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

Anubhab Khan[1,2,3,6], Yulianto Yulianto[4], Sabhrina Gita Aninta[1], Wirdateti Wirdateti[5]

[1]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Denmark
[2]School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, UK
[3]Department of Biology, Pwani University, Kenya
[4]Research Center for Applied Zoology, National Research and Innovation Agency (BRIN), Cibinong, West Java, Indonesia
[5]Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, West Java, Indonesia
[6]anubhabkhan@gmail.com

Abstract

Big cats are of conservation concern throughout their range. Genetic tools are often employed to study them for various purposes. However, there are several difficulties in using genetic tools for big cat conservation which may be resolved by modern methods of DNA sequencing. Recent reports of discovery of Javan tigers in West Java, Indonesia highlights this. We reanalysed the data of the original reports and find that the results were unreliable. However, resequencing of the DNA extracts confirm that the sighting could have been that of a tiger, but the subspecies cannot be confirmed. The work highlights the urgency for development of high throughput sequencing infrastructure in the tropics and the need for reliable databases for studies of big cats.

Introduction

Big cats are rare and elusive. While they attract a huge amount conservation funding and have been studied across their range by various researchers, they still remain a mystery. Poaching, livestock depredation, trafficking, *ex-situ* breeding and dispersal events are a few examples where identifying big cats properly is of prime importance. Genetics tools are often used for identifying big cats but the classical tools maybe inadequate. A recent study by Wirdateti et al. (2024) where presence of Javan tigers have been reported from West Java is one such example. Javan tigers were classified as extinct by IUCN in 2008 (Jackson and Nowell 2008) and their presence has not been detected since 1990s. However, in 2019 villagers in West Java had seen an alleged tiger and collected a hair. Wirdateti et al (2024) collected samples of Javan and Sumatran tigers from museum and extracted DNA from the hair strand and the museum samples. They sequenced Cytochrome B (cytB) region and performed comparative phylogenetic analysis with other previously published cytB sequences of tigers and leopards to conclude that the hair belongs to a Javan tiger. Here, we have reanalysed those sequences and highlight the difficulties of studying big cat genetics in the wild and some potential solutions.

Results

We reanalysed the sequences generated by Wirdateti et al (2024) alongside a nuclear copy of cytB pseudogene (NuMt) sequence of Bengal tiger (*Panthera tigris tigris*, AF053053.1) and the cytB sequences of several other tiger subspecies. The clustering of the putative Javan tiger sequence with the NuMt sequence revealed that the sequences generated for the samples were NuMts and not cytB region that they were being compared to (Figure1).

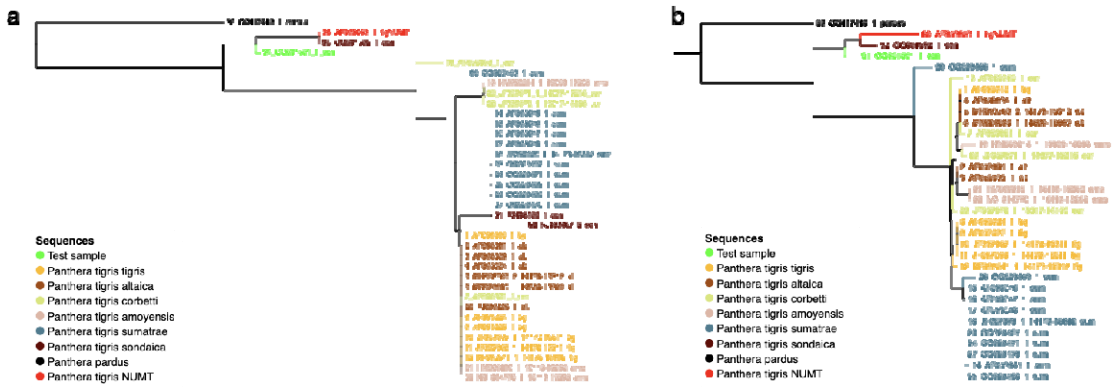

Figure1: Neighbour joining tree of the test sample (green) and other tiger sequences using (a) 36 sequences and 252 sites and (b) 33 sequences and 937 sites. A sequence of nuclear copy of cytB pseudogene (NuMt, red) was included in the analysis.

We further added NuMt sequences of other Pantherine cats and dog as an outgroup to see the possibility of NuMts as species identifier (Figure2). Comparisons of the available annotated cytB NuMt sequences of Pantherine cats on NCBI demonstrates that NuMt sequences can be highly divergent within the Panthera lineage (Figure 2) with *Panthera leo* NuMt sequence being more diverged from other cats and also from a Canidae cytB NuMt..

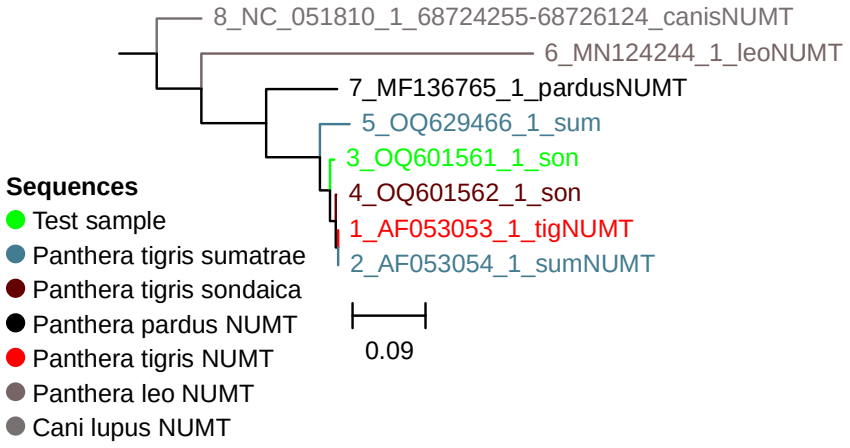

Figure2: Neighbour joining tree of the test sample and cytB NuMt sequences from other big cats and a dog.

We re-sequenced DNA extract remains from the putative Javan tiger hair strands and the specimen of Javan tiger form the museum, and retained sequences closely related to the cytB sequences of the other tigers (Figure 3). However, the polytomy that was made

alongside other Sumatran tiger showed that mitochondrial cytB DNA does not have the power to distinguish between Sumatran tigers and Javan tigers (Figure 3).

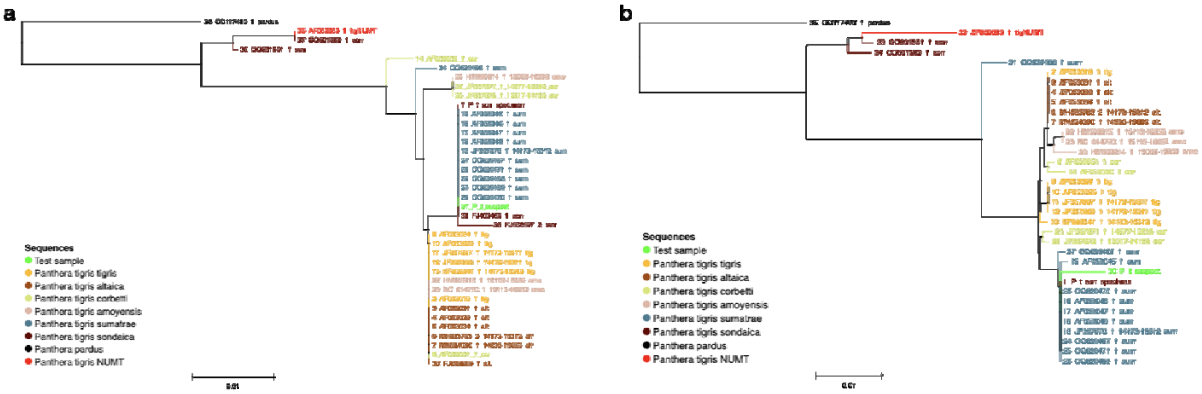

Figure3: Neighbour joining tree of the re-sequenced test sample and other tiger sequences using (a) 38 sequences and 252 sites and (b) 35 sequences and 869 sites.

Discussions

Mitochondrial genomes are used commonly for delimiting species and subspecies. However, they have several limitations especially for big cats. In big cats, there were nuclear copies of mitochondrial pseudogene have been a common problem for genetic investigations (Kim et al. 2006, Morgan et al. 2021). These regions, often called NuMts, evolve independently of the mitochondrial genome, have different rate of evolution, and should not be compared directly. However, since they have sequence similarities, primers intended to amplify the mitochondrial copy of a gene, may accidently amplify the nuclear copy and provide misleading results.

Additionally, since mitochondrial DNA is inherited matrilineally, analysis of only mitochondrial sequences limits detection of admixture events. For example, mitochondrial DNA would reveal the subspecies of the mother of a tiger sample but says 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.

Mitochondrial DNA is cheap to obtain and is better preserved in non-invasive samples due to its abundance in the cells compared to nuclear DNA. Since, mitochondrial DNA has been used for a long time, the protocols are more familiar to most researchers. However, ancestry informative SNP panels (Khan et al. 2022), low depth sequencing (Fuentes-Pardo and Ruzzante 2017) and multiplex PCR panels (Natesh et al. 2019) and pooled sequencing (Fuentes-Pardo and Ruzzante 2017) are cheap alternatives that overcome the limitations of mitochondrial DNA from non-invasive samples. Presently, the need for computational infrastructure and expertise, high start-up costs, high import cost of reagents and access to high-throughput sequencers are a major hinderance in the tropical countries.

Though we can confirm that the hair sample observed in West Java is a tiger we are unable to assign it a subspecies. We may state it belongs to the Sundaland tigers which is reassuring. This inability is partially because of a lack of database of what we think are extinct tiger lineages. There are several specimens of Java, Bali, and Caspian tigers in the

museum (Yamaguchi et al. 2013), but they are understudied their genetics are understudied and hence genetic resources for their study 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 demonstrate the inability of short DNA sequences to resolve the tiger subspecies while whole genome sequencing studies (Liu et al. 2018, Armstrong et al. 2020) have been successful in doing so.

We would like to highlight few important challenges in big cat genetics revealed by the study in Wirdateti et al. (2024):

A) Good quality samples of big cats are difficult to come by and non-invasive samples are the norm. Due to abundance of mitochondrial DNA compared to nuclear DNA in cells, it has been common practise to analyse mitochondrial sequences. However, NuMts are common in big cats leading to misrepresentation in analysis.
B) Despite there being methods for analysing whole genomes from non-invasive samples like shed hair and faeces (Khan et al. 2020, Tyagi et al. 2022) lack of resources in terms of sequencing facilities, computational infrastructure, DNA enrichment reagents or SNP panel makes it an inaccessible tool.
C) There is an urgent need to develop expertise in next-generation sequencing in biodiversity rich tropical countries otherwise we will miss out on discovering species. This is especially important now in the on-going mass extinction of the Anthropocene.
D) Big cats and their parts are often the subject of forensic investigation in cases of trafficking, livestock depredation or like in the case of Wirdateti et al (2024). All such investigations could be affected by NuMts if analysis is only restricted to the mitogenome. There is an urgent need for cost effective SNP panels for ancestry determination and population assignment (for example Khan et al. 2022)
E) Databases like GenBank and the journals need to insist and verify that the sequences reported are archived properly and made available. This would make it easier for the peers to build on the science for big cats. For example, despite their being so many genetic studies on tigers, the sequences are difficult to retrieve or use because of improper annotations or not archiving (for example sequences from Wilting et al. (2015) need to better annotations and genome assemblies from Armstrong et al (2021) and Zhang et al. (2023) need to be released and many more).

Methods

Phylogenetic tree reconstruction. We reanalysed the Javan tiger sequences in Wirdateti et al. (2024) by making a phylogenetic tree including additional cytB and NuMt sequences downloaded from NCBI (see Supplementary Table 1). The sequences were aligned using MAFFT (Katoh et al. 2002). Three batches of analysis were done based on the length of sequences analyzed and the type of sequences used: (1) sequences MH290773, AB211408 to AB211411, and FJ403465 were removed from analysis due to excess of missing data. All regions with any missing data were trimmed using Jalview (Waterhouse et al. 2009). This retained 36 sequences with 265 bp data for analysis. (2) Sequences AB211408 to AB211411, FJ895266, FJ403466, FJ403467, MH290773, and FJ403465 were removed from analysis due to excess of missing data and all regions with any missing data were trimmed using Jalview

(Waterhouse et al. 2009). This retained 33 sequences with 971 bp data for analysis. (3) CytB NuMt sequences were aligned using MAFFT and sites with missing data were trimmed using Jalview retaining 453 sites with 8 sequences.

Neighbour joining tree were built using https://mafft.cbrc.jp. Default option for chosen for multiple sequence alignment. For building the tree, the option of *Conserved sites* was chosen which retained 252 sites for dataset (1), 937 sites for dataset (2), and 240 sites for dataset (3). Raw differences were used for *Substitution model* and bootstrap resampling was performed 1000 times.

DNA extract re-sequencing. The NuMt sequences got amplified due to stochastic binding of the primers to the NuMt region instead of mitochondrial cytB in Wirdateti et al. (2024); we expect this can be rectified by repeated PCR and sequencing. Some amount of the extracted DNA from the hair strand and the Javan tiger specimen from Museum Zoologicum Bogoriense (Wirdateti et al. 2024) remained in the tubes. We re-amplified and re-sequenced these samples several times a described in Wirdateti et al. (2024). One of the sequences batches yielded an approximately 900bp long DNA fragment from the test hair strand and the museum sample. These new sequences were aligned to the sequences listed in Supplementary table 1 using MAFFT and trimmed using Jalview. Two batches of analysis were done due to the trimming. One that retained 38 sequences and 264 bases and the other retained 35 sequences and 907 sites.

References

Armstrong, E.E., Khan, A., Taylor, R.W., Gouy, A., Greenbaum, G., Thiéry, A., Kang, J.T., Redondo, S.A., Prost, S., Barsh, G. and Kaelin, C., 2021. Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. *Molecular Biology and Evolution*, *38*(6), pp.2366-2379.

Fuentes-Pardo, A.P. and Ruzzante, D.E., 2017. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular ecology*, *26*(20), pp.5369-5406.

Jackson, P. & Nowell, K. 2008. Panthera tigris ssp. sondaica. The IUCN Red List of Threatened Species 2008: e.T41681A10509194. https://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T41681A10509194.en.

Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, *30*(14), pp.3059-3066.

Khan, A., Krishna, S.M., Ramakrishnan, U. and Das, R., 2022. Recapitulating whole genome based population genetic structure for Indian wild tigers through an ancestry informative marker panel. *Heredity*, *128*(2), pp.88-96.

Khan, A., Patel, K., Bhattacharjee, S., Sharma, S., Chugani, A.N., Sivaraman, K., Hosawad, V., Sahu, Y.K., Reddy, G.V. and Ramakrishnan, U., 2020. Are shed hair genomes the most effective noninvasive resource for estimating relationships in the wild?. *Ecology and Evolution*, *10*(11), pp.4583-4594.

Kim, J.H., Antunes, A., Luo, S.J., Menninger, J., Nash, W.G., O'Brien, S.J. and Johnson, W.E., 2006. Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. *Gene*, *366*(2), pp.292-302.

Liu, Y.C., Sun, X., Driscoll, C., Miquelle, D.G., Xu, X., Martelli, P., Uphyrkina, O., Smith, J.L., O'Brien, S.J. and Luo, S.J., 2018. Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Current Biology*, *28*(23), pp.3840-3849.

Morgan, K.I., Ewart, K.M., Nguyen, T.Q., Sitam, F.T., Ouitavon, K., Lightson, A.L., Kotze, A. and McEwing, R., 2021. Avoiding common numts to provide reliable species identification for tiger parts. *Forensic Science International: Reports*, *3*, p.100166.

Natesh, M., Taylor, R.W., Truelove, N.K., Hadly, E.A., Palumbi, S.R., Petrov, D.A. and Ramakrishnan, U., 2019. Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, *10*(6), pp.853-859.

Tyagi, A., Khan, A., Thatte, P. and 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*(5), pp.648-659.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), pp.1189-1191.

Wilting, A., Courtiol, A., Christiansen, P., Niedballa, J., Scharf, A.K., Orlando, L., Balkenhol, N., Hofer, H., Kramer-Schadt, S., Fickel, J. and Kitchener, A.C., 2015. Planning tiger recovery: understanding intraspecific variation for effective conservation. *Science advances*, *1*(5), p.e1400175.

Wirdateti, W., Yulianto, Y., Raksasewu, K. and Adriyanto, B., 2024. Is the Javan tiger Panthera tigris sondaica extant? DNA analysis of a recent hair sample. *Oryx*, pp.1-6. doi:10.1017/S0030605323001400

Yamaguchi, N., Driscoll, C.A., Werdelin, L., Abramov, A.V., Csorba, G., Cuisin, J., Fernholm, B., Hiermeier, M., Hills, D., Hunter, L. and Itakura, H., 2013. Locating specimens of extinct tiger (Panthera tigris) subspecies: Javan tiger (PT sondaica), Balinese tiger (PT balica), and Caspian tiger (PT virgata), including previously unpublished specimens. *Mammal Study*, *38*(3), pp.187-198.

Zhang, L., Lan, T., Lin, C., Fu, W., Yuan, Y., Lin, K., Li, H., Sahu, S.K., Liu, Z., Chen, D. and Liu, Q., 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*(2), pp.330-347.