Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Whole-genome resequencing reveals strong purifying selection and ribosome pathway enrichment in a multi-generational captive Sunda pangolin (Manis javanica) population

Xiaobing Guo

guoxiaobin@gxlky.com.cn

Guangxi Forestry Research Institute

Shanghua Xu

Guangxi Forestry Research Institute

Miaomiao Jia

Guangxi Forestry Research Institute

Yong Pan

Guangxi Forestry Research Institute

Yuan Lin

Guangxi Forestry Research Institute

Xinyue Li

Guangxi Forestry Research Institute

Haitao Qiu

Guangxi Forestry Research Institute

Shijin Lei

Guangxi Forestry Research Institute

Dingyu Yan

Guangxi Forestry Research Institute

Wenhui Liang

Guangxi Forestry Research Institute

Research Article

Keywords: Manis javanica (Sunda pangolin), whole-genome resequencing, Single nucleotide polymorphisms (SNPs), Insertions/deletions (InDels), Captive population

Posted Date: July 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10354155/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
No competing interests reported.

Table 1 and 2 are available in the Supplementary Files section.

Original Article

Whole-genome resequencing reveals strong purifying selection and ribosome pathway enrichment in a multi-generational captive Sunda pangolin (*Manis javanica*) population

Xiaobing Guo[1,2 *], Shanghua Xu[1,2], Miaomiao Jia[1,2], Yong Pan[1,2], Yuan Lin[1,2], Xinyue Li[1,2], Haitao Qiu[1], Shijin Lei[1], Dingyu Yan[1*], Wenhui Liang[1,2*]

[1]Guangxi Forestry Research Institute, Nanning 530002, P.R. China

[2]Guangxi Laboratory of Forestry, Nanning 530002, P.R. China

[*] = **Corresponding authors:**

Xiaobing Guo, Guangxi Forestry Research Institute, Nanning, China

Email: guoxiaobin@gxlky.com.cn

Dingyu Yan, Guangxi Forestry Research Institute, Nanning, China

Email: yandy6@126.com

Wenhui Liang, Guangxi Forestry Research Institute, Nanning, China

Email: l.wenhui@163.com

Abstract

Background: The Sunda pangolin (*Manis javanica*), a critically endangered species, faces severe population decline due to poaching and illegal trade, making ex situ conservation breeding a critical safeguard. However, a lack of genomic resources hinders the genetic management and long-term viability of captive populations. We performed whole-genome resequencing of 12 captive *M. javanica* individuals from the Pangolin Rescue and Breeding Centre of the Guangxi Forestry Research Institute, China. This cohort represented a multi-generational captive population spanning F0 to F3 generations. Genome-wide single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) were identified, annotated, and subjected to functional

enrichment analyses.

Results: A total of 380.62 Gb of high-quality clean data were generated, with an average sequencing depth of 6.57–15.78×. Each individual harbored 5.2–7.8 million SNPs and approximately 0.6–1.0 million InDels. Most variants were located in intergenic and intronic regions, whereas coding regions contained relatively fewer variants. The distinct length distribution of coding-sequence InDels, together with the predominance of synonymous over nonsynonymous SNPs, suggested strong purifying selection in functional genomic regions. Functional enrichment analyses revealed convergent enrichment of both SNP- and InDel-containing genes in ribosome-related pathways, while additional variant-associated genes were involved in reproductive development, gamete generation, and DNA metabolism.

Conclusions: This study provides the first comprehensive genome-wide variant map of a multi-generational captive Sunda pangolin population in Guangxi, China. These findings offer candidate molecular markers for genetic monitoring, kinship assessment, inbreeding avoidance, and breeding-pair optimization, thereby supporting evidence-based management of reproductive difficulty and genetic decline in captive conservation programs.

Keywords: *Manis javanica* (Sunda pangolin); whole-genome resequencing; Single nucleotide polymorphisms (SNPs); Insertions/deletions (InDels); Captive population

Background

Pangolins (order Pholidota), the world's only scaled mammals, comprise eight extant species belonging to African and Asian genera[1]. Three of the four Asian pangolin species, including the Philippine pangolin (*M. culionensis*), Sunda pangolin (*M. javanica*), and Chinese pangolin (*M. pentadactyla*), are listed as Critically Endangered on the IUCN Red List[2-4]. Among them, the Sunda pangolin (*M. javanica*) has experienced particularly severe population declines due to intensive poaching and illegal trade for its scales and meat [5]. The *M. javanica* is distributed across mainland Southeast Asia, including Thailand, Malaysia, Myanmar, Laos, Cambodia, and Vietnam, as well as throughout the Indonesian archipelago with its northern range extending into Yunnan provinces of southern China[5-7]. Historical data from regions within its range indicate large-scale offtake; for example, annual harvests of pangolins in China during the latter half of the 20th century reached hundreds of thousands of individuals, underscoring the immense historical pressure that has contributed to the species' precarious status[8, 9]. In response, ex situ conservation breeding has

been established as a critical safeguard[10, 11].

However, the success of captive assurance colonies is hindered by persistent biological and management challenges. High mortality rates, reproductive difficulties, and health disorders linked to dietary mismatches are commonly reported[10, 12, 13]. These phenotypic problems may not be fully explained by husbandry conditions alone, but may also reflect underlying functional genomic variation, reduced adaptive potential, or the accumulation of deleterious mutations, collectively referred to as genetic load [14]. In small and closed captive populations, the erosion of genomic diversity, increased inbreeding, and genetic drift can further accelerate the exposure and accumulation of such harmful variants[15-17]. Consequently, dietary and environmental optimization alone may be insufficient to ensure the long-term viability of captive populations. A functional genomic assessment is therefore urgently needed to clarify the molecular basis of captive health and reproductive challenges and to guide evidence-based genetic management [14]. Notably, in our study, the Pangolin Rescue and Breeding Centre of the Guangxi Forestry Research Institute has successfully established a multi-generational captive breeding population of Sunda pangolins, with continuous reproduction from founder individuals (F0) to the third filial generation (F3). This rare F0–F3 breeding cohort provides an exceptional model for investigating genome-wide variation, genetic stability, and functional genomic features associated with captive adaptation and reproductive success.

Advances in conservation genomics provide a powerful framework for addressing the genetic risks associated with captive breeding. However, previous genomic studies on pangolins have mainly focused on phylogenetic relationships, adaptive evolution, demographic history, and the forensic tracing of illegally traded individuals [18, 19]. Moreover, existing studies have generally involved limited sample sizes or incomplete geographic representation, which restricts their applicability to the genetic management of ex situ breeding colonies [18, 20]. There is a pronounced lack of comprehensive, genome-wide characterization of functional genetic variation—specifically single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels)—within captive Sunda pangolin populations[21, 22]. This knowledge gap hinders accurate assessment of the genetic health of captive populations, impedes the formulation of evidence-based breeding strategies, and may ultimately result in genetic deterioration in captive colonies.

Leveraging a unique multi-generational breeding pedigree (F0 to F3), we performed whole-genome

resequencing of 12 Sunda pangolins from a multi-generational breeding population spanning F0 to F3 generations at the Pangolin Rescue and Breeding Centre of the Guangxi Forestry Research Institute, China. As a key methodological approach in conservation genomics, whole-genome resequencing enables high-resolution surveys of genome-wide variation, including SNPs and InDels, thereby allowing precise estimation of genetic diversity, inbreeding, and mutational load[23-25]. We systematically identified genome-wide SNPs and InDels, characterized their genomic distribution patterns and selective constraints, and explored key genes and biological pathways potentially associated with captive adaptation and reproductive capacity. By constructing the first comprehensive genome-wide variant map for a captive Sunda pangolin population, this study provides an essential molecular marker resource for genetic monitoring, kinship assessment, inbreeding avoidance, and breeding-pair selection. These findings are expected to improve the precision of genetic management and enhance the long-term conservation value of ex situ Sunda pangolin breeding programs.

Methods

Experimental animals

Saliva samples were collected from 12 adult *M. javanica* at the Pangolin Rescue and Breeding Centre (PRBC) of the Guangxi Forestry Research Institute in Nanning, China (**Table S1**). This cohort represented a multi-generational captive breeding population spanning founder individuals (F0) to the third filial generation (F3). To our knowledge, this population is among the few captive Sunda pangolin populations worldwide that have achieved stable reproduction to the F3 generation. According to the available pedigree records, C31, CC7, CC18, and CC19 were offspring of M6, among which CC7 and CC18 shared the same parents. CCC6 was the offspring of CC7. No close kinship was recorded among the remaining individuals.

Captive breeding and management system for *M. javanica*

The ex situ conservation and breeding program for *M. javanica* at the PRBC of the Guangxi Forestry Research Institute established a specialized husbandry and housing protocol to support the species' long-term survival and eventual reintroduction. The core management strategy centered on replicating the species' natural ecological niche and addressing its specific physiological needs within a controlled captive environment.

A primary focus was the design of individual enclosures that provided environmental choice and thermal comfort. Each unit was compartmentalized into a primary activity area and two distinct, insulated dens—one designated for winter and an underground burrow for summer. This structure allowed individuals to self-regulate their micro-environment according to seasonal and daily temperature fluctuations, a critical consideration for a species native to tropical Southeast Asia. To safeguard animal health, a rigorous hygiene protocol was implemented, featuring regular cleaning of cage surfaces and periodic high-temperature disinfection of the resting dens.

Nutritional management constituted another cornerstone of the program. A bespoke, nutritionally complete semi-liquid diet was formulated, composed primarily of insect-derived proteins (including black ant, silkworm pupae, and mealworm powders) supplemented with plant-based protein, minerals, and essential vitamins. This standardized formula was delivered daily, with fresh water provided ad libitum, and all foodstuffs were stored under conditions preventing pathogen contamination.

DNA extraction, library construction, and sequencing

Genomic DNA was extracted from all saliva samples using the Magnetic Universal Genomic DNA Kit (TIANGEN, China; Cat. No. DP705). DNA quality was assessed by evaluating degradation and contamination via 1% agarose gel electrophoresis, determining purity based on the OD260/280 ratio measured with a Nanodrop spectrophotometer, and quantifying concentration using the Qubit® DNA Assay Kit on a Qubit® 3.0 Fluorometer (Invitrogen, USA).

For library preparation, 0.2 μg of DNA per sample served as input. Sequencing libraries were constructed with the NEB Next® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) according to the manufacturer's instructions, with unique index codes assigned to each sample. In brief, genomic DNA was sheared by sonication to an average size of 350 bp. The fragments were subsequently end-repaired, A-tailed, and ligated to Illumina full-length adapters, followed by PCR amplification. Amplified products were purified with the AMPure XP system (Beckman Coulter, USA), and library concentration was measured using the Qubit® 3.0 Fluorometer. Size distribution was assessed on an Agilent 2100 Bioanalyzer, and final quantification was performed by real-time PCR (all libraries > 2 nM).

Indexed libraries were clustered on a cBot Cluster Generation System with the Illumina PE Cluster Kit (Illumina, USA) following the standard protocol. After cluster generation, paired-end sequencing

(2 × 150 bp) was carried out on an Illumina platform.

Sequencing data processing and alignment

Raw sequencing data in FASTQ format were generated from the original fluorescence images through base calling on the Illumina platform. Prior to alignment, quality control was performed to remove artifacts including adapter contamination, low-quality bases, and ambiguous nucleotides (N), which could otherwise compromise downstream analyses. Paired-end reads were filtered if either read met any of the following criteria: containing adapter sequence (>10-nt aligned, ≤10% mismatch), consisting of >10% uncertain bases, or having >50% of bases with Phred quality score <5.

High-quality reads were aligned to the reference genome using the Burrows–Wheeler Aligner (BWA-MEM) with parameters "-t 4 -k 32 -M". The resulting alignments were stored in BAM format, and PCR duplicates were removed using SAMtools (rmdup) and Picard tools.

Variant calling and functional annotation

Variant calling was performed using the HaplotypeCaller in GATK (v4.0). Raw single-nucleotide polymorphisms (SNPs) and small insertions/deletions (InDels) were filtered with the following criteria: sequencing depth at the variant site >4× and mapping quality >20. Genetic diversity parameters were calculated based on this high-confidence filtered variant set. For each diploid individual, a heterozygous SNP site was defined based on standard diploid genotype calling (0/1) with a minor allele frequency (MAF) filtering across the cohort to exclude potential sequencing errors. Observed heterozygosity (H_{obs}) for each sample was calculated as the ratio of the number of heterozygous SNPs to the total number of callable variant sites:

$$H_{obs} = \frac{N_{hets}}{N_{callable}}$$

where N_{hets} is the number of heterozygous SNPs meeting the MAF criterion, and $N_{callable}$ is the total number of high-confidence callable SNP sites. Heterozygosity estimates were generated using VCFtools (v0.1.16) with the --hetflag, and results were cross-validated using bcftools (v1.17) stats. All transition/transversion (Ti/Tv) and heterozygosity calculations were restricted to biallelic SNPs; InDels were excluded from these metrics to avoid potential alignment artifacts. The Ti/Tv ratio, used to assess variant calling fidelity and characterize the mutational spectrum of the host genome, was calculated as the ratio of transitions (purine-to-purine [A↔G] or pyrimidine-to-pyrimidine [C↔T]

substitutions, N_{Ti}) to transversions (all other nucleotide substitution types, N_{Tv}):

$$Ti/Tv = \frac{N_{Ti}}{N_{Tv}}$$

Ti/Tv values were extracted directly from VCFtools --TsTvoutputs and validated using bcftools stats. Functional classification of variants (including synonymous, non-synonymous, stop-gain, stop-loss, and splice-site altering variants) was performed using ANNOVAR with gene model annotations. To ensure annotation consistency, a random 10% subset of variants was re-annotated using SnpEff (v5.1), with a concordance rate exceeding 98%.

Functional annotation of the filtered variants was carried out with ANNOVAR, with gene and regional annotations based on the UCSC known gene database.

To investigate the potential biological impact, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed on genes harboring the annotated SNPs and InDels using the online tool KOBAS 3.0. Results were visualized using the ggplot2 package in R.

Results

Data quality and genomic mapping efficiency

Sequencing generated a total of 386.97 Gb of raw data, yielding 380.62 Gb of high-quality clean data after filtering. The raw data per sample ranged from 24.24 to 53.84 Gb. The effective rate ranged from 97.18% to 98.96%, indicating that most raw reads were retained after removal of low-quality sequences and adapter contamination. All samples showed consistently low base error rates (0.01%) and high sequencing quality, with Q20 values ranging from 98.10% to 98.77% and Q30 values ranging from 95.16% to 96.65%. The consistently high Q30 values indicated that the majority of bases were called with high confidence. In addition, the GC content was stable across samples, ranging from 41.35% to 42.70%, suggesting a normal and uniform base composition without obvious GC bias (**Table 1**). These results confirm that all samples met the required data volume, possessed high sequencing quality, showed normal GC distribution, and were successfully sequenced, thereby ensuring the reliability of data for downstream analyses.

Clean reads from each sample were aligned to the reference genome (*Manis javanica*, assembly ManJav1.0, GCF_001685135.1; genome size: 2.55 Gb) using BWA-MEM with parameters “mem -t 4 -k 32 -M”. PCR duplicates were marked and removed using SAMtools rmdup. The mapping rates

ranged from 57.19% to 96.42% across samples. Most samples showed mapping rates above 80%, whereas C22 (57.19%) and CCC6 (68.75%) had relatively lower mapping rates. Because saliva samples were collected from animals maintained in a captive environment, the lower mapping rates in some samples may reflect the presence of oral or environmental microbial DNA, or partial DNA degradation. Nevertheless, the mapped reads still provided sufficient genomic coverage and sequencing depth to support downstream genome-wide variant detection. After duplicate removal, the average sequencing depth (excluding N-regions) varied between 6.57× and 15.78×. The fraction of the reference genome covered by at least one read (1× coverage) was greater than 97.18% in all samples, with coverage of at least 4× ranging from 73.75% to 96.38% (**Table 2**). These results indicate that the sequencing data provided sufficient depth and breadth of coverage for reliable downstream variant detection.

Genomic variant landscape and genetic diversity

Genome-wide variant calling identified a large number of SNPs and InDels across the 12 captive Sunda pangolin individuals. The total number of SNPs ranged from 5,241,332 to 7,842,901 per individual, with an average of 6,269,502 SNPs (**Table S2**). Most SNPs were located in non-coding regions, particularly intergenic and intronic regions, indicating that genetic variation was predominantly distributed outside protein-coding sequences. In contrast, exonic SNPs accounted for 0.79% of the total variants (**Table S2**), suggesting a relatively conserved coding genome and indicating strong conservation of core genomic functions.

Within the coding regions, synonymous SNPs were more abundant than nonsynonymous SNPs. On average, synonymous variants accounted for approximately 51.32% of coding SNPs, whereas nonsynonymous variants accounted for 37.30%. Stop-gain and stop-loss mutations are relatively rare (0.38% and 0.08%, respectively), while the remaining unknowns account for 10.92% (**Table S2**). The proportion of nonsynonymous SNPs in coding regions was markedly higher than that of other coding variant categories excluding synonymous substitutions, potentially indicating an accumulation of functionally relevant mutations, which may reflect genetic drift in small captive populations.

For InDels, each individual carried 595,094 to 1,032,402 variants, with an average of 741,753 InDels (**Table S3**). Similar to SNPs, most InDels were located in intergenic and intronic regions, whereas exonic InDels were relatively rare. Within exons, frameshift and non-frameshift deletions were more

frequent than insertions. Notably, a small number of stop-gain and stop-loss variants were detected in all samples. Variants in upstream and downstream regulatory regions were also abundant. Across all genomic categories, the number of deletions consistently exceeded insertions.

Analysis of SNPs mutation spectra showed that the six major mutation categories were consistently represented across individuals (**Figure 1**). Among them, C>T and T>C transitions were the predominant mutation types, whereas transversion categories were less frequent. The transition/transversion ratio (Ti/Tv) ranged from 2.26 to 2.31, with an average value of 2.28 (**Table S4**). This relatively stable Ti/Tv ratio among samples supports the overall reliability of SNPs detection and indicates similar mutational patterns across the captive cohort.

The six substitution classes are shown on the x-axis, and SNP counts are shown on the y-axis. C:G>T:A and T:A>C:G transitions dominate across all individuals, indicating a conserved mammalian mutational pattern characterized by a strong transition bias.

Genetic diversity estimates further showed that the 12 individuals maintained detectable levels of heterozygosity (**Table S4**). The number of heterozygous SNPs ranged from 59,175 to 141,800 per individual, with an average of 94,251 heterozygous SNPs. SNP heterozygosity ranged from 1.13% to 1.81%, with an average of 1.48%. InDel heterozygosity was much lower, ranging from 0.10‰ to 0.18‰, with an average of 0.14‰. When stratified by generation, mean SNP heterozygosity was 1.406% in wild-caught founders (F0), and remained relatively stable in subsequent captive-born generations (F1: 1.478%; F2: 1.505%), with a higher value observed in the F3 generation (1.756%). This pattern suggests that the captive breeding program at the Pangolin Rescue and Breeding Centre (PRBC) has effectively maintained genome-wide genetic diversity across multiple generations.

Together, these results indicate that the captive Sunda pangolin population retained measurable genetic diversity at the genome-wide level, while the lower heterozygosity observed for SNPs and InDels suggest stronger selective constraints acting on structurally disruptive variation.

Strong purifying selection in coding sequences (CDS)

Based on the non-random genomic distribution of variants observed above, we further investigated the functional patterns of mutations within coding sequences. Within CDS regions, SNPs functional annotation showed a clear predominance of synonymous over nonsynonymous substitutions. As summarized in **Table S5**, synonymous SNPs accounted for 51.32% of coding variants, whereas nonsynonymous SNPs represented 37.30%. Stop-gain and stop-loss mutations were rare (0.36% and

0.08%, respectively), further suggesting that protein-altering mutations are efficiently constrained in the coding genome.

Analysis of InDel functional categories revealed a similar pattern of selective constraint. Frameshift mutations were more frequent than non-frameshift mutations; however, overall InDels affecting coding sequences remained limited (**Table S5)**. Importantly, InDels exhibited a strong length-dependent distribution bias. As shown in **Figure 2A**, genome-wide InDel length distribution displayed a broader spectrum, with detectable longer InDels across non-coding regions. 1-bp insertions and deletions overwhelmingly dominated CDS InDels, whereas longer InDels (⩾2 bp and ⩾3 bp) were extremely rare (**Figure 2B**). This marked contrast between CDS and genome-wide patterns suggests stronger purifying selection acting against longer and potentially disruptive InDels in coding regions. Together, the enrichment of synonymous over nonsynonymous SNPs, combined with the strong depletion of long InDels and the dominance of 1-bp InDels in CDS, indicates that protein-coding regions in this captive Sunda pangolin population are under strong purifying selection, effectively removing potentially deleterious functional variation.

Convergent enrichment of variants in the Ribosome pathway

KEGG enrichment analysis revealed that SNPs-associated genes were significantly enriched in ribosome biogenesis in eukaryotes ($p < 0.05$) (**Figure 3A**). In contrast, the metabolic pathways term, while containing the largest number of enriched genes, did not show significant statistical enrichment. Other pathways, such as neuroactive ligand-receptor interaction and tight junction, were identified but showed weak enrichment signals and did not reach statistical significance.

Similarly, InDel-associated genes displayed a highly consistent enrichment profile, with ribosome-related pathways again ranking among the top enriched categories (**Figure 3B**). This highlights a potential role for InDel variations in influencing translational machinery. Other pathways showing enrichment included vascular smooth muscle contraction, lysosome, and RNA transport. In contrast, metabolic pathways, while containing a high number of input genes, did not show statistically significant enrichment. This pattern suggests that InDel variations are more specifically associated with core cellular processes and specialized functions rather than general metabolism.

Importantly, several ribosome-associated genes were consistently identified in both SNPs and InDels enrichment datasets (**Table S6**), including *GNL3, NXF1B, RBM28, MRPL24,* and *RPSA*, which are involved in ribosome biogenesis and RNA-binding processes. These shared genes provide strong

evidence for a convergent functional signal targeting the protein synthesis machinery.

Together, these results demonstrate that both SNPs and InDels consistently converge on ribosome-related pathways, suggesting that the protein synthesis apparatus represents a central functional target of genome-wide variation in this captive Sunda pangolin population.

Variants associated with reproductive development and gamete generation

Gene Ontology analysis of the identified SNPs and InDels revealed a consistent signal of reproductive-related functional enrichment across both variant types (**Figure 4**). In the SNPs-based analysis, enriched terms were predominantly associated with reproductive development and gamete-related processes, including sexual reproduction, gamete generation, reproductive development, as well as DNA recombination and RNA processing (**Figure 4A**), suggesting that SNPs-associated variation is broadly distributed across meiosis-linked genetic processes and transcriptional regulation pathways that are fundamental to germline development. The InDels-based analysis showed a functionally focused enrichment pattern, in which the reproduction biological process was clearly enriched and accompanied by strong signals in cell cycle checkpoint and DNA damage checkpoint pathways (**Figure 4B**), suggesting a potential involvement of InDels in genes related to genomic integrity maintenance and cell-cycle regulation during reproductive cell development.

Go analysis results show that both variant types consistently converged on reproduction as a significantly enriched biological process, suggesting that genetic variation is non-randomly distributed in genes associated with reproductive function in this multi-generational captive population. Notably, SNP-associated signals appear to be more closely linked to meiotic recombination and early germline developmental regulation, whereas InDel-associated genes are more enriched in downstream processes ensuring cell-cycle fidelity and DNA repair, indicating complementary roles of different variant types in reproductive biology.

Functional annotation of reproductive pathway-associated genes (**Table S7**) further identified a set of candidate loci linking genomic variation to reproductive regulation in the captive Sunda pangolin population. These genes converged on a limited number of key regulatory axes. Among them, ADCY8 and ADCY2B**,** encoding adenylate cyclase enzymes, were consistently associated with hormone-responsive reproductive pathways, including GnRH-related signaling and cyclic AMP-mediated processes, which are central to gonadotropin signal transduction and oocyte maturation. These genes may therefore reflect variation affecting upstream endocrine regulation of

gametogenesis. In addition, *FBXW11A* and *FBXW11B*, two paralogous genes encoding F-box and WD repeat domain-containing protein 11, were detected within oocyte meiosis-related functional categories. As a core substrate recognition subunit of the SCF E3 ubiquitin ligase complex, FBXW11 mediates the ubiquitination and proteasomal degradation of key cell cycle regulators, and is essential for spindle assembly, chromosome segregation, and DNA damage checkpoint control during mammalian oocyte meiosis. Variants in these genes may therefore influence the fidelity of germ cell division and the stability of gamete development. The presence of two FBXW11 paralogs likely reflects gene duplication events or annotation fragmentation in the current *Manis javanica* reference genome, rather than functionally distinct genes.

Overall, these candidate genes provide convergent evidence that reproductive regulation in this captive population may be influenced by genetic variation affecting both endocrine signaling (ADCY family) and ubiquitin-mediated cell-cycle control (FBXW11 axis), supporting the GO-level enrichment patterns observed in **Figure 4**.

Discussion

This study provides the first comprehensive whole-genome resequencing analysis of a captive *Manis javanica* population, constructing the first genome-wide variant map for a multi-generational ex situ breeding cohort. The sequencing data quality in the present study (Q30 ≥ 95.16%) and the scale of variant detection are comparable to previously reported genomic studies of captive Chinese pangolin populations [23], thereby supporting the reliability and robustness of the dataset. Based on these high-quality data, our study further reveals genome-wide patterns of genetic variation, selective constraints, and functional distribution of variants in a multi-generational captive breeding system, providing a foundational framework for future genetic management and conservation breeding strategies in this species.

The genomic distribution of SNPs and InDels observed shows a strong enrichment in intergenic and intronic regions, with only a minor fraction located in exonic sequences. This overall pattern is consistent with findings reported in the Chinese pangolin genome [23] as well as in other mammalian species [26], suggesting that such a distribution represents a conserved feature of mammalian genomes under evolutionary constraint. The Ti/Tv ratio across all samples averaged 2.12, which falls within the typical range observed in mammalian genomes and further supports the accuracy and reliability of SNPs detection in this study [27, 28]. The predominance of C:G>T:A and T:A>C:G

transitions among SNPs types is a common mutational signature observed across many species, often attributed to cytosine deamination and other endogenous mutational processes [29-31]. The InDel profile, characterized by a strong bias toward 1-bp variants in coding sequences, together with the relatively higher proportion of synonymous versus nonsynonymous substitutions in SNPs (ratio = 1.38), provides evidence consistent with purifying selection acting on protein-coding regions. This pattern suggests that deleterious functional mutations are likely being efficiently removed from the population, which may contribute to maintaining the genetic integrity of this captive breeding cohort. Overall, this non-random distribution of genomic variation indicates that functionally important genomic regions are under stronger evolutionary constraint.

The KEGG analyses reveal the ribosome pathway is consistently enriched in both SNP- and InDel-associated gene sets. This cross-variant concordance strongly suggests that ribosome-related genetic variation may represent a central and non-random feature of the genomic architecture in this captive *Manis javanica* population, potentially playing a key role in physiological adaptation under ex situ breeding conditions. Ribosomes constitute the core machinery of cellular protein synthesis, and genetic variation in ribosomal components may directly influence translational efficiency, fidelity, and cellular stress responsiveness [32, 33]. In the context of captive environments, pangolins are exposed to multiple artificial stressors, including dietary substitution, temperature fluctuation, and increased population density, all of which can impose substantial metabolic demands on protein synthesis systems [13, 34, 35]. Variants in ribosome-related genes may therefore affect the capacity to maintain translational homeostasis under these conditions, potentially contributing to commonly observed captive phenotypes such as reduced growth performance and impaired immune system [10, 32]. Notably, previous studies have demonstrated that ribosomal protein gene variation is closely associated with stress response regulation and adaptive evolution across diverse organisms [30, 36, 37], further supporting the functional relevance of this signal. In addition to ribosome-related pathways, enrichment of genes involved in sulfur amino acid metabolism may also reflect species-specific biological adaptation. Given the specialized myrmecophagous diet of pangolins, which is characterized by insect consumption with unique amino acid compositions, variation in sulfur amino acid metabolic pathways may be linked to dietary adaptation and metabolic efficiency [38, 39]. Moreover, considering that keratin is the primary structural component of pangolin scales and is rich in sulfur-containing amino acids, such variation may also have indirect implications for scale

formation and structural integrity [40]. Notably, reproduction- and gamete generation-related genes were significantly enriched in both SNP- and InDel-associated gene sets, further suggesting a potential link between these variants and reproductive capacity in this captive population. Taken together, these functional enrichments highlight a coordinated pattern of variation affecting protein synthesis, metabolic adaptation, and reproductive biology, providing a potential genomic basis for understanding physiological constraints in captive *M. javanica* populations.

From a conservation genetics perspective, this catalog of genome-wide variants represents a valuable genomic resource for understanding genetic variation in a critically endangered species under captive management. The present dataset may contribute to evaluating genetic diversity, population structure, and long-term viability of ex situ breeding populations. The relatively low heterozygosity observed in InDel loci across all individuals (generally <0.2%) suggests a reduced level of genetic diversity within this captive population. This pattern may be attributable to founder effects and the cumulative impact of genetic drift in a small, closed breeding system [41]. Such reductions in genetic variability are of particular concern in conservation programs, as they may increase susceptibility to disease, reduce adaptive potential, and ultimately compromise long-term population persistence [42, 43]. Therefore, maintaining and monitoring genome-wide diversity should be considered a central objective in the genetic management of captive *M. javanica* populations [44]. In this context, the present variant dataset provides a foundational resource for the development of practical conservation tools, including molecular markers for genetic diversity assessment, kinship inference, and breeding pair optimization [45]. Integration of these genomic resources into routine captive management may facilitate more informed breeding decisions and help mitigate further loss of genetic diversity, thereby enhancing the sustainability of ex situ conservation programs for this species.

Several limitations of the current study should be acknowledged. First, the analysis is based on a relatively small, captive cohort from a single breeding center. This limits the ability to generalize findings to the wider, wild *M. javanica* population across Southeast Asia, which is known to harbor significant genetic structure[22]. Second, while functional enrichment points to biologically relevant pathways, these are in silico predictions. The actual phenotypic consequences of the identified variants, particularly those in the ribosome pathway, require experimental validation[46, 47]. Future research should prioritize expanding genomic surveys to wild populations with known geographical

origins. This would enable robust population genetic analyses, estimation of genetic diversity and inbreeding, and the identification of evolutionarily significant units for conservation[24, 48]. Furthermore, integrating genomic data with phenotypic and ecological data could help unveil the adaptive significance of the identified variants [48], particularly those enriched in pathways related to metabolism, reproduction, and environmental response. Importantly, the genome-wide SNPs and InDels dataset generated in this study provides a valuable resource for developing low-cost SNP panels for routine genetic monitoring in captive breeding programs, enabling efficient kinship inference and pedigree management to support informed conservation breeding strategies [49].

Conclusion

In conclusion, this study provides the first detailed genome-wide map of single nucleotide polymorphisms and insertion-deletion variants for the Sunda pangolin (*Manis javanica*) based on whole-genome resequencing. We systematically characterized the spectrum and distribution of SNPs and InDels and demonstrated that coding-region variants are subject to strong purifying selection, indicating that deleterious mutations in functionally important regions are likely being efficiently constrained in this captive breeding system. Functional analyses further revealed that genes associated with ribosome-related processes and reproductive functions are significantly enriched for SNPs and InDels, suggesting that these core biological systems may be important components of genomic constraint and adaptation in ex situ conditions. Importantly, this study establishes a transition in pangolin conservation research from phenotype-based observation toward genome-level monitoring, providing a foundational genomic framework for future conservation genetic management. These findings not only contribute to the foundational genomic knowledge of a critically endangered mammal but also deliver a vital toolkit of molecular markers which can support the development of precise genetic management tools for captive breeding programs, including monitoring of genetic diversity and optimization of breeding strategies, thereby enhancing the long-term sustainability of ex situ conservation efforts for this highly endangered and heavily trafficked species.

References

1. Philippe, G., R. W. John, P. H. Sean and J. G. Timothy. "Phylogeny and systematics." In *Pangolins science, society and conservation*. D. S. CHALLENDER, HELENC.NASH and C. WATERMAN. P. J. NYHUS. 2. London: Elsevier, 2020, 25–37.
2. Schoppe, S., L. Katsis and L. Lagrada. "Manis culionensis." *The IUCN Red List of Threatened Species* (2019): e.T136497A123586862. https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T136497A123586862.en.
3. Challender, D., D. H. A. Willcox, E. Panjang, N. Lim, Nash, H, S. Heinrich and J. Chong. "Manis javanica." *The IUCN Red List of Threatened Species* (2019): e.T12763A123584856. https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T12763A123584856.en.
4. Challender, D., S. Wu, P. Kaspal, A. Khatiwada, A. Ghose, N. Ching-Min Sun, R. K. Mohapatra and T. Laxmi Suwal. "Manis pentadactyla." *The IUCN Red List of Threatened Species* (2019): e.T12764A168392151. https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T12764A168392151.en.
5. Guo, X., S. Xu, W. Liang, M. Jia, Y. Pan, Y. Lin and X. Li. "Identification of the complete mitochondrial genome of the malayan pangolin (manis javanica demarest, 1822) and its evolutionary relationship with other pangolin species." *Genes* 17 (2026): 10.3390/genes17050498.
6. Chong, J. L., E. Panjang, D. Willcox, H. C.Nash, G. Semiadi, W. Sodsai, N. T.-L. Lim, L. Fletcher, A. Kurniawan and S. Cheema. "Sunda pangolin manis javanica (desmarest, 1822)." In *Pangolins science, society and conservation*. D. S. CHALLENDER, HELENC.NASH and C. WATERMAN. P. J. NYHUS. 6. London: Elsevier, 2020, 90–104.
7. Cen, P., J. Sun, Q. Wang, F. Zhang, L. Mo, A. Mahmood, J. Li, Y. Wang and S. Wu. "Only sunda and chinese pangolin (pholidota) are naturally distributed in china." *Integr Zool* 18 (2023): 704–09. 10.1111/1749-4877.12697. https://www.ncbi.nlm.nih.gov/pubmed/36519580.
8. Zhang, F., S. Wu and P. Cen. "The past, present and future of the pangolin in mainland china." *Global Ecology and Conservation* 33 (2022): 10.1016/j.gecco.2021.e01995.

9. Jing-Yang, H., H. Zi-Qian, F. Laurent, W. Shi-Fang, C. Wu, J. Yun-Fang, W. Hong, K. Wei-Min, L. Haipeng, Z. Ya-Ping*, et al.* "Genomic consequences of population decline in critically endangered pangolins and their demographic histories %j national science review." 7 (2020): 798–814. 10.1093/nsr/nwaa031. https://link.cnki.net/doi/10.1093/nsr/nwaa031.

10. Yan, D., X. Zeng, M. Jia, X. Guo, S. Deng, L. Tao, X. Huang, B. Li, C. Huang, T. Que*, et al.* "Successful captive breeding of a malayan pangolin population to the third filial generation." *Communications Biology* 4 (2021): 10.1038/s42003-021-02760-4.

11. Lacher, T. E., S. H. M. Butchart, R. Gumbs, B. Long, C. Lopez-Gallego, D. Raimondo, A. T. Simkins, S. Sunarto and M. Hoffmann. "The status, threats and conservation of critically endangered species." *Nature Reviews Biodiversity* 1 (2025): 421–38. 10.1038/s44358-025-00059-4. https://doi.org/10.1038/s44358-025-00059-4.

12. Hua, L., S. Gong, F. Wang, W. Li, Y. Ge, X. Li and F. Hou. "Captive breeding of pangolins: Current status, problems and future prospects." *ZooKeys* 507 (2015): 99–114. 10.3897/zookeys.507.6970.

13. Wang, X.-M., G. P. J. Janssens, C.-G. Xie, B.-W. Xie, Z.-G. Xie, H.-J. He, Y.-N. Wang and J. Xu. "To save pangolins: A nutritional perspective." *Animals* 12 (2022): 10.3390/ani12223137.

14. Speak, S. A., T. Birley, C. Bortoluzzi, M. D. Clark, L. Percival-Alwyn, H. E. Morales and C. van Oosterhout. "Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations." *Mol Ecol Resour* 24 (2024): e13967. 10.1111/1755-0998.13967. https://www.ncbi.nlm.nih.gov/pubmed/38727721.

15. Hoffmann, A. A., A. D. Miller and A. R. Weeks. "Genetic mixing for population management: From genetic rescue to provenancing." *Evolutionary Applications* 14 (2020): 634–52. 10.1111/eva.13154.

16. Robinson, J., C. C. Kyriazis, S. C. Yuan and K. E. Lohmueller. "Deleterious variation in natural populations and implications for conservation genetics." *Annual Review of Animal Biosciences* 11 (2023): 93–114. 10.1146/annurev-animal-080522-093311.

17. Frankham, R. "Genetics and conservation biology." *Comptes Rendus. Biologies* 326 (2003): 22–29. 10.1016/s1631-0691(03)00023-4.

18. Choo, S. W., M. Rayko, T. K. Tan, R. Hari, A. Komissarov, W. Y. Wee, A. A. Yurchenko, S. Kliver, G. Tamazian, A. Antunes*, et al.* "Pangolin genomes and the evolution of mammalian scales and immunity." *Genome Research* 26 (2016): 1312–22. 10.1101/gr.203521.115.

19. Nash, H. C., Wirdateti, G. W. Low, S. W. Choo, J. L. Chong, G. Semiadi, R. Hari, M. H. Sulaiman, S. T. Turvey, T. A. Evans*, et al.* "Conservation genomics reveals possible illegal trade routes and admixture across pangolin lineages in southeast asia." *Conservation Genetics* 19 (2018): 1083–95. 10.1007/s10592-018-1080-9. https://doi.org/10.1007/s10592-018-1080-9.

20. Frankham, R. "Genetics and extinction." *Biological conservation* 126 (2005): 131–40.

21. Banterng, N., K. Ewart and R. Ogden. "Genomic population structure of the sunda pangolin (manis javanica) in thailand: Implications for conservation and wildlife forensics." (2025): 2025.10.06.680665. 10.1101/2025.10.06.680665 %J bioRxiv. https://www.biorxiv.org/content/biorxiv/early/2025/10/06/2025.10.06.680665.full.pdf.

22. Heighton, S. P., R. Allio, J. Murienne, J. Salmona, H. Meng, C. Scornavacca, A. D. S. Bastos, F. Njiokou, D. W. Pietersen, M.-K. Tilak*, et al.* "Pangolin genomes offer key insights and resources for the world's most trafficked wild mammals." *Molecular Biology and Evolution* 40 (2023): 10.1093/molbev/msad190.

23. Wang, Q., T. Lan, H. Li, S. K. Sahu, M. Shi, Y. Zhu, L. Han, S. Yang, Q. Li, L. Zhang*, et al.* "Whole-genome resequencing of chinese pangolins reveals a population structure and provides insights into their conservation." *Communications Biology* 5 (2022): 10.1038/s42003-022-03757-3.

24. Hohenlohe, P. A., W. C. Funk and O. P. Rajora. "Population genomics for wildlife conservation and management." *Molecular Ecology* 30 (2020): 62–82. 10.1111/mec.15720.

25. Meuwissen, T. H. E., A. K. Sonesson, G. Gebregiwergis and J. A. Woolliams. "Management of genetic diversity in the era of genomics." *Frontiers in Genetics* 11 (2020): 10.3389/fgene.2020.00880.

519 26. Davidson, S., A. Starkey and A. MacKenzie. "Evidence of uneven selective pressure on
520 different subsets of the conserved human genome; implications for the significance of intronic
521 and intergenic DNA." *BMC Genomics* 10 (2009): 10.1186/1471-2164-10-614.
522 27. Wang, J., L. Raskin, D. C. Samuels, Y. Shyr and Y. Guo. "Genome measures used for quality
523 control are dependent on gene function and ancestry." *Bioinformatics* 31 (2015): 318–23.
524 10.1093/bioinformatics/btu668.
525 28. Rosenberg, M. S., S. Subramanian and S. Kumar. "Patterns of transitional mutation biases
526 within and among mammalian genomes." *Mol Biol Evol* 20 (2003): 988–93.
527 10.1093/molbev/msg113.
528 29. Jiang, C. and Z. Zhao. "Mutational spectrum in the recent human genome inferred by single
529 nucleotide polymorphisms." *Genomics* 88 (2006): 527–34.
530 https://doi.org/10.1016/j.ygeno.2006.06.003.
531 https://www.sciencedirect.com/science/article/pii/S0888754306001741.
532 30. Abou Elela, S., M. Catala and M. M. Ghulam. "Differential expression of duplicated
533 ribosomal protein genes modifies ribosome composition in response to stress." *Nucleic Acids*
534 *Research* 48 (2020): 1954–68. 10.1093/nar/gkz1183.
535 31. Helleday, T., S. Eshtad and S. Nik-Zainal. "Mechanisms underlying mutational signatures in
536 human cancers." *Nat Rev Genet* 15 (2014): 585–98. 10.1038/nrg3729.
537 32. Rodnina, M. V. "The ribosome in action: Tuning of translational efficiency and protein
538 folding." In *Protein sci*. 25. United States: 2016, 1390–406.
539 33. Byrne, M. E. "A role for the ribosome in development." In *Trends plant sci*. 14. England:
540 2009, 512–9.
541 34. Liu, C., J. Hu, Y. Wu, D. M. Irwin, W. Chen, Z. Zhang and L. Yu. "Comparative study of gut
542 microbiota from captive and confiscated-rescued wild pangolins." *J Genet Genomics* 48
543 (2021): 825–35. 10.1016/j.jgg.2021.07.009.
544 https://www.ncbi.nlm.nih.gov/pubmed/34474998.
545 35. Bedoya Duque, M. A., W. R. Thomas, D. K. N. Dechmann, J. Nieland, C. Baldoni, D. von
546 Elverfeldt, M. Muturi, A. P. Corthals and L. M. Davalos. "Gene expression comparisons

between captive and wild shrew brains reveal captivity effects." In *Biol lett*. 21. England: 2025, 20240478.

36. Welfer, G. A., R. A. Brady, S. K. Natchiar, Z. L. Watson, E. J. Rundlet, J. L. Alejo, A. P. Singh, N. K. Mishra, R. B. Altman and S. C. Blanchard. "Impacts of ribosomal rna sequence variation on gene expression and phenotype." *Philosophical Transactions of the Royal Society B: Biological Sciences* 380 (2025): 10.1098/rstb.2023.0379.

37. Matsuki, Y., Y. Matsuo, Y. Nakano, S. Iwasaki, H. Yoko, T. Udagawa, S. Li, Y. Saeki, T. Yoshihisa, K. Tanaka*, et al.* "Ribosomal protein s7 ubiquitination during er stress in yeast is associated with selective mrna translation and stress outcome." *Scientific Reports* 10 (2020): 10.1038/s41598-020-76239-3.

38. Ma, J. E., H. Y. Jiang, L. M. Li, X. J. Zhang, G. Y. Li, H. M. Li, X. J. Jin and J. P. Chen. "The fecal metagenomics of malayan pangolins identifies an extensive adaptation to myrmecophagy." *Front Microbiol* 9 (2018): 2793. 10.3389/fmicb.2018.02793. https://www.ncbi.nlm.nih.gov/pubmed/30532742.

39. Cheng, S. C., C. B. Liu, X. Q. Yao, J. Y. Hu, T. T. Yin, B. K. Lim, W. Chen, G. D. Wang, C. L. Zhang, D. M. Irwin*, et al.* "Hologenomic insights into mammalian adaptations to myrmecophagy." In *Natl sci rev*. 10. China: 2023, nwac174.

40. Yan, D., X. Luo, J. Tang, S. Xu, K. Huang, X. Wang, T. Feng, T. Que, M. Jia, X. Guo*, et al.* "High-quality genomes of pangolins: Insights into the molecular basis of scale formation and adaption to myrmecophagous diet." In *Mol biol evol*. 40. United States: 2023,

41. Purohit, D., S. Manu, M. S. Ram, S. Sharma, H. C. Patnaik, P. J. Deka, G. Narayan and G. Umapathy. "Genetic effects of long-term captive breeding on the endangered pygmy hog." *PeerJ* 9 (2021): e12212. 10.7717/peerj.12212.

42. Russello, M. A., J. R. Willoughby, J. A. Ivy, R. C. Lacy, J. M. Doyle and J. A. DeWoody. "Inbreeding and selection shape genomic diversity in captive populations: Implications for the conservation of endangered species." *Plos One* 12 (2017): 10.1371/journal.pone.0175996.

43. Bosse, M. and S. van Loon. "Challenges in quantifying genome erosion for conservation." In *Front genet*. 13. Switzerland: 2022, 960958.

44. Wangmo, L. K., G. Jabin, A. Ghosh, P. Priyambada, S. Dalui, B. D. Joshi, V. K. Singh, S. Dolker, R. K. Mohapatra, M. V. Nair*, et al.* "Genetic assessment of captive breeding program of indian pangolin: Implications for conservation and management." In *Mol biol rep*. 51. Netherlands: 2024, 136.

45. Willoughby, J. R., J. A. Ivy, R. C. Lacy, J. M. Doyle and J. A. DeWoody. "Inbreeding and selection shape genomic diversity in captive populations: Implications for the conservation of endangered species." In *Plos one*. 12. United States: 2017, e0175996.

46. Sendrowski, J., T. Bataillon and G. P. Ramstein. "In silico prediction of variant effects: Promises and limitations for precision plant breeding." *Theor Appl Genet* 138 (2025): 193. 10.1007/s00122-025-04973-1.

47. Lynch, M. "Rate, molecular spectrum, and consequences of human mutation." *Proc Natl Acad Sci U S A* 107 (2010): 961–8. 10.1073/pnas.0912629107.

48. Flanagan, S. P., B. R. Forester, E. K. Latch, S. N. Aitken and S. Hoban. "Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation." *Evolutionary Applications* 11 (2017): 1035–52. 10.1111/eva.12569.

49. Montanari, S., C. Deng, E. Koot, N. V. Bassil, J. D. Zurn, P. Morrison-Whittle, M. L. Worthington, R. Aryal, H. Ashrafi, J. Pradelles*, et al.* "A multiplexed plant–animal snp array for selective breeding and species conservation applications." *G3 Genes|Genomes|Genetics* 13 (2023): 10.1093/g3journal/jkad170. https://doi.org/10.1093/g3journal/jkad170.

Abbreviations

SNPs: Single Nucleotide Polymorphisms; InDels: Insertions/Deletions; PRBC: Pangolin Rescue and Breeding Centre; MAF: Minor Allele Frequency; Ti/Tv: Transition/Transversion; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; CDS: Coding Sequences; NCBI: National Center for Biotechnology Information; BWA: BurrowsWheeler Aligner; ; GATK: Genome Analysis Toolkit;

Supplementary Information

Additional file 1: Table S1. Sunda Pangolin (*M. javanica*) sample information sheet.

Additional file 2: Table S2. SNP detection and annotation results.

Additional file 3: Table S3. InDel detection and annotation results.

Additional file 4: Table S4. Genetic diversity parameters of 12 individuals.

Additional file 5: Table S5. Functional classification statistics of SNPs and InDel in the Coding Region.

Additional file 6: Table S6. Core Genes Co-enriched in Ribosome Pathways.

Additional file 7: Table S7. Core Genes Co-enriched in Reproduction-Related Pathways

Acknowledgments

My profound personal thanks go to my wife and daughter, whose unwavering motivation and selfless support have been my greatest source of strength. I extend my sincere thanks to the members of our wildlife research team for their invaluable collaboration in sample preparation and data analysis. Finally, we acknowledge the anonymous reviewers for their constructive and insightful feedback, which significantly improved the quality of this manuscript.

Author' contributions

X.B.G., W.H.L., and D.Y.Y., initiated and designed this project. X.B.G., H.T.Q., and S.J.L., collected the samples. S.H.X., and M.M.J., analyzed the data. X.B.G., S.H.X., and W.H.L., interpreted the data and drafted the manuscript. Y.P., Y.L., and X.Y.L., critically reviewed the manuscript for important intellectual content. All authors approved the final version to be published and agree to be accountable for all aspects of their work.

Funding

This work was supported by Guangxi Natural Science Foundation under Grant No.2025GXNSFBA069419. The authors gratefully acknowledge the Guangxi Science and Technology Department for the financial support.

Availability of data and materials

The whole-genome sequencing data supporting the findings of this study are openly available in the

NCBI Sequence Read Archive (SRA) at https://dataview.ncbi.nlm.nih.gov/?search=SUB16310216&archive=sra, under BioProject accession [PRJNA1491543] and SRA run accessions SRR39585984–SRR39585995. The data are released to the public as of July 12, 2026.

Each sample corresponds to a unique BioSample accession: [SAMN1463813–SAMN1463824].

Declarations

Ethics approval and consent to participate

This project was approved by the Forestry Department of Guangxi Zhuang Autonomous Region, Guangxi Province (domestication and breeding permit of national key protected wild animals; Permit Number: A2016008). All works and protocols were approved by the Biology Ethics Committee of the Guangxi Forestry Research Institute [Reference Number: GXFI (A2016006)].

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

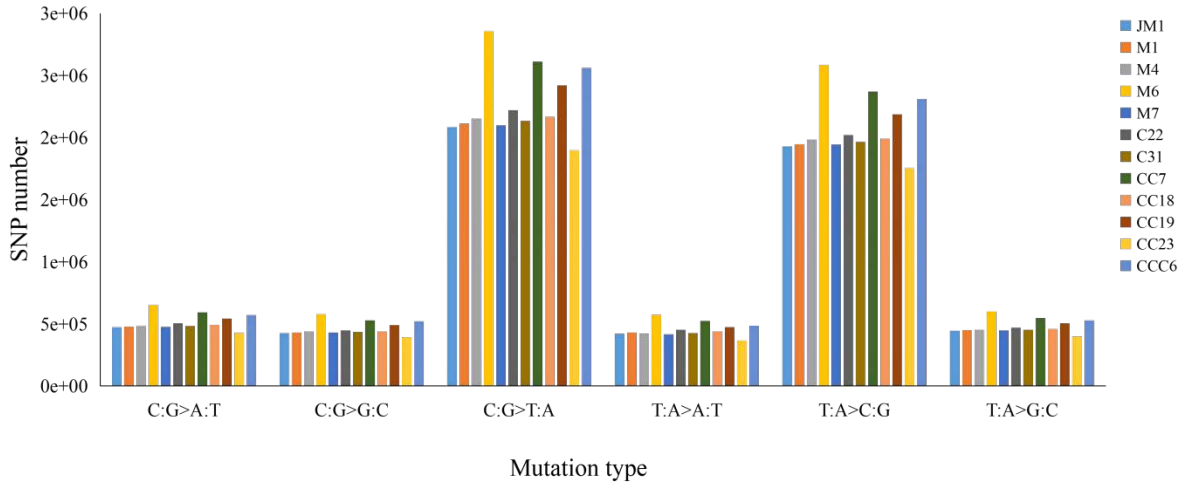


Figure 1. SNP mutation spectrum across 12 captive *Manis javanica* individuals

The six substitution classes are shown on the x-axis, and SNP counts are shown on the y-axis. C:G>T:A and T:A>C:G transitions dominate across all individuals, indicating a conserved mammalian mutational pattern characterized by a strong transition bias.

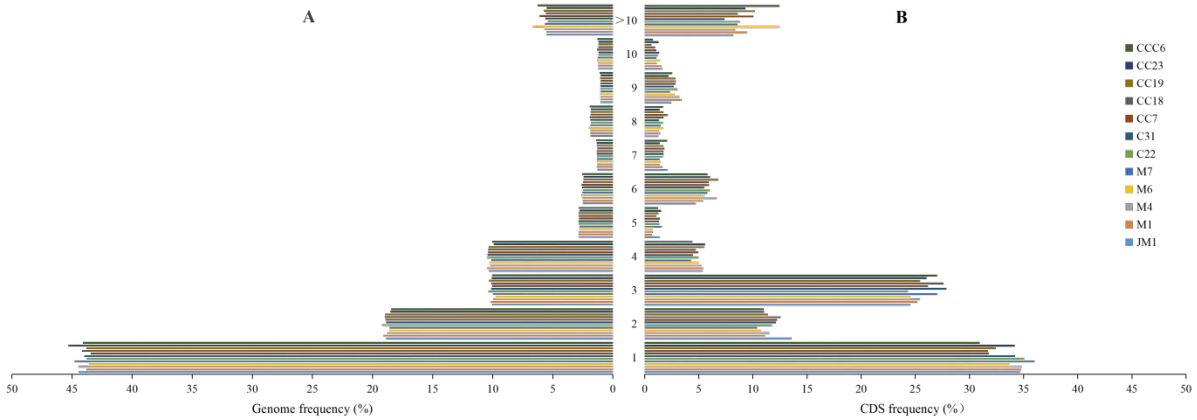


Figure 2. Length distribution of InDels in coding and genome-wide regions

Length distribution of insertion-deletion (InDel) variants across (A) the whole genome and (B) coding sequence (CDS) regions in 12 captive *Manis javanica* individuals.

The bar chart illustrates the frequency distribution of InDels (insertions and deletions) across different length categories. Bars extending to the left represent the percentage of InDels in the genomic regions, while bars extending to the right represent those in the coding sequence (CDS) regions. Each color corresponds to a different biological sample or replicate. The vertical axis denotes the InDel length in base pairs (bp), grouped into intervals. The horizontal axis indicates the percentage frequency. This comparative visualization highlights the

differential abundance of short versus long InDels in non-coding versus coding regions.

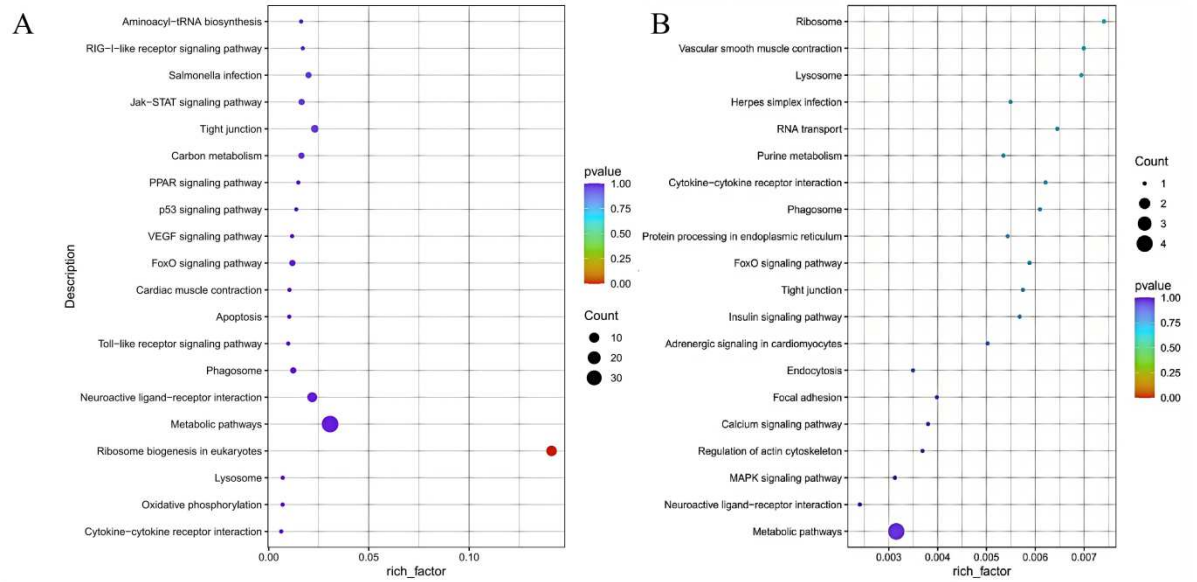


Figure 3. KEGG pathway enrichment of SNP- and InDel-associated genes

KEGG pathway enrichment analysis of genes harboring (A) SNPs and (B) InDels in 12 captive *Manis javanica* individuals. The x-axis represents the rich factor, while point size indicates the number of genes assigned to each pathway and color reflects statistical significance.

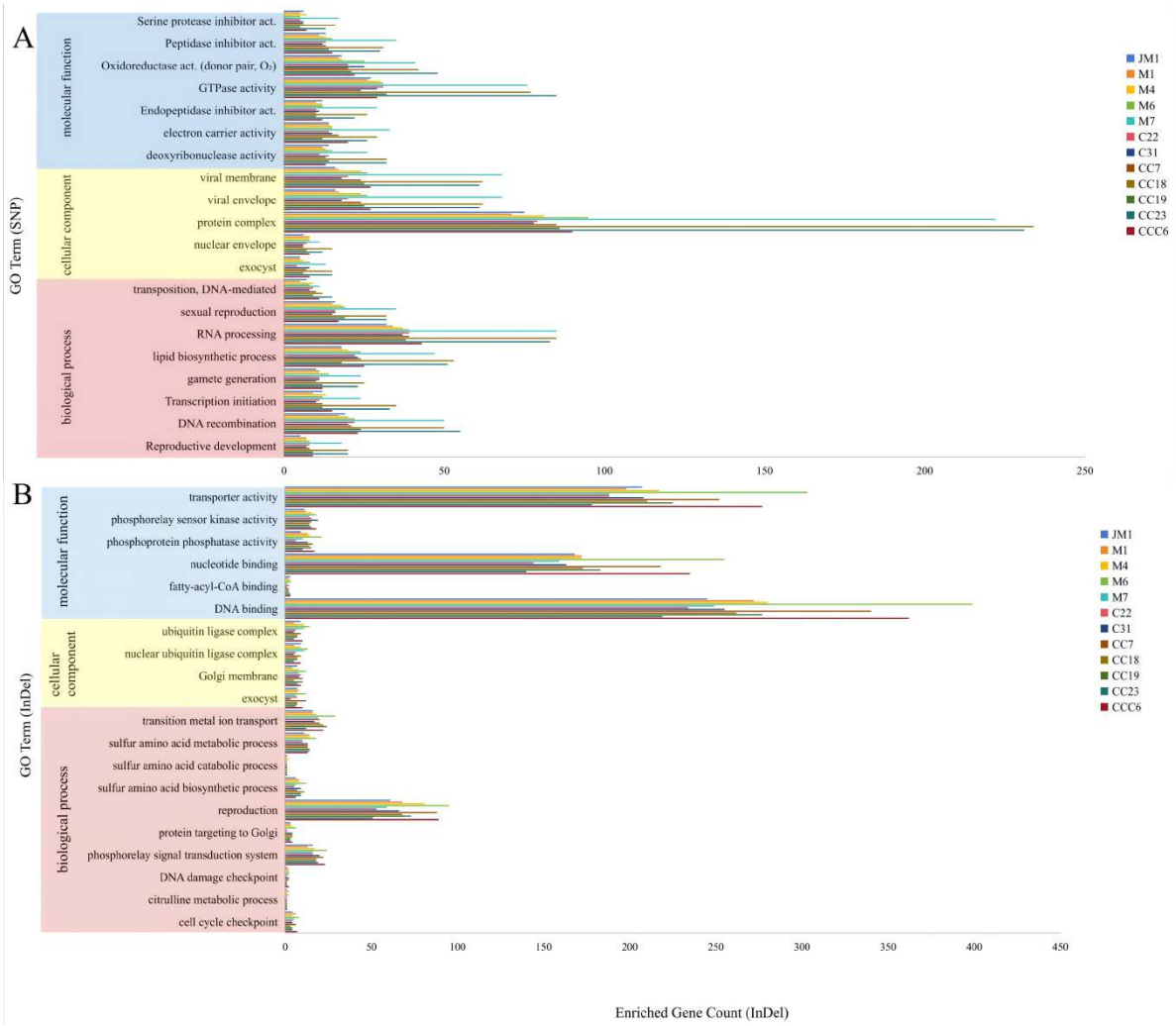


Figure 4. GO enrichment analysis of SNP- and InDel-associated genes

GO biological process enrichment of genes harboring (A) SNPs and (B) InDels in 12 captive *Manis javanica* individuals. The x-axis represents the number of enriched genes, and colors indicate sample contribution across individuals.

Supplementary Files

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

- Tables.xlsx
- TableS1S7.xlsx