA AVAILABILITY

The raw-read files produced by GRAS-Di analysis were deposited in the DDBJ Sequence Read Archive (DRA), with Accession No. DRA010234. The mtDNA sequences were deposited in the DDBJ/EMBL/NCBI nucleotide database, with Accession No. LC548011-LC548056. Please see Table S1 for details.

REFERENCES

- Arnason U, Gullberg A, Xu X (1996) A complete mitochondrial DNA molecule of the white-handed gibbon, *Hylobates lar*, and comparison among individual mitochondrial genes of all hominoid genera. *Hereditas* 124:185–189
- Asensio N, José-Domínguez JM, Kongrit C, Brockelman WY (2017) The ecology of white-handed and pileated gibbons in a zone of overlap and hybridization in Thailand. *Am J Phys Anthropol* 163:716–728
- Baiz MD, Tucker PK, Cortés-Ortiz L (2019) Multiple forms of selection shape reproductive isolation in a primate hybrid zone. *Mol Ecol* 28:1056–1069
- Barelli C, Matsudaira K, Wolf T, Roos C, Heistermann M, Hodges K et al. (2013) Extra-pair paternity confirmed in wild white-handed gibbons. *Am J Primatol* 75:1185–1195
- Bergey CM, Pozzi L, Disotell TR, Burrell AS (2013) A new method for genome-wide marker development and genotyping holds great promise for molecular primatology. *Int J Primatol* 34:303–314
- Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH (2019) Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol* 3:286–292
- Bond M, Tejedor MF, Campbell KEJ, Chornogubsky L, Novo N, Goin F (2015) Eocene primates of South America and the African origins of New World monkeys. *Nature* 520:538–541
- Boubli JP, Byrne H, da Silva MNF, Silva-Júnior J, Araújo RC, Bertuol F et al. (2019) On a new species of titi monkey (Primates: *Plecturocebus* Byrne et al., 2016), from Alta Floresta, southern Amazon, Brazil. *Mol Phylogenet Evol* 132:117–137
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D et al. (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457
- Brockelman WY, Gittins SP (1984) Natural hybridization in the *Hylobates lar* species group: implications for speciation in gibbons. In: Brockelman WY, Preuschoft H, Chivers DJ, Creel N (eds) *The lesser apes: evolutionary and behavioural biology*. Edinburgh University Press, Edinburgh, p 498–532
- Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* 418:145–51
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Mol Biol Evol* 29:1917–1932
- Cahill JA, Stirling I, Kistler L, Salamzade R, Ersmark E, Fulton TL et al. (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol Ecol* 24:1205–1217
- Cannon CH, Morley RJ, Bush ABG (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proc Natl Acad Sci U S A* 106:11188–11193
- Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, Huddleston J et al. (2014) Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513:195–201
- Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Pei KJ-C et al. (2010) Mitochondrial genome sequences effectively reveal the phylogeny of *Hylobates* gibbons. *PLoS ONE* 5:e14419
- Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Pei KJ-C et al. (2013) Inferring the evolutionary histories of divergences in *Hylobates* and *Nomascus* gibbons through multilocus sequence data. *BMC Evol Biol* 13:82
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:7
- Chattopadhyay B, Garg KM, Soo YJ, Low GW, Frechette JL, Rheindt FE (2019) Conservation genomics in the fight to help the recovery of the critically endangered Siamese crocodile *Crocodylus siamensis*. *Mol Ecol* 28:936–950
- Chivers DJ (1977) The lesser apes. In: Prince Rainier III of Monaco HSH, Bourne GH (eds) *Primate conservation*. Academic Press, New York, NY, p 539–598
- Cros E, Chattopadhyay B, Garg KM, Ng NSR, Tomassi S, Benedick S et al. (2020) Quaternary land bridges have not been universal conduits of gene flow. *Mol Ecol* 29:2692–2706
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- de Manuel M, Kuhlwilim M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J et al. (2016) Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354:477–481
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Enoki H, Takeuchi Y (2018) New genotyping technology, GRAS-Di, using next generation sequencer. In: *Proceedings of the Plant and Animal Genome Conference XXVI*. San Diego, CA. <https://pag.confex.com/pag/xxvi/meetingapp.cgi/Paper/29067>
- Fan P-F, He K, Chen X, Ortiz A, Zhang B, Zhao C et al. (2017) Description of a new species of *Hoolock* gibbon (Primates: Hylobatidae) based on integrative taxonomy. *Am J Primatol* 79:e22631
- Fan Z, Zhou A, Osada N, Yu J, Jiang J, Li P et al. (2018) Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Mol Phylogenet Evol* 127:376–386

- Finstermeier K, Zinner D, Brameier M, Meyer M, Kreuz E, Hofreiter M et al. (2013) A mitogenomic phylogeny of living primates. *PLoS ONE* 8:e69504
- Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D et al. (2016) The genetic history of Ice Age Europe. *Nature* 534:200–205
- Geissmann T (1991) Sympatry between white-handed gibbons (*Hylobates lar*) and pileated gibbons (*H. pileatus*) in Southeastern Thailand. *Primates* 32:357–363
- Geissmann T (1993) Evolution of communication in gibbons (Hylobatidae). Ph.D. thesis, Zürich University
- Gorog AJ, Sinaga MH, Engstrom MD (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf murine rodents (*Maxomys surifer*, *Leopoldamys sabanus* and *Maxomys whiteheadi*). *Biol J Linn Soc* 81:91–109
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M et al. (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722
- Groves CP (1972) Systematics and phylogeny of gibbons. In: Rumbaugh DM (ed) *Gibbon and Siangang: a series of volumes on the lesser apes*, vol 1. Karger, Basel, p 1–89
- Groves CP (2001) *Primate taxonomy*. Smithsonian Institution Press, Washington, DC
- Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Clim Change* 19:53–61
- Hirai H, Hayano A, Tanaka H, Mootnick AR, Wijayanto H, Perwitasari-Farajallah D (2009) Genetic differentiation of agile gibbons between Sumatra and Kalimantan in Indonesia. In: Lappan S, Whittaker DJ (eds) *The Gibbons: new perspectives on small ape socioecology and population biology*. Springer, New York, NY, p 37–50
- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A* 92:532–536
- Hosoya S, Hirase S, Kikuchi K, Nanjo K, Nakamura Y, Kohno H et al. (2019) Random PCR-based genotyping by sequencing technology GRAS-Di (genotyping by random amplicon sequencing, direct) reveals genetic structure of mangrove fishes. *Mol Ecol Resour* 19:1153–1163
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267
- Ishida T, Yamamoto K, Ishimoto G, Shotake T, Takenaka O, Nozawa K et al. (1985) A field study of infection with human T-cell leukemia virus among Asian primates. *Microbiol Immunol* 29:839–46
- Ito H, Nakajima N, Onuma M, Murayama M (2020) Genetic diversity and genetic structure of the wild Tsushima leopard cat from genome-wide analysis. *Animals* 10:1375
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589
- Kelley J (2002) The hominoid radiation in Asia. In: Hartwig W (ed) *The primate fossil record*. Cambridge University Press, Cambridge, p 369–384
- Kim SK, Carbone L, Becquet C, Mootnick AR, Li DJ, de Jong PJ et al. (2011) Patterns of genetic variation within and between gibbon species. *Mol Biol Evol* 28:2211–2218
- Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol* 50:913–925
- Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. <https://arxiv.org/abs/1303.3997v2>. [q-bio.GN]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–9
- Liedigk R, Roos C, Brameier M, Zinner D (2014) Mitogenomics of the Old World monkey tribe Papionini. *BMC Evol Biol* 14:176
- Lim HC, Gawin DF, Shakya SB, Harvey MG, Rahman MA, Sheldon FH (2017) Sundaland's east-west rain forest population structure: variable manifestations in four polytypic bird species examined using RAD-Seq and plumage analyses. *J Biogeogr* 44:2259–2271
- Malukiewicz J, Guschanski K, Gratiol AD, Oliveira MAB, Ruiz-Miranda CR, Stone AC (2017) Application of PE-RADSeq to the study of genomic diversity and divergence of two Brazilian marmoset species (*Callithrix jacchus* and *C. penicillata*). *Am J Primatol* 79:e22587
- Marshall J, Sugardjito J (1986) *Gibbon systematics*. In: Swindler DR, Erwin J (eds) *Comparative primate biology, volume I: systematics, evolution, and anatomy*. Alan R. Liss, New York, NY, p 137–185
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12
- Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol* 32:244–257
- Mason VC, Li G, Minx P, Schmitz J, Churakov G, Doronina L et al. (2016) Genome analysis reveals hidden biodiversity within colugos, the sister group to primates. *Sci Adv* 2:e1600633
- Mason VC, Helgan KM, Murphy WJ (2019) Comparative phylogeography of forest-dependent mammals reveals paleo-forest corridors throughout Sundaland. *J Hered* 110:158–172
- Mather R (1992) A field study of hybrid Gibbons in Central Kalimantan, Indonesia. Ph.D. thesis, Cambridge University
- Matsudaira K, Ishida T (2010) Phylogenetic relationships and divergence dates of the whole mitochondrial genome sequences among three gibbon genera. *Mol Phylogenet Evol* 55:454–459
- Matsudaira K, Reichard UH, Malaivijitnond S, Ishida T (2013) Molecular evidence for the introgression between *Hylobates lar* and *H. pileatus* in the wild. *Primates* 54:33–37
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–303
- Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195
- Müller S, Neusser M, Weinberg J (2002) Towards unlimited colors for fluorescence *in situ* hybridization (FISH). *Chromosome Res* 10:223–232
- Nakayama K, Ishida T (2006) Alu-mediated 100-kb deletion in the primate genome: the loss of the agouti signaling protein gene in the lesser apes. *Genome Res* 16:485–490
- Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, Desai T et al. (2017) Morphometric, behavioral, and genomic evidence for a new orangutan species. *Curr Biol* 27:1–12
- Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* 76:5269–5273
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274
- Nye J, Laayouni H, Kuhlwil M, Mondal M, Marques-Bonet T, Bertranpetit J (2018) Selection in the introgressed regions of the chimpanzee genome. *Genome Biol Evol* 10:1132–1138
- Oka T, Takenaka O (2001) Wild gibbons' parentage tested by non-invasive DNA sampling and PCR-amplified polymorphic microsatellites. *Primates* 42:67–73
- Okonechnikov K, Conesa A, García-Alcalde F (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294
- Osada N, Matsudaira K, Hamada Y, Malaivijitnond S (2021) Testing sex-biased admixture origin of macaque species using autosomal and X-chromosomal genomic sequences. *Genome Biol Evol* 13:evaa209
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093
- Petr M, Pääbo S, Kelso J, Vernot B (2019) Limits of long-term selection against Neandertal introgression. *Proc Natl Acad Sci U S A* 116:1639–1644
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967
- Pozzi L, Hodgson JA, Burrell AS, Sterner KN, Raaum RL, Disotell TR (2014) Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Mol Phylogenet Evol* 75:165–183
- Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B et al. (2013) Great ape genetic diversity and population history. *Nature* 499:471–475
- R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904
- Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461:489–494
- Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G et al. (2019) The comparative genomics and complex population history of *Papio* baboons. *Sci Adv* 5:eaau6947
- Roos C (2016) Phylogeny and classification of gibbons (Hylobatidae). In: Reichard UH, Hirai H, Barelli C (eds) *Evolution of Gibbons and Siangang: phylogeny, morphology, and cognition*. Springer, New York, NY, p 151–164
- Rylands AB, Mittermeier RA (2014) *Primate taxonomy: species and conservation*. *Evol Anthropol* 23:8–10
- Scally A, Yngvadottir B, Xue Y, Ayub Q, Durbin R, Tyler-Smith C (2013) A genome-wide survey of genetic variation in gorillas using reduced representation sequencing. *PLoS ONE* 8:e65066
- Senut B, Pickford M, Gommery D, Mein P, Cheboi K, Coppens Y (2001) First hominid from the Miocene (Lukeino formation, Kenya). *C R Acad Sci* 332:137–144
- Smith J, Kronfrost MR (2013) Do *Heliconius* butterfly species exchange mimicry alleles? *Biol Lett* 9:20130503
- Stange M, Sánchez-Villagra MR, Salzburger W, Matschner M (2018) Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Syst Biol* 67:681–699
- Stevens NJ, Seiffert ER, O'Connor PM, Roberts EM, Schmitz MD, Krause C (2013) Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature* 497:611–614

- Suwanvecho U, Brockelman WY (2012) Interspecific territoriality in gibbons (*Hylobates lar* and *H. pileatus*) and its effects on the dynamics of interspecies contact zones. *Primates* 53:97–108
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Thinh VN, Mootnick AR, Geissmann T, Li M, Ziegler T, Agil M et al. (2010a) Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. *BMC Evol Biol* 10:74
- Thinh VN, Rawson B, Hallam C, Kenyon M, Nadler T, Walter L et al. (2010b) Phylogeny and distribution of crested gibbons (genus *Nomascus*) based on mitochondrial cytochrome b gene sequence data. *Am J Primatol* 72:1047–1054
- van der Valk T, Gonda CM, Silegowa H, Almanza S, Sifuentes-Romero I, Hart TB et al. (2020) The genome of the endangered Dryas monkey provides new insights into the evolutionary history of the vervets. *Mol Biol Evol* 37:183–194
- Whittaker DJ, Morales JC, Melnick DJ (2007) Resolution of the *Hylobates* phylogeny: congruence of mitochondrial D-loop sequences with molecular, behavioral, and morphological data sets. *Mol Phylogent Evol* 45:620–628
- Woodruff DS (2010) Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodivers Conserv* 19:919–941
- Wurster CM, Rifai H, Zhou B, Haig J, Bird MI (2019) Savanna in equatorial Borneo during the late Pleistocene. *Sci Rep.* 9:6392
- Xu X, Arnason U (1996) The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. *J Mol Evol* 43:431–437
- Xue C, Raveendran M, Harris RA, Fawcett GL, Liu X, White S et al. (2016) The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome Res* 26:1651–1662
- Zichello JM (2018) Look in the trees: hylobatids as evolutionary models for extinct hominins. *Evol Anthropol* 27:142–146
- Zinner D, Wertheimer J, Liedigk R, Groeneveld LF, Roos C (2013) Baboon phylogeny as inferred from complete mitochondrial genomes. *Am J Phys Anthropol* 150:133–140

ACKNOWLEDGEMENTS

This study was supported by JSPS KAKENHI [Grant Numbers JP18H02516 and JP19K16240]. We are grateful to the zoos for their cooperation.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00452-7>.

Correspondence and requests for materials should be addressed to K.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.