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 few of the difficulties in big cat genetics. We reanalysed the data of the original reports and found 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 of conservation funding and have been studied across their range by various researchers, they remain a mystery. Poaching, livestock depredation, trafficking, *ex-situ* breeding and dispersal events are a few examples of the importance of identifying big cats properly. Genetics tools are often used for identifying big cats but the classical tools may be inadequate. A recent study by Wirdateti et al. (2024) where the presence of a Javan tiger has 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 the 1990s. However, in 2019, a local resident had seen an alleged tiger near a village in West Java and one of the authors of Wirdateti et al. (2024) collected a hair (Wirdateti et al. 2024).

To determine the possibility of finding the extinct javan tiger, museum samples of Javan and Sumatran tigers were also collected, and DNA was extracted from the hair strand and the museum samples (Wirdateti et al. 2024). 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 repeated a few experiments to highlight the difficulties of studying big cat genetics in the wild and some potential solutions.

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 analysed 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

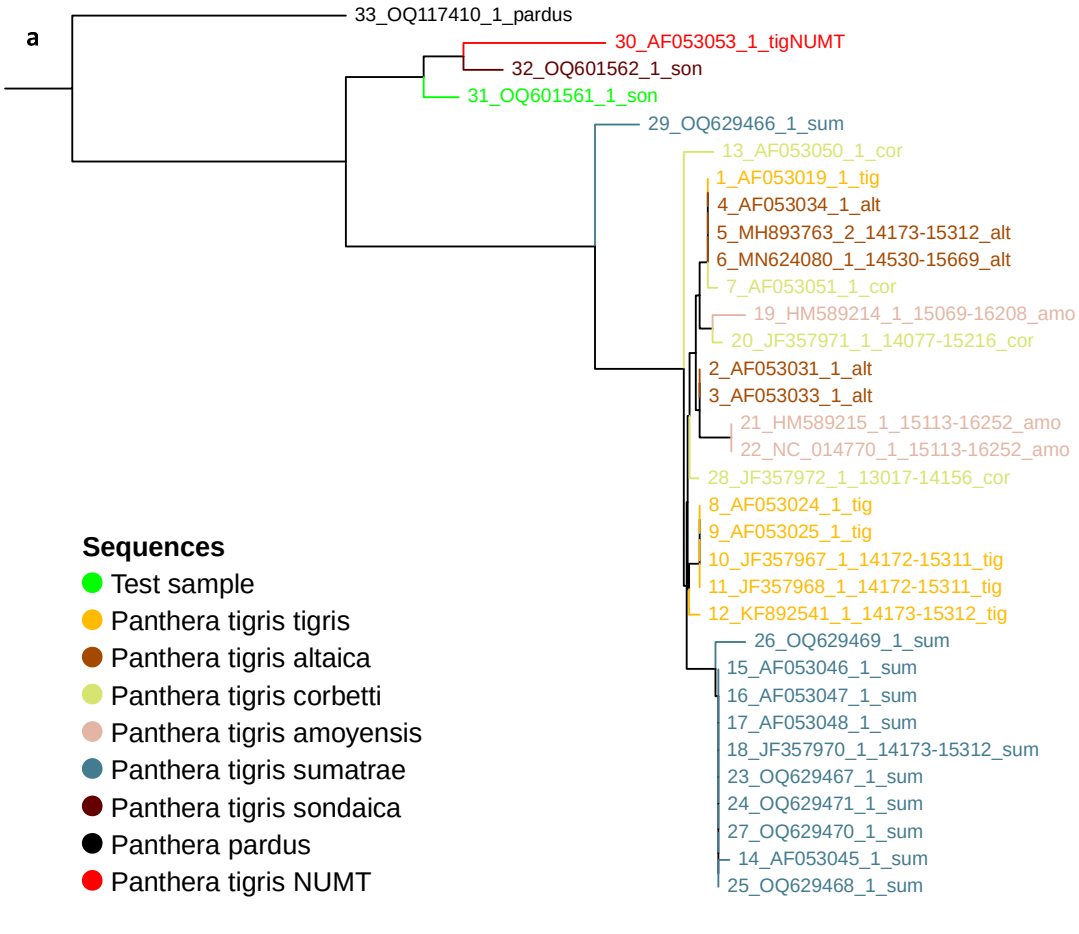
to excess of missing data and all regions with any missing data were trimmed using Jalview (Waterhouse et al. 2009). This retained 33 sequences with 971 bp data for analysis. (3) cytB NuMt sequences were aligned using MAFFT and sites with missing data were trimmed using Jalview retaining 453 sites with 8 sequences.

Neighbour joining trees were built using MAFFT online service (Katoh et al, 2017). Default option was 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.

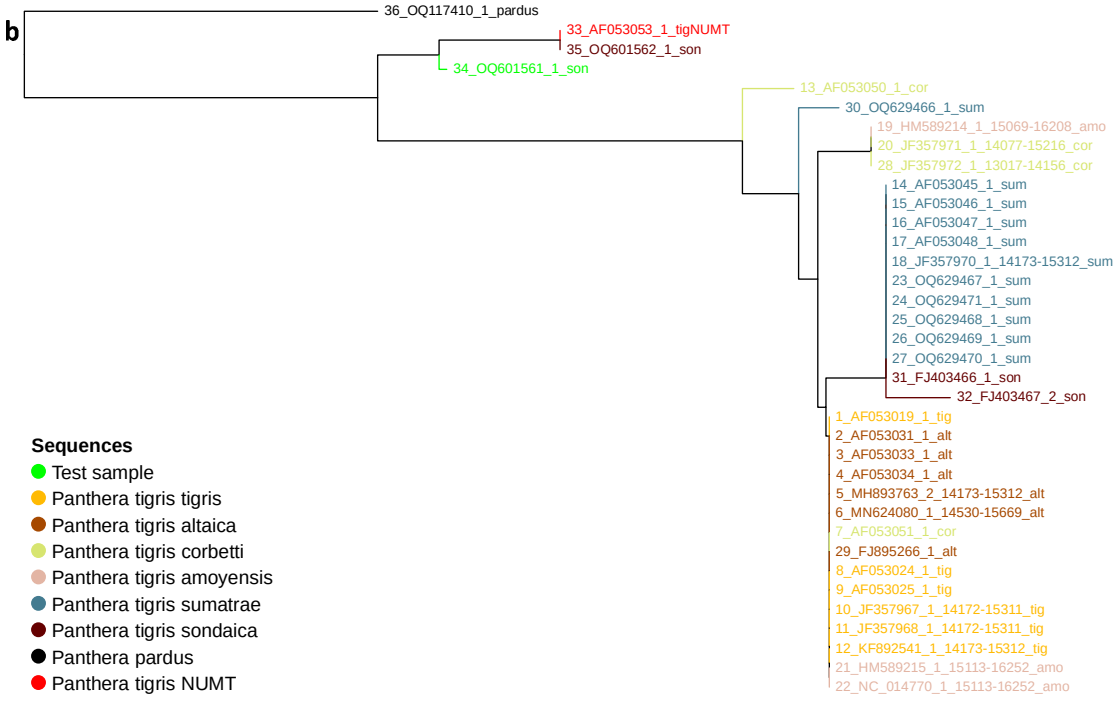
***DNA extract re-sequencing*.** The NuMt sequences got amplified due to stochastic binding of the primers to the NuMt region instead of mitochondrial cytB in Wirdateti et al. (2024); we expect this can be rectified by repeated PCR and sequencing. Some amount of the extracted DNA from the hair strand 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 as described in Wirdateti et al. (2024). One of the sequencing batches yielded an approximately 900 bp 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.

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 ("Test sample" in Figure 1) and the museum sample of Javan tiger ("OQ601562" Figure 1) with the NuMt sequence revealed that the sequences generated for the samples were NuMts and not cytB region that they were being compared to (Figure 1).



83



84

Figure1: Neighbour joining tree of the test sample (hair strand of putative Javan tiger, green), the javan tiger museum sample (OQ601562_1_son, brown) 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 (Figure 2). 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). It is also observed that the *Panthera leo* cytB NuMt sequence seems more diverged from other cats and from a canid cytB NuMt.

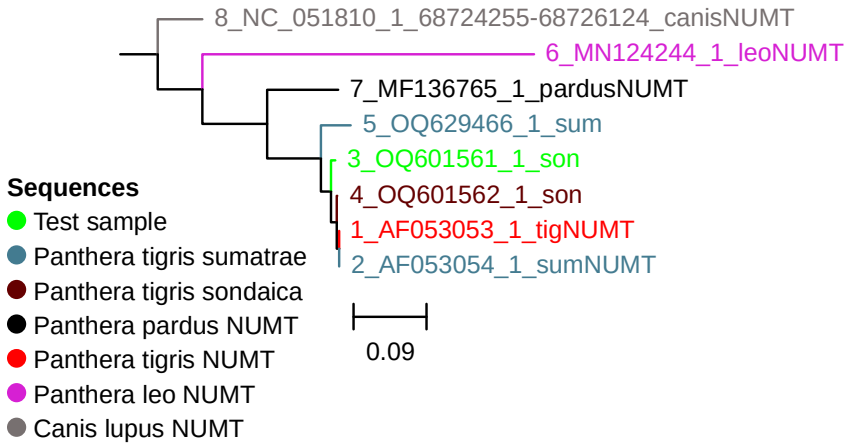


Figure2: Neighbour joining tree of the test sample (hair strand of putative Javan tiger, green) and cytB NuMt sequences from other big cats and a dog.

We re-sequenced DNA extract remains from the putative Javan tiger hair strand 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 tigers revealed that mitochondrial cytB DNA does not have the power to distinguish between Sumatran tigers and Javan tigers (Figure 3).

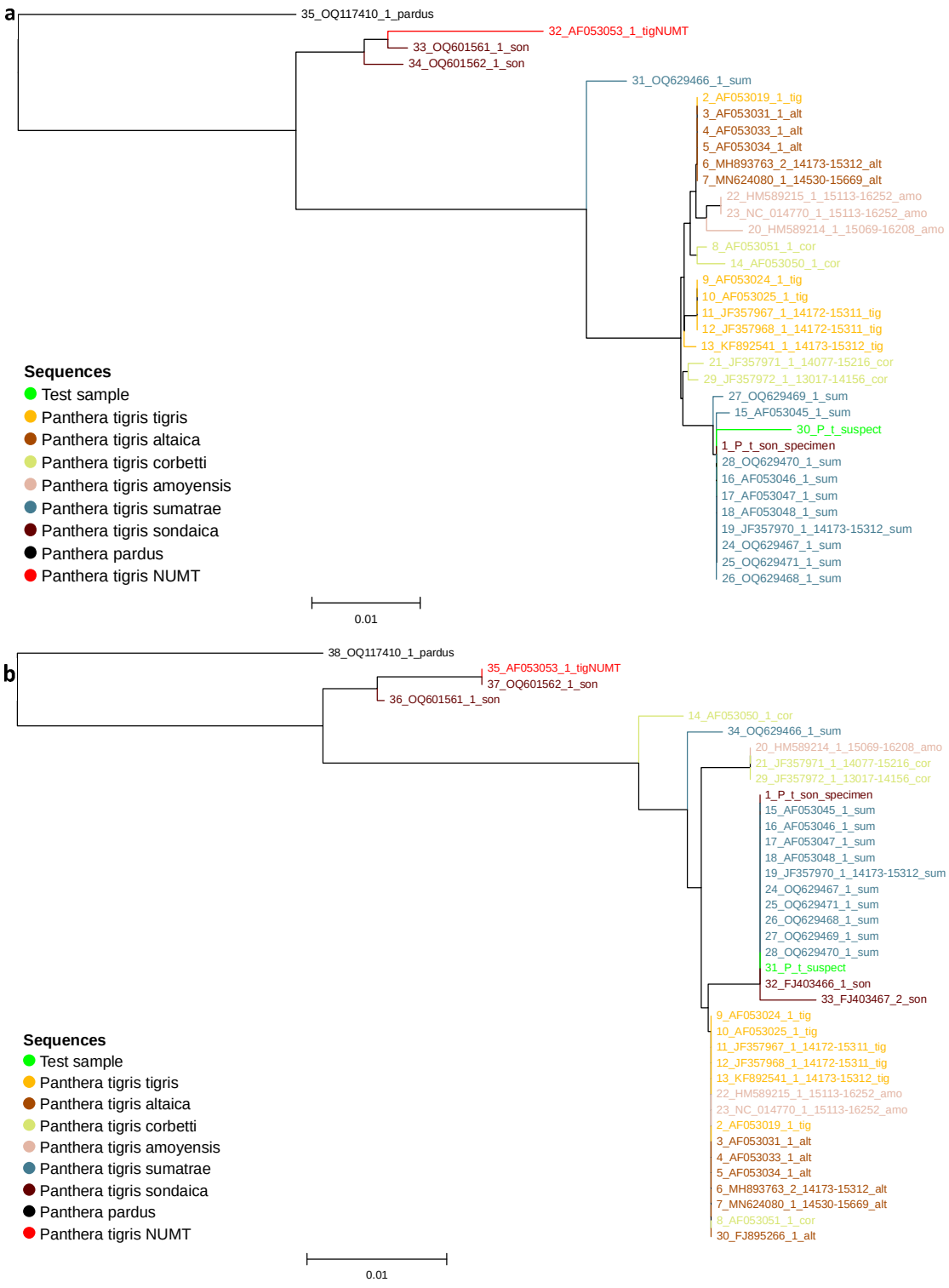


Figure3: Neighbour joining tree of the re-sequenced test sample (hair strand of putative javan sample, P_t_suspect, green), museum sample of javan tiger (OQ601562_1_son,

brown) 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, 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 independent of the mitochondrial genome, have different rates 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, and 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), 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 hindrance in the tropical countries (Khan and Tyagi 2021).

Presently, we can confirm that the hair sample observed in West Java nests within the clade of Sundaland tigers but we are unable to assign it to a subspecies. This is partially because of a lack of database of extinct tiger lineages. There are several specimens of Java, Bali, and

136 Caspian tigers in the museum (Yamaguchi et al. 2013), but they are understudied and hence
137 genetic resources from these lineages are lacking. This has limited our ability to explore the
138 possibility of shared haplotypes in extant tigers and to determine the similarities and
139 differences between extinct and extant lineages. Wilting et al. (2015) for example
140 demonstrate the inability of short DNA sequences to resolve the tiger subspecies while
141 whole genome sequencing studies (Liu et al. 2018, Armstrong et al. 2020) have been
142 successful in doing so at least for the extant lineages.

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

145 A) Good quality samples of big cats are difficult to come by and non-invasive samples
146 are the norm. Due to the abundance of mitochondrial DNA compared to nuclear DNA
147 in cells, it has been common practice to analyse mitochondrial sequences. However,
148 NuMts are common in big cats and can lead to misrepresentation in analysis.
149 B) Despite there being methods for analysing whole genomes from non-invasive
150 samples like shed hair and faeces (Khan et al. 2020, Tyagi et al. 2022, Khan 2023) lack
151 of resources in terms of sequencing facilities, computational infrastructure, DNA
152 enrichment reagents or SNP panel makes it an inaccessible tool (Khan and Tyagi
153 2021).
154 C) There is an urgent need to develop expertise in next-generation sequencing in
155 biodiversity rich tropical countries otherwise we will miss out on discovering species.
156 This is especially important now in the on-going mass extinction of the
157 Anthropocene.
158 D) Big cats and their parts are often the subject of forensic investigation in cases of
159 trafficking, livestock depredation or like in the case of Wirdateti et al (2024). All such
160 investigations could be affected by NuMts if analysis is only restricted to the
161 mitogenome. There is an urgent need for cost effective SNP panels for ancestry
162 determination and population assignment (for example Khan et al. 2022)
163 E) Databases like GenBank and the journals need to insist and verify that the sequences
164 reported are archived properly and made available. This would make it easier for the
165 peers to build on the science for big cats. For example, despite their being so many
166 genetic studies on tigers, the sequences are difficult to retrieve or use because of

167 improper annotations or not archiving (for example sequences from Wilting et al.
168 (2015) need to better annotations and genome assemblies from Armstrong et al
169 (2021) and Zhang et al. (2023) need to be released and many more).

170

171 **References**

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

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

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

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

185 Katoh, K., Rozewicki, J., and Yamada, K.D., 2019. MAFFT online service: multiple sequence
186 alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4),
187 pp.1160–1166.

188 Khan, A., 2023. The year of the tiger and the year of tiger genomes!. *Molecular Ecology*
189 *Resources*, *23*(2), pp.327-329.

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

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

197 Khan, A. and Tyagi, A., 2021. Considerations for initiating a wildlife genomics research
198 project in south and south-east Asia. *Journal of the Indian Institute of Science*, *101*(2),
199 pp.243-256.

200 Kim, J.H., Antunes, A., Luo, S.J., Menninger, J., Nash, W.G., O'Brien, S.J. and Johnson, W.E.,
201 2006. Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome
202 of the Panthera genus species. *Gene*, *366*(2), pp.292-302.

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

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

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

212 Tyagi, A., Khan, A., Thatte, P. and Ramakrishnan, U., 2022. Genome-wide single nucleotide
213 polymorphism (SNP) markers from fecal samples reveal anthropogenic impacts on
214 connectivity: Case of a small carnivore in the central Indian landscape. *Animal*
215 *Conservation*, *25*(5), pp.648-659.

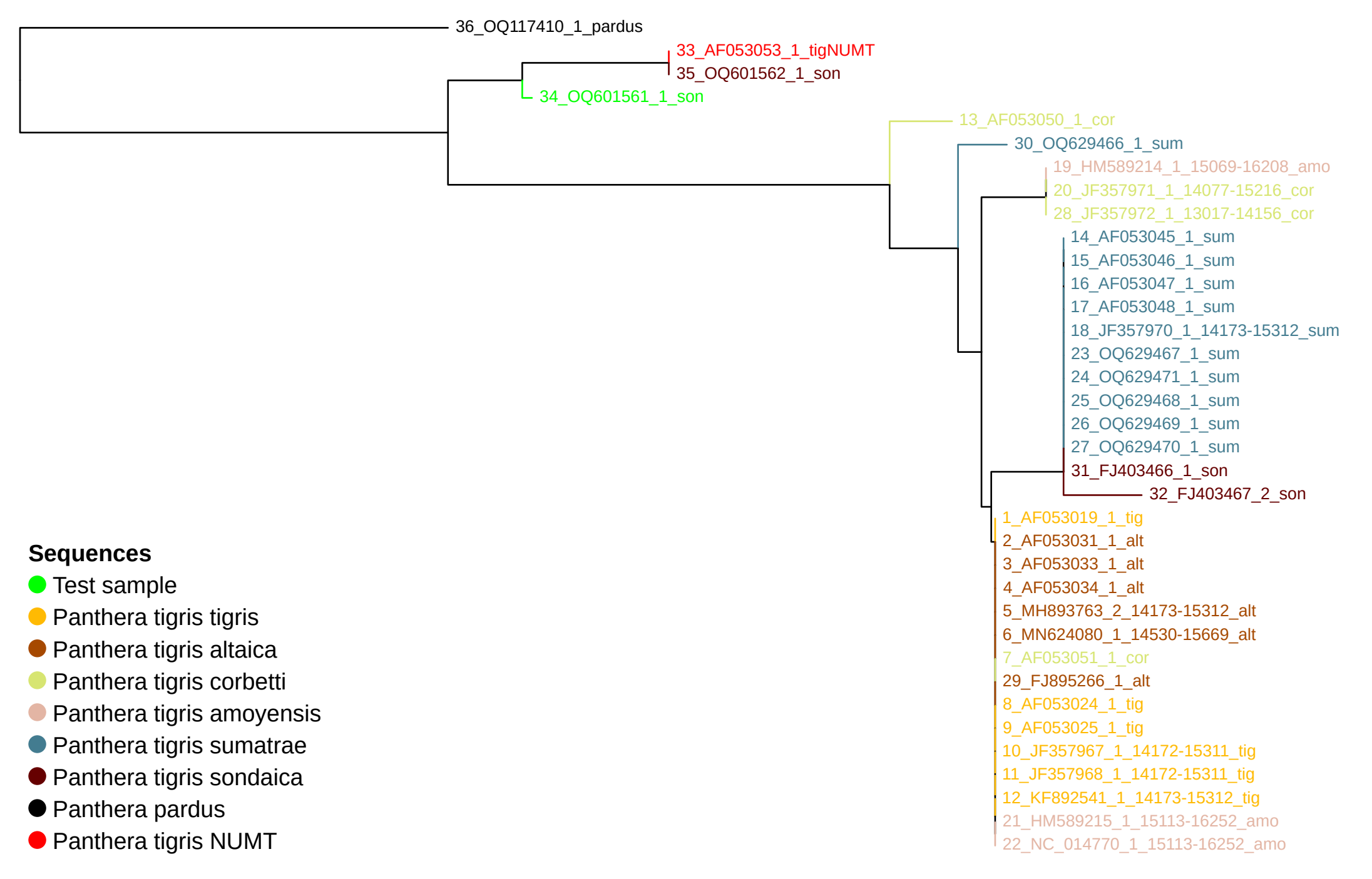
216 Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J., 2009. Jalview
217 Version 2—a multiple sequence alignment editor and analysis
218 workbench. *Bioinformatics*, *25*(9), pp.1189-1191.

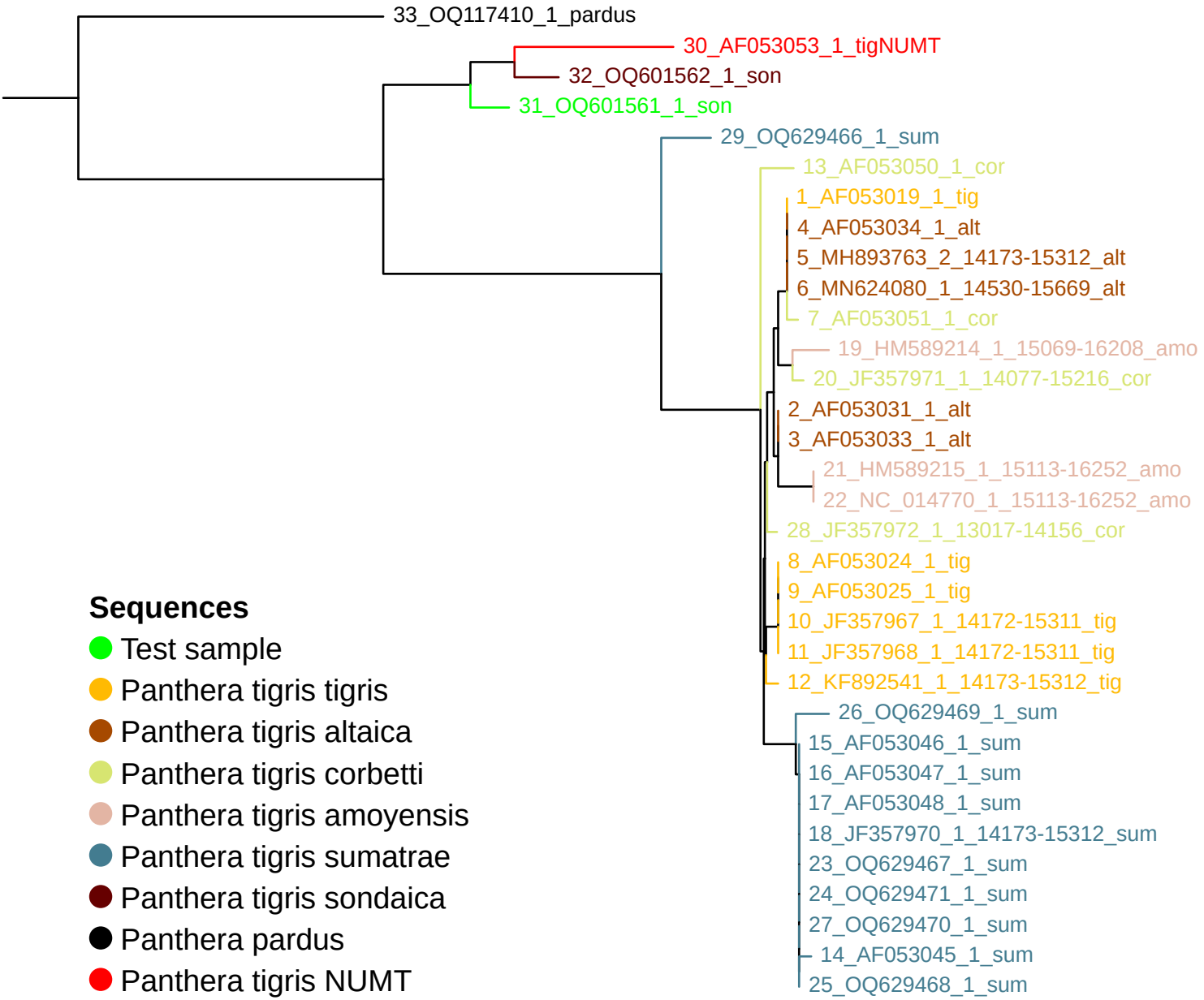
219 Wilting, A., Courtiol, A., Christiansen, P., Niedballa, J., Scharf, A.K., Orlando, L., Balkenhol, N.,
220 Hofer, H., Kramer-Schadt, S., Fickel, J. and Kitchener, A.C., 2015. Planning tiger recovery:
221 understanding intraspecific variation for effective conservation. *Science advances*, *1*(5),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

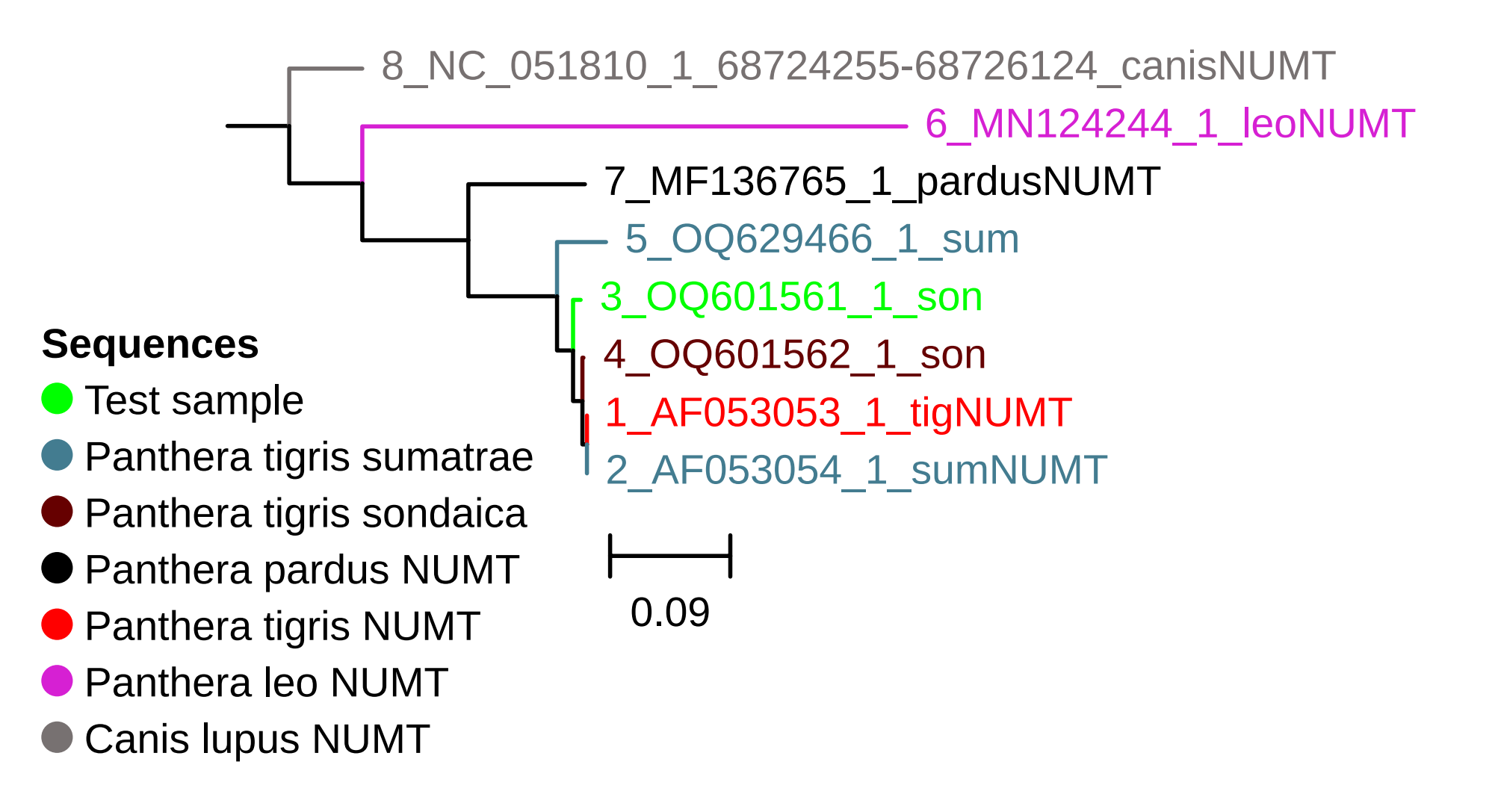
223 Wirdateti, W., Yulianto, Y., Raksasewu, K. and Adriyanto, B., 2024. Is the Javan tiger Panthera
224 tigris sondaica extant? DNA analysis of a recent hair sample. *Oryx*, pp.1-6.
225 doi:10.1017/S0030605323001400

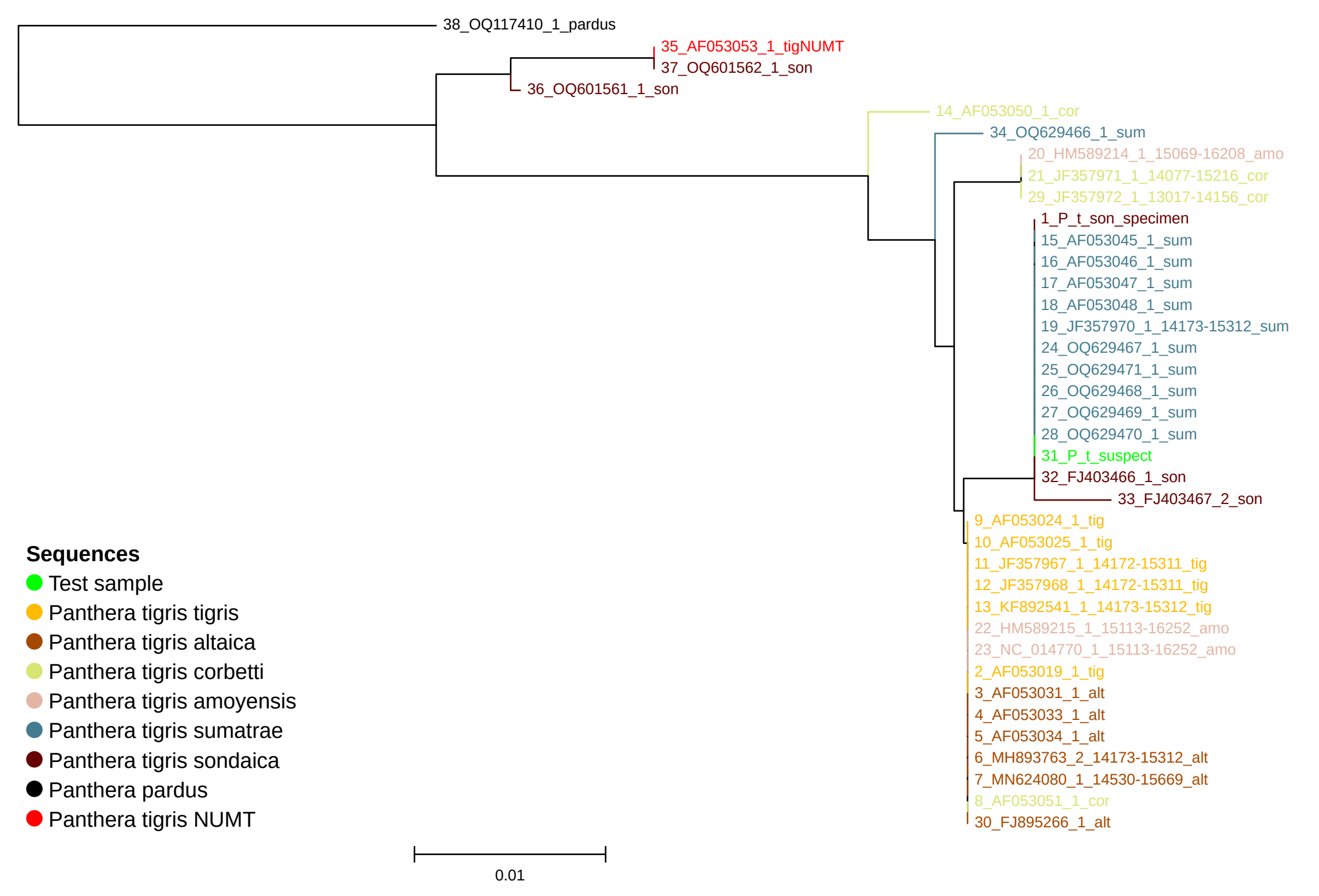
226 Yamaguchi, N., Driscoll, C.A., Werdelin, L., Abramov, A.V., Csorba, G., Cuisin, J., Fernholm, B.,
227 Hiermeier, M., Hills, D., Hunter, L. and Itakura, H., 2013. Locating specimens of extinct tiger
228 (Panthera tigris) subspecies: Javan tiger (PT sondaica), Balinese tiger (PT balica), and Caspian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

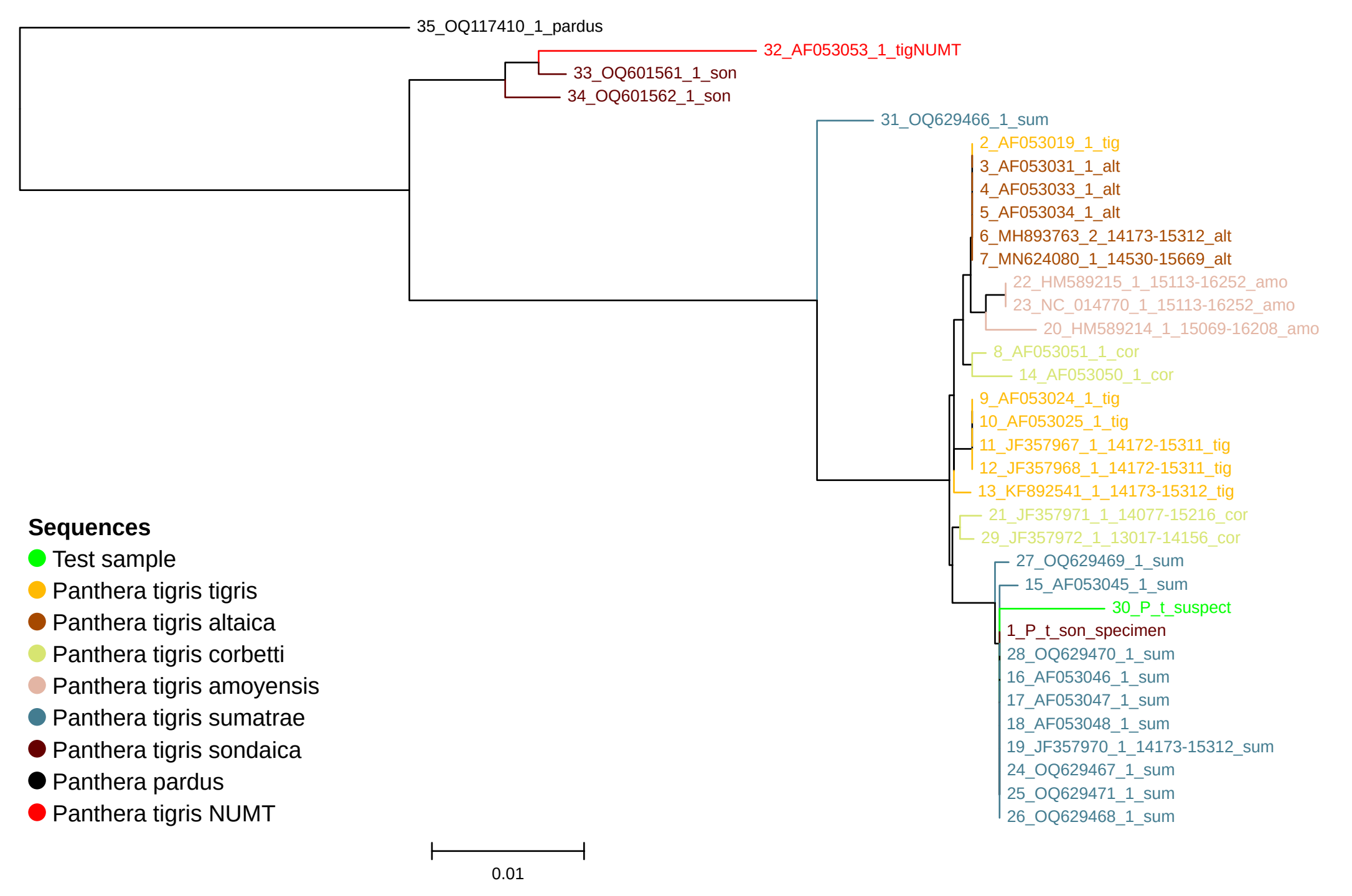
231 Zhang, L., Lan, T., Lin, C., Fu, W., Yuan, Y., Lin, K., Li, H., Sahu, S.K., Liu, Z., Chen, D. and Liu, Q.,
232 2023. Chromosome-scale genomes reveal genomic consequences of inbreeding in the South
233 China tiger: A comparative study with the Amur tiger. *Molecular Ecology Resources*, *23*(2),
234 pp.330-347.


36_OQ117410_1_pardus
33_AF053053_1_tigNUMT
35_OQ601562_1_son
34_OQ601561_1_son
13_AF053050_1_cor
30_OQ629466_1_sum
19_HM589214_1_15069-16208_amo
20_JF357971_1_14077-15216_cor
28_JF357972_1_13017-14156_cor
14_AF053045_1_sum
15_AF053046_1_sum
16_AF053047_1_sum
17_AF053048_1_sum
18_JF357970_1_14173-15312_sum
23_OQ629467_1_sum
24_OQ629471_1_sum
25_OQ629468_1_sum
26_OQ629469_1_sum
27_OQ629470_1_sum
31_FJ403466_1_son
32_FJ403467_2_son
1_AF053019_1_tig
2_AF053031_1_alt
3_AF053033_1_alt
4_AF053034_1_alt
5_MH893763_2_14173-15312_alt
6_MN624080_1_14530-15669_alt
7_AF053051_1_cor
29_FJ895266_1_alt
8_AF053024_1_tig
9_AF053025_1_tig
10_JF357967_1_14172-15311_tig
11_JF357968_1_14172-15311_tig
12_KF892541_1_14173-15312_tig
21_HM589215_1_15113-16252_amo
22_NC_014770_1_15113-16252_amo
Sequences
Test sample
Panthera tigris tigris
Panthera tigris altaica
Panthera tigris corbetti
Panthera tigris amoyensis
Panthera tigris sumatrae
Panthera tigris sondaica
Panthera pardus
Panthera tigris NUMT


33_OQ117410_1_pardus
30_AF053053_1_tigNUMT
32_OQ601562_1_son
31_OQ601561_1_son
29_OQ629466_1_sum
13_AF053050_1_cor
1_AF053019_1_tig
4_AF053034_1_alt
5_MH893763_2_14173-15312_alt
6_MN624080_1_14530-15669_alt
7_AF053051_1_cor
19_HM589214_1_15069-16208_amo
20_JF357971_1_14077-15216_cor
2_AF053031_1_alt
3_AF053033_1_alt
21_HM589215_1_15113-16252_amo
22_NC_014770_1_15113-16252_amo
28_JF357972_1_13017-14156_cor
8_AF053024_1_tig
9_AF053025_1_tig
10_JF357967_1_14172-15311_tig
11_JF357968_1_14172-15311_tig
12_KF892541_1_14173-15312_tig
26_OQ629469_1_sum
15_AF053046_1_sum
16_AF053047_1_sum
17_AF053048_1_sum
18_JF357970_1_14173-15312_sum
23_OQ629467_1_sum
24_OQ629471_1_sum
27_OQ629470_1_sum
14_AF053045_1_sum
25_OQ629468_1_sum
Sequences
Test sample
Panthera tigris tigris
Panthera tigris altaica
Panthera tigris corbetti
Panthera tigris amoyensis
Panthera tigris sumatrae
Panthera tigris sondaica
Panthera pardus
Panthera tigris NUMT


8_NC_051810_1_68724255-68726124_canisNUMT
6_MN124244_1_leoNUMT
7_MF136765_1_pardusNUMT
5_OQ629466_1_sum
3_OQ601561_1_son
4_OQ601562_1_son
1_AF053053_1_tigNUMT
2_AF053054_1_sumNUMT
0.09
Sequences
Test sample
Panthera tigris sumatrae
Panthera tigris sondaica
Panthera pardus NUMT
Panthera tigris NUMT
Panthera leo NUMT
Canis lupus NUMT


38_OQ117410_1_pardus
35_AF053053_1_tigNUMT
37_OQ601562_1_son
36_OQ601561_1_son
14_AF053050_1_cor
34_OQ629466_1_sum
20_HM589214_1_15069-16208_amo
21_JF357971_1_14077-15216_cor
29_JF357972_1_13017-14156_cor
1_P_t_son_specimen
15_AF053045_1_sum
16_AF053046_1_sum
17_AF053047_1_sum
18_AF053048_1_sum
19_JF357970_1_14173-15312_sum
24_OQ629467_1_sum
25_OQ629471_1_sum
26_OQ629468_1_sum
27_OQ629469_1_sum
28_OQ629470_1_sum
31_P_t_suspect
32_FJ403466_1_son
33_FJ403467_2_son
9_AF053024_1_tig
10_AF053025_1_tig
11_JF357967_1_14172-15311_tig
12_JF357968_1_14172-15311_tig
13_KF892541_1_14173-15312_tig
22_HM589215_1_15113-16252_amo
23_NC_014770_1_15113-16252_amo
2_AF053019_1_tig
3_AF053031_1_alt
4_AF053033_1_alt
5_AF053034_1_alt
6_MH893763_2_14173-15312_alt
7_MN624080_1_14530-15669_alt
8_AF053051_1_cor
30_FJ895266_1_alt
Sequences
Test sample
Panthera tigris tigris
Panthera tigris altaica
Panthera tigris corbetti
Panthera tigris amoyensis
Panthera tigris sumatrae
Panthera tigris sondaica
Panthera pardus
Panthera tigris NUMT
0.01


35_OQ117410_1_pardus
32_AF053053_1_tigNUMT
33_OQ601561_1_son
34_OQ601562_1_son
31_OQ629466_1_sum
2_AF053019_1_tig
3_AF053031_1_alt
4_AF053033_1_alt
5_AF053034_1_alt
6_MH893763_2_14173-15312_alt
7_MN624080_1_14530-15669_alt
22_HM589215_1_15113-16252_amo
23_NC_014770_1_15113-16252_amo
20_HM589214_1_15069-16208_amo
8_AF053051_1_cor
14_AF053050_1_cor
9_AF053024_1_tig
10_AF053025_1_tig
11_JF357967_1_14172-15311_tig
12_JF357968_1_14172-15311_tig
13_KF892541_1_14173-15312_tig
21_JF357971_1_14077-15216_cor
29_JF357972_1_13017-14156_cor
27_OQ629469_1_sum
15_AF053045_1_sum
30_P_t_suspect
1_P_t_son_specimen
28_OQ629470_1_sum
16_AF053046_1_sum
17_AF053047_1_sum
18_AF053048_1_sum
19_JF357970_1_14173-15312_sum
24_OQ629467_1_sum
25_OQ629471_1_sum
26_OQ629468_1_sum
Sequences
Test sample
Panthera tigris tigris
Panthera tigris altaica
Panthera tigris corbetti
Panthera tigris amoyensis
Panthera tigris sumatrae
Panthera tigris sondaica
Panthera pardus
Panthera tigris NUMT
0.01