Reanalysis of sequences of alleged Javan tiger highlights the difficulties in studying big cats and the need for high-throughput sequencing

ANUBHAB KHAN*[1,2,3] ⓘD, YULIANTO YULIANTO[4]
SABHRINA GITA ANINTA[2] ⓘD and WIRDATETI WIRDATETI[5] ⓘD

Abstract Big cats are of conservation concern throughout their range, and genetic tools are often employed to study them for various purposes. However, there are several difficulties in using genetic tools for big cat conservation that could be resolved by modern methods of DNA sequencing. Recent reports of the sighting of a putative Javan tiger *Panthera tigris sondaica* in West Java, Indonesia, highlight some of the difficulties of studying the genetics of big cats. We reanalysed the data of the original reports and found that the conclusions were drawn based on incorrect copies of the genes. Specifically, the nuclear copy of the mitochondrial gene was analysed with the mitochondrial sequence, leading to discordance in the results. However, re-sequencing of the remaining DNA confirms that the sighting could have been that of a tiger, but the subspecies cannot be confirmed. This work highlights the urgency of developing high-throughput sequencing infrastructure in the tropics and the need for reliable databases for the study of big cats.

Keywords Capacity building, DNA analysis, extinct, Javan tiger, Lazarus species, next-generation sequencing, Sanger sequencing, shed hair

Introduction

Big cats are rare and elusive. Although they attract a significant amount of conservation funding and have been studied across their range, much remains unknown about them. Accurate species identification of big cats is important in research on matters such as poaching and trafficking, livestock depredation, ex situ breeding and dispersal events. Genetic tools are often used to identify species of large felids, but classical tools may be inadequate for this task. A recent study by Wirdateti et al. (2024) that reported the

presence of a Javan tiger *Panthera tigris sondaica* in West Java is an example of the inadequacy of these classical genetic tools. The Javan tiger was categorized as Extinct in 2008 (Jackson & Nowell, 2008), and has not been detected since the 1990s. However, in 2019 a local resident reported seeing a tiger near a village in West Java, and one of the authors of Wirdateti et al. (2024) collected a hair sample from the sighting location.

To determine whether this sighting could be the extinct Javan tiger, museum samples of Javan and Sumatran tigers *Panthera tigris sumatrae* were also collected and DNA was extracted from the hair and the museum samples (Wirdateti et al., 2024). The authors sequenced the cytochrome B (cytB) region of the samples and performed comparative phylogenetic analysis with previously published cytB sequences of tigers and leopards, concluding that the hair belonged to a Javan tiger. However, concerns were raised regarding the study (Emont, 2024; Sui et al., 2024). Here we reanalyse the sequences and repeat some of the experiments, to highlight the difficulties of studying big cat genetics and some potential solutions to these issues.

Phylogenetic tree reconstruction

We reanalysed the Javan tiger sequences in Wirdateti et al. (2024) by making a phylogenetic tree that included additional cytB and nuclear copies of mitochondrial pseudogene (Numt) sequences downloaded from the National Center for Biotechnology Information database (Table 1). We aligned the sequences using MAFFT (Katoh et al., 2002). We conducted three batches of analysis based on the length of sequences analysed and the type of sequences used. For Dataset 1 we removed sequences MH290773, AB211408–AB211411 and FJ403465 from the analysis because of the excess of missing data. We trimmed all regions with any missing data using *Jalview* (Waterhouse et al., 2009). This retained 36 sequences with 265 bp of data for analysis. For Dataset 2 we removed sequences AB211408–AB211411, FJ895266, FJ403466, FJ403467, MH290773 and FJ403465 from the analysis because of the excess of missing data. We trimmed all regions with any missing data using *Jalview*, retaining 33 sequences with 971 bp of data for analysis. For Dataset 3 we aligned the cytB Numt sequences using MAFFT. We trimmed sites with missing data using *Jalview*. This retained 453 sites with eight sequences.

*Corresponding author, anubhabkhan@gmail.com
[1]Centre for Ecological Sciences, Indian Institute of Science, Bangalore, India
[2]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark
[3]Department of Biology, Pwani University, Kilifi, Kenya
[4]Research Center for Applied Zoology, National Research and Innovation Agency (BRIN), Cibinong, Indonesia
[5]Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, Indonesia

TABLE 1 List of sequences (of domestic dog *Canis familiaris*, leopard *Panthera pardus*, lion *Pathera leo* and six subspecies of tiger *Panthera tigris*) from the National Center for Biotechnology Information database used in this analysis.

Accession numbers	Species	Comment
NC_051810.1: 68724255–68726124	*Canis familiaris*	Numt
AB048589.1	*Canis familiaris*	Numt
MF136765.1	*Panthera pardus*	Numt
OQ117410.1	*Panthera pardus*	
MN124244.1	*Panthera leo*	Numt
AB211408.1	*P. tigris altaica*	Large amount of missing data
AB211409.1	*P. tigris altaica*	
AB211410.1	*P. tigris altaica*	
AB211411.1	*P. tigris altaica*	Large amount of missing data
AF053031.1	*P. tigris altaica*	
AF053033.1	*P. tigris altaica*	
AF053034.1	*P. tigris altaica*	
FJ895266.1	*P. tigris altaica*	
MH893763.2:14173–15312	*P. tigris altaica*	
MN624080.1:14530–15669	*P. tigris altaica*	
AF053053.1	*P. tigris tigris*	Numt
AF053019.1	*P. tigris tigris*	
AF053024.1	*P. tigris tigris*	
AF053025.1	*P. tigris tigris*	
JF357967.1:14172–15311	*P. tigris tigris*	
JF357968.1:14172–15311	*P. tigris tigris*	
KF892541.1:14173–15312	*P. tigris tigris*	
AF053050.1	*P. tigris corbetti*	
AF053051.1	*P. tigris corbetti*	
JF357971.1:14077–15216	*P. tigris corbetti*	
JF357972.1:13017–14156	*P. tigris corbetti*	
FJ403465.1	*P. tigris sondaica*	Large amount of missing data
FJ403466.1	*P. tigris sondaica*	
FJ403467.2	*P. tigris sondaica*	
MH290773.1	*P. tigris sondaica*	Large amount of missing data
OQ601561.1	*P. tigris sondaica*	
OQ601562.1	*P. tigris sondaica*	
HM589214.1:15069–16208	*P. tigris amoyensis*	
HM589215.1:15113–16252	*P. tigris amoyensis*	
NC_014770.1:15113–16252	*P. tigris amoyensis*	
AF053054.1	*P. tigris sumatrae*	Numt
AF053045.1	*P. tigris sumatrae*	
AF053046.1	*P. tigris sumatrae*	
AF053047.1	*P. tigris sumatrae*	
AF053048.1	*P. tigris sumatrae*	
JF357970.1:14173–15312	*P. tigris sumatrae*	
OQ629466.1	*P. tigris sumatrae*	
OQ629467.1	*P. tigris sumatrae*	
OQ629468.1	*P. tigris sumatrae*	
OQ629469.1	*P. tigris sumatrae*	
OQ629470.1	*P. tigris sumatrae*	
OQ629471.1	*P. tigris sumatrae*	

We built neighbour joining trees using MAFFT. We chose the default option for multiple sequence alignment. For building the tree we chose the conserved sites option, which retained 252 sites for Dataset 1, 937 sites for Dataset 2 and 240 sites for Dataset 3. We used raw differences for the substitution model and performed bootstrap resampling 1,000 times.

DNA extract re-sequencing

The Numt sequences were amplified because of stochastic binding of the primers to the Numt region instead of the mitochondrial cytB region (Sui et al., 2024) in Wirdateti et al. (2024); we attempted to rectify this through repeated PCR and sequencing (Fig. 1). In Round 1 we amplified the DNA extract using 3, 4, 5 and 7 μl templates in a 30 μl PCR, using the conditions described in Wirdateti et al. (2024). The amplicons from these reactions were not visible in agarose gel electrophoresis. Thus, we further amplified the 3 and 5 μl products from all four PCRs (henceforth referred to as 'amplicon templates') in the previous step using the same PCR conditions in a 30 μl volume (leading to a total of eight reactions in this Round 2). Five of the eight reactions in Round 2 yielded bands on the agarose gel, four of which were from the 5 μl amplicon template and in the PCR in which the 7 μl original template was used. The 3 μl amplicon template in the second PCR also yielded bands. We amplified all templates on different days to avoid cross-contamination.

The reaction with the 5 μl original template in Round 1 and the 5 μl amplicon template in Round 2 of PCR yielded the c. 900 bp-long DNA fragment from the test hair strand sequence that is presented here. We aligned these new sequences to the sequences listed in Table 1 using MAFFT
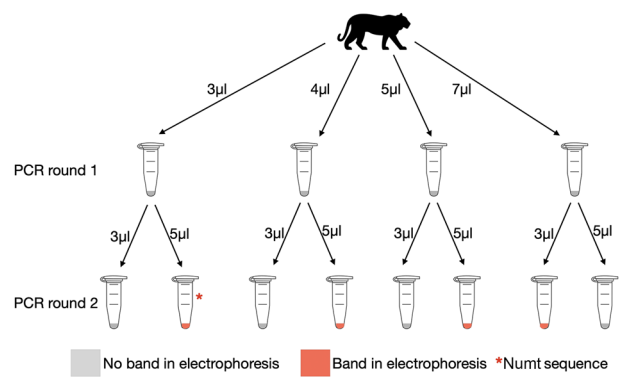

FIG. 1 DNA amplification strategy used in this study. We performed two rounds of PCR. In Round 1 we used DNA extracts from samples as templates, and in Round 2 we used the amplicons from Round 1 as templates. The numbers on the arrows indicate the volumes of the templates used for each step. (Readers of the printed journal are referred to the online article for a colour version of this figure.)
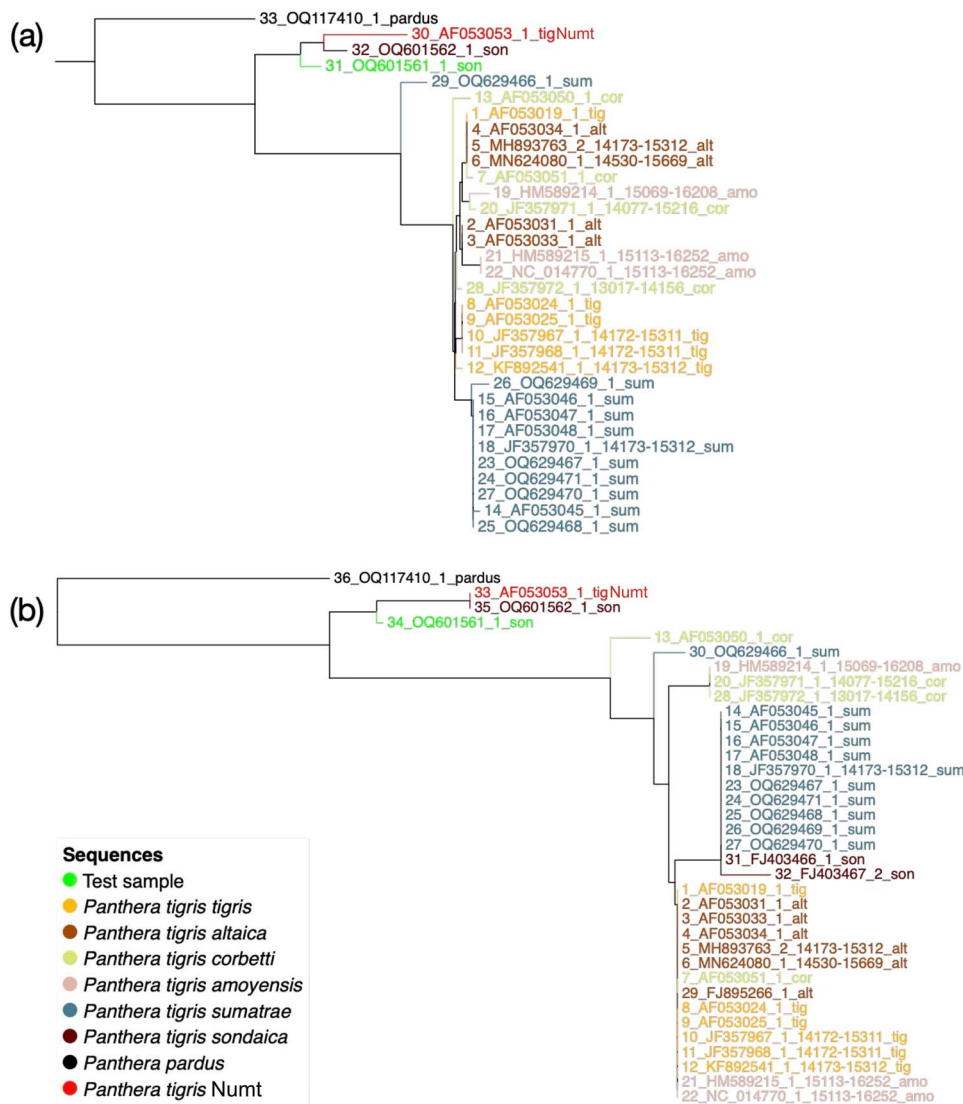
FIG. 2 Neighbour joining tree of the test sample (hair strand of putative Javan tiger *Panthera tigris sondaica*), the Javan tiger museum sample (OQ601562_1_son) and other tiger sequences using (a) 36 sequences and 252 sites and (b) 33 sequences and 937 sites. We included a sequence of the nuclear copy of the cytochrome B pseudogene (Numt) in the analysis. (Readers of the printed journal are referred to the online article for a colour version of this figure.)

and trimmed them using *Jalview*. We conducted two batches of analysis on the basis of this trimming: one that retained 38 sequences and 264 bases and another that retained 35 sequences and 907 bases.

We reanalysed the sequences generated by Wirdateti et al. (2024) alongside a nuclear copy of the cytB pseudogene (Numt) sequence of the Bengal tiger *Panthera tigris tigris* (AF053053.1) and the cytB sequences of several other tiger subspecies. The clustering of the putative Javan tiger sequence (test sample in Fig. 2) and the museum sample of the Javan tiger (OQ601562 in Fig. 2) with the Numt sequence revealed that the sequences generated for the samples were Numts and not the cytB regions that they were being compared to (Fig. 2).

We further added Numt sequences of other *Panthera* species and the domestic dog *Canis familiaris* (as an outgroup) to test the potential of Numts for species identification (Fig. 3). Comparisons of the available annotated cytB Numt sequences of *Panthera* species in the National Center for Biotechnology

Information database demonstrate that Numt sequences can be highly divergent within the *Panthera* genus (Fig. 3). We also observed that the lion *Panthera leo* cytB Numt sequence is more diverged from other cats and from a canid cytB Numt.

We re-sequenced DNA amplicons from the putative Javan tiger hair strand and the museum specimen of Javan tiger, and we retained sequences closely related to the cytB sequences of the other tigers (Fig. 4). The 3 µl template in Round 1 and the 5 µl amplicon template in Round 2 yielded the Numt sequence (Fig. 1). However, the mitochondrial cytB DNA does not have the power to distinguish between Sumatran and Javan tigers (Fig. 4).

Discussion

Mitochondrial genomes are often used to delimit species and subspecies. However, they have several limitations, especially for big cats. In big cats nuclear copies of mitochondrial pseudogenes are a common problem in genetic
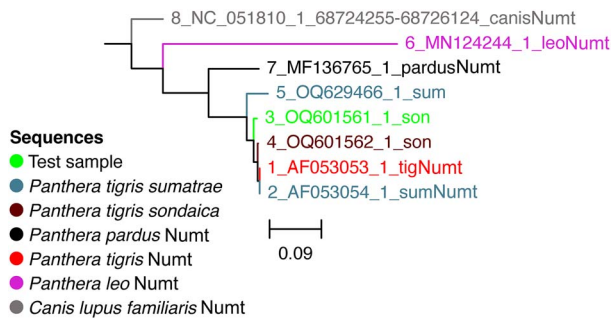
FIG. 3 Neighbour joining tree of the test sample (hair strand of putative Javan tiger) and cytochrome B Numt sequences from other big cats and a domestic dog *Canis familiaris*. (Readers of the printed journal are referred to the online article for a colour version of this figure.)

compared directly. However, as they have sequence similarities, primers intended to amplify the mitochondrial copy of a gene might accidently amplify the nuclear copy and thus provide misleading results (Sui et al., 2024).

Additionally, as mitochondrial DNA is inherited matrilineally, analysis of only mitochondrial sequences limits the ability to detect admixture events. For example, mitochondrial DNA could reveal the subspecies of the mother of a tiger sample but would indicate nothing about the father. This is especially relevant if samples of an admixed tiger are obtained for forensic analysis or captive breeding or for tracing the origins of a sample.

Extracting mitochondrial DNA is low cost, and this DNA is well preserved in non-invasive samples because of its abundance in cells compared to nuclear DNA. As mitochondrial DNA has been analysed experimentally for a long time, the protocols are familiar to most researchers. However, ancestry-informative single nucleotide
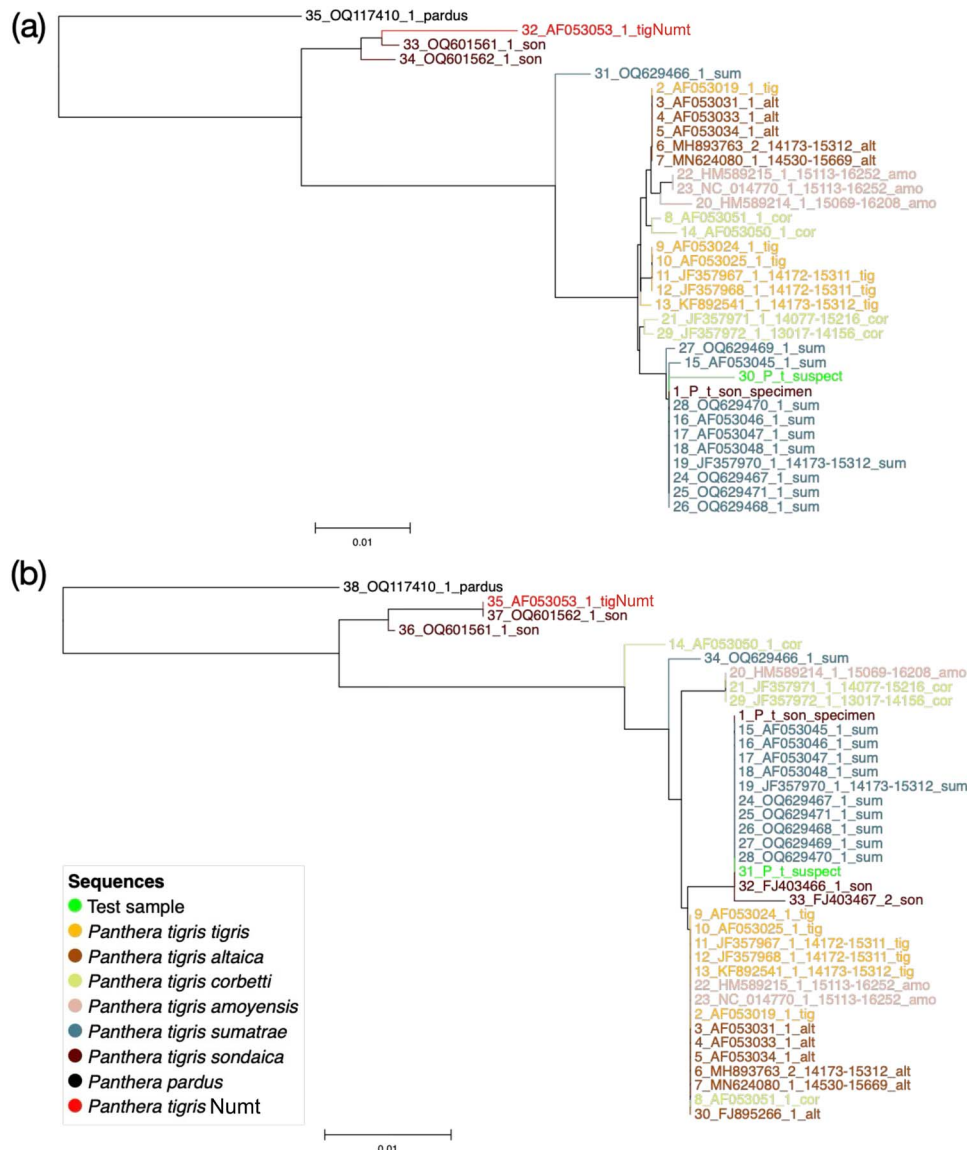
investigations (Kim et al., 2006; Morgan et al., 2021). These regions evolve independently of the mitochondrial genome, have different rates of evolution and should not be

FIG. 4 Neighbour joining tree of the re-sequenced test sample (hair strand of putative Javan tiger, P_t_suspect), a museum sample of a Javan tiger (OQ601562_1_son) and other tiger sequences using (a) 38 sequences and 252 sites and (b) 35 sequences and 869 sites. (Readers of the printed journal are referred to the online article for a colour version of this figure.)

polymorphism (SNP) panels (Khan et al., 2022), low-depth sequencing (Fuentes-Pardo & Ruzzante, 2017), multiplex PCR panels (Natesh et al., 2019) and pooled sequencing (Fuentes-Pardo & Ruzzante, 2017) are low-cost alternatives that overcome the limitations of mitochondrial DNA from non-invasive samples. At present, the need for computational infrastructure and expertise, high start-up costs, high import costs of reagents and lack of access to high-throughput sequencers are major barriers to using these alternatives in many tropical countries (Khan & Tyagi, 2021).

In this study, we confirm that the hair sample collected in West Java nests within the clade of Sundaland tigers, but we are unable to assign it to a subspecies. This is partially because there is no database of extinct tiger lineages. Although Sanger sequencing techniques are commonly used in several lower- and middle-income countries, the lack of databases remains a challenge. There are several specimens of the extinct Javan, Bali and Caspian *Panthera tigris virgata* tigers in museums (Yamaguchi et al., 2013), but genetic resources from these lineages are lacking. This has limited our ability to explore the possibility of shared haplotypes in extant tigers and to determine the similarities and differences between extinct and extant lineages. Wilting et al. (2015), for example, demonstrated the inability of short DNA sequences to resolve tiger subspecies, whereas whole-genome sequencing studies (Liu et al., 2018; Armstrong et al., 2021) have been successful in doing so, at least for extant lineages.

We highlight some of the challenges in studies of big cat genetics (Wirdateti et al., 2024): (1) There are few good-quality samples of big cats and use of non-invasive samples is the norm. Because of the abundance of mitochondrial DNA compared to nuclear DNA in cells, it has been common practice to analyse mitochondrial sequences. However, Numts are common in big cats and can lead to misrepresentation in analyses. (2) Despite the availability of methods for analysing whole genomes from non-invasive samples such as shed hair and faeces (Khan et al., 2020; Tyagi et al., 2022; Khan 2023), the lack of sequencing facilities, computational infrastructure, DNA enrichment reagents or SNP panels render these tools inaccessible in some countries (Khan & Tyagi, 2021). (3) There is a need to develop expertise in next-generation sequencing in biodiversity-rich tropical countries to support the discovery, conservation and management of the species in these regions. (4) Big cats and their parts are often subjects of forensic investigation in cases of trafficking or livestock depredation or in cases such as that of Wirdateti et al. (2024). All such investigations could be affected by Numts if their analyses are only restricted to the mitogenome. There is a need for cost-effective SNP panels to be developed for ancestry determination and population assignment (Khan et al., 2022). (5) Databases such as GenBank, and research journals, need to insist and verify that the sequences they report are archived properly and made available, as the lack of genetic data hampers further studies in this field. For example, despite there being numerous genetic studies of tigers, these sequences are difficult to retrieve or use because of improper annotations or a lack of archiving; e.g. the sequences from Wilting et al. (2015) need better annotation, and the genome assemblies from Armstrong et al. (2021) and Zhang et al. (2023) need to be released, along with those of many other studies. Making these data available would facilitate progress in the science of large felid genetics.

References

ARMSTRONG, E.E., KHAN, A., TAYLOR, R.W., GOUY, A., GREENBAUM, G., THIÉRY, A. et al. (2021) Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. *Molecular Biology and Evolution*, 38, 2366–2379.

EMONT, J. (2024) A late-night sighting, and a single hair, rekindle hopes that an extinct tiger lives on. *The Wall Street Journal*, published 4 April 2024. wsj.com/science/javan-tiger-extinct-dna-2cf84d7d [accessed October 2024].

FUENTES-PARDO, A.P. & RUZZANTE, D.E. (2017) Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. *Molecular Ecology*, 26, 5369–5406.

JACKSON, P. & NOWELL, K. (2008) *Panthera tigris* ssp. *sondaica*. In *The IUCN Red List of Threatened Species* 2008.

KATOH, K., MISAWA, K., KUMA, K.I. & MIYATA, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30, 3059–3066.

KATOH, K., ROZEWICKI, J. & YAMADA, K.D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20, 1160–1166.

KHAN, A. (2023) The year of the tiger and the year of tiger genomes! *Molecular Ecology Resources*, 23, 327–329.

KHAN, A. & TYAGI, A. (2021) Considerations for initiating a wildlife genomics research project in south and South-East Asia. *Journal of the Indian Institute of Science*, 101, 243–256.

KHAN, A., KRISHNA, S.M., RAMAKRISHNAN, U. & DAS, R. (2022) Recapitulating whole genome based population genetic structure for Indian wild tigers through an ancestry informative marker panel. *Heredity*, 128, 88–96.

KHAN, A., PATEL, K., BHATTACHARJEE, S., SHARMA, S., CHUGANI, A.N., SIVARAMAN, K. et al. (2020) Are shed hair genomes the most effective noninvasive resource for estimating relationships in the wild? *Ecology and Evolution*, 10, 4583–4594.

KIM, J.H., ANTUNES, A., LUO, S.J., MENNINGER, J., NASH, W.G., O'BRIEN, S.J. & JOHNSON, W.E. (2006) Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the *Panthera* genus species. *Gene*, 366, 292–302.

LIU, Y.C., SUN, X., DRISCOLL, C., MIQUELLE, D.G., XU, X., MARTELLI, P. et al. (2018) Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Current Biology*, 28, 3840–3849.

MORGAN, K.I., EWART, K.M., NGUYEN, T.Q., SITAM, F.T., OUITAVON, K., LIGHTSON, A.L. et al. (2021) Avoiding common *numts* to provide reliable species identification for tiger parts. *Forensic Science International: Reports*, 3, 100166.

NATESH, M., TAYLOR, R.W., TRUELOVE, N.K., HADLY, E.A., PALUMBI, S.R., PETROV, D.A. & RAMAKRISHNAN, U. (2019) Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, 10, 853–859.

SUI, Z.-Y., YAMAGUCHI, N., LIU, Y.-C., XUE, H.-R., SUN, X., NYHUS, P. & LUO, S.-J. (2024) No reliable evidence supports the presence of the Javan tiger: data issues related to the DNA analysis of a recent hair sample. *Oryx*, published online 6 December 2024.

TYAGI, A., KHAN, A., THATTE, P. & RAMAKRISHNAN, U. (2022) Genome-wide single nucleotide polymorphism (SNP) markers from fecal samples reveal anthropogenic impacts on connectivity: case of a small carnivore in the central Indian landscape. *Animal Conservation*, 25, 648–659.

WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M., CLAMP, M. & BARTON, G.J. (2009) *Jalview* version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189–1191.

WILTING, A., COURTIOL, A., CHRISTIANSEN, P., NIEDBALLA, J., SCHARF, A.K., ORLANDO, L. et al. (2015) Planning tiger recovery: understanding intraspecific variation for effective conservation. *Science Advances*, 1, e1400175.

WIRDATETI, W., YULIANTO, Y., RAKSASEWU, K. & ADRIYANTO, B. (2024) Is the Javan tiger *Panthera tigris sondaica* extant? DNA analysis of a recent hair sample. *Oryx*, 58, 468–473.

YAMAGUCHI, N., DRISCOLL, C.A., WERDELIN, L., ABRAMOV, A.V., CSORBA, G., CUISIN, J. et al. (2013) Locating specimens of extinct tiger (*Panthera tigris*) subspecies: Javan tiger (*P. T. sondaica*), Balinese tiger (*P. T. balica*), and Caspian tiger (*P. T. virgata*), including previously unpublished specimens. *Mammal Study*, 38, 187–198.

ZHANG, L., LAN, T., LIN, C., FU, W., YUAN, Y., LIN, K. et al. (2023) Chromosome-scale genomes reveal genomic consequences of inbreeding in the south China tiger: a comparative study with the Amur tiger. *Molecular Ecology Resources*, 23, 330–347.