1 **Divergence and introgression in small apes, the genus**
2 ***Hylobates*, revealed by reduced representation sequencing**

3

4 Authors: Kazunari Matsudaira^{1*}, Takafumi Ishida¹

5

6 Affiliation:

7 ¹ Department of Biological Sciences, School of Science, The University of
8 Tokyo, Tokyo 113-0033, Japan

9

10 Corresponding Author:

11 Kazunari Matsudaira

12 Department of Biological Sciences, School of Science, The University of
13 Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

14 Tel.: +81-3-5841-4496

15 kaz.matsudaira@gmail.com

16

17 Running title: Introgression among *Hylobates* gibbons

18 Word count for main text: 5,498

Introgression among *Hylobates* gibbons

19 **Abstract**

20 Gibbons in the genus *Hylobates*, which live in Southeast Asia, show
 21 great diversity, comprising seven to nine species. Natural hybridisation has
 22 been observed in the species contact zones, although the roles played by
 23 hybridisation and introgression in the evolution of these species remain
 24 unclear. To uncover the divergence history and the contributions of
 25 hybridisation and introgression to the evolution of *Hylobates*, random
 26 amplicon sequencing-direct (GRAS-Di) analysis was employed to genotype
 27 47 gibbons, representing eight species from three genera. After quality
 28 filtering, over 300,000 autosomal single-nucleotide variant (SNV) sites were
 29 identified. The SNV-based autosomal phylogeny, together with the
 30 mitochondrial phylogeny, supported a divergence pattern beginning
 31 approximately 4.3 million years ago. First, the mainland species, *H. pileatus*
 32 and *H. lar*, consecutively diverged from the Sundaic island species. Second,
 33 *H. moloch*, in Java (and likely *H. klossii*, in the Mentawai Islands) diverged
 34 from the other species. Third, *H. muelleri*, in Borneo, and *H. agilis*/*H.*
 35 *albibarbis*, in Sumatra and southwestern Borneo, diverged. Lastly, *H. agilis*
 36 and *H. albibarbis* diverged from each other. The Patterson's D-statistics
 37 indicated significant introgression between *H. lar* and *H. pileatus*, between
 38 *H. lar* and *H. agilis*, and between *H. albibarbis* and *H. muelleri*, and weak
 39 introgression was identified between *H. moloch* and *H. albibarbis*, and

Introgression among *Hylobates* gibbons

40 between *H. moloch* and *H. muelleri abbotti*, suggesting incomplete
 41 reproductive barriers among *Hylobates* species and that hybridisation and
 42 introgression occur whenever the distribution ranges contact. Some
 43 candidates for introgressed genomic regions were detected, and the
 44 functions of these would be revealed by further genome-wide studies.

45

46 Keywords:

47 gibbons, GRAS-Di, hybridisation, introgression

48

Introgression among *Hylobates* gibbons

49 **Introduction**

50 Primates is a highly diversified mammalian order, with an
 51 estimated 500 (and a minimum of 350) extant species (Groves 2001;
 52 Rylands and Mittermeier 2014). Recently, genome-wide studies of
 53 non-human primates have become possible, revealing evolutionary
 54 processes, such as speciation, and the demographic histories of various
 55 taxa (Prado-Martinez et al. 2013; Xue et al. 2016; Nater et al. 2017).
 56 Historical hybridisation and introgression have been detected in many cases
 57 (e.g., chimpanzees and bonobos [de Manuel et al. 2016], macaques [Fan et
 58 al. 2018], baboons [Rogers et al. 2019], and Dryas and green monkeys [van
 59 der Valk et al. 2020]); thus, these processes are recognised as common
 60 phenomena.

61 Moreover, the potential roles played by hybridisation and
 62 introgression in the evolution of non-human primates, such as adaptive
 63 introgression, have been widely discussed (Nye et al. 2018; Rogers et al.
 64 2019; van der Valk et al. 2020), although further studies remain necessary
 65 to reveal how hybridisation occurs and the degree to which introgression
 66 has affected primate evolution. Among primates, Hylobatidae, which is a
 67 family of small apes or gibbons, is considered to be a good model for
 68 studying the contributions of divergence and hybridisation/introgression to
 69 evolution, especially as an analogy for evolution in hominins (Zichello

Introgression among *Hylobates* gibbons

70 2018).

71 Gibbons are small apes that live in the Indomalaya realm and show
72 great diversity among four extant genera, which each have different
73 numbers of chromosomes (*Hoolock* [2n = 38], *Hylobates* [2n = 44],
74 *Symphalangus* [2n = 50], and *Nomascus* [2n = 52]), comprising 15 to 20
75 species (Roos 2016; Fan et al. 2017). Among these four genera, *Hylobates*
76 shows the highest taxonomic diversity, followed by *Nomascus*. *Hylobates* is
77 distributed throughout both the mainland and islands of Southeast Asia,
78 and seven allopatric species are recognised, and characterised by various
79 pelage colours and vocalisations: *Hylobates pileatus* (pileated gibbons), *H.*
80 *lar* (white-handed gibbons), *H. agilis* (agile gibbons), *H. albibarbis* (Bornean
81 white-bearded gibbons), *H. muelleri* (Müller's gibbons), *H. moloch* (Javan
82 gibbons), and *H. klossii* (Kloss's gibbons) (Fig. 1; Groves 2001).

83 Some studies have proposed to elevate three subspecies of *H.*
84 *muelleri*, *H. muelleri muelleri* (Müller's gibbons), *H. muelleri abbotti*
85 (Abbot's gibbons) and *H. muelleri funereus* (North Bornean grey gibbons),
86 to independent species (Thinh et al. 2010a; Roos 2016); however, no
87 consensus has been reached, thus far. We used the recognised
88 seven-species classification, in this study. In addition, in this study, we refer
89 to *H. pileatus* and *H. lar* as mainland species for simplicity, although some *H.*
90 *lar* is known to be distributed in northern Sumatra, and refer to the other

Introgression among *Hylobates* gibbons

91 species as Sundaic species.

92 The phylogeny of the *Hylobates* species has been studied
 93 according to morphology (Groves 1972; Creel and Preuschoft 1984;
 94 Geissmann 2002), vocalisation (Geissmann 2002), mitochondrial DNA
 95 (mtDNA) (Garza and Woodruff 1992; Hayashi et al. 1995; Takacs et al. 2005;
 96 Chatterjee 2006; Whittaker et al. 2007; Chan et al. 2010; Matsudaira and
 97 Ishida 2010; Thinh et al. 2010a), and nuclear DNA (Israfil et al. 2011; Kim et
 98 al. 2011; Chan et al. 2012, 2013). A whole-mitochondrial genome
 99 (mitogenome) study successfully resolved the mtDNA phylogeny of
 100 *Hylobates*, which demonstrated that *H. pileatus* diverged first and *H. lar*
 101 diverged second, followed by the divergence between proto-*H. agilis*/*H.*
 102 *muelleri* and proto-*H. moloch*/*H. klossii* (Chan et al. 2010 and its correction).
 103 This study suggested that divergence began among mainland species and
 104 was followed by the Sundaic species, which was supported a scenario in
 105 which the common ancestor of *Hylobates* originated on the mainland and
 106 that the distribution later expanded to the islands of Southeast Asia
 107 (Groves 1972; Chivers 1977; Whittaker et al. 2007). The same and other
 108 studies estimated that mtDNA divergence began approximately 3.5 to 3.9
 109 million years ago (MYA) and was completed in a relatively short time (Chan
 110 et al. 2010; Matsudaira and Ishida 2010; Thinh et al. 2010a).

111 In contrast with the mitogenome studies, nuclear DNA markers

Introgression among *Hylobates* gibbons

112 have yielded ambiguous results on species phylogeny (Kim et al. 2011;
 113 Chan et al. 2013), likely due to the rapid divergence that occurred among
 114 *Hylobates* species, as suggested by mtDNA, resulting in the incomplete
 115 lineage sorting of many nuclear genetic markers. In addition, because
 116 hybridisation has been observed in the contact zones between *Hylobates*
 117 species (Brockelman and Gittins 1984; Marshall and Sugardjito 1986), many
 118 historical hybridisations and introgression may have occurred after the
 119 initial divergence of each species; thus, their nuclear genomes may have
 120 become admixed. In a previous study, using 14 autosomal loci, covering
 121 approximately 11.5 kb, the presence of historical introgression between
 122 some *Hylobates* species was detected (Chan et al. 2013). However, the
 123 numbers of nuclear DNA markers used in previous studies were likely too
 124 few to assess the complicated evolutionary history of the *Hylobates* species.
 125 Genome-wide studies that utilise larger numbers of nuclear markers are
 126 expected to overcome these potential issues, revealing the detailed
 127 evolution of *Hylobates* gibbons.

128 To date, several approaches have been used to conduct
 129 genome-wide studies of non-model organisms (Beichman et al. 2018). The
 130 whole-genome sequencing (WGS) of several tens of individuals in each
 131 species can cover all of the information in the genomes and can account for
 132 genetic diversity in each species (e.g., de Manuel et al. 2016; Nater et al.

Introgression among *Hylobates* gibbons

2017). Although the cost of WGS has been dropping rapidly, studying multiple individuals of many species, such as of *Hylobates*, using WGS remains expensive and is also costly in terms of computational resources.

Reduced-representation sequencing methods, such as restriction-site-associated DNA sequencing (RAD-seq) and genotype by sequencing (GBS), are alternative methods that limit the number of sequencing loci across genomes, facilitating the effective acquisition of sufficient numbers of genetic markers from large numbers of samples, supporting a variety of genome-wide studies (Andrews et al. 2016). Genotyping by random amplicon sequencing-direct (GRAS-Di) is an alternative method 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 a variety of fish species and was able to detect several thousands of single nucleotide variants (SNVs) in each species (Hosoya et al. 2019).

In this study, we conducted a GRAS-Di analysis of gibbon species, to uncover the species phylogeny and the evolutionary history of divergence and hybridisation/introgression events among *Hylobates* gibbons.

Introgression among *Hylobates* gibbons

154 **Materials and Methods**

155 **Samples**

156 DNA was extracted from 46 Epstein Barr virus-transformed B
157 lymphoblastoid cell lines and one muscle tissue sample, using the QIAamp
158 DNA Mini Kit (Qiagen, Germany). The DNA samples consisted of two
159 samples from *H. agilis*, three from *H. albibarbis*, 20 from *H. lar*, one from *H.*
160 *moloch*, four from *H. muelleri* (consisting of two from *H. muelleri muelleri*
161 and two from *H. muelleri abbotti*), 10 from *H. pileatus*, one from *Nomascus*
162 *leucogenys* (northern white-cheeked gibbons), and six from *Symphalangus*
163 *syndactylus* (siamangs) (Supplementary Table S1). The samples were
164 derived from captive zoo gibbons, as described in previous studies (Ishida
165 et al. 1985; Nakayama and Ishida 2006).

166 At the time of sample collection, the species status and key
167 external characteristics were not well-recorded for some individuals.
168 Therefore, we confirmed the maternal species status of all 47 gibbons by
169 comparing the *cytochrome b* (*cytb*) sequence from mtDNA with publicly
170 available data (see details below). To minimise the inclusion of
171 unidentifiable hybrid individuals, we analysed multiple individuals from
172 each species/subspecies, when possible.

173

174 **Mitochondrial DNA sequencing and analysis**

Introgression among *Hylobates* gibbons

We determined the *cytb*, tRNA-Thr, tRNA-Pro, and partial D-loop sequences (approximately 1.9 kb) of all 47 gibbons and determined the whole-mitogenome sequences (approximately 16.5 kb) of a subset of individuals (n = 6). Approximately 8-kb (covering the *cytb* to D-loop region) and 10-kb fragments of mtDNA were PCR amplified, using the following primers: 5'-GGCTTTCTCAACTTTTAAAGGATA-3' and 5'-TGTCCTGATCCAACATCGAG-3' (8-kb fragment), and 5'-CCGTGCAAAGGTAGCATAATC-3' and 5'-TTACTTTTATTGGAGTTGCACCA-3' (10-kb fragment) (Finstermeier et al. 2013).

PCR was performed in a 25 µl mixture, consisting of 0.5 units KOD FX DNA polymerase (Toyobo, Japan), 1× PCR Buffer for KOD FX, 0.4 mM of each dNTP, 7.5 pmol of each primer, and 1 µl DNA solution. The PCR thermal cycle conditions were as follows: the initial incubation was performed at 94°C, for 2 min, followed by 35 cycles consisted of a denaturation step at 98°C, for 10 s, and an annealing/extension step at 68°C for 8 min (8-kb fragment) or an annealing step at 63°C for 30 s and extension at 68°C for 8 min (10-kb fragment), followed by a final extension step at 68°C for 10 min. PCR amplification was confirmed by agarose gel electrophoresis.

The PCR products were purified using a FavorPrep PCR Clean-Up Mini Kit (Favorgen, Taiwan) and used as templates for sequencing reactions.

Introgression among *Hylobates* gibbons

196 Sanger sequencing was outsourced to Fasmac (Japan), with internal
197 primers, including three forward primers (5'-CCACGACCAATGATACGAAA-3';
198 5'-AGACAACGCCACACTCACAC-3'; and 5'-CTTCACCCTCAGCACCCAAAGC-3'
199 [Andayani et al. 2001]) and four reverse primers
200 (5'-ATGAGGAGGAGGAGAAATAGTCCTAG-3';
201 5'-GGGGTGGAAGGTAATTTTGTG-3'; 5'-CGGCTTACAAGACCGGTGT-3'; and
202 5'-AAGACAGATACTGCGACATAGG-3' [Matsudaira et al. 2013]), to determine
203 the *cytb* to D-loop region. Other internal primer sequences are available, on
204 reasonable request, from the authors. Sequences were called from
205 spectrograms, using MEGA 6 (Tamura et al. 2013).

206 To confirm the maternal species status, we assessed the mtDNA
207 phylogenetic positions of all 47 *cytb* sequences (1,140 bases) by comparing
208 them with the published *cytb* sequences of 175 gibbons (Arnerson et al.
209 1996; Chan et al. 2010; Matsudaira and Ishida 2010; Thinh et al. 2010a;
210 Thinh et al. 2010b; Finstermeier et al. 2013; Fan et al. 2017) and eight other
211 primate species: *Homo sapiens*, *Pan troglodytes*, *Pan paniscus*, *Gorilla*
212 *gorilla*, *Pongo pygmaeus*, *Pongo abelli*, *Macaca mulatta*, and *Papio anubis*
213 (Horai et al. 1995; Xu and Arnason 1996; Zinner et al. 2013; Liedigk et al.
214 2014) (Supplementary Table S2). We constructed a maximum likelihood
215 (ML) tree, using IQ-TREE (Nguyen et al. 2015). The TN+F+I+G4 substitution
216 model was selected by ModelFinder (Kalyaanamoorthy et al. 2017). An

Introgression among *Hylobates* gibbons

217 ultrafast bootstrap approximation test (Minh et al. 2013) was performed,
218 using 1,000 iterations.

219 Previous whole-mitogenome studies did not analyse *H. albibarbis*
220 (Chan et al. 2010; Matsudaira and Ishida 2010). Thus, we newly determined
221 the whole-mitogenome sequences of one *H. albibarbis* and five other
222 *Hylobates* gibbons. We analysed these six sequences, together with the
223 published whole-mitogenome sequences of 44 gibbons (which included
224 one *S. syndactylus* in the GRAS-Di data set [Sample ID: G75]) (Arneron et al.
225 1996; Chan et al. 2010; Matsudaira and Ishida 2010; Finstermeier et al.
226 2013; Fan et al. 2017) and eight outgroup species, similar to the *cytb*
227 analysis (Supplementary Table S3). After aligning the sequences using
228 MUSCLE (Edgar 2004), which was implemented in MEGA, we extracted the
229 sequences of 13 protein-coding genes and two rRNA-coding regions, which
230 resulted in 13,798-bp sequences. The ML tree was constructed using
231 IQ-TREE. Among the 15 identified genes, seven partitions with different
232 substitution models were selected by ModelFinder (Supplementary Table
233 S4). An ultrafast bootstrap test was conducted, for 1,000 iterations.

234 Divergence time estimation was conducted using BEAST 2.6.0.
235 (Bouckaert et al. 2014). The relaxed, lognormal clock model and birth-death
236 model tree priors were selected. The partition scheme and the substitution
237 models used in the ML tree inference were also applied. Three divergence

Introgression among *Hylobates* gibbons

time priors were used, assuming normal distribution: (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), (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 once in every 1,000 generations. The conversions of the results were assessed by Tracer 1.7.1. (Rambaut et al. 2018). The first 25% samples from each run were discarded as burn-in, and the remaining results were merged, using LogCombiner in BEAST. A consensus tree was constructed from 75,004 trees, using TreeAnnotator in BEAST, and was visualised using FigTree v1.4.4 (<https://github.com/rambaut/figtree/releases>).

251

252 GRAS-Di

The 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). We measured the DNA concentrations of the samples using a Qubit 4 fluorometer, with the Qubit dsDNA BR Assay Kit (Thermo Scientific), and prepared 100 ng of DNA (> 15 ng/μl) per sample. Library preparation and

Introgression among *Hylobates* gibbons

sequencing using the GRAS-Di method were outsourced to GeneBay, Inc. (Japan). The protocols for the library preparation were those described by Hosoya et al. (2019), except that the PCR reaction volume was 25 µl during the first PCR run and 50 during the second PCR run, instead of 10 µl. The sequences of the 64 primers used to obtain amplicons during the first PCR run were the same as those described in Hosoya et al. (2019). The library was sequenced on a HiSeq 4000 (Illumina, USA), with 151-bp paired-end reads.

267

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 CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -A CTGTCTCTTATACACATCTGACGCTGCCGACGA --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 Burrows-Wheeler Aligner (BWA-MEM) 0.7.17 (Li 2013). Generated sam files were converted into bam files and sorted using SortSam command of Genome Analysis Toolkit (GATK) 4.1.0.0. (McKenna et al. 2010). After adding read group information, using GATK AddOrReplaceReadGroups,

Introgression among *Hylobates* gibbons

each bam file was indexed by samtools 1.9 (Li et al. 2009). GATK HaplotypeCaller with --emit-ref-confidence and GVCF options, was used to call variants in each individual. The generated 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.

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 5× depth of coverage for each gibbon; (2) biallelic (i.e., no more than two alleles); and (3) variable among the 47 gibbons. Last, we extracted only those SNV sites located on the 25 autosomal long scaffolds of nomLeu3. We used GATK VariantFiltration, vcftools 0.1.17 (Danecek et al. 2011), and vcflib (<https://github.com/vcflib/vcflib>) to perform the filtering. The generated dataset consisted of 358,142 SNV sites.

297

Species tree and network construction

We constructed a phylogenetic tree of the autosomal SNVs, using ML and Bayesian frameworks. The ML tree was constructed using IQ-TREE.

Introgression among *Hylobates* gibbons

301 Because the dataset consisted of only SNV sites, we used an ascertainment
302 bias correction (ASC) model (Lewis 2001) for the ML tree construction. In
303 the ASC model, heterozygous and invariant sites were ignored; thus,
304 196,186 of 358,142 SNV sites were used for the analysis. The
305 TVM+F+ASC+R2 model was selected by ModelFinder, and an ultrafast
306 bootstrap approximation test was performed, with 1,000 iterations.

307 A Bayesian tree was constructed, using MrBayes 3.2.7a (Ronquist
308 et al. 2012). The GTR+G model was selected by MrModeltest 2.4 (Nylander
309 2004). Four independent MCMC runs were performed, consisting of one
310 heat chain and three cold chains. Each run comprised 10,000,000
311 generations, and the parameters were sampled every 5,000 generations.
312 Convergence among the four runs was assessed by Tracer. The first 25% of
313 samples from each run were discarded as burn-in, and the results of the
314 four runs (i.e., 6,004 samples) were merged. The tree was visualised using
315 FigTree.

316 To assess the reticulate evolution among *Hylobates*, a phylogenetic
317 network of autosomal SNVs was generated, using the NeighborNet
318 algorithm in SplitsTree 4.15.1 (Huson and Bryant 2006). To clarify the
319 results, we only included *Hylobates* gibbons in this analysis.

320

321 Allele-sharing distance

Introgression among *Hylobates* gibbons

322 To uncover the genetic relationships among individuals, we
323 calculated the allele-sharing distances (ASD) (Bowcock et al. 1994) between
324 each pair of individuals, using asd v1.0.0 (<https://github.com/szpiech/asd>).
325 Based on the ASD matrix, multi-dimensional scaling (MDS) plots
326 (two-dimensional plots) were drawn, using R 3.6.2 (R Core Team 2019).

327

328 Detection of introgression

329 The presence of introgression among *Hylobates* was assessed by
330 Patterson's D-statistics tests (also known as ABBA-BABA tests) (Green et al.
331 2010; Patterson et al. 2012). Because the phylogenetic relationships among
332 *Hylobates* species/subspecies have been inferred from both mitogenomes
333 and autosomal SNVs and are well-supported and consistent, (except for one
334 mitogenome from *H. agilis*; see Results), we used the autosomal tree
335 topology and calculated the D-statistics by setting the outgroup as either *N.*
336 *leucogenys* or *S. syndactylus*. The D-statistics were calculated for both
337 combinations of one individual per species/subspecies and the
338 combinations of population data for species/subspecies. The D-statistics
339 were calculated using AdmixTools (Patterson et al. 2012), with the R
340 package admixr (Petr et al. 2019).

341

342 Potentially introgressed loci

Introgression among *Hylobates* gibbons

When an introgression event occurs, most genomic regions from other species are expected to be removed, due to incompatibility and deleteriousness. However, some regions may remain as introgressed loci because of adaptive advantages. Potentially introgressed regions can be detected by comparing D-statistics and genetic divergence in each segment of the genome (Smith and Kronfrost 2013). Because introgressed regions should have shorter coalescent times than average, the regions with high absolute D-statistics and low genetic divergence can be considered candidates for introgressed regions.

To detect the potentially introgressed loci among *Hylobates* species/subspecies, we calculated the corrected D-statistics, f_d , (Martin et al. 2015), and genetic divergence d_{xy} (Smith and Kronfrost 2013), using ABBABABAWindows.py and popgenWindows.py (https://github.com/simonhmartin/genomics_general), respectively. Because of limitations in the number of SNV sites in our data set, we calculated the values for regions with a large window size of 1,000,000 bases, with at least 100 SNVs per window.

360

Results

Genotyping

We obtained a mean $\pm$ standard deviation (SD) of 8.6 million (M) $\pm$

Introgression among *Hylobates* gibbons

0.87 (min = 6.6 M; max = 10.2 M) raw read pairs, per sample. Adapter trimming and quality filtering by cutadapt yielded $8.5 \text{ M} \pm 0.85$ (min = 6.5 M; max = 10.0 M) read pairs, per sample. In the nomLeu3 genome, 3.6% to 5.2% (95,030,378 to 139,035,756 sites) of the 2,654,007,897 autosomal sites were mapped by at least one read in each sample, and 1.2% to 1.8% (31,448,837 to 47,600,791 sites) were mapped by at least five reads in each sample. 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 358,142 biallelic, autosomal, SNV sites that were successfully genotyped in all 47 gibbons was generated.

We calculated the nucleotide diversity (Nei and Li 1979) of each species/subspecies across the 358,142 autosomal SNV sites (Table 1). *H. pileatus* showed the lowest nucleotide diversity (0.0198 ± 0.0811). The three Bornean species/subspecies (*H. albibarbis*, *H. muelleri muelleri*, and *H. muelleri abbotti*) showed the highest nucleotide diversity (0.0598–0.0694). The other three *Hylobates* species (*H. moloch*, *H. lar*, and *H. agilis*), *S. syndactylus*, and *N. leucogenys* were in the middle (0.0371–0.0511).

Phylogenetic tree, divergence time, ASD, and phylogenetic network

Phylogenetic relationships among *Hylobates* were well resolved, both in the mitogenome and the autosomal SNV trees (Fig. 2), and the

Introgression among *Hylobates* gibbons

385 topologies of species/subspecies relationships were consistent, except for
 386 one mitogenome sequence from *H. agilis* (Sample ID: G80). This individual
 387 clustered with the other *H. agilis* individuals in the autosomal SNV tree (Fig.
 388 2B) but clustered with *H. lar* individuals and separated from other *H. agilis*
 389 individuals in the mitogenome tree (Fig. 2A; labelled as *H. agilis* [group2]).
 390 In the *cytb* tree, which consisted of the reference sequences of four *H. lar*
 391 subspecies, this *H. agilis* individual tightly clustered with the sequences
 392 from *H. lar vestitus* (Sumatran white-handed gibbons) (Supplementary Fig.
 393 S1). Because the mitogenome represents the genealogy of only a single
 394 locus, we considered the autosomal SNV tree to represent the true species
 395 tree and used the autosomal SNV tree for all further analyses.

396 Both the autosomal and mitogenome phylogenetic trees
 397 demonstrated that the divergence of *Hylobates* species started in the
 398 mainland species, followed by the Sundaic species/subspecies. Based on
 399 the estimated divergence time indicated by the mitogenomes (Fig. 2A,
 400 Supplementary Table S5), among *Hylobates*, divergence first occurred
 401 between *H. pileatus* and the common ancestor of the other *Hylobates*
 402 species, approximately 4.3 MYA (95% highest posterior density credibility
 403 interval [HPD CI]: 3.7 to 5.0). Next, *H. lar* diverged from the Sundaic species,
 404 approximately 3.8 MYA (3.3 to 4.4). Then, *H. moloch* (and *H. klossii*)
 405 diverged from the Sumatran/Bornean species, approximately 3.5 MYA (3.0

Introgression among *Hylobates* gibbons

406 to 4.0). *H. muelleri* and *H. agilis*/*H. albibarbis* diverged approximately 3.2
407 MYA (2.8 to 3.7). *H. agilis* and *H. albibarbis* diverged approximately 1.4 MYA
408 (1.2 to 1.7). *H. muelleri muelleri* and *H. muelleri abbotti* diverged
409 approximately 1.4 MYA (1.1 to 1.7).

410 MDS plots of ASD also revealed similar relationships among the
411 *Hylobates* species/subspecies. In the MDS plot containing all *Hylobates*
412 gibbons (n = 40), three clusters appeared. *H. pileatus* formed one cluster, *H.*
413 *lar* formed another cluster, and the Sundaic *Hylobates* species/subspecies
414 formed the final cluster (Fig. 3A). In addition, among the Sundaic *Hylobates*
415 species/subspecies, *H. moloch* separated from the others, which is
416 consistent with the autosomal and mitogenome phylogeny results. When
417 we plotted only the Sundaic *Hylobates* species/subspecies (n = 10), the
418 species/subspecies were all separated from each other (Fig. 3B). *H. agilis*, *H.*
419 *albibarbis*, and *H. muelleri* (*muelleri* and *abbotti* subspecies) were plotted
420 on a line, with *H. agilis* at one end and *H. muelleri* at the other end. *H.*
421 *albibarbis* was located between the two species, closer to *H. agilis* than to
422 *H. muelleri*. Furthermore, *H. muelleri muelleri* (n = 2) and *H. muelleri*
423 *abbotti* (n = 2) were slightly separated from each other in the plot, which
424 indicated the presence of genetic structure-divergence between the two
425 subspecies.

426 The phylogenetic network (Supplementary Fig. S2) showed

Introgression among *Hylobates* gibbons

consistent relationships among the *Hylobates* species/subspecies, as shown in the phylogenetic trees and the MDS plots of ASD. A clear reticulation, which indicates the introgression, was observed in *H. albibarbis* and *H. muelleri* (both *muelleri* and *abbotti* subspecies). In addition, the two *H. agilis* individuals (Sample ID: G80 and G83) formed a reticulation, which may be related to differences in their mitogenomes (Fig. 2A). Furthermore, the network highlighted the presence of a population structure within *H. lar*, in which 14 of 20 individuals were tightly clustered together. This result was similar to the *cytb* tree, in which the same 14 individuals and one additional individual were clustered more tightly compared with the other five individuals (Supplementary Fig. S1).

438

Introgression

The D-statistics calculated for species quartets showed signals of introgression (i.e., absolute Z-score > 3) between *H. albibarbis* and *H. muelleri* (both *muelleri* and *abbotti* subspecies), between *H. pileatus* and *H. lar*, between *H. lar* and *H. agilis*, between *H. albibarbis* and *H. moloch*, and between *H. muelleri abbotti* and *H. moloch* (Table 2, Supplementary Table S6). These results were consistent, regardless of the outgroup species (*N. leucogenys* and *S. syndactylus*) used for the analysis. No other signals of introgression were detected among the other testable quartets of

Introgression among *Hylobates* gibbons

448 *Hylobates* species/subspecies. The D-statistics calculated for individual
 449 quartets also showed signals of introgression in the same pairs of
 450 species/subspecies (and several cases between the *H. muelleri abbotti* and
 451 *H. lar* individuals) (Fig. 4, Supplementary Fig. S3, Supplementary Table S7).
 452 However, the differences in the genetic makeups of the two *H. agilis*
 453 individuals (Sample ID: G80 and G83) affected the D-statistics and Z-scores,
 454 and the signals could be both higher and lower than the threshold
 455 (absolute Z-score > 3), in some cases, such as between *H. albibarbis* and *H.*
 456 *moloch*.

457

Potentially introgressed loci

459 Calculations of f_d and d_{xy} were performed for the pairs of species
 460 whose introgression was detected by D-statistics. We considered autosomal
 461 regions with the top 5% f_d values and the bottom 5% d_{xy} values to be
 462 potentially introgressed loci (Supplementary Table S8). Among those, three
 463 loci (chr7b: 50,000,001-51,000,000; chr7b: 51,000,0001-52,000,000; and
 464 chr10: 68,000,0001-69,000,000) were detected as candidates between *H.*
 465 *lar* and *H. pileatus*, in all of the tested quartets (Fig. 5, Supplementary Table
 466 S8). Furthermore, two loci (chr4: 151,000,001-152,000,000; chr16:
 467 55,000,001-56,000,000) were detected between *H. agilis* and *H. lar*, in
 468 several tested quartets (Supplementary Table S8). Note that the genomic

Introgression among *Hylobates* gibbons

positions shown here are those for *N. leucogenys* ($2n = 52$) and, thus, differ from the exact genomic positions found in *Hylobates* ($2n = 44$).

471

472 Discussion

473 The autosomal phylogeny identified in this study was consistent
474 with the previously reported mitogenome phylogeny (Chan et al. 2010 and
475 its correction) and the mitogenome phylogeny identified in this study,
476 which supports the following scenario: the most recent common ancestor
477 of *Hylobates* originated on the mainland in Southeast Asia, expanded their
478 distribution southward and emerged in Sundaland during the Pliocene era
479 (approximately 3.0 to 4.3 MYA), and speciated due to subsequent isolations
480 (Whittaker et al. 2007; Tinh et al. 2010a; Chan et al. 2013).

481 The early divergence of *H. moloch* on Java, relative to the other
482 Sundaic species living in Sumatra and Borneo, which was previously
483 uncovered by the earlier mitogenome study (Chan et al. 2010) was also
484 supported by the present autosomal phylogeny study. We could not include
485 *H. klossii* on the Mentawai islands in the GRAS-Di analysis; thus, the close
486 relationship between *H. moloch* and *H. klossii* that has been suggested by
487 mtDNA phylogeny (Takacs et al. 2005; Chan et al. 2010) and vocalisation
488 analysis (Geissmann 2002) should be tested in the future.

489 Both the mitogenome and the autosomal trees suggested that

Introgression among *Hylobates* gibbons

490 soon after the divergence of *H. moloch*, *H. agilis*/*H. albibarbis* and *H.*
 491 *muelleri* diverged from each other. ASD also suggests that *H. albibarbis* has
 492 a closer relationship with *H. agilis* than with *H. muelleri* (*muelleri* and
 493 *abbotti* subspecies). This result is consistent with a previous
 494 population-based genetic study (Hirai et al. 2009), a *cytb* phylogeny analysis
 495 (Thinh et al. 2010a), and a previous taxonomy classification, which included
 496 *H. albibarbis* as a subspecies of *H. agilis*, based on vocalisations (Marshall
 497 and Sugardjito 1986).

498 At present, *H. albibarbis* and *H. muelleri* are distributed in Borneo,
 499 whereas *H. agilis* is distributed in Sumatra. The current distribution may
 500 have occurred as follows; the common ancestor of *H. agilis* and *H.*
 501 *albibarbis* was first distributed only in Sumatra, and approximately 1.4 MYA,
 502 the proto-*H. albibarbis* extended their distribution to southwestern Borneo
 503 and either replaced or admixed with *Hylobates* gibbons (likely *H. muelleri*)
 504 that once lived in that region (Thinh et al. 2010a). Lower sea levels and the
 505 appearance of land bridges in the Sundaic region during glacial periods
 506 (Woodruff 2010) likely made the migration between Sumatra and Borneo
 507 possible.

508 Alternatively, the common ancestor of *H. agilis* and *H. albibarbis*
 509 may have widely distributed, from Sumatra to southwestern Borneo, after
 510 the divergence from proto-*H. muelleri*, approximately 3.2 MYA, and 1.4

Introgression among *Hylobates* gibbons

511 MYA may represent the time point at which the forests of Sumatra and
512 southwestern Borneo became disconnected. Interestingly, the estimated
513 mtDNA divergence time between *H. muelleri muelleri* and *H. muelleri*
514 *abbotti* was also approximately 1.4 MYA. Thus, the fragmentation of forests
515 and the isolation of gibbon populations in the Sundaic region may have
516 become more significant during this period than before.

517 Both the mtDNA divergence time and the ASD of autosomal SNVs
518 between *H. muelleri muelleri* and *H. muelleri abbotti* (0.0697 to 0.0701; n =
519 4 pairs) were similar to (or slightly larger than) those between *H. agilis* and
520 *H. albibarbis* (0.0671–0.0693; n = 6 pairs). Therefore, these results suggest
521 that these two taxa may be better classified as different species (i.e., *H.*
522 *muelleri* and *H. abbotti*), which was proposed by Thinh et al. (2010a).
523 However, these results are not conclusive, because the sample size of this
524 study was limited and only two individuals in each subspecies were
525 analysed. In addition, *H. muelleri funereus* was not analysed in this study.
526 The analysis of a larger number of samples, including *H. muelleri funereus*,
527 remains necessary to determine whether our results successfully captured
528 the genetic diversity among *H. muelleri* and to draw conclusions regarding
529 their taxonomic statuses.

530 The GRAS-Di analysis detected introgression signals between
531 several pairs of *Hylobates* species/subspecies in Patterson's D-statistics. The

Introgression among *Hylobates* gibbons

532 signals of introgression between *H. lar* and *H. pileatus*, between *H.*
 533 *albibarbis* and *H. muelleri* (both *muelleri* and *abbotti* subspecies), and
 534 between *H. lar* and *H. agilis* were detected in every individual of each
 535 species. Interestingly, these three species pairs have been known to form
 536 contact/hybrid zones at the species boundaries (Brockelman and Gittins
 537 1984); however, at present, the contact/hybrid zones are limited to small
 538 areas. Because every individual of the species demonstrated signals of
 539 introgression, the limited hybridisation and introgression that has recently
 540 been observed in the contact/hybrid zones appears insufficient to explain
 541 the results. Following the glacial cycles that formed the Sundaland and
 542 fragmented the forests (Woodruff 2010), the shapes of species boundaries
 543 and hybrid zones are likely to have changed over time.

544 The degree of admixed ancestry that can be identified in
 545 individuals from the same species/subspecies may not be uniform across
 546 their distribution. Two individuals of *H. agilis* (Sample ID: G80 and G83)
 547 showed different levels of D-statistic and Z-score values in several
 548 combinations, which likely reflects differences in the amounts of *H.*
 549 *lar*-derived genetic components in their genetic makeups. Individual G80,
 550 which has a mitogenome that is closely related to *H. lar* (*vestitus*
 551 subspecies), showed higher absolute D-statistics in tests with *H. lar*. This
 552 type of mtDNA has been observed in several *H. agilis* individuals from

Introgression among *Hylobates* gibbons

Sumatra, in a previous study (Hirai et al. 2009). Thus, the mtDNA of G80 likely introgressed from *H. lar* to *H. agilis* in the past, similar to the observed mtDNA introgression from *H. pileatus* to *H. lar* (Matsudaira et al. 2013). The frequency and distribution of this mtDNA group among *H. agilis* are not known, but the differences in the mtDNA and the D-statistics may reflect the distance from the species boundary between *H. lar* and *H. agilis*, which is located around Lake Toba. Because differences in the genetic makeup between G80 and G83 affected the D-statistic values, future studies of *Hylobates* gibbons should be aware of potential hidden genetic structures within a species.

Although the introgression signals were not robust among every individual, introgression between *H. albibarbis* and *H. moloch* and between *H. muelleri abbotti* and *H. moloch* were also suggested. For these two species pairs, the current distributions are separated by seas; however, these distributions may have contacted each other on land bridges, when the sea level was low. These results suggested that whenever the distribution range of two *Hylobates* species formed contact regions, hybrid zones were formed and that the reproductive barriers between *Hylobates* species, other than geographical barriers, are incomplete. However, the degrees of introgression may vary among hybrid zones, as suggested by previous observations. Although the degree of introgression may be limited

Introgression among *Hylobates* gibbons

574 between *H. lar* and *H. pileatus* within the hybrid zone at Khao Yai, Thailand
575 (Brockelman and Gittins 1984), a hybrid swarm (a stable hybrid/admixed
576 population) has formed between *H. albibarbis* and *H. muelleri* within a
577 hybrid zone at the Barito headwaters, central Borneo (Mather 1992).

578 Previous studies have suggested possible factors that may limit
579 hybridisation and introgression in these contact/hybrid zones, such as
580 positive assortative mating and the reduced fitness of hybrid gibbons
581 (Brockelman and Gittins 1984, Mather 1992, Suwanvecho and Brockelman
582 2012; Asensio et al. 2017). Further genetic studies that focus on such
583 hybrid zones are expected to reveal factors that affect the frequency of
584 hybridisation and the degree of introgression among species.

585 Some potentially introgressed regions were detected by assessing
586 their f_d and d_{xy} values; however, this analysis was unable to detect
587 particular genes or sets of introgressed SNVs because the density of SNVs
588 available in this study was low; thus, we were only able to conduct the test
589 using large window sizes. Intensive sequencing efforts, such as the WGS of
590 *Hylobates* species, will reveal additional details regarding the introgressed
591 regions detected in this study, and will likely detect additional regions
592 across the genome.

593 Overall, the GRAS-Di analysis, which successfully identified
594 hundreds of thousands of autosomal SNVs, effectively resolved the species

Introgression among *Hylobates* gibbons

595 phylogeny of *Hylobates* gibbons (except *H. klossii*) and revealed the
 596 presence of hybridisation and introgression throughout the evolution of
 597 *Hylobates*. Importantly, this study suggested that the phenotypic diversity
 598 of *Hylobates* gibbons that is currently observed, such as in pelage colour
 599 and vocalisation, represent the products of both divergence and
 600 hybridisation/introgression. This study supports the current view of
 601 hybridisation/introgression as driving factors in the evolution of primates
 602 and other taxa. Future studies of *Hylobates* gibbons that include more
 603 samples with known origins and WGS studies will likely uncover additional
 604 details regarding the contributions of genetic changes that have occurred
 605 during divergence and introgression to species/subspecies differences and
 606 diversity among *Hylobates*.

607

608 **Acknowledgements**

609 This study was supported by JSPS KAKENHI [Grant Numbers JP18H02516
 610 and JP19K16240].

611

612 **Conflict of interest**

613 The authors declare that they have no conflicts of interest.

614

615 **Data archiving**

Introgression among *Hylobates* gibbons

616 The raw read files produced by the GRAS-Di analysis were deposited in
617 DDBJ Sequence Read Archive (DRA), with BioProject Accession No.
618 PRJDB9913. The mtDNA sequences were deposited in the
619 DDBJ/EMBL/NCBI nucleotide database, with Accession No.
620 LC548011-LC548056. Please see Supplementary Table S1 for the details.

621

622 **References**

623 Andayani N, Morales JC, Forstner MRJ, Supriatna J, Melnick DJ (2001)
624 Genetic variability in mtDNA of the silvery gibbon: implications for the
625 conservation of a critically endangered species. *Conserv Biol* **15**:
626 770-775.

627 Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016)
628 Harnessing the power of RADseq for ecological and evolutionary
629 genomics. *Nat Rev Genet* **17**: 81-92.

630 Arnason U, Gullberg A, Xu X (1996) A complete mitochondrial DNA
631 molecule of the white-handed gibbon, *Hylobates lar*, and comparison
632 among individual mitochondrial genes of all hominoid genera.
633 *Hereditas* **124**: 185-189.

634 Asensio N, José-Domínguez JM, Kongrit C, Brockelman WY (2017) The
635 ecology of white-handed and pileated gibbons in a zone of overlap and
636 hybridization in Thailand. *Am J Phys Anthropol* **163**: 716-728.

Introgression among *Hylobates* gibbons

- 637 Beichman AC, Huerta-Sanchez E, Lohmueller KE (2018) Using genomic data
638 to infer historic population dynamics of nonmodel organisms. *Annu Rev*
639 *Ecol Evol Syst* **49**: 433-456.
- 640 Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH
641 (2019) Direct estimation of mutations in great apes reconciles
642 phylogenetic dating. *Nat Ecol Evol* **3**: 286-292.
- 643 Bond M, Tejedor MF, Campbell KEJ, Chornogubsky L, Novo N, Goin F (2015)
644 Eocene primates of South America and the African origins of New
645 World monkeys. *Nature* **520**: 538-541.
- 646 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D et al. (2014)
647 BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS*
648 *Comput Biol* **10**: e1003537.
- 649 Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza
650 LL (1994) High resolution of human evolutionary trees with
651 polymorphic microsatellites. *Nature* **368**: 455-457.
- 652 Brockelman WY, Gittins SP (1984) Natural hybridization in the *Hylobates lar*
653 species group: implications for speciation in gibbons. In: Brockelman
654 WY, Preuschoft H, Chivers DJ, Creel N (eds) *The Lesser Apes:*
655 *Evolutionary and Behavioural Biology*, Edinburgh University Press:
656 Edinburgh. pp 498–532.
- 657 Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D et al. (2002) A

Introgression among *Hylobates* gibbons

- 658 new hominid from the Upper Miocene of Chad, Central Africa. *Nature*
659 **418**: 145-51.
- 660 Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B,
661 Huddleston J et al. (2014) Gibbon genome and the fast karyotype
662 evolution of small apes. *Nature* **513**: 195-201.
- 663 Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Pei KJ-C et al.
664 (2010) Mitochondrial genome sequences effectively reveal the
665 phylogeny of *Hylobates* gibbons. *PLoS ONE* **5**: e14419.
- 666 Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Pei KJ-C et al.
667 (2013) Inferring the evolutionary histories of divergences in *Hylobates*
668 and *Nomascus* gibbons through multilocus sequence data. *BMC Evol*
669 *Biol* **13**: 82.
- 670 Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Vigilant L (2012) A
671 comparative analysis of Y chromosome and mtDNA phylogenies of the
672 *Hylobates* gibbons. *BMC Evol Biol* **12**: 150.
- 673 Chatterjee HJ (2006) Phylogeny and biogeography of gibbons: a
674 dispersal-vicariance analysis. *Int J Primatol* **27**: 699-712.
- 675 Chivers DJ (1977) The lesser apes. In: Prince Rainier III of Monaco HSH,
676 Bourne GH (eds) *Primate Conservation*, Academic Press: New York. pp
677 539-598.
- 678 Creel N, Preuschoft H (1984) Systematics of the lesser apes: a quantitative

Introgression among *Hylobates* gibbons

- 679 taxonomic analysis of craniometric and other variables. In: Brockelman
680 WY, Preuschoft H, Chivers DJ, Creel N (eds) *The Lesser Apes:*
681 *Evolutionary and Behavioural Biology*, Edinburgh University Press:
682 Edinburgh. pp 562–613.
- 683 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al.
684 (2011) The variant call format and VCFtools. *Bioinformatics* **27**:
685 2156-2158.
- 686 de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J
687 et al. (2016) Chimpanzee genomic diversity reveals ancient admixture
688 with bonobos. *Science* **354**: 477-481.
- 689 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy
690 and high throughput. *Nucleic Acids Res* **32**: 1792-1797.
- 691 Enoki H, Takeuchi Y (2018) New genotyping technology, GRAS-Di, using next
692 generation sequencer. Proceedings of the Plant and Animal genome
693 conference XXVI. San Diego, CA.
694 <https://pag.confex.com/pag/xxvi/meetingapp.cgi/Paper/29067>
- 695 Fan P-F, He K, Chen X, Ortiz A, Zhang B, Zhao C et al. (2017) Description of a
696 new species of *Hoolock* gibbon (Primates: Hylobatidae) based on
697 integrative taxonomy. *Am J Primatol* **79**: e22631.
- 698 Fan Z, Zhou A, Osada N, Yu J, Jiang J, Li P et al. (2018) Ancient hybridization
699 and admixture in macaques (genus *Macaca*) inferred from whole

Introgression among *Hylobates* gibbons

- 700 genome sequences. *Mol Phylogenet Evol* **127**: 376-386.
- 701 Finstermeier K, Zinner D, Brameier M, Meyer M, Kreuz E, Hofreiter M et al.
- 702 (2013) A mitogenomic phylogeny of living primates. *PLoS ONE* **8**:
- 703 e69504.
- 704 Garza JC, Woodruff D (1992) A phylogenetic study of the gibbons
- 705 (*Hylobates*) using DNA obtained noninvasively from hair. *Mol*
- 706 *Phylogenet Evol* **1**: 202-210.
- 707 Geissmann T (2002) Taxonomy and evolution of gibbons. *Evol Anthropol*
- 708 *Supple 1*: 28-31.
- 709 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M et al. (2010) A
- 710 draft sequence of the Neandertal genome. *Science* **328**: 710-722.
- 711 Groves CP (1972) Systematics and phylogeny of gibbons. In: Rumbaugh DM
- 712 (ed) *Gibbon and Siamang: A Series of Volumes on the Lesser Apes Vol 1*,
- 713 Karger: Basel. pp. 1-89.
- 714 Groves CP (2001) *Primate Taxonomy*, Smithsonian Institution Press:
- 715 Washington, DC.
- 716 Hayashi S, Hayasaka K, Takenaka O, Horai S (1995) Molecular phylogeny of
- 717 gibbons inferred from mitochondrial DNA sequences: preliminary
- 718 report. *J Mol Evol* **41**: 359-365.
- 719 Hirai H, Hayano A, Tanaka H, Mootnick AR, Wijayanto H,
- 720 Perwitasari-Farajallah D (2009) Genetic differentiation of agile gibbons

Introgression among *Hylobates* gibbons

- 721 between Sumatra and Kalimantan in Indonesia. In: Lappan S, Whittaker
722 DJ (eds) *The Gibbons: New Perspectives on Small Ape Socioecology and*
723 *Population Biology*, Springer: New York. pp 37-50.
- 724 Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African
725 origin of modern humans revealed by complete sequences of
726 hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* **92**: 532-536.
- 727 Hosoya S, Hirase S, Kikuchi K, Nanjo K, Nakamura Y, Kohno H et al. (2019)
728 Random PCR-based genotyping by sequencing technology GRAS-Di
729 (genotyping by random amplicon sequencing, direct) reveals genetic
730 structure of mangrove fishes. *Mol Ecol Resour* **19**: 1153-1163.
- 731 Huson DH, Bryant D (2006) Application of phylogenetic networks in
732 evolutionary studies. *Mol Biol Evol* **23**: 254-267.
- 733 Israfil H, Zehr SM, Mootnick AR, Ruvolo M, Steiper ME (2011) Unresolved
734 molecular phylogenies of gibbons and siamangs (Family: Hylobatidae)
735 based on mitochondrial, Y-linked, and X-linked loci indicate a rapid
736 Miocene radiation or sudden vicariance event. *Mol Phylogenet Evol* **58**:
737 447-455.
- 738 Ishida T, Yamamoto K, Ishimoto G, Shotake T, Takenaka O, Nozawa K et al.
739 (1985) A field study of infection with human T-cell leukemia virus
740 among Asian primates. *Microbiol Immunol* **29**: 839-46.
- 741 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017)

Introgression among *Hylobates* gibbons

- 742 ModelFinder: fast model selection for accurate phylogenetic estimates.
743 *Nat Methods* **14**: 587-589.
- 744 Kelley J (2002) The hominoid radiation in Asia. In: Hartwig W (ed) *The*
745 *Primate Fossil Record*, Cambridge University Press: Cambridge. pp
746 369–384.
- 747 Kim SK, Carbone L, Becquet C, Mootnick AR, Li DJ, de Jong PJ et al. (2011)
748 Patterns of genetic variation within and between gibbon species. *Mol*
749 *Biol Evol* **28**: 2211-2218.
- 750 Lewis PO (2001) A likelihood approach to estimating phylogeny from
751 discrete morphological character data. *Syst Biol* **50**: 913-925.
- 752 Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs
753 with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].
- 754 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The
755 Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:
756 2078-9.
- 757 Liedigk R, Roos C, Brameier M, Zinner D (2014) Mitogenomics of the Old
758 World monkey tribe Papionini. *BMC Evol Biol* **14**: 176.
- 759 Marshall J, Sugardjito J (1986) Gibbon systematics. In: Swindler DR, Erwin J
760 (eds) *Comparative Primate Biology, Volume I: Systematics, Evolution,*
761 *and Anatomy*, Alan R. Liss: New York. pp 137-185.
- 762 Martin M (2011) Cutadapt removes adapter sequences from

Introgression among *Hylobates* gibbons

- 763 high-throughput sequencing reads. *EMBnet.journal* **17**: 10-12.
- 764 Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA–BABA
765 statistics to locate introgressed loci. *Mol Biol Evol* **32**: 244-257.
- 766 Mather R (1992) A Field Study of Hybrid Gibbons in Central Kalimantan,
767 Indonesia. PhD thesis, Cambridge University.
- 768 Matsudaira K, Ishida T (2010) Phylogenetic relationships and divergence
769 dates of the whole mitochondrial genome sequences among three
770 gibbon genera. *Mol Phylogenet Evol* **55**: 454-459.
- 771 Matsudaira K, Reichard UH, Malaivijitnond S, Ishida T (2013) Molecular
772 evidence for the introgression between *Hylobates lar* and *H. pileatus* in
773 the wild. *Primates* **54**: 33-37.
- 774 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A et al.
775 (2010) The Genome Analysis Toolkit: a MapReduce framework for
776 analyzing next-generation DNA sequencing data. *Genome Res* **20**:
777 1297-303.
- 778 Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast approximation for
779 phylogenetic bootstrap. *Mol Biol Evol* **30**: 1188-1195.
- 780 Nakayama K, Ishida T (2006) Alu-mediated 100-kb deletion in the primate
781 genome: the loss of the agouti signaling protein gene in the lesser apes.
782 *Genome Res* **16**: 485-490.
- 783 Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M,

Introgression among *Hylobates* gibbons

- 784 Desai T et al. (2017) Morphometric, behavioral, and genomic evidence
785 for a new orangutan species. *Curr Biol* **27**: 1-12.
- 786 Nei M, Li W-H (1979) Mathematical model for studying genetic variation in
787 terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**:
788 5269-5273.
- 789 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast
790 and effective stochastic algorithm for estimating maximum-likelihood
791 phylogenies. *Mol Biol Evol* **32**: 268-274.
- 792 Nye J, Laayouni H, Kuhlwilm M, Mondal M, Marques-Bonet T, Bertranpetit J
793 (2018) Selection in the introgressed regions of the chimpanzee genome.
794 *Genome Biol Evol* **10**: 1132-1138.
- 795 Nylander, JAA (2004) MrModeltest v2. Program distributed by the author.
796 Evolutionary Biology Centre, Uppsala University.
797 <https://github.com/nylander/MrModeltest2>
- 798 Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y et al. (2012)
799 Ancient admixture in human history. *Genetics* **192**: 1065-1093.
- 800 Petr M, Vernot B, Kelso J (2019) admixr-R package for reproducible analyses
801 using ADMIXTOOLS. *Bioinformatics* **35**: 3194-3195.
- 802 Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B et
803 al. (2013) Great ape genetic diversity and population history. *Nature*
804 **499**: 471-475.

Introgression among *Hylobates* gibbons

- 805 R Core Team (2019) R: A language and environment for statistical
806 computing. R Foundation for Statistical Computing, Vienna, Austria.
807 <http://www.R-project.org/>
- 808 Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior
809 summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* **67**:
810 901-904.
- 811 Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G et
812 al. (2019) The comparative genomics and complex population history
813 of *Papio* baboons. *Sci Adv* **5**: eaau6947.
- 814 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S et al.
815 (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and
816 model choice across a large model space. *Syst Biol* **61**: 539–542.
- 817 Roos C (2016) Phylogeny and classification of gibbons (Hylobatidae). In:
818 Reichard UH, Hirai H, Barelli C (eds) *Evolution of Gibbons and Siamang:*
819 *Phylogeny, Morphology, and Cognition*, Springer: New York. pp
820 151–164.
- 821 Rylands AB, Mittermeier RA (2014) Primate taxonomy: species and
822 conservation. *Evol Anthropol* **23**: 8-10.
- 823 Senut B, Pickford M, Gommery D, Mein P, Cheboi K, Coppens Y (2001). First
824 hominid from the Miocene (Lukeino formation, Kenya). *C R Acad Sci*
825 **332**: 137–144.

Introgression among *Hylobates* gibbons

- 826 Smith J, Kronfrost MR (2013) Do *Heliconius* butterfly species exchange
827 mimicry alleles? *Biol Lett* **9**: 20130503.
- 828 Stevens NJ, Seiffert ER, O'Connor PM, Roberts EM, Schmitz MD, Krause C
829 (2013) Palaeontological evidence for an Oligocene divergence between
830 Old World monkeys and apes. *Nature* **497**: 611-614.
- 831 Suwanvecho U, Brockelman WY (2012) Interspecific territoriality in gibbons
832 (*Hylobates lar* and *H. pileatus*) and its effects on the dynamics of
833 interspecies contact zones. *Primates* **53**: 97-108.
- 834 Takacs Z, Morales JC, Geissmann T, Melnick DJ (2005) A complete
835 species-level phylogeny of the Hylobatidae based on mitochondrial
836 ND3–ND4 gene sequences. *Mol Phylogenet Evol* **36**: 456-467.
- 837 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6:
838 molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**:
839 2725-2729.
- 840 Thinh VN, Mootnick AR, Geissmann T, Li M, Ziegler T, Agil M et al. (2010a)
841 Mitochondrial evidence for multiple radiations in the evolutionary
842 history of small apes. *BMC Evol Biol* **10**: 74.
- 843 Thinh VN, Rawson B, Hallam C, Kenyon M, Nadler T, Walter L et al. (2010b)
844 Phylogeny and distribution of crested gibbons (genus *Nomascus*) based
845 on mitochondrial cytochrome b gene sequence data. *Am J Primatol* **72**:
846 1047-1054.

Introgression among *Hylobates* gibbons

- 847 van der Valk T, Gonda CM, Silegowa H, Almanza S, Sifuentes-Romero I, Hart
848 TB et al. (2020) The genome of the endangered Dryas monkey provides
849 new insights into the evolutionary history of the vervets. *Mol Biol Evol*
850 **37**: 183-194.
- 851 Whittaker DJ, Morales JC, Melnick DJ (2007) Resolution of the *Hylobates*
852 phylogeny: congruence of mitochondrial D-loop sequences with
853 molecular, behavioral, and morphological data sets. *Mol Phylogent Evol*
854 **45**: 620-628.
- 855 Woodruff DS (2010) Biogeography and conservation in Southeast Asia: how
856 2.7 million years of repeated environmental fluctuations affect today's
857 patterns and the future of the remaining refugial-phase biodiversity.
858 *Biodivers Conserv* **19**: 919-941.
- 859 Xu X, Arnason U (1996) The mitochondrial DNA molecule of Sumatran
860 orangutan and a molecular proposal for two (Bornean and Sumatran)
861 species of orangutan. *J Mol Evol* **43**: 431-437.
- 862 Xue C, Raveendran M, Harris RA, Fawcett GL, Liu X, White S et al. (2016)
863 The population genomics of rhesus macaques (*Macaca mulatta*) based
864 on whole-genome sequences. *Genome Res* **26**: 1651-1662.
- 865 Zichello JM (2018) Look in the trees: Hylobatids as evolutionary models for
866 extinct hominins. *Evol Anthropol* **27**: 142-146.
- 867 Zinner D, Wertheimer J, Liedigk R, Groeneveld LF, Roos C (2013) Baboon

Introgression among *Hylobates* gibbons

868 phylogeny as inferred from complete mitochondrial genomes. *Am J*

869 *Phys Anthropol* **150**: 133-140.

870

Introgression among *Hylobates* gibbons

871

872 Fig. 1 Distribution of extant gibbons. Adapted from Thinh et al. 2010a,
873 2010b; Fan et al. 2017.

874

Introgression among *Hylobates* gibbons

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. (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.

Introgression among *Hylobates* gibbons

884 Fig. 3 MDS plot of ASD. (A) A two-dimensional plot of ASD for seven
885 *Hylobates* species/subspecies (n = 40). (B) A two-dimensional plot of ASD
886 for five Sundaic *Hylobates* species/subspecies (n = 10).
887

888

Introgression among *Hylobates* gibbons

889

890 Fig. 4 The ranges of Z-scores for Patterson's D-statistics among *Hylobates*

891 species/subspecies. *N. leucogenys* was using as an outgroup. Species

892 quartets are shown as (P1, P2; P3, Outgroup). Positive Z-scores (> 3)

893 indicate introgression between P1 and P3. Negative Z-scores (< -3) indicate

894 introgression between P2 and P3. Hagi: *H. agilis*, Halb: *H. albibarbis*, Hlar: *H.*

895 *lar*, Hmab: *H. muelleri abbotti*, Hmuu: *H. muelleri muelleri*, Hmol: *H.*

896 *moloch*, Hpil: *H. pileatus*, Nleu: *N. leucogenys*.

897

Introgression among *Hylobates* gibbons

898

899 Fig. 5 Potentially introgressed genomic regions between *H. lar* and *H.*

900 *pileatus*. (A) f_d (Martin et al. 2015) and (B) d_{xy} were calculated for 1 M

901 base-windows, containing more than 100 SNV sites. Genomic regions with

902 the top 5% of f_d values are shown as black dots, whereas all others are

903 shown as grey dots. The genomic regions are presented using the reference

904 genome of *N. leucogenys* (nomLeu3; $2n = 52$) and, thus, do not match the

905 chromosomal conditions of *Hylobates* ($2n = 44$).

906

Introgression among *Hylobates* gibbons

907 Table 1. Nucleotide diversity (π) across the 358 142 SNV sites.

Species	Sample size	π ($\pm$ SD)
<i>H. pileatus</i>	10	0.0198 $\pm$ 0.0811
<i>H. lar</i>	20	0.0489 $\pm$ 0.1240
<i>H. moloch</i>	1	0.0371 $\pm$ 0.1891
<i>H. agilis</i>	2	0.0511 $\pm$ 0.1605
<i>H. albibarbis</i>	3	0.0646 $\pm$ 0.1543
<i>H. muelleri muelleri</i>	2	0.0598 $\pm$ 0.1718
<i>H. muelleri abboti</i>	2	0.0694 $\pm$ 0.1814
<i>S. syndactylus</i>	6	0.0415 $\pm$ 0.1151
<i>N. leucogenys</i>	1	0.0462 $\pm$ 0.2100

908

909

Introgression among *Hylobates* gibbons

910 Table 2. Patterson's D-statistics by species. *N. leucogenys* was used as an
911 outgroup. Positive Z-scores (> 3) indicate the introgression between P1 and
912 P3. Negative Z-scores (< -3) indicate the introgression between P2 and P3.

P1	P2	P3	Outgroup	D	SE	Z
<i>H. moloch</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.11	0.01	-9.28
<i>H. agilis</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.11	0.01	-9.22
<i>H. albibarbis</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.11	0.01	-9.89
<i>H. m. muelleri</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.11	0.01	-9.88
<i>H. m. abbotti</i>	<i>H. lar</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.11	0.01	-9.68
<i>H. agilis</i>	<i>H. moloch</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.01	0.01	1.00
<i>H. albibarbis</i>	<i>H. moloch</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.00	0.01	-0.01
<i>H. m. muelleri</i>	<i>H. moloch</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.00	0.01	0.02
<i>H. m. abbotti</i>	<i>H. moloch</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.00	0.01	0.26
<i>H. albibarbis</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.02	0.01	-1.66
<i>H. m. muelleri</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.02	0.01	-1.29
<i>H. m. abbotti</i>	<i>H. agilis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	-0.01	0.01	-0.99
<i>H. m. muelleri</i>	<i>H. albibarbis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.00	0.01	0.05
<i>H. m. abbotti</i>	<i>H. albibarbis</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.00	0.01	0.41
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. pileatus</i>	<i>N. leucogenys</i>	0.00	0.01	-0.37
<i>H. agilis</i>	<i>H. moloch</i>	<i>H. lar</i>	<i>N. leucogenys</i>	0.07	0.01	5.86
<i>H. albibarbis</i>	<i>H. moloch</i>	<i>H. lar</i>	<i>N. leucogenys</i>	0.01	0.01	1.07

Introgression among *Hylobates* gibbons

<i>H. m. muelleri</i>	<i>H. moloch</i>	<i>H. lar</i>	<i>N. leucogenys</i>	0.01	0.01	0.87
<i>H. m. abbotti</i>	<i>H. moloch</i>	<i>H. lar</i>	<i>N. leucogenys</i>	0.02	0.01	1.66
<i>H. albibarbis</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.08	0.01	-7.96
<i>H. m. muelleri</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.07	0.01	-6.34
<i>H. m. abbotti</i>	<i>H. agilis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.06	0.01	-5.59
<i>H. m. muelleri</i>	<i>H. albibarbis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	0.00	0.01	-0.17
<i>H. m. abbotti</i>	<i>H. albibarbis</i>	<i>H. lar</i>	<i>N. leucogenys</i>	0.01	0.01	0.97
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. lar</i>	<i>N. leucogenys</i>	-0.01	0.01	-1.14
<i>H. albibarbis</i>	<i>H. agilis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	0.04	0.01	4.07
<i>H. m. muelleri</i>	<i>H. agilis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	0.03	0.01	2.29
<i>H. m. abbotti</i>	<i>H. agilis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	0.05	0.01	3.88
<i>H. m. muelleri</i>	<i>H. albibarbis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.01	0.01	-0.94
<i>H. m. abbotti</i>	<i>H. albibarbis</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	0.01	0.01	1.04
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. moloch</i>	<i>N. leucogenys</i>	-0.02	0.01	-2.06
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. agilis</i>	<i>N. leucogenys</i>	0.02	0.01	1.97
<i>H. m. muelleri</i>	<i>H. m. abbotti</i>	<i>H. albibarbis</i>	<i>N. leucogenys</i>	0.02	0.01	2.06
<i>H. agilis</i>	<i>H. albibarbis</i>	<i>H. m. muelleri</i>	<i>N. leucogenys</i>	-0.13	0.01	-15.43
<i>H. agilis</i>	<i>H. albibarbis</i>	<i>H. m. abbotti</i>	<i>N. leucogenys</i>	-0.13	0.01	-15.70