Patterns of Genetic Variation Within and Between Gibbon Species

Sung K. Kim,†,[1] Lucia Carbone,†,[2,3] Celine Becquet,[1] Alan R. Mootnick,[4] David Jiang Li,[2] Pieter J. de Jong,[2] and Jeffrey D. Wall*,[1,5]

[1] Institute for Human Genetics, University of California San Francisco
[2] Children's Hospital of Oakland Research Institute
[3] Department of Behavioral Neuroscience, Oregon Health and Science University
[4] Gibbon Conservation Center
[5] Department of Epidemiology and Biostatistics, University of California San Francisco
†These authors contributed equally to the work.
*Corresponding author: E-mail: wallj@humgen.ucsf.edu.
Associate editor: Anne Stone

Abstract

Gibbons are small, arboreal, highly endangered apes that are understudied compared with other hominoids. At present, there are four recognized genera and approximately 17 species, all likely to have diverged from each other within the last 5–6 My. Although the gibbon phylogeny has been investigated using various approaches (i.e., vocalization, morphology, mitochondrial DNA, karyotype, etc.), the precise taxonomic relationships are still highly debated. Here, we present the first survey of nuclear sequence variation within and between gibbon species with the goal of estimating basic population genetic parameters. We gathered ~60 kb of sequence data from a panel of 19 gibbons representing nine species and all four genera. We observe high levels of nucleotide diversity within species, indicative of large historical population sizes. In addition, we find low levels of genetic differentiation between species within a genus comparable to what has been estimated for human populations. This is likely due to ongoing or episodic gene flow between species, and we estimate a migration rate between *Nomascus leucogenys* and *N. gabriellae* of roughly one migrant every two generations. Together, our findings suggest that gibbons have had a complex demographic history involving hybridization or mixing between diverged populations.

Key words: population history, chromosomal rearrangements, genetic diversity, gibbon, population genetics.

Introduction

Gibbons are small apes native to the forests of Southeast, South, and East Asia. They belong to the same superfamily as humans and other great apes (Hominoidea), and their common ancestor was the first to branch off from the other hominoids roughly 16–20 Ma (e.g., Matsudaira and Ishida 2010; Van Ngoc, Mootnick, Li et al. 2010). Most gibbon species are considered "endangered" or "critically endangered" (IUCN 2009), and the Hainan gibbon (*Nomascus hainanus*), with approximately 20 extant individuals, is the rarest primate in the world (Mootnick et al. 2007). Despite their conservation importance and despite their distinct characteristics such as accelerated karyotype evolution (cf. Muller et al. 2003; Carbone et al. 2006; Misceo et al. 2008) and high species diversity (Mootnick 2006; IUCN 2009; Van Ngoc, Mootnick, Li et al. 2010), gibbons have mostly been neglected by population genetic studies. We do not have even basic data on levels of nuclear sequence diversity within or divergence between gibbon species.

There are four currently recognized gibbon genera (*Hylobates*, *Nomascus*, *Symphalangus*, and *Hoolock*), which are defined by their different karyotypes: Their diploid chromosome counts vary from $2n = 38–52$. Overall, 17 gibbon species have been identified; however, the global taxonomy for gibbons remains very controversial (Mootnick 2006; Misceo et al. 2008; Van Ngoc, Mootnick, Geissmann et al. 2010; Van Ngoc, Mootnick Li et al. 2010). Gibbons are known for their high rates of chromosomal rearrangements, estimated to be roughly 10–20 times faster than the standard mammalian rate (Misceo et al. 2008). This accelerated rate can be seen in the large number of rearrangements separating the Hylobatidae common ancestor from other hominoids (Muller et al. 2003; Carbone et al. 2006), the numerous rearrangements that separate different gibbon species, as well as the rearrangements that are polymorphic within a species (van Tuinen et al. 1999; Carbone, Mootnick, et al. 2009). In contrast, humans and the other great apes are separated by just two interchromosomal rearrangements—a fusion of two chromosomes that formed human chromosome 2 and a reciprocal translocation that occurred on the gorilla lineage (Dutrillaux et al. 1973). This makes gibbons a very good model to study factors responsible for chromosomal instability in primate genomes.

Previous phylogenetic and taxonomic studies have been conducted for gibbon species using different traits (i.e., vocalization, morphology, mitochondrial DNA [mtDNA],

Open Access

Table 1. Specimen Information.

Symbol	Scientific name	Common name	ISIS#	Institution	Distribution
NLE1	N. leucogenys	northern white-cheeked gibbon	NL605	GCC	S Yunnan, N Lao PDR, NW Vietnam
NLE2	N. leucogenys	northern white-cheeked gibbon	982103	Columbus Zoo	S Yunnan, N Lao PDR, NW Vietnam
NLE3	N. leucogenys	northern white-cheeked gibbon	NL600	GCC	S Yunnan, N Lao PDR, NW Vietnam
NLE4	N. leucogenys	northern white-cheeked gibbon	92	Gladys Porter Zoo	S Yunnan, N Lao PDR, NW Vietnam
NLE5	N. leucogenys	northern white-cheeked gibbon	NL607	GCC	S Yunnan, N Lao PDR, NW Vietnam
NLE6	N leucogenys	northern white-cheeked gibbon	NL606	GCC	S Yunnan, N Lao PDR, NW Vietnam
NLE7	N. leucogenys	northern white-cheeked gibbon	UNK	Kunming Institute	S Yunnan, N Lao PDR, NW Vietnam
NGA1	N. gabriellae	southern buff-cheeked gibbon	96070	Los Angeles Zoo	S Vietnam, S Lao PDR, NE Cambodia
NGA2	N. gabriellae	southern buff-cheeked gibbon	96075	Los Angeles Zoo	S Vietnam, S Lao PDR, NE Cambodia
HLE	Hoolock leucodenys	eastern hoolock gibbon	HL305	GCC	E Myanmar, SW Yunnan, NE India
SSY	S. syndactylus	siamang	SS901	GCC	NW & C Malay Peninsula, Sumatra
HPI1	H. pileatus	pileated gibbon	HP103	GCC	W Cambodia, SE Thailand, SW Lao PDR
HPI2	H. pileatus	pileated gibbon	8097	Gladys Porter Zoo	W Cambodia, SE Thailand, SW Lao PDR
HAGI1	H. agilis	agile gibbon	94135	Fort Wayne Zoo	Sumatra, Malay Peninsula
HAGI2	H. agilis	agile gibbon	94137	Fort Wayne Zoo	Sumatra, Malay Peninsula
HAGI3	H. agilis	agile gibbon	15353	Henry Doorly Zoo	Sumatra, Malay Peninsula
HMO	H. moloch	Javan gibbon	HMO894	GCC	W and C Java
HMU	H. muelleri	Mueller's gibbon	86	Gladys Porter Zoo	Borneo, except for the southwest
HLA	Hylobates lar	lar gibbon	9087	Gladys Porter Zoo	SW Yunnan, N, C, & S Thailand, SE Myanmar, C & S Malay Peninsula, N Sumatra

NOTE.—List of the species, individuals, institution from which each sample was received and distribution of the species. Each gibbon species is abbreviated by the first letter of its genus and the two first letters of its species.

karyotype, etc.) but with conflicting results. Initial genetic variation studies in gibbons have focused on mtDNA (e.g., Hayashi et al. 1995; Takacs et al. 2005; Monda et al. 2007; Whittaker et al. 2007; Matsudaira and Ishida 2010; Van Ngoc, Mootnick, Li et al. 2010), in part due to the difficulty in obtaining high-quality DNA samples from rare and endangered gibbon species. However, the mitochondrion makes up a very small fraction of the genome. A fuller picture of genetic diversity within and between gibbon species and the evolutionary relationships among gibbon species can only be achieved by a comprehensive study of nuclear sequence variation.

In this study, we investigate levels of nuclear genetic diversity in a panel of 19 unrelated individuals from nine different gibbon species, including representatives from all four genera. Because interspecific hybridization has been linked both to speciation and to genome reshuffling (O'Neil et al. 1998; Fontdevila 2005), we are interested in the possibility of detecting gene flow between gibbon species. Our study generated the largest autosomal genetic data set in gibbons, which was made possible, thanks to a unique sample collection of high-quality genomic DNA from captive individuals.

Materials and Methods

Specimen Identification

The species identification of the Hylobatidae in this study (table 1) was based on visual and auditorial examination by Alan Mootnick of the gibbons housed at the Gibbon Conservation Center (GCC) between 1976 and 2011 and the examination of Hylobatidae skins by Alan Mootnick at the following museums: American Museum of Natural History,

New York; British Museum (Natural History), London; Field Museum of Natural History, Chicago; Institute of Ecology and Biological Resources, Hanoi; Harvard Museum of Comparative Zoology, Cambridge; Muséum National d'Histoire Naturelle, Paris; Museum Zoologicum Borgoriense, Bogor; National Museum of Natural History, Washington, DC; Zoological Museum, Vietnam National University, Hanoi; and Zoological Reference Collection, National University of Singapore, Singapore. Additional criteria for determining taxonomic identification followed (Groves 1972; 2001; Marshall and Sugardjito 1986; Geissmann 1995; Mootnick 2006; Van Ngoc, Mootnick, Geissmann, 2010; Mootnick and Fan 2011). Vocalizations of live specimens were compared with Marshall et al. (1984), Marshall JT and Marshall ER (1976, 1978), and Marshall and Sugardjito (1986).

The Hylobatidae in this study that were housed at the GCC, Fort Wayne Children's Zoo, Gladys Porter Zoo, Henry Doorly Zoo, and Los Angeles Zoo were observed in person to confirm their identification. The northern white-cheeked gibbon at the Columbus Zoo was identified by photographs of his sire and dam housed at the Columbus Zoo. The cell line from the Kunming Institute of Zoology was a wild-born N. leucogenys, and this specimen's species identification was also confirmed by Christian Roos.

Sequencing

Genomic DNA was extracted from gibbon whole blood using Gentra Puregene DNA extraction kit (distributed by Qiagen). We used about 100 ng of DNA for each long-range polymerase chain reaction (PCR) reaction. Long-range PCR products were obtained using standard protocols as suggested by Takara Bio, Inc. The specific primers used are available from the authors upon request. We pooled

together the 20 different amplicons produced for each gibbon, and to assure equal representation of all the targeted regions, we normalized the PCR product based on their concentration. We thus obtained 19 different individual-specific pools, which were converted into 19 Illumina paired-end libraries following the standard Illumina protocol. To minimize the number of sequencing reactions, we used the Illumina indexing strategy and created two pools made up of ten indexed libraries each (supplementary table S1, Supplementary Material online). Each of the two pools was sequenced on a single lane of the Illumina flow cell.

Estimating the Sequencing Error Rate

We obtained six reference bacterial artificial chromosomes (BACs) from the gibbon genomic library that was generated using genomic DNA from one of the gibbon individual from our sample set (NLE1). We then compared the BAC sequences with the amplicon sequences generated during this study for NLE1 to estimate the Illumina sequencing error rate. We filtered the NLE1 sequences for bases that both contained an alternate allele and had $>100\times$ coverage. Assuming a binomial distribution, we considered any alternate alleles with a P value < 0.001 to be sequencing errors rather than evidence for polymorphisms. We find that these criteria often select bases with an alternate allele whose coverage ranges from 1–10$\times$. We estimate a sequencing error rate of 0.5% per base, and our results are unchanged regardless of the inclusion or exclusion of repeats (supplementary table S2, Supplementary Material online).

Generating Consensus Sequences for the 20 Amplicons

We implemented a simple decision tree algorithm to generate the final consensus data set for use in our study. For each gibbon and for each base, we required a minimum of 20$\times$ coverage with a minimum averaged PHRED score (cf. Ewing et al. 1998) of 20. Heterozygotes were called if the alternate allele had a coverage rate of 20% and passed the minimum averaged quality score threshold. These criteria were chosen to maximize the sequence length while minimizing the sequencing error rate. This filtering resulted in an average of 60 kb (51 kb lay outside of repeats, cf. supplementary table S3, Supplementary Material online) of genomic sequence data per sample with a total of 2,578 unique single nucleotide polymorphisms (SNPs) located outside repeat regions (supplementary tables S1 and S4 and supplementary fig. S1, Supplementary Material online).

Divergence and Differentiation

To estimate divergence, we considered two measures of nucleotide diversity, π and θ_w. π measures the average number of pairwise differences between any two sequences (Nei and Li 1979; Tajima 1983), whereas θ_w reflects the proportion of sites that are polymorphic in the sample (Watterson 1975). To account for varying sample sizes at each base position, we computed π and θ_w for each SNP separately and normalize by the total sequence length used. We report the

estimates of π and θ_w for each genus, species, and amplicon in supplementary table S7, Supplementary Material online. We calculated F_{ST} as a measure of differentiation, or genetic distance, within our sample (Malécot 1948; Wright 1951; Hudson et al. 1992). F_{ST} is calculated as $1 - \pi_w/\pi_b$, where π_w is the average π for each within-population comparison and π_b is the average number of pairwise differences for sequences sampled from different populations.

Estimating the Parameters of the Isolation–Migration Model Using MIMAR

We wanted to estimate parameters of the isolation–migration model (supplementary fig. S2, Supplementary Material online; Hey and Nielsen 2004) for pairs of species for which we had polymorphism data (i.e., ≥ 2 individuals) and focused on the pair N. leucogenys (NLE) and N. gabriellae (NGA) for which we had the most data. The parameters of interest are m, T, N_A, N_1, and N_2, where m is the migration rate and T is the time when the two populations split. N_A, N_1, and N_2 are the effective population size for the ancestral and newly formed populations, respectively. The data set for this analysis was constructed by first discarding the NLE individual with the most missing data per amplicon and then considering only the sites with data in the remaining six NLE and two NGA gibbon individuals. The final data set consisted of 16 amplicons of ~2.5 kb each (for a total of ~40 kb) and 540 SNPs (supplementary table S5, Supplementary Material online).

To estimate the parameters of the isolation–migration model between NLE and NGA, we applied a modified version of MIMAR to our data set (MIMAR_noanc version 03/19/2010, available at http://przeworski.uchicago.edu/cbecquet/download.html; Becquet and Przeworski 2007). MIMAR uses Markov chain Monte Carlo to estimate the posterior distributions of the parameters from polymorphism data sampled in two recently diverged species at multiple recombining genomic regions. More specifically, MIMAR relies on summaries of the polymorphisms that we calculated for the amplicons: the number of polymorphisms unique to the samples from populations 1 and 2 (S_1 and S_2, respectively) and the number of shared and fixed polymorphisms between the two population samples (S_s and S_f). In addition, we specified the number of sites with data, L, and the inheritance scalar (1 for autosomal and 3/4 for X linked) for each amplicon. To allow for mutation rate variation, we also specified a mutation rate scalar for each amplicon to the ratio of observed and expected divergence with a human sequence (for details, see Becquet and Przeworski 2007). We considered the population recombination rate as a nuisance parameter: that is, at each MCMC step and for each amplicon, the ratio c/μ is sampled from an exponential prior with mean 0.6. c is the generational crossover rate per base pair (assuming the genomic average $c = 1.2 \times 10^{-8}$, cf. Kong et al. 2002), and μ is the assumed mutation rate per base pair per generation ($\mu = 1 \times 10^{-8}$, cf. Roach et al. 2010; 1000 Genomes Project Consortium et al. 2010). With this parameterization, the amplicon-specific population recombination rate is calculated by

$vq_1(L - 1) \times c/\mu$, where $q_1 = 4N_1\mu$ is the value of the population mutation rate per base pair for the MCMC step and $v = w(L_t - 1)/(L - 1)$ is the recombination scalar specified for the amplicon. w is 1 or ½ for an autosomal or X-linked amplicon, and L_t is the total length of the locus before filtering out gaps and missing data.

We ran MIMAR with two different seeds and varied the number of burn-ins and the width of the uniform priors on the parameters to assure convergence of the algorithm. We used the mode of the smoothed marginal posterior distributions estimated by MIMAR as our point estimates of the isolation–migration parameters. We assumed 10 years per generation and $\mu = 1 \times 10^{-8}$ to rescale the time of divergence in years and obtained estimates of the effective populations sizes. To test whether the model estimated by MIMAR provides an adequate fit to the data, we used the posterior predictive probabilities as described in Becquet and Przeworski (2007, 2009). Specifically, we calculated the sums over the 16 amplicons of S_1, S_2, S_s, S_f as well as the means of F_{ST} and π and Tajima's D (Tajima 1989) in each species. To gauge how well the estimated isolation–migration model fits the "observed" data, we calculated the sums or means of these statistics for 10,000 data sets of 16 amplicons simulated by sampling the parameters of the isolation–migration model from the posterior distributions estimated by MIMAR. We calculated the probability of obtaining the "observed" statistic or a more extreme value from the distributions of these statistics expected under the estimated model. Because some of the statistics are discrete, we calculated the randomized probabilities, P^R, as described by Becquet and Przeworski (2009) and used these Bayesian posterior predictive P values as we would regular P values, conservatively considering the model to be a poor fit if the observed value of a statistic falls in the 2.5th percentile tails of the distribution (Meng 1994). Results are shown in supplementary figure S5 (Supplementary Material online). In general, the model provides a reasonable fit to the data.

Results

We used long-range PCR to amplify 20 genomic regions in a panel of 19 individuals, followed by Illumina multiplex sequencing (see Methods). See table 1 for description of the gibbons used for analysis. After quality control filters, we obtained an average of 60 kb of sequence from each sample (supplementary table S1 and S4, Supplementary Material online).

High Genetic Diversity in Gibbon Species

We measured nucleotide diversity for each species using π (Nei and Li 1979; Tajima 1983) and θ_w (Watterson 1975), two standard estimators of the population mutation rate (fig. 1A and supplementary table S7, Supplementary Material online). We observed high levels of nucleotide diversity in Nomascus and Hylobates species (π ranging from 0.23% to 0.26% and 0.24% to 0.47%, respectively) comparable to what has been observed in Bornean ($\pi = 0.27\%$) and Sumatran ($\pi = 0.35\%$) Orangutans, the most diverse of the

great apes (Fischer et al. 2006). The levels of diversity from Symphalangus syndactylus and Hoolock leuconedys were a little lower ($\pi = 0.15\%$ and 0.19%), more in line with levels of diversity observed in Chimpanzees and Gorillas (Fischer et al. 2006), but still higher than levels of genetic diversity in human populations (e.g., Wall et al. 2008). Estimates of the effective population size (N_e) for gibbon species range from 37,500 to 117,500, somewhat larger than comparable estimates for humans and great apes.

Low Levels of Genetic Differentiation between Gibbon Species

We quantified genetic differentiation between species and genera using F_{ST}, which measures allele frequency differences between two sets of samples, and π_b, the expected sequence divergence between single sequences from two species (Malécot 1948; Wright 1951; Hudson et al. 1992 and see Holsinger and Weir 2009 for a review of F_{ST}). We observed high levels of differentiation between gibbon genera (mean $F_{ST} = 0.80$ and mean $\pi_b = 1.50\%$, cf. table 2) similar to the levels of divergence between humans and chimpanzees ($F_{ST} = 0.89$ and $\pi_b = 1.19\%$, cf. Fischer et al. 2006). This is consistent with estimates of divergence between gibbon genera (7–8 Ma, cf. Matsudaira and Ishida 2010; Van Ngoc, Mootnick, Li et al. 2010) that are slightly older than estimates of the human–chimpanzee divergence time (5–7 Ma, e.g., Glazko and Nei 2003; Kumar et al. 2005). In contrast, we observed low levels of differentiation between gibbon species within the same genus, especially among some of the Hylobates species (see fig. 2 and table 2). For example, $F_{ST} = 0.18$ between Hylobates agilis and H. muelleri (95% confidence interval [CI]: 0.10–0.25) and $F_{ST} = -0.10$ between H. moloch and H. muelleri (i.e., no evidence for any genetic differentiation between the two samples). CIs based on bootstrap resampling (supplementary table S8, Supplementary Material online) suggest that these low F_{ST} values are not just the result of small sample size but rather reflect the genetic similarity between species. For comparison, these values are smaller than published estimates of F_{ST} between Eastern and Western chimpanzees (0.32, cf. Fischer et al. 2006) and between different human populations (0.28 between San and Melanesians, cf. Wall et al. 2008). This suggests that divergence between species within a genus is recent and/or that there has been ongoing gene flow between species.

Incomplete Lineage Sorting Among Gibbon Species

We used a unweighted pair group method with arithmetic mean clustering algorithm implemented by the program PHYLIP to generate phylogenetic trees for the whole data as well as for each amplicon separately (Felsenstein 1989) (figs. 1B and supplementary fig. S3, Supplementary Material online). Although each genus is consistently monophyletic, the relative ordering of the four genera changes depending on which amplicon is analyzed (see fig. 1C and D). Furthermore, the genealogical relationships of species within the same genus vary from region to region, and both N. leucogenys and H. agilis are not consistently monophyletic. This pattern is not
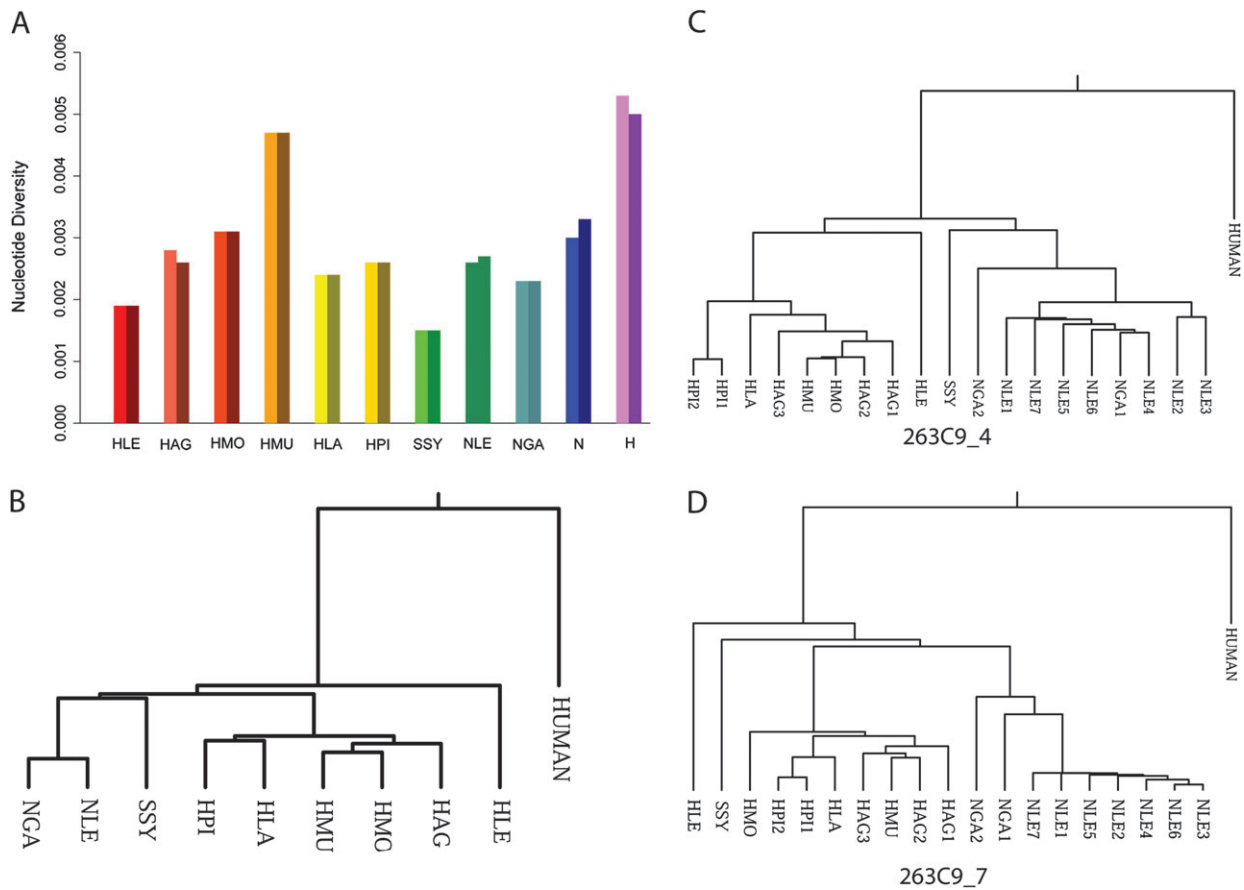
Fig. 1. (A) Estimates of π (left) and θ_w (right) averaged across all amplicons. N and H denote all *Nomascus* and *Hylobates* individuals, respectively. (B) Phylogenetic dendrogram of the nine species used in our study. See table 1 for a list of species abbreviations used. (C–D) The phylogenetic relationships between species for two different amplicons.

particularly surprising, given the conflicting phylogenies previously published for gibbons (e.g., Muller et al. 2003; Takacs et al. 2005; Van Ngoc, Mootnick, Li, et al. 2010).

Phylogenies that are inconsistent across different genomic regions, sometimes referred to as incomplete lineage sorting, result when species have recent split times, moderate levels of gene flow, and/or large N_e values. In the absence of natural selection, the average time from the introduction of a neutral mutation to its fixation is $4N_e g$, where g is the average generation time (e.g., Crow and Kimura 1970). Split times that are near $4N_e g$ years ago or more recent will often lead to incomplete lineage sorting because

there will be a substantial number of polymorphisms shared across species. Older split times, coupled with recurrent gene flow, can also lead to the same result.

Gene Flow Detected between Gibbon Species

To assess the possibility of gene flow between closely related gibbon species, we estimated demographic parameters using a modified version of the program MIMAR (Becquet and Przeworski 2007; see Methods). MIMAR is a Markov chain Monte Carlo–based approach to estimating parameters of a simple isolation–migration model (supplementary fig. S2, Supplementary Material online) that is widely used in

Table 2. Estimates of Averaged F_{ST} and π_b (above and below the diagonal, respectively) for All Pairwise Comparisons.

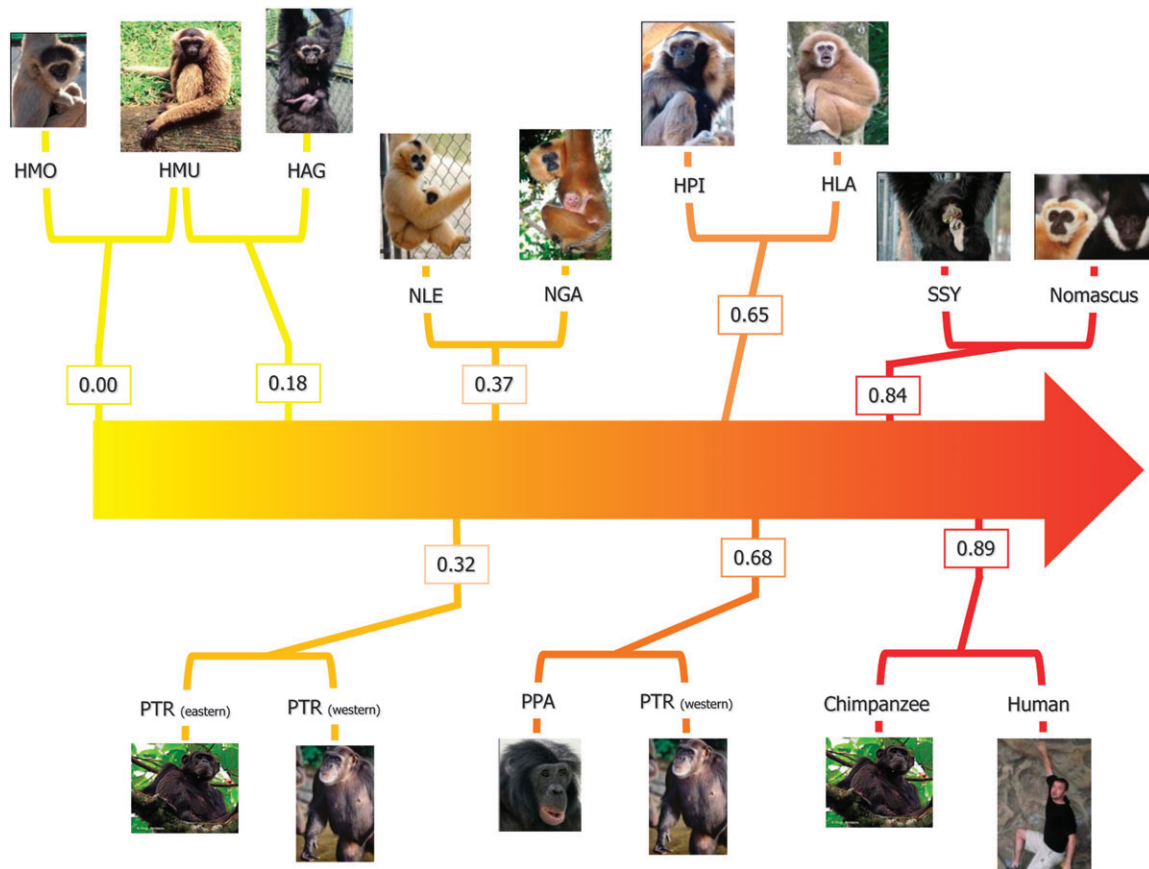| π_b (%)	F_{ST}	*Nomascus*		*Symphalangus*	*Hoolock*	*Hylobates*				
		NLE	NGA	SSY	HLE	HAG	HMO	HPI	HLA	HMU
Nomascus	NLE		0.372	0.839	0.846			0.704		
	NGA	0.390								
Symphalangus	SSY	1.411			0.895			0.756		
Hoolock	HLE	1.594		1.612				0.770		
	HAG						0.385	0.658	0.632	0.180
	HMO					0.483		0.575	0.497	0.000
Hylobates	HPI	1.410		1.404	1.564	0.785	0.672		0.651	0.552
	HLA					0.703	0.549	0.709		0.528
	HMU					0.459	0.359	0.815	0.753	

Fig. 2. Comparative illustration of F_{ST} (a measure of differentiation) for gibbons and other primates. The HMO–HMU comparison was rounded up to 0.

population genetics (e.g., Nielsen and Wakeley 2001; Hey and Nielsen 2004). We applied this method to the *N. leucogenys* and *N. gabriellae* data, the only pair of closely related species with a sample size large enough for our analyses. The two *Nomascus* species share 28.5% of their single nucleotide polymorphisms, which suggests some degree of shared history between the two. MIMAR estimates a split time of approximately 6 Ma (95% credibility interval: 2.8–11.1 Ma; see supplementary table S6 and supplementary figs. S4 and S5, Supplementary Material online), somewhat older than previous divergence time estimates obtained from mtDNA data (e.g., Van Ngoc, Mootnick, Li et al. 2010) and a large ancestral population size with mode 892,000 (95% credibility interval: 165,000–923,000). Only part of this difference can be explained by different assumptions regarding the mutation rate μ. More interestingly, the estimate of the scaled migration rate is $M = 0.44$ (95% credibility interval: 0.26–0.80) equivalent to a rate of about one migrant every two generations. Because we can strongly reject a model with no migration (i.e., $M = 0$), we conclude that the observed genetic similarity between *N. leucogenys* and *N. gabriellae* is due to ongoing or episodic gene flow between the two species.

Discussion

Our goal in this study was to investigate levels of genetic variation and differentiation in gibbon species. Previous studies of DNA sequence variation in gibbons have generally focused on small (generally 1 kb or less) subsets of mtDNA (e.g., Takacs et al. 2005; Monda et al. 2007; Whittaker et al. 2007; Van Ngoc, Mootnick, Geissmann et al. 2010; Van Ngoc, Mootnick, Li et al. 2010). The ability of these studies to make robust conclusions about population history is compromised because the mitochondrion forms a single linkage group, influenced both by chance and by natural selection. The best way to learn about population history in any species is to systematically assess genetic variation across many (unlinked) regions of the genome. Our study of roughly 60 kb of sequence from a panel of 19 gibbon individuals and nine different species is a first step toward that goal. Though modest in size by today's standards, our data were still sufficient to make several key observations—that levels of diversity within species are generally high (relative to other apes), that levels of differentiation between species are often low, and that much more data will be needed to accurately ascertain the phylogenetic relationships between all gibbon species. We discuss the implications of the observations in greater detail below.

Although the primary focus of our study was not on the evolutionary relationships between different gibbon species, our result of discordant phylogenies from different genomic regions suggests that the previously published mtDNA-based gibbon phylogenies must be considered provisional. Due to incomplete lineage sorting, the "gene

tree" differs from one genomic region to another. The true phylogenetic or species tree can only be obtained by combining information from hundreds (or thousands) of separate gene trees. This is especially true for sorting out the branching order for the four gibbon genera because all four taxa seem to have diverged at roughly the same time.

We also observed low levels of differentiation between gibbon species within the same genus (as measured by F_{ST}) and estimated substantial gene flow between *N. leucogenys* and *N. gabriellae*. Migration rates were estimated under a simple isolation–migration model, which assumes a constant symmetric migration rate after the initial split into two daughter populations (see supplementary fig. S2, Supplementary Material online). Biogeographic evidence suggests that this migration may have been episodic instead. Throughout the Pliocene and Pleistocene, gibbons' forest habitats alternatingly expanded and contracted as periodic glaciations led to dramatic changes in climate and sea levels (e.g., Morley and Flenley 1987; Meijaard and Groves 2006). This may have led to a cycle of isolation followed by secondary contact for different gibbon populations, which might explain the nonzero migration rates that we estimate. We also note that hybridization between diverged populations may contribute to genome instability (Fontdevila 2005; Brown and O'Neill 2010), which provides an intriguing possible explanation for the high rate of chromosomal rearrangements in gibbons.

The accelerated rate of chromosomal rearrangements in gibbons is difficult to explain—at present, we do not know if it is a result of a higher (rearrangement) mutation rate within gibbons, a higher fixation rate of segregating rearrangements within gibbons, or both. A priori, one might expect most rearrangements to be deleterious either due to direct disruption of coding or regulatory sequences or due to reduced fertility in heterokaryotypic matings (because single crossovers can lead to unbalanced gametes). Because of this purifying selection, it has been suggested that chromosomal rearrangements are more likely to fix in species with small effective population sizes (N_e) because the effects of genetic drift (relative to selection) will be stronger (e.g., Bush et al. 1977). The observation of high levels of genetic variation in gibbon species, and thus large N_e values, implies that the accelerated rate of chromosomal rearrangements in gibbons cannot be explained by enhanced genetic drift of slightly deleterious rearrangements. Because there is no a priori reason to believe that selection coefficients for chromosomal rearrangements are different for gibbons than they are for other mammals, we posit that the rate at which new chromosomal rearrangements arise is higher in gibbons than in other mammals. This hypothesis is supported by recent observations about the unique molecular and epigenetic structure of gibbon chromosomal breakpoints (Carbone, Harris, et al. 2009).

In conclusion, we present here the first evolutionary analysis of gibbon species based on genomic data. We were able to detect gene flow and observe high genetic diversity. Thanks to the lowering costs of sequencing, we plan to extend these observations by gathering whole-genome sequencing data on more unrelated individuals. The main hurdle for population genetic projects on endangered species, as gibbons, is the availability of DNA samples from enough unrelated individuals. We are exploring whether this obstacle can be overcome by using whole-genome amplification of low-quality DNA accessible with less-invasive approaches (hair, feces, and museum skins).

Supplementary Material

Supplementary tables S1–S8 and supplementary figs. S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

References

1000 Genomes Project Consortium Durbin RM, Abecasis GR, et al. (542 co-authors). 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073.

Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* 17:1505–1519.

Becquet C, Przeworski M. 2009. Learning about modes of speciation by computational approaches. *Evolution* 63:2547–2562.

Brown JD, O'Neill RJ. 2010. Chromosomes, conflict and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet.* 11:291–316.

Bush GL, Case SM, Wilson AC, Patton JL. 1977. Rapid speciation and chromosomal evolution in mammals. *Proc Natl Acad Sci USA.* 74:3942–3945.

Carbone L, Harris RA, Vessere GM, et al. (12 co-authors). 2009. Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet.* 5:e1000538.

Carbone L, Mootnick AR, Nadler T, Moisson P, Ryder O, Roos C, de Jong PJ. 2009. A chromosomal inversion unique to the northern white-cheeked gibbon. *PLoS One.* 4:e4999.

Carbone L, Vessere GM, ten Hallers BF, et al. (14 co-authors). 2006. A high-resolution map of synteny disruptions in gibbon and human genomes. *PLoS Genet.* 2:e223.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. Minneapolis, MN: Burgess Publishing.

Dutrillaux B, Rethore MO, Prieur M, Lejeune J. 1973. Analysis of the structure of chromatids of Gorilla gorilla. Comparison with Homo sapiens and Pan troglodytes. *Humangenetik* 20:343–354.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175–185.

Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). *Cladistics* 5:164–166.

Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S. 2006. Demographic history and genetic differentiation in apes. *Curr Biol.* 16:1133–1138.

Fontdevila A. 2005. Hybrid genome evolution by transposition. *Cytogenet Genome Res.* 110:49–55.

Geissmann T. 1995. Gibbon systematics and species identification. *Int Zoo News.* 42:467–501.

Glazko GV, Nei M. 2003. Estimation of divergence times for major lineages of primate species. *Mol Biol Evol.* 20:424–434.

Groves CP. 1972. Systematics and phylogeny of gibbon. In: Rumbaugh DM, editor. Gibbon and Siamang, evolution, ecology behavior, and captive maintenance. Basel (Switzerland): S. Karger. Vol. 1p. 1–89.

Groves CP. 2001. Primate taxonomy. Washington (DC): Smithsonian Institution Press.

Hayashi S, Hayasaka K, Takenaka O, Horai S. 1995. Molecular phylogeny of gibbons inferred from mitochondrial DNA sequences: preliminary report. *J Mol Evol.* 41:359–365.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. *Genetics* 167:747–760.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nat Rev Genet.* 10:639–650.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589.

IUCN. 2009. IUCN red list of threatened species 2009.2. 2009. [cited 2010 Oct 1]. Available from: http://www.iucnredlist.org

Kong A, Gudbjartsson D, Sainz J, et al. (16 co-authors). 2002. A high-resolution recombination map of the human genome. *Nat Genet.* 31:241–247.

Kumar S, Filipski A, Swarna V, Walker A, Hedges SB. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc Natl Acad Sci USA.* 102:18832–18847.

Malécot G. 1948. Les mathématiques de l'hérédié. Paris (France): Masson.

Marshall JT, Marshall ER. 1976. Gibbons and their territorial songs. *Science* 193:235–237.

Marshall JT, Marshall ER. 1978. The Gibbons. Gainesville (FL): ARA records. (Long playing phonograph disc)

Marshall JT, Sugardjito J. 1986. Gibbon systematics. In: Swindler DR, Erwin J, editors. Comparative primate biology. New York: AR Liss. p. 137–185.

Marshall JT, Sugardjito J, Markaya M. 1984. Gibbons of the lar group: relationships based on voice. In: Preuschoft H, Chivers DJ, Brockelman WY, Creel N, editors. The Lesser apes. Edinburgh (UK): Edinburgh University Press. p. 533–541.

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.

Meijaard E, Groves CP. 2006. The geography of mammals and rivers in mainland southeast Asia. In: Primate Biogeography: progress and prospects. New York: Springer. p. 305–329.

Meng XL. 1994. Posterior predictive p-values. *Ann Stat.* 22:1142–1160.

Misceo D, Capozzi O, Roberto R, Dell'oglio MP, Rocchi M, Stanyon R, Archidiacono N. 2008. Tracking the complex flow of chromosome rearrangements from the Hominoidea Ancestor to extant *Hylobates* and *Nomascus* Gibbons by high-resolution synteny mapping. *Genome Res.* 18:1530–1537.

Monda K, Simmons RE, Kressirer P, Su B, Woodruff DS. 2007. Mitochondrial DNA hypervariable region-1 sequence variation and phylogeny of the concolor gibbons, Nomascus. *Am J Primatol.* 69:1–22.

Mootnick AR. 2006. Gibbon (Hylobatidae) species identification recommended for rescue or breeding centers. *Primate Conserv.* 21:103–138.

Mootnick AR, Fan P-F. 2011. A comparative study of crested gibbons (*Nomascus*). *Am J Primatol.* 73(2):135–154.

Mootnick AR, Xiaoming W, Moisson P, Chan BPL, Fellowes JR, Nadler T. 2007. Hainan gibbon, *Nomascus hainanus* (Thomas, 1892). In: Mittermeier RA, Ratsimbazafy J, Rylands JB, et al., editors. Primates in peril: the World's 25 most endangered primates, 2006–2008. Arlington (TX): IUCN/SSC Primate Specialist group. p. 16–17.

Morley RJ, Flenley JR. 1987. Late Cainozoic vegetational and environmental changes in the Malay Archipelago. In: Whitmore TC, editor. Biogeographical Evolution of the Malay Archipelago. Oxford: Clarendon Press. p. 50–59.

Muller S, Hollatz M, Wienberg J. 2003. Chromosomal phylogeny and evolution of gibbons (Hylobatidae). *Hum Genet.* 113:493–501.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 76:5269–5273.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–896.

O'Neil RJ, O'Neil MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodeling in an interspecific mammalian hybrid. *Nature* 393:68–72.

Roach JC, Glusman G, Smit AF, et al. (15 co-authors). 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–639.

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

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Takacs Z, Morales JC, Geissmann T, Melnick DJ. 2005. A complete species-level phylogeny of the Hylobatidae based on mitochondrial ND3-ND4 gene sequence. *Mol Phylogenet Evol.* 36:456–467.

Van Ngoc T, Mootnick AR, Geissmann T, Vu Ngoc T, Nadler T, Roos C. 2010. A new species of crested gibbon from the central Annamite mountain range. *Vietnam J Primatol.* 4:1–12.

Van Ngoc T, Mootnick AR, Li M, Ziegler T, Agil M, Moisson P, Nadler T, Walter L, Roos C. 2010. Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. *BMC Evol Biol.* 10:74.

van Tuinen P, Mootnick AR, Kingswood SC, Hale DW, Kumamoto AT. 1999. Complex, compound inversion/translocation polymorphism in an ape: presumptive intermediate stage in the karyotypic evolution of the agile gibbon *Hylobates agilis*. *Am J Phys Anthropol.* 110:129–142.

Wall JD, Cox MP, Mendez FL, Woerner A, Severson T, Hammer MF. 2008. A novel DNA sequence database for analyzing human demographic history. *Genome Res.* 18:1354–1361.

Watterson G. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.

Whittaker DJ, Morales JC, Melnick DJ. 2007. Resolution of the *Hylobates* phylogeny: congruence of mitochondrial D-loop sequences with molecular, behavioral, and morphological datasets. *Mol Phylogenet Evol.* 45:620–628.

Wright S. 1951. The genetical structure of populations. *Ann Eugen.* 15:323–354.