

Unexpected occurrence of the western hoolock gibbon (*Hoolock hoolock*) in southwestern China suggests human-mediated translocation along the Southwestern Silk Road

Xia Huang

Guangzhou University

Qingjian Chen

Guangzhou University

Zhenyun Huang

Guangzhou University

Alejandra Ortiz

New York University

Xiaoyun Wang

Guangzhou University

Wenyi Du

Guangzhou University

Xian Wang

Kunming Institute of Zoology

Kairui Chen

Guangzhou University

Jiazhao Xie

Guangzhou University

Yuhan He

Guangzhou University

Song Li

Kunming Institute of Zoology

Dongdong Wu

Kunming Institute of Zoology

Xiaoming Zhang

Kunming Institute of Zoology

Pengfei Fan

Sun Yat-sen University

Zhong Cao

Guangzhou University

Kai He

hekai@gzhu.edu.cn

Guangzhou University

Article

Keywords: Hoolock hoolock, human-mediated translocation, museum genomics, AI-assisted dental identification, Southwestern Silk Road

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11194159/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Human activities have reshaped the distributions of animals for thousands of years, yet historical movements of wild species across biogeographic barriers often remain undetected. We investigated two hoolock gibbon specimens collected in 1960 from Datang, Tengchong, southwestern China, within the recognized range of *Hoolock tianxing* but far outside the known range of *H. hoolock*. We evaluated their taxonomic identity and geographic origin using curatorial records, deep-learning-based identification of upper-molar images, pelage morphology, mitochondrial phylogenetics, and genome-wide autosomal SNP analyses. The dental model assigned the available upper molar of the skull specimen to *H. hoolock*, and the pelage of the skin specimen was also consistent with this identification. Genomic analyses placed both specimens within *H. hoolock*, whereas mitochondrial data associated them with the southern mitochondrial lineage from central and southern Assam, northeastern India. Genome-wide analyses detected no substantial excess allele sharing with geographically adjacent *H. tianxing*, and both specimens showed relatively high heterozygosity, providing little support for an origin in a long-isolated natural population in western Yunnan. The occurrence of these specimens in Tengchong is therefore best explained by historical human-mediated translocation. The Southwestern Silk Road provides a plausible transport corridor, although the exact time of translocation remains unresolved. This study demonstrates how museum collections, genomics, and deep-learning-based identification can reveal undocumented human-driven changes in animal distributions.

Introduction

Human-mediated animal translocation offers a critical framework for assessing anthropogenic impacts on biodiversity¹. However, the detectability of these movements varies markedly among taxa. For small commensals and livestock, which are rarely documented in premodern written records, human-assisted dispersal is primarily tracked through population-level genetic and zooarchaeological signatures^{2–5}. Conversely, the movement of large or culturally prestigious mammals, often exchanged as tribute or diplomatic gifts or displayed in elite menageries, is more commonly documented in historical texts and specialized archaeological assemblages. In historical China, this is exemplified by Southeast Asian elephants, East African giraffes at the Ming court, and the diverse fauna—including giant pandas, rhinoceroses, tapirs, gaurs, and primates—excavated from imperial burial pits^{6–9}. Together, these records show that prestigious mammals were routinely captured, transported over long distances, and maintained in elite settings.

Gibbons occupied a prominent position among these prized mammals, symbolizing refinement and moral virtue in elite Chinese culture^{10–12}. Multiple lines of evidence indicate that these primates were integrated into elite networks through capture, captivity, and long-distance transport. Zooarchaeological findings, notably the recovery of the extinct *Nomascus (Junzi) imperialis* from a royal tomb, confirm their presence in imperial menageries^{13,14}. Historical archives, including court paintings such as the Xuande Emperor's *Gibbons at Play*^{10,11} and Ming dynasty texts¹⁵, also document the presentation of gibbons

from Guangxi as imperial tribute. However, while these records verify anthropogenic translocation, they lack the genetic and biogeographic resolution required to reconstruct the precise geographic origins or dispersal pathways of specific captive individuals.

Genetic provenance inference provides a robust framework for resolving such cases when populations exhibit strong geographic genetic structure. Mitochondrial and genome-wide data can be used to assess whether a specimen is genetically compatible with its recorded locality^{16,17}. *Hoolock* gibbons represent a suitable system for this approach. The distributions of the three currently recognized species—*Hoolock hoolock*, *H. leuconedys*, and *H. tianxing*—are sharply delimited by major river systems that have acted as long-term biogeographic barriers. *H. hoolock* occurs primarily west of the Chindwin River, *H. leuconedys* between the Chindwin and Irrawaddy–Nmai Hka systems, and *H. tianxing* east of the Irrawaddy–Nmai Hka system, including western Yunnan and the Gaoligong region¹⁸ (**Fig. 1B**). Thus, a specimen recovered outside its expected geographic range is unlikely to reflect natural dispersal and is best explained by human-mediated translocation.

Figure 1 | Mitochondrial phylogeny and geographic distribution of the genus *Hoolock*. (A) Maximum-likelihood tree of *Hoolock* mitochondrial sequences. The focal museum specimens are highlighted in red; reference individuals from known localities are in bold. Nodal circles indicate bootstrap support; (B) Geographic ranges and sampling localities. Shaded regions denote the recognized distributions of *H. hoolock* (pink), *H. leuconedys* (blue), and *H. tianxing* (green). Open and solid symbols represent museum and field specimens, respectively. Solid diamond colors correspond to the *H. hoolock* mitochondrial lineages defined in (A). The red star marks the anomalous collection locality of the two focal specimens in western Yunnan, China.

We recently investigated two museum *Hoolock* specimens collected in 1960 from Lengshuihe, Datang, Tengchong, Yunnan, China, a locality situated within the currently recognized range of *H. tianxing*. Preliminary examination unexpectedly assigned both specimens to *H. hoolock*, whose accepted natural range lies several hundred kilometers to the west. To resolve their taxonomic identity and evaluate the cause of this geographic discrepancy, we combined deep-learning-assisted species identification based on dental morphology with mitochondrial and genome-wide analyses. We then tested two alternative hypotheses: that these specimens represent either a previously unrecognized natural occurrence of *H. hoolock* in western Yunnan or a historical case of human-mediated translocation.

Methods

“ *H. hoolock* ” specimens from Lengshuihe, Datang, Tengchong, Yunnan

The two museum specimens, KIZ011338 (represented by a skull, including the cranium and mandible) and KIZ000166 (represented by a skin), were housed at the Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences (KIZ, CAS). Both were collected from Lengshuihe, Datang, Tengchong, Yunnan Province, China, and were preliminarily identified as *H.*

hoolock (**Fig. 1B**; 98°41'06"E, 25°41'06"N). Both specimens are accompanied by original field collection records dating to 1960 (**Fig. S1**). This locality lies outside the currently recognized distribution of *H. hoolock* but within the known range of *H. tianxing*, making the taxonomic assignment unexpected.

Specimen photography and image processing

Our comparative dental dataset comprised 75 *Hoolock* specimens representing all three recognized species (**Table S1**) and housed across five institutions: the American Museum of Natural History (AMNH), the Institute of Zoology, Chinese Academy of Sciences (IOZ, CAS), KIZ and the Museum of Comparative Zoology (MCZ, Harvard University). Specimen photography was conducted by A.O. and K.H. using a digital single-lens reflex (DSLR) camera following a protocol detailed in¹⁹.

Individual upper molars (M1, M2, and M3) were manually segmented from the raw photographs and treated as independent image units. These tooth-level images were subjected to rigorous quality-control screening, which excluded images exhibiting incomplete occlusal surfaces, out-of-focus targets, glare or shadows that obscured diagnostic features, or focal axes deviating > 15° from the perpendicular to the occlusal plane. To standardize the assessment of occlusal wear, tooth wear was scored according to the degree-of-wear categories described by Molnar²⁰, and only molars in wear categories 1–5 were retained; severely worn teeth in categories 6–8 were excluded because diagnostic occlusal morphology was substantially obscured or lost. Post-filtering, the dataset comprised 340 high-quality tooth images: 151 from *H. hoolock* (M1 = 55; M2 = 61; M3 = 35), 173 from *H. leuconedys* (M1 = 54; M2 = 70; M3 = 49), and 16 from *H. tianxing* (M1 = 10; M2 = 6) (**Table S1**). Given its limited sample size, *H. tianxing* was not treated as an independent classification category. Instead, its images were merged with those of its sister taxon, *H. leuconedys*, to form a composite analytical class (hereafter, the *leuconedys* group). In KIZ011338, only the M1 was available for analysis (two qualifying photographs), as the M2 was partially erupted and the M3 was fully unerupted, failing the complete occlusal surface requirement.

For deep-learning preprocessing, the entire occlusal surface of each tooth was manually isolated in Adobe Photoshop® using a tight rectangular bounding box around the crown outline. The cropped images were subsequently resized to fit within a 600 × 600-pixel bounding frame while strictly maintaining their original aspect ratios. Zero-value padding was applied symmetrically to center each tooth against a black background. These standardized images constituted the final input data for deep-learning-based species identification and downstream morphological interpretation.

Image dataset partitioning and data augmentation

To prevent information leakage arising from images of the same specimen being distributed across multiple data subsets, the dataset was partitioned at the individual rather than image level. All tooth images derived from a given specimen were assigned to the same subset. Of the 75 individuals included in our sample, 60 individuals (258 images) were assigned to the training dataset, whereas the remaining 15 individuals (82 images) were assigned to the testing dataset (**Table S2**).

Because the dataset is small and susceptible to overfitting, we applied a combined offline and online augmentation strategy to the training set. Following the individual-level dataset partition described above, augmentation was applied only to the training dataset. During the offline stage, each retained original training image was used to generate five augmented images by randomly sampling five transformations without replacement from eight predefined colour-augmentation operations: *bright_sat* (brightness factor, 1.1–1.3; saturation factor, 1.2–1.4), *cool* (HSV hue shift, – 15 to + 15; brightness factor, 0.9–1.1), *dark_con* (brightness factor, 0.7–0.9; contrast factor, 1.2–1.4), *hicon* (contrast factor, 1.3–1.5; sharpness factor, 1.2–1.4), *muted* (saturation factor, 0.6–0.8; brightness factor, 0.95–1.05), *soft* (contrast factor, 0.8–0.95; saturation factor, 0.85–0.95), *vivid* (saturation factor, 1.3–1.5; contrast factor, 1.1–1.2), and *warm* (HSV hue shift, – 15 to + 15; saturation factor, 1.1–1.3). These operations were implemented using a custom Python script based on Pillow and OpenCV. In the online stage, conservative spatial and colour perturbations were applied in real time during training, with parameters selected to preserve morphologically informative features, including the occlusal surface, tooth cusps, and fissure systems. Online augmentation comprised direct square resizing to each model’s native input size, horizontal flipping ($p = 0.50$), rotation up to 5° ($p = 0.50$), and brightness jitter (amplitude 0.10). In addition, museum-style augmentation was performed to account for differences in image appearance among museum collections. We converted images to tensors and normalized them using the ImageNet channel means [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225]²¹, consistent with standard transfer-learning protocols based on ImageNet-pretrained weights. Data augmentation yielded a final training set of 1,984 images (~ 7.7-fold expansion), consisting of 966 images from the *hoo-lock* group and 1,018 from the *leuconedys* group (**Table S2**).

Deep learning model selection, training, and performance evaluation

We evaluated 31 pre-trained convolutional neural networks and Transformer-based models, all of which were trained and had test set metrics generated. These included EfficientNet²² and ConvNeXt²³, among others. Definitions of all models, references, training details, and evaluation metrics are listed in **Table S3**.

We used identical training configurations across all models, with the following exceptions: the backbone of MobileNetV5-300M was frozen and only its classification head was fine-tuned because of GPU memory constraints; EfficientNet-B7 and MobileNetV5-300M used a physical micro-batch of 4 with gradient accumulation to retain a logical batch size of 16; and SwinV2-Large was trained on two GPUs with DataParallel. The models’ training hyperparameters were: AdamW optimizer²⁴ combined with non-adaptive SAM updates²⁵ ($\rho = 0.05$), with an initial learning rate 3×10^{-4} and weight decay 0.01; cosine learning-rate schedule²⁶; logical batch size 16; maximum 200 training epochs; early-stopping patience of 30 epochs, with macro-F1 monitored on the test set because no separate validation split was used; and weighted cross-entropy loss with class weights set inversely proportional to class frequency, to address class imbalance. We implemented all models in PyTorch 2.10.0 with CUDA 12.8 and cuDNN 9.10.2, and

trained each model on NVIDIA GeForce RTX 4090 GPUs with 24 GB of memory²⁷. We evaluated model performance on the test set using accuracy, macro-F1 and balanced accuracy.

To assess whether models had learned biologically meaningful dental features rather than image artefacts, we generated Grad-CAM activation maps for all original training-set and test-set images²⁸. Activation maps were also generated for the original training images to verify that the observed patterns were not artifacts introduced by offline colour augmentation. Each activation map was then manually inspected, with reference to the dental anatomical terminology/morphology²⁹, to determine whether high-activation regions corresponded to recognized dental features and whether these patterns were consistent within each analytical group and differed between the two groups.

DNA extraction, library preparation and sequencing

Specimens KIZ011338 and KIZ000166 were additionally subjected to whole-genome resequencing to resolve their genetic identity. Genomic DNA was extracted from the museum specimens following a previously published protocol³⁰ optimized for degraded DNA. Single-stranded libraries were prepared without uracil-DNA glycosylase (UDG) treatment³¹ to maximize the recovery of short endogenous DNA fragments. The resulting libraries were sequenced on the MGI-DNBSEQ-T7 platform using 2 × 150 bp paired-end sequencing. For samples KIZ011338 and KIZ000166, a total of 87.8 Gb and 77.4 Gb of raw sequencing data were generated, respectively. The mapping rate, average depth, and genome coverage statistics are reported in **Table S4**.

Whole-genome read processing, mapping, and variant calling

For comparative analyses, whole-genome sequencing data from all three recognized species in the genus *Hoolock* were downloaded from public repositories (**Table S4**). Raw reads were processed using fastp v0.23.4 to remove adapter sequences and low-quality reads³². Clean reads were aligned to an *H. leuconedys* reference genome (GCA_047372625.1_UCONN_HLeu1) using the BWA-MEM algorithm implemented in BWA³³. Alignment files were sorted and indexed using SAMtools v1.21³⁴, and PCR duplicates were removed with Picard MarkDuplicates v3.4.0. Despite being collected in the 1960s, both museum specimens yielded high-quality genomic data. Reads mapped efficiently to the *H. leuconedys* reference genome, with mapping rates of 99.2% and 99.0% for KIZ000166 and KIZ011338, respectively. Average sequencing depths reached 12.25× and 14.21×, and genome coverage exceeded 97.0% in both specimens (**Table S4**). Across all samples, mapping rates ranged from 97.4% to 99.6%, indicating overall high alignment quality.

Because the museum libraries were prepared without UDG treatment, post-mortem DNA damage was characterized using mapDamage v2.2.2³⁵. MapDamage analyses revealed elevated frequencies of C-to-T substitutions at the 5' ends and G-to-A substitutions at the 3' ends of sequencing reads, consistent with post-mortem DNA damage patterns expected for historical museum specimens (**Fig. S2**).

Variant calling was performed using the Genome Analysis Toolkit (GATK v4.6.1.0)³⁶. Per-sample gVCFs were generated using GATK HaplotypeCaller in gVCF mode and subsequently combined using CombineGVCFs. Joint genotyping was performed using GenotypeGVCFs, and high-confidence SNPs were extracted using SelectVariants. Variants were filtered using VariantFiltration with the following criteria: $QUAL > 30.0$, $QD \geq 2.0$, $MQ \geq 40.0$, $FS \leq 60.0$, $SOR \leq 3.0$, $MQRankSum \geq -12.5$, and $ReadPosRankSum \geq -8.0$. Genotypes with depth < 5 or genotype quality < 20 were treated as missing. Sites with more than 20% missing genotypes were excluded using VCFtools v0.1.16³⁷. Biallelic autosomal SNPs were then retained by excluding multiallelic sites and sites located on sex chromosomes. The resulting filtered autosomal SNP dataset was used as the starting dataset for analyses. Linkage disequilibrium pruning was performed separately for analyses that required LD-independent markers.

Nuclear phylogenomic and population-genomic analyses

Phylogenomic reconstruction

For phylogenomic reconstruction, we used the filtered biallelic autosomal SNP dataset without linkage disequilibrium pruning. The dataset was converted to PHYLIP format using vcf2phylip v2.8³⁸. Maximum-likelihood phylogenetic trees were reconstructed using IQ-TREE v2.1.4³⁹ under the GTR + G substitution model, and branch support was assessed using 1,000 ultrafast bootstrap replicates. The genome of *Nomascus gabriellae* (SAMC4603452) was included as the outgroup. Trees were visualized using tvBOT⁴⁰.

Principal component analysis

For principal component analysis, the filtered genus-wide biallelic autosomal SNP dataset was pruned for linkage disequilibrium using PLINK v1.9 with the `--indep-pairwise 50 5 0.2` parameter⁴¹. This procedure resulted in 617,086 autosomal SNPs. Missing genotypes were imputed to the sample allele mean using PLINK v1.9. The first 20 principal components were extracted, and the components explaining the largest proportions of genetic variation were visualized.

Pairwise genomic relatedness

To assess pairwise genetic relatedness within *H. hooock*, we constructed a species-specific autosomal SNP dataset from the joint-called VCF before applying the cross-species site-level missingness filter. The *H. hooock* samples were extracted, and genotype- and site-level quality filters were reapplied within this species-specific subset. Biallelic autosomal SNPs with genotype depth ≥ 5 , genotype quality ≥ 20 , and site missingness $\leq 20\%$ were retained. Linkage disequilibrium pruning was then performed using PLINK v1.9 with the `--indep-pairwise 50 5 0.2` parameter, resulting in 3,753,960 *H. hooock*-specific autosomal SNP markers. These markers were used to estimate pairwise identity-by-descent sharing using the

PLINK --genome function⁴¹ and pairwise kinship coefficients using KING v2.2.7⁴². Hierarchical clustering was performed based on the pairwise relatedness matrices to visualize genetic relatedness among individuals.

Observed heterozygosity based on autosomal transversion SNPs

Observed heterozygosity was estimated from the filtered biallelic autosomal SNP dataset without linkage disequilibrium pruning. Genotypes were filtered using QUAL > 30.0, genotype depth ≥ 5 , genotype quality ≥ 20 , and site missingness $\leq 20\%$. To minimize the potential influence of post-mortem DNA damage in museum specimens, only transversion SNPs were used. For each individual, missing genotypes were excluded, and SNP-based observed heterozygosity was calculated as the number of heterozygous transversion SNP genotypes divided by the total number of called transversion SNP genotypes. Values were reported as proportions rather than percentages.

Tests of allele sharing and historical gene flow

To investigate the presence and relative magnitude of historical gene flow among *Hoolock* species, we analyzed a high-quality SNP dataset from 22 samples using Dsuite v0.5⁴³ without linkage disequilibrium pruning. We first filtered the SNP dataset to retain sites with $\leq 20\%$ missing data using VCFtools (--max-missing 0.8). We then used the Dtrios program in Dsuite to calculate D-statistics and f₄-ratio statistics for all relevant population trios. *N. gabriellae* was used as the outgroup. The f-branch statistic was further calculated to summarize patterns of excess allele sharing across the species tree, and the results were visualized using the dtools.py script.

Mitochondrial genome assembly and phylogenetic analyses

For mitochondrial analyses, raw paired-end reads were processed using leeHom to trim sequencing adapters and merge overlapping read pairs. Filtered reads were mapped against the *Hoolock* mitochondrial reference genome (NC_033885) using BWA v0.7.17. Alignment files were converted to BAM format, sorted, and indexed using SAMtools v1.18, and PCR duplicates were removed with Picard MarkDuplicates. Local realignment around indels was performed using GATK v3.8 with RealignerTargetCreator and IndelRealigner. Reads with mapping quality scores below 25 and unmapped reads were excluded. Consensus mitochondrial sequences were generated and annotated in Geneious Prime® 2024.0.5.

Two datasets were analyzed. First, we reconstructed mitogenomic phylogenies using 12 mitochondrial protein-coding genes and two ribosomal RNA genes from representatives of *Hoolock*; all available mitochondrial genomes were included (**Table S4**). Transfer RNA genes, the NADH dehydrogenase subunit 6 (ND6) gene, and the highly variable control region (D-loop) were excluded because of their heterogeneous evolutionary rates and alignment uncertainty. This dataset comprises 44 mitochondrial genomes, including 38 genomes from *H. hoolock* individuals. Second, to assess the likely geographic

affinity of KIZ011338 and KIZ000166 within *H. hoolock*, we performed additional phylogenetic analyses using cytochrome c oxidase subunit I (COX1) and control region (D-loop) sequences of western hoolock gibbons with known sampling localities. Multiple sequence alignments were generated using MAFFT v7.490⁴⁴. Maximum-likelihood phylogenetic analyses were conducted in IQ-TREE v1.6.8⁴⁵ as implemented in PhyloSuite v1.2.2⁴⁶. All the resulting trees were visualized with tvBOT.

Results

Deep-learning-assisted dental analysis and pelage morphology support the identification of the Datang specimens as *H. hoolock*

Model performance varied among the tested architectures, with classification accuracy ranging from 64.6% to 92.7% (**Table S3**). The best-performing models were ConvNeXt-Base (224×224), EfficientNet-X-B5 (448×448), EfficientNet-B5 (448×448), and EfficientNetV2-S (300×300), with accuracies of 92.7%, 91.5%, 89.0%, and 87.8%, respectively. To evaluate whether high classification performance reflected biologically meaningful dental morphology rather than image artefacts or background information, we examined class activation map (CAM) heatmaps for the top-performing models. These models generally focused on the tooth region, especially the occlusal surface and crown structures, indicating that they relied on anatomically relevant features for group classification. However, CAM inspection also revealed background spillover in all EfficientNet models, with portions of the model attention extending beyond the tooth region (**Fig. S3**). We therefore selected ConvNeXt as the final model because it combined high classification accuracy with more precise localization on the tooth crown.

The ConvNeXt model achieved classification accuracies of 97.1%, 90.0%, and 88.9% for M1, M2, and M3, respectively (**Figs. 2A–C**). For specimen KIZ011338 from Datang, ConvNeXt assigned both M1 images to *H. hoolock* with high confidence (predicted probabilities: 99.7% and 96.2%). Grad-CAM analyses indicated that the classification of M1 images was mainly associated with the buccal notch and cusp configuration. Qualitative inspection supported this interpretation (**Figs. 2D, E**). In the *H. leuconedys* group, the two buccal cusps (paracone and metacone) tend to be closely spaced, the postparacrista and premetacrista are steep, and the buccal notch forms a narrow V-shaped indentation. Correspondingly, the trigon basin enclosed by the paracone, metacone, and protocone is relatively small but somewhat deep. In contrast, the paracone and metacone are more widely separated, the buccal notch forms a broader angle, and the trigon basin is more open and larger in *H. hoolock*. Among hylobatids (and other hominoids), the buccal cusps are positioned more mesially than their lingual counterparts, such that the paracone lies mesial to the protocone and the metacone is mesial to the hypocone; this feature is less pronounced in *H. hoolock* than in *H. leuconedys*. Taken together, these characteristics indicate that the M1 of KIZ011338 conforms more closely to the *H. hoolock*-like condition.

Although the incomplete dental eruption of KIZ011338 (partially erupted M2 and unerupted M3) precluded analysis of its posterior molars, we nevertheless examined the Grad-CAM activation patterns across M2 and M3 in the broader comparative sample (**Fig. S4**). The morphological differences

observed between *H. hoolock* and the *H. leuconedys* group in M1 apply equally to M2. Despite variation in M3 shape, the *H. leuconedys* group was distinguished from *H. hoolock* by its distinct crista obliqua, well-developed hypocone and posthypocrista, and narrower trigon basin, in contrast to the reduced crests and broader basin observed in *H. hoolock*.

Figure 2 | ConvNeXt-Based dental classification performance and feature visualization. (A–C)

Confusion matrices for isolated upper molars M1 (A), M2 (B), and M3 (C) from the test set. Rows and columns represent true and predicted classes, respectively, with classes ordered as *H. hoolock* and *H. leuconedys* in both dimensions. Cells display image counts and accuracies. (D) Representative Grad-CAM activation maps, where warmer colors indicate regions contributing more strongly to model predictions. The teeth shown in (D) and (E) are M1s; corresponding M2 and M3 images are provided in Fig. S4.

The second Datang specimen, KIZ000166, was represented by skin rather than dental material and therefore could not be evaluated by the dental classification model. However, its closely joined white eyebrow markings are consistent with the adult male pelage pattern of *H. hoolock* (Fig. S1). Taken together, the dental classification of KIZ011338 and the pelage morphology of KIZ000166 support identification of the Datang museum specimens as *H. hoolock*.

Genome-wide SNPs and mitogenomes confirm the Datang specimens as *H. hoolock*

We next examined the phylogenetic positions of the two Datang specimens using genome-wide autosomal SNPs and complete mitochondrial genomes. Maximum-likelihood phylogenies reconstructed from both datasets consistently recovered KIZ011338 and KIZ000166 within the *H. hoolock* clade with strong support (BS > 90; Fig. 3A; Fig. S5). In the genome-wide SNP-based PCA plot, the first two principal components clearly separated the three recognized *Hoolock* species into distinct clusters (Fig. 3B). All sampled *H. hoolock* individuals, including the two Datang museum specimens, occupied the rightmost region along PC1, whereas *H. leuconedys* and *H. tianxing* occupied the leftmost region. These genomic clusters corroborate the identification of the Datang specimens as *H. hoolock*.

Figure 3 | Genomic affinities and population structure of the Datang museum specimens. (A) Maximum-likelihood phylogeny inferred from autosomal single-nucleotide polymorphisms (SNPs). Focal historical specimens (KIZ000166 and KIZ011338) are highlighted in bold. Nodal circles indicate bootstrap support. **(B)** Principal component analysis (PCA) of autosomal SNPs. Colors indicate species assignments. Focal specimens are explicitly labeled. **(C)** *f*-branch statistics visualizing historical gene flow among *Hoolock* species; color intensity reflects the magnitude of excess allele sharing. **(D)** Hierarchically clustered heatmap of pairwise genetic relatedness (PI_HAT) among *H. hoolock* individuals. Darker colors indicate higher relatedness. **(E)** Individual observed heterozygosity based on autosomal transversion SNPs.

Mitochondrial phylogenies nest the Datang specimens within the Assam *H. hoolock* lineage

To investigate the maternal geographic affinity of the Datang specimens, we reconstructed phylogenetic trees using mitochondrial COX1 and D-loop sequences from *H. hoolock* individuals with known sampling localities. The analyses recovered four major mitochondrial lineages within *H. hoolock*, corresponding to the Center, South, North, and Waro populations, the latter three of which were reported previously^{47,48} (**Fig. 1A**). Both Datang specimens were nested within the South clade from central and southern Assam, northeastern India.

Genome-wide evidence does not support long-term local residence of the Datang *H. hoolock* specimens in western Yunnan

We next evaluated whether the Datang specimens exhibited genomic patterns expected under long-term local residence in western Yunnan. If *H. hoolock* had persisted naturally near Tengchong for an extended period, we would expect detectable allele sharing with the geographically adjacent *H. tianxing* population, especially given the close proximity of their inferred ranges. However, our Dsuite analysis did not support substantial excess allele sharing between *H. hoolock* and *H. tianxing* (**Fig. 3C**). The strongest *f*-branch signal occurred between *H. hoolock* and *H. leuconedys*, with an *fb* value of approximately 0.05, whereas *f*-branch values involving *H. tianxing* were low or close to zero. This pattern indicates limited detectable introgression or excess shared ancestry involving *H. tianxing*.

If the Datang specimens had instead originated from a long-established, small, and isolated population, we would expect reduced genetic diversity and elevated inbreeding. However, SNP-based heterozygosity analyses show that both individuals fall toward the higher end of the *H. hoolock* distribution (KIZ000166: 0.05; KIZ011338: 0.04; **Fig. 3E**). Further, if the Datang individuals represented a small natural population, strong relatedness among individuals would be expected. However, our IBD analysis yielded a PI_HAT value of 0.091 and a KING kinship coefficient of -0.005 for KIZ000166 and KIZ011338 (**Fig. S6**), providing no evidence of a uniquely close pedigree relationship between the two specimens. This result suggests that the two individuals did not originate from a single, closely related source population. Taken together, our results do not support the hypothesis that the Datang specimens originated from a long-established, isolated *H. hoolock* population in western Yunnan. Instead, they suggest a recent arrival of *H. hoolock* individuals into the Tengchong region, plausibly through human-mediated movement rather than long-term natural persistence.

Discussion

Datang specimens are *H. hoolock*

The combined morphological, mitochondrial, and genome-wide evidence strongly supports the identification of the Datang specimens (KIZ000166 and KIZ011338) as *H. hoolock* rather than the locally native *H. tianxing* (**Figs. 1A, 2, 3A, 3B; Fig. S4**). This genetic identity establishes an extra-limital occurrence of *H. hoolock* deep within the modern geographic range of *H. tianxing* in western Yunnan, supporting an anthropogenic origin rather than a natural range overlap.

Furthermore, genomic analyses placed the Datang specimens within *H. hoolock* and detected no significant excess allele sharing with *H. tianxing*. Although genus-wide SNP-based ordination and phylogenomics resolve *H. hoolock*, *H. leuconedys*, and *H. tianxing* as distinct evolutionary units^{14,18} (**Figs. 3A, 3B**), Dsuite analysis revealed no significant excess allele sharing between the Datang *H. hoolock* specimens and sympatric *H. tianxing* (**Fig. 3C**). The strongest *f*-branch signal observed exclusively between *H. hoolock* and *H. leuconedys* reflects baseline incomplete lineage sorting or ancient divergence dynamics within the genus. These results provide no genomic evidence that the Datang individuals were hybrids or recent backcrosses, making human-mediated translocation the most plausible explanation for their presence in western Yunnan.

Deep-learning-based classification of isolated upper molars

An unexpected finding of this study is the high classification accuracy achieved by the ConvNeXt-Base dental morphometric model when applied to single isolated upper molars. For the three tooth positions, the model achieved per-position accuracies of 97.1% (M1), 90.0% (M2), and 88.9% (M3) in distinguishing *H. hoolock* from the *H. leuconedys* group. Previous approaches to species-level identification from primate dental material have generally relied on relatively complete craniodental assemblages or multiple teeth per individual, in which the aggregate morphological signal is substantially larger^{49–52}. In the present case, classification was performed on isolated upper molars, a substantial proportion of which exhibited moderate wear (wear categories 4–5²⁰), resulting in partial alteration or obscuration of conventional anatomical landmarks. Despite this, the ConvNeXt model maintained robust classification performance across all three molars, suggesting that worn occlusal surfaces retain sufficient morphological information for taxonomic discrimination. This finding is consistent with previous work showing that morphologically informative signals can be extracted from worn primate teeth⁵³.

Inspection of Grad-CAM activation maps provided further insight into the biological basis of this classification. Across the three tooth positions, the Grad-CAM maps revealed consistent regions of interest distinguishing *H. hoolock* from the *leuconedys* group, confirming that classification relied on biologically meaningful crown morphology rather than imaging artifacts or non-diagnostic wear patterns. These results indicate that CNN-based approaches can capture taxonomically informative variation in tooth morphology without relying on homologous points. Consequently, single-tooth classification can be effective even for degraded historical museum specimens. Broader application to fossil primate assemblages, where isolated teeth are common and species-level identification is often challenging, warrants future investigation. Given that the present study was limited to a small dataset and a single genus, however, the generalizability of this approach to other taxa will require validation with broader comparative datasets.

Specimen provenance and museum record integrity

Before evaluating the biogeographic implications of this identification, it is necessary to establish that the recorded locality for KIZ000166 and KIZ011338 accurately reflects their collection at Datang in 1960. First, the original specimen labels on both the skin (KIZ000166) and the skull (KIZ011338) show clear signs of age, including yellowing of the paper and the exclusive use of Traditional Chinese characters, which were largely replaced by Simplified Chinese after 1964 (**Fig. S1**), consistent with a 1960 collection date. Second, the museum holds no other specimens of *H. hoolock* from any locality, making label-swapping or mislabeling less likely among conspecifics. Third, the skin and skull are housed separately in the skin and osteology collection rooms, respectively, rendering simultaneous mislabeling of both specimens implausible. Fourth, re-measured body and hindfoot lengths of the skin specimen (KIZ000166) directly match those recorded on the original field collection label, providing a direct physical link between the specimen and its documented provenance. Finally, museum accession records systematically document the captive origin of all zoo-derived specimens; the absence of such annotations provides no evidence for a modern captive origin. Together, these convergent lines of evidence rule out curatorial error, specimen mixing, and captive origin, confirming the authenticity and reliability of the 1960 Datang locality record.

Evidence against a natural occurrence in western Yunnan

Having authenticated the 1960 Datang collection records in the previous section, the anomalous presence of *H. hoolock* in western Yunnan can be attributed to either natural range expansion or human-mediated translocation. Gibbons (Hylobatidae) are predominantly arboreal and rarely move terrestrially, making dispersal between forest blocks dependent on continuous forest habitat⁵⁴. Within the genus *Hoolock*, the Chindwin and Irrawaddy–Nmai Hka river systems coincide with the recognized geographic boundaries between *H. hoolock*, *H. leuconedys*, and *H. tianxing*¹⁸. An eastward expansion of *H. hoolock* into western Yunnan would therefore require movement across several major landscape barriers, making recent human-mediated translocation a more parsimonious explanation than natural colonization.

Even if such dispersal were conceivable, a naturally established population of *H. hoolock* in western Yunnan would be expected to show a relatively continuous distribution along the Gaoligong mountains in a north–south direction, which represents the most plausible overland corridor into the region. However, extensive field surveys of gibbon populations in western Yunnan have provided no confirmed records for the occurrence of either *H. leuconedys* or *H. hoolock*^{18,55,56}.

Crucially, the genetic profiles of the Datang specimens robustly corroborate these ecological limitations. If these individuals represented a relict population naturally established east of the Irrawaddy, their genomes would reflect prolonged geographic isolation through deep lineage differentiation, alongside detectable introgression from the immediately parapatric *H. tianxing*¹⁸. Our data demonstrate the exact opposite: an unadmixed genomic profile firmly nested within the extant southern Assam lineage^{47,48} (**Fig. 1A**). The absence of detectable localized divergence and significant allele sharing provides no support for a long-term evolutionary history in western Yunnan. Synthesizing these ecological, observational, and genomic constraints, we conclude that the available evidence does not support the

natural occurrence hypothesis. The anomalous presence of these specimens is most parsimoniously explained by a recent, human-mediated translocation.

The Southwestern Silk Road as a plausible translocation corridor

Confirming anthropogenic translocation necessitates identifying both a plausible source population and a viable transport corridor. Phylogeographic nesting of the Datang specimens within the southern (Assam) lineage of *H. hoolock* is consistent with northeastern India as a possible source region. A direct geographic connection between Assam and western Yunnan is provided by the Southwestern Silk Road—an extensive network of overland trade corridors linking Sichuan and Yunnan with northern Myanmar, eastern South Asia, and mainland Southeast Asia⁵⁷ (图4A). Active since at least the Han Dynasty^{9,58}, this route system offers a historically verified mechanism through which taxa originating west of the Irrawaddy–Nmai Hka barrier could be transported into the Tengchong–Gaoligong region.

At a finer geographic scale, three principal corridors of the Southwestern Silk Road traverse the Gaoligong range (Fig. 4B): (1) a northern route via Beizhaigongfang (图4B); (2) a central route via Baihualing–Nanzhaigongfang (图4B–图4C); and (3) a southern route via the Fenshuiguan/Mopanshi–Pumansao–Dahaoping corridor (图4C/图4D–图4E–图4F)⁵⁹. These corridors remained functional into the mid-20th century and were utilized by Chinese expeditionary forces during the Burma Campaign (1942–1945)^{59,60}. Crucially, the specimen collection locality, Lengshuihe, Datang, is situated only 12 km in a straight line from the Beizhaigongfang corridor. This spatial proximity identifies the northern Gaoligong corridor as a plausible transport route for these translocated animals, although it remains unclear whether these individuals arrived through accidental escape during transport or deliberate release into the local environment.

Although the transport history remains unrecorded, developmental age at collection and genomic data are consistent with a translocation before 1960, possibly during the 1940s–1950s. Long-term persistence of an isolated local group is not supported by the available genomic evidence; prolonged reproductive isolation alongside *H. tianxing* would produce severe inbreeding and close kinship, whereas these specimens show neither. Furthermore, both individuals were subadults at their 1960 collection date—evidenced by subadult body metrics⁶¹ (**Fig. S1**), incomplete M2/M3 eruption (**Fig. S7**), and negligible occlusal wear. The co-occurrence of two non-kin subadults strongly points to recent removal from a broad source population shortly prior to 1960, making survival from an earlier introduction less likely.

Conclusions

The convergent morphological, mitochondrial, and genome-wide nuclear evidence identifies the two Datang museum specimens as *H. hoolock* rather than the locally distributed *H. tianxing*. Their placement within the southern Assam mitochondrial lineage, together with the absence of substantial excess allele

sharing with *H. tianxing* and the lack of close pairwise relatedness, does not support long-term residence of an isolated *H. hoolock* population in western Yunnan.

The most parsimonious explanation is that the two animals were transported into the Tengchong region by human activities before their collection in 1960. The Southwestern Silk Road provides a historically plausible corridor, but the exact transport route, date, and human context cannot be established from the available evidence. Additional sampling of *H. hoolock* populations and historical specimens would help refine the geographic source and temporal history of this translocation.

Data Availability

The whole-genome sequencing data generated in this study have been deposited in the Genome Sequence Archive under accession number CRA046348. The assembled mitochondrial genome sequences are available from the Science Data Bank at <https://doi.org/10.57760/sciencedb.43078>. The code used for the deep-learning analyses is available at <https://github.com/CjQkJ/hoolock-gibbon-dental-classification>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the National Natural Science Foundation of China (Grant No. 32170452), the NSFC Regional Innovation and Development Joint Fund (Key Project, Grant No. U23A20161), and the Position of Bioclassonomist of the Chinese Academy of Sciences (CAS-TAX-24-055). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contribution

Conceptualization, Z.C and K.H; methodology, X.H, Q.C., Z.H., A.O, X.Y.W, X.W, Z.C, and K.H; software, X.H, Q.C., and Z.H; formal analysis, X.H and Q.C.; investigation, X.H, A.O, X.Y.W, W.D., X.W, K.C., J.X., Y.H., S.L, D.W., and X.Z; resources, A.O, S.L, D.W., X.Z., P.F., Z.C, and K.H; visualization, X.H and Q.C.; writing—original draft preparation, X.H, Q.C., and K.H; writing—review and editing, X.H, Q.C., A.O, P.F., Z.C, and K.H; supervision, K.H; project administration, P.F., Z.C, and K.H; funding acquisition, S.L and K.H; All authors read and approved the final manuscript.

Acknowledgement

We thank the curators and collection managers of the American Museum of Natural History, the Institute of Zoology, Chinese Academy of Sciences, the Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, and the Museum of Comparative Zoology, Harvard University, for providing access to museum specimens and associated collection records.

Data Availability

The whole-genome sequencing data generated in this study have been deposited in the Genome Sequence Archive under accession number CRA046348. The assembled mitochondrial genome sequences are available from the Science Data Bank at <https://doi.org/10.57760/sciencedb.43078>. The code used for the deep-learning analyses is available at <https://github.com/CjQkJ/hoolock-gibbon-dental-classification>.

References

1. Boivin NL, Zeder MA, Fuller DQ, Crowther A, Larson G, Erlandson JM, Denham T, Petraglia MD. Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proc. Natl Acad. Sci. USA* 113, 6388–6396 <https://doi.org/10.1073/pnas.1525200113> (2016).
2. Matisoo-Smith E, Robins JH. Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *Proc. Natl Acad. Sci. USA* 101, 9167–9172 <https://doi.org/10.1073/pnas.0403120101> (2004).
3. Puckett EE, Orton D, Munshi-South J. Commensal rats and humans: Integrating rodent phylogeography and zooarchaeology to highlight connections between human societies. *BioEssays*. 2020;42:e1900160. <https://doi.org/10.1002/bies.201900160>.
4. Yu H, et al. Palaeogenomic analysis of black rat (*Rattus rattus*) reveals multiple European introductions associated with human economic history. *Nat Commun*. 2022;13:2399. <https://doi.org/10.1038/s41467-022-30009-z>.
5. Stanton DWG, et al. Genomic and morphometric evidence for Austronesian-mediated pig translocation in the Pacific. *Science*. 2026;391:eadv4963. <https://doi.org/10.1126/science.adv4963>.
6. Schafer EH. *The Golden Peaches of Samarkand: A Study of T'ang Exotics*. University of California Press; 1936.
7. Church SK. The giraffe of Bengal: A medieval encounter in Ming China. *Medieval Hist J*. 2004;7:1–37. <https://doi.org/10.1177/097194580400700101>.
8. Wade G. *Ming China and Southeast Asia in the 15th century: A reappraisal*. Asia Research Institute Working Paper Series 28 National University of Singapore, (2004).

9. Wang XL. Preliminary excavation of the Han Nanling accompanying burial pit and related issues concerning the unearthed giant panda skull and rhinoceros remains. *Cult Relics (Wenwu)*. 1981;11:24–9. in Chinese).
10. van Gulik RH. *The Gibbon in China: An Essay in Chinese Animal Lore (猩猩)*. Leiden: E. J. Brill; 1967.
11. Geissmann T. Gibbon paintings in China, Japan, and Korea: Historical distribution, production rate and context. *Gibbon J*. 2008;4:1–38.
12. Zhang P. Good gibbons and evil macaques: A historical review on cognitive features of non-human primates in Chinese traditional culture. *Primates*. 2015;56:215–25. <https://doi.org/10.1007/s10329-015-0472-4>.
13. Turvey ST, Bruun K, Ortiz A, Hansford J, Hu S, Ding Y, Zhang T, Chatterjee HJ. New genus of extinct Holocene gibbon associated with humans in Imperial China. *Science*. 2018;360:1346–9. <https://doi.org/10.1126/science.aao4903>.
14. Wang S, et al. Genome sequences of extant and extinct gibbons reveal their phylogeny, demographic history, and conservation status. *Cell*. 2026;189:34–51. <https://doi.org/10.1016/j.cell.2025.10.016>.
15. Wang SZ. *Chibei Outan (猩猩)*. Edited by S. Jin (Zhonghua Shuju, Beijing, 1982; in Chinese).
16. Manel S, Gaggiotti OE, Waples RS. Assignment methods: Matching biological questions with appropriate techniques. *Trends Ecol Evol*. 2005;20:136–42. <https://doi.org/10.1016/j.tree.2004.12.004>.
17. Ogden R, Linacre A. Wildlife forensic science: A review of genetic geographic origin assignment. *Forensic Sci Int Genet*. 2015;18:152–9. <https://doi.org/10.1016/j.fsigen.2015.02.008>.
18. Fan PF, He K, Chen X, Ortiz A, Zhang B, Zhao C, Li YQ, Zhang HB, Kimock C, Wang WZ, Groves C, Turvey ST, Roos C, Helgen KM, Jiang XL. Description of a new species of *Hoolock* gibbon (Primates: Hylobatidae) based on integrative taxonomy. *Am J Primatol*. 2017;79:e22631. <https://doi.org/10.1002/ajp.22631>.
19. He K, Li A, Martinez Q, Wang X, Chen Z, He S, Xie S, Zeng Z, Wang K, Ye Z, Ruan H, Liu S, Lu Q, Zheng X, Luo J, Song W, Schwermann A, Yu H, Yu W, Springer MS, Liu S, Li S, Tu F, Cao Z, Campbell K. L. Unraveling hidden species diversity of talpid moles using phylogenomics and skull-based deep learning. *Commun Biol*. 2026. <https://doi.org/10.1038/s42003-026-10392-9>.
20. Molnar S. Human tooth wear, tooth function and cultural variability. *Am J Phys Anthropol*. 1971;34:175–89. <https://doi.org/10.1002/ajpa.1330340204>.
21. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L, ImageNet. A large-scale hierarchical image database. In *Proc. 2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, 248–255 <https://doi.org/10.1109/CVPR.2009.5206848> (2009).
22. Tan M, Le Q, EfficientNet. Rethinking model scaling for convolutional neural networks. In *Proc. 36th Int. Conf. Mach. Learn.* 97, 6105–6114 <https://doi.org/10.48550/arXiv.1905.11946> (2019).
23. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In *Proc. 2022 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 11966–11976 <https://doi.org/10.1109/CVPR52688.2022.01167> (2022).

24. Loshchilov I, Hutter F. Decoupled weight decay regularization. In *International Conference on Learning Representations* (2019).
25. Foret P, Kleiner A, Mobahi H, Neyshabur B. Sharpness-aware minimization for efficiently improving generalization. In *International Conference on Learning Representations* <https://doi.org/10.48550/arXiv.2010.01412> (2021).
26. Loshchilov I, Hutter FSGDR. Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations* <https://doi.org/10.48550/arXiv.1608.03983> (2017).
27. Paszke A et al. PyTorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 32 (2019).
28. Selvaraju RR, Cogswell A, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proc. IEEE Int. Conf. Comput. Vis.*, 618–626 <https://doi.org/10.1109/ICCV.2017.74> (2017).
29. Zhang YQ, et al. Fossil gibbons (Mammalia, Hylobatidae) from the Pleistocene of Chongzuo, Guangxi Zhuang Autonomous Region, China. *Vertebr PalAsiat.* 2017;55:183–94. <https://doi.org/10.19615/j.cnki.1000-3118.180403>.
30. Rohland N, Glocke I, Aximu-Petri A, Meyer M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat Protoc.* 2018;13:2447–61. <https://doi.org/10.1038/s41596-018-0050-5>.
31. Gansauge MT, Aximu-Petri A, Nagel S, Meyer M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat Protoc.* 2020;15:2279–300. <https://doi.org/10.1038/s41596-020-0338-0>.
32. Chen S, Zhou Y, Chen Y, Gu J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>. & 1000 Genome Project Data Processing Subgroup
35. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* 2013;29:1682–4. <https://doi.org/10.1093/bioinformatics/btt193>.
36. McKenna A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303. <https://doi.org/10.1101/gr.107524.110>.
37. Danecek P, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.

38. Ortiz EM. vcf2phyliip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo (2019).
39. Minh BQ, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4. <https://doi.org/10.1093/molbev/msaa131>.
40. Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. Tree Visualization By One Table (tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 2023;51:W587–92. <https://doi.org/10.1093/nar/gkad359>.
41. Purcell S, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75. <https://doi.org/10.1086/519795>.
42. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010;26:2867–73. <https://doi.org/10.1093/bioinformatics/btq559>.
43. Malinsky M, Matschiner M, Svoldal H, Dsuite. Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour.* 2021;21:584–95. <https://doi.org/10.1111/1755-0998.13265>.
44. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
45. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–74. <https://doi.org/10.1093/molbev/msu300>.
46. Zhang D, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 2020;20:348–55. <https://doi.org/10.1111/1755-0998.13096>.
47. Trivedi M, Manu S, Balakrishnan S, Biswas J, Asharaf NVK, Umapathy G. Understanding the phylogenetics of Indian hoolock gibbons: *Hoolock* and *H. leuconedys*. *Int J Primatol.* 2021;42:463–77. <https://doi.org/10.1007/s10764-021-00212-8>.
48. Trivedi M, Sree KD, Biswas J, Umapathy G. Genetic diversity and population structure in western hoolock (*Hoolock hoolock*) gibbons in India. *Int J Primatol.* 2022;44:1–5. <https://doi.org/10.1007/s10764-022-00301-2>.
49. Gilbert CC. Cladistic analysis of extant and fossil African papionins using craniodental data. *J Hum Evol.* 2013;64:399–433. <https://doi.org/10.1016/j.jhevol.2013.01.013>.
50. Nova Delgado M, Galbany J, Perez-Perez A. Morphometric variation of extant platyrrhine molars: Taxonomic implications for fossil platyrrhines. *PeerJ* 4, e1967 <https://doi.org/10.7717/peerj.1967> (2016).
51. Gamarra B, Nova Delgado M, Romero A, Galbany J, Perez-Perez A. Phylogenetic signal in molar dental shape of extant and fossil catarrhine primates. *J Hum Evol.* 2016;94:13–27. <https://doi.org/10.1016/j.jhevol.2016.01.005>.
52. Monson TA, Armitage DW, Hlusko LJ. Using machine learning to classify extant apes and interpret the dental morphology of the chimpanzee-human last common ancestor. *PaleoBios* 35, 1–20

- <https://doi.org/10.5070/P9351040776> (2018).
53. Ungar PS, M'Kirera F. A solution to the worn tooth conundrum in primate functional anatomy. *Proc. Natl Acad. Sci. USA* 100, 3874–3877 <https://doi.org/10.1073/pnas.0637016100> (2003).
 54. Chivers DJ. *The Siamang in Malaya: A Field Study of a Primate in Tropical Rain Forest*. Contributions to Primatology 4S. Karger, (1974).
 55. Yang JH, Huang XY, Jin SH, Chan BPL. Filling a longstanding knowledge gap: Population size and conservation status of the endangered Gaoligong hoolock gibbon (*Hoolock tianxing*) in Houqiao Town, Yunnan. *Glob Ecol Conserv.* 2020;24:e01347. <https://doi.org/10.1016/j.gecco.2020.e01347>.
 56. Zhang L, Guan Z, Fei H, Yan L, Turvey ST, Fan P. Influence of traditional ecological knowledge on conservation of the skywalker hoolock gibbon (*Hoolock tianxing*) outside nature reserves. *Biol Conserv.* 2020;241:108267. <https://doi.org/10.1016/j.biocon.2019.108267>.
 57. Duan Y. Early external communications of southwestern China: The Southern Silk Road during the pre-Qin and Han periods (先秦——西汉时期). *Lishi Yanjiu (Historical Research)* 2009(1), 20–23 (2009; in Chinese).
 58. Yang B. *Between Winds and Clouds: The Making of Yunnan (Second Century BCE to Twentieth Century CE)*. New York: Columbia University; 2009.
 59. Ai HS. Xiangyou Qiannian: Gaoligongshan Baohu de Qiannian Licheng. 湘西: 湘西州民族出版社 (Intellectual Property Publishing House, Beijing,; 2010. in Chinese).
 60. Yu G. 1944: *Longling Huizhan* (1944: 侏儒猴) (SDX Joint Publishing Company, Beijing, 2017; in Chinese).
 61. Sati JP, Alfred JRB. Hoolock gibbon (*Bunopithecus hoolock*). *ENVIS Bulletin: Wildl Protected Areas* 1(1) (2001).

Figures

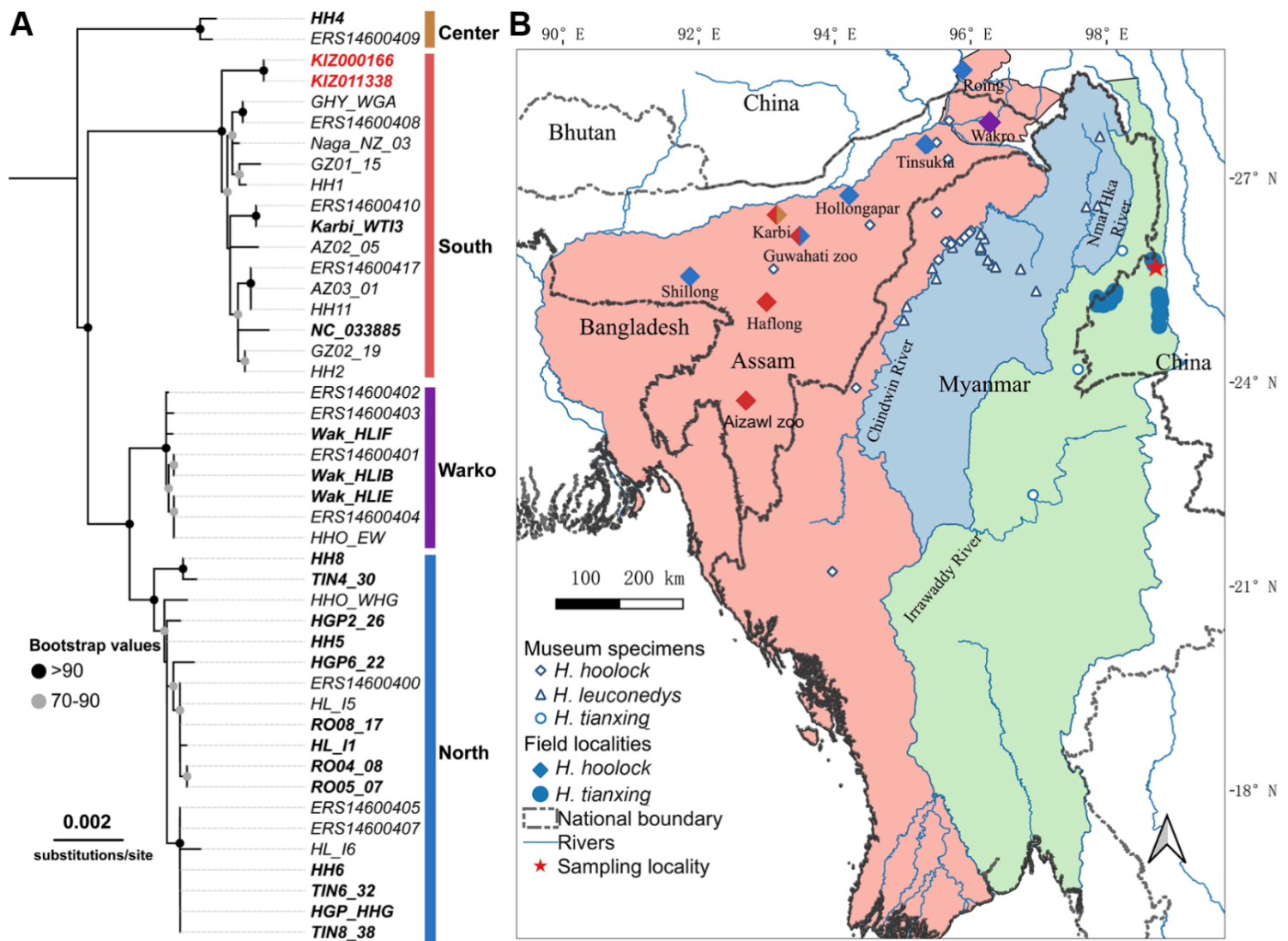


Figure 1

Mitochondrial phylogeny and geographic distribution of the genus *Hoolock*. (A) Maximum-likelihood tree of *Hoolock* mitochondrial sequences. The focal museum specimens are highlighted in red; reference individuals from known localities are in bold. Nodal circles indicate bootstrap support; (B) Geographic ranges and sampling localities. Shaded regions denote the recognized distributions of *H. hoolock* (pink), *H. leuconedys* (blue), and *H. tianxing* (green). Open and solid symbols represent museum and field specimens, respectively. Solid diamond colors correspond to the *H. hoolock* mitochondrial lineages defined in (A). The red star marks the anomalous collection locality of the two focal specimens in western Yunnan, China.

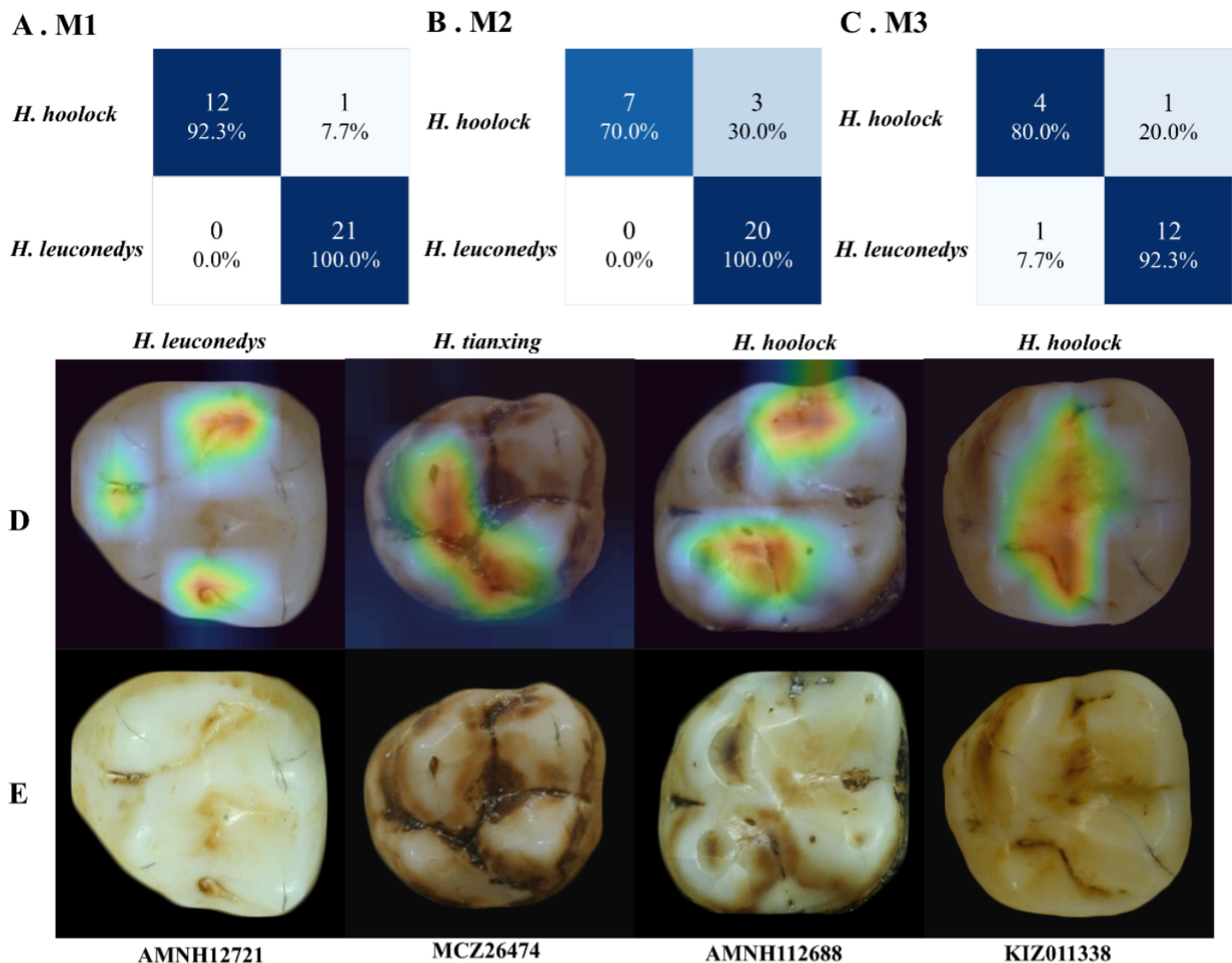


Figure 2

ConvNeXt-Based dental classification performance and feature visualization. (A–C) Confusion matrices for isolated upper molars M1 (A), M2 (B), and M3 (C) from the test set. Rows and columns represent true and predicted classes, respectively, with classes ordered as *H. hoolock* and *H. leuconedys* in both dimensions. Cells display image counts and accuracies. (D) Representative Grad-CAM activation maps, where warmer colors indicate regions contributing more strongly to model predictions. The teeth shown in (D) and (E) are M1s; corresponding M2 and M3 images are provided in Fig. S4.

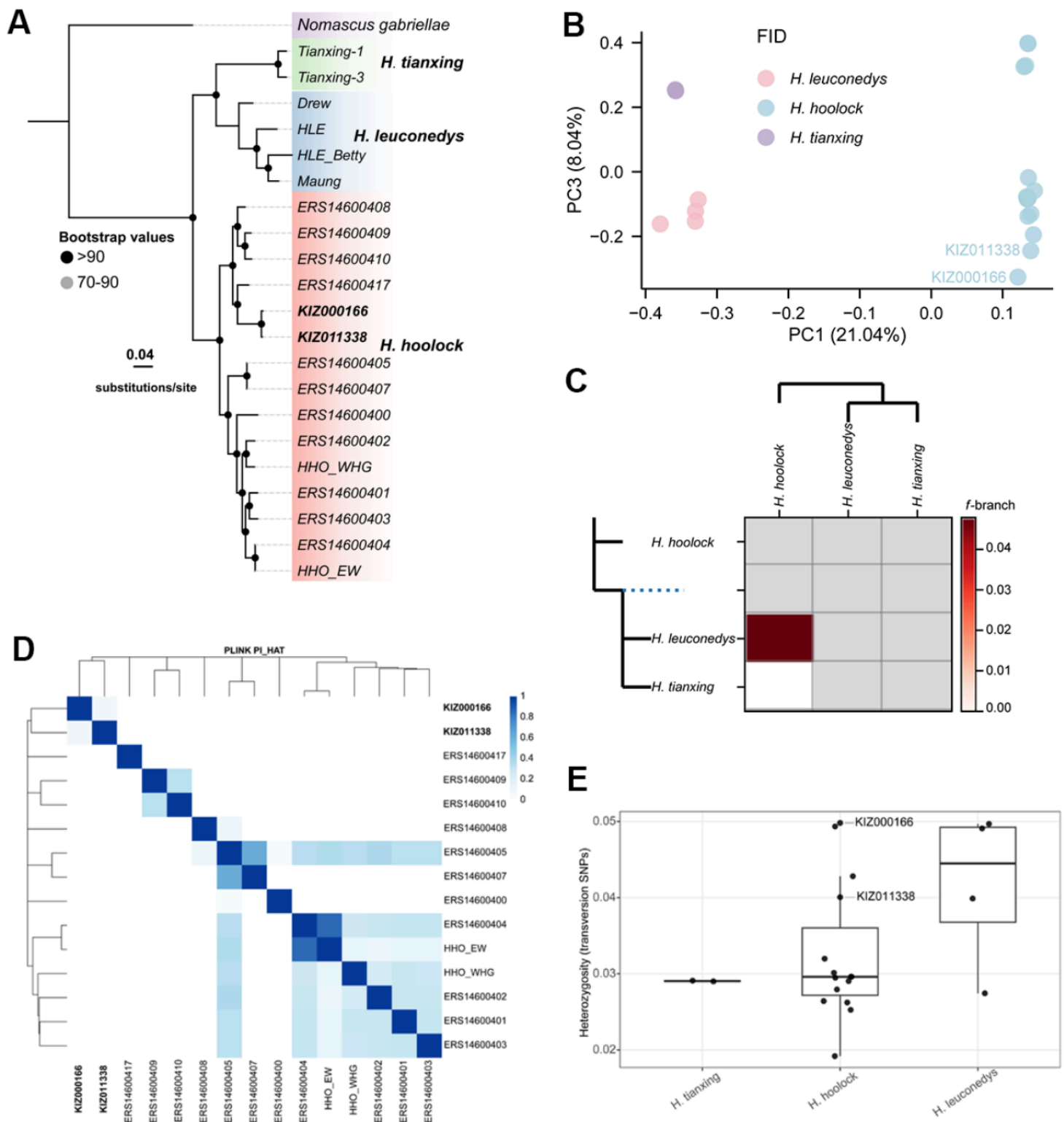


Figure 3

Genomic affinities and population structure of the Datang museum specimens. (A) Maximum-likelihood phylogeny inferred from autosomal single-nucleotide polymorphisms (SNPs). Focal historical specimens (KIZ000166 and KIZ011338) are highlighted in bold. Nodal circles indicate bootstrap support. (B) Principal component analysis (PCA) of autosomal SNPs. Colors indicate species assignments. Focal specimens are explicitly labeled. (C) *f*-branch statistics visualizing historical gene flow among *Hoolock*

species; color intensity reflects the magnitude of excess allele sharing. **(D)** Hierarchically clustered heatmap of pairwise genetic relatedness (PI_HAT) among *H. hoolock* individuals. Darker colors indicate higher relatedness. **(E)** Individual observed heterozygosity based on autosomal transversion SNPs.

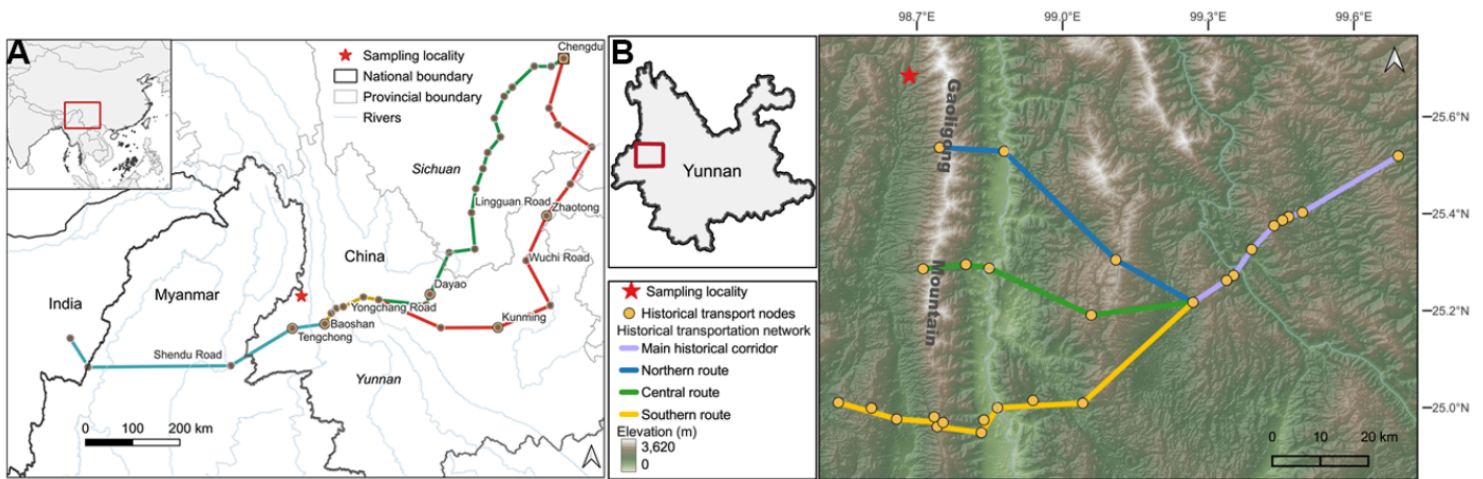


Figure 4

Geographic context and schematic reconstruction of historical transport corridors. **(A)** Regional overview of the Southwestern Silk Road network (including the Shendu and Yongchang roads) connecting South Asia, Myanmar, and China. **(B)** Map of the Gaoligong Mountain and the three principal routes of the Southwestern Silk Road traversing the mountain range: the northern (blue), central (green), and southern (yellow) routes. In both panels, the red star marks the sample locality (Lengshuihe, Datang) of the focal *H. hoolock* specimens.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfigures.docx](#)
- [Supplementarytables.xlsx](#)