# Near-chromosomal *de novo* assembly of Bengal tiger genome reveals genetic hallmarks of apex-predation

Harsh Shukla[1,*], Kushal Suryamohan[2,3,*], Anubhab Khan[1*], Krishna Mohan[4], Rajadurai C. Perumal[4], Oommen K. Mathew[4], Ramesh Menon[5], Mandumpala Davis Dixon[4], Megha Muraleedharan[4], Boney Kuriakose[4], Saju Michael[4], Sajesh P. Krishnankutty[4], Arun Zachariah[3,6], Somasekar Seshagiri[3,@], and Uma Ramakrishnan[1,@]

[1]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India; [2]Research and Development Department, MedGenome Inc., Foster City, CA 94404, USA; [3]SciGenom Research Foundation, 3rd Floor, Narayana Nethralaya Building, Narayana Health City, #258/A, Bommasandra, Hosur Road, Bangalore, Karnataka, 560099, India; [4]AgriGenome Labs Private Ltd, Kochi, Kerala 682030, India; [5]MedGenome Labs Ltd., 3rd Floor, Narayana Nethralaya Building, Narayana Health City, #258/A, Bommasandra, Hosur Road, Bangalore, Karnataka 560099, India; [6]Wayanad Wildlife Sanctuary, Sultan Bathery, Kerala 673592, India.

* equal contribution
[@] corresponding author (SS: sekar@sgrf.org; UR: uramakri@ncbs.res.in)

**Abstract**

The tiger, a poster child for conservation, remains an endangered apex predator. Continued survival and recovery will require a comprehensive understanding of their genetic diversity and the use of such information for population management. A high-quality tiger genome assembly will be an important tool for conservation genetics, especially for the Indian tiger, the most abundant subspecies in the wild. Here, we present high-quality near-chromosomal genome assemblies of a female and a male wild Indian tiger (*Panthera tigris tigris*). Our assemblies had a scaffold N50 of >140 Mb, with 19 scaffolds, corresponding to the 19 numbered chromosomes, containing 95% of the genome. Our assemblies also enabled detection of longer stretches of runs of homozygosity compared to previous genome assemblies which will improve estimates of genomic inbreeding. Comprehensive genome annotation identified 26,068 protein-coding genes, including several gene families involved in key morphological features such as the teeth, claws, vision, olfaction, taste and body stripes. We also identified 301 microRNAs, 365 small nucleolar RNAs, 632 tRNAs and other noncoding RNA elements, several of which are predicted to regulate key biological pathways that likely contribute to tiger's apex predatory traits. We identify signatures of positive selection in the tiger genome that are consistent with the *Panthera* lineage. Our high-quality genome will enable use of non-invasive samples for comprehensive assessment of genetic diversity, thus supporting effective conservation and management of wild tiger populations.

## Introduction

Tigers are among the most iconic and recognizable species in the world. Despite being apex predators, they are among the most endangered animals with an estimated global population of 3,900 tigers remaining in the wild compared to over 100,000 at the turn of the 20th century (https://www.worldwildlife.org/species/tiger). Historically, wild tigers roamed large swaths of the planet that included a range spanning present-day Armenia in Eastern Asia to Indonesia in Southeast Asia and from the Russian Far East to the Southern tip of India (Luo et al., 2019). Anthropogenic activities such as hunting, urbanization, expansion of agriculture and deforestation, leading to both loss of habitat and prey have resulted in a >95% decline in the wild tiger populations (Ripple et al., 2014).

The Bengal tiger or the Indian tiger, *Panthera tigris tigris*, is a subspecies native to the Indian subcontinent and is highly endangered. It is estimated that over 50,000 tigers inhabited the unbroken forests of India (Mondol et al., 2009). Much like global tiger population decline, Bengal tiger population in India also began to decline more than a century ago, leaving fewer than 2,000 tigers in the wild by 1970. In 1973, India declared the tiger its national animal and set up 'Project Tiger' to conserve this majestic animal (Task Force, 1972). With over 50 tiger reserves, the project has given the tiger a chance for continued survival. Today there are ~2,900 wild tigers in these reserves in India, accounting for ~60% of the global wild tiger population (Jhala et al., 2020). The Bengal tiger population is estimated to be genetically the most diverse and hence represents the best gene pool reservoir important for conservation (Armstrong et al., 2021; Luo et al., 2019). While the genetic diversity for Bengal tigers continues to increase, small and fragmented reserves averaging <1,500 square kilometers in India will likely lead to isolated population groups susceptible to inbreeding depression and genetic bottlenecks, limiting the success of conservation efforts (Armstrong et al., 2021; Khan et al., 2021; Luo et al., 2019). In this context, continued monitoring of the genetic diversity of these populations will be important for conservation. Obtaining high quality samples from the animals in the wild to conduct genetic studies is not practical. A high-quality reference genome will allow the use of DNA from non-invasive samples scat samples to study, monitor and manage wild populations (Khan et al., 2020).

The evolution of genomics technologies over the last decade (Giani et al., 2020; Phillippy, 2017) enable high-quality *de novo* genome assemblies of endangered and threatened species (Fan et al., 2019; Ge et al., 2019; Murchison et al., 2012; Suryamohan et al., 2020; Sutton et al., 2018). These include short- and long-read sequencing, optical mapping and

chromosome conformation capture sequencing technologies that together allow for the generation of near-chromosomal genome assembly.

The Bengal tiger is a diploid organism with a chromosome number 2n = 38, which includes 16 pairs of metacentric/submetacentric autosomes, two pairs of acrocentric autosomes, a meta-centric X, and Y sex chromosome. To date, none of the published tiger genomes are near-chromosomal assemblies. The Amur tiger (*Panthera tigris altaica*) genome assembly, generated using short-read data was highly fragmented (Cho et al., 2013). Recently, a genome assembly for a captive-wild-caught Malayan tiger (*Panthera tigris jacksoni*) using 10x linked-reads has been reported (Armstrong et al., 2021). Additionally, although several tiger resequencing studies have produced additional sequence data (Khan et al., 2020; Liu et al., 2018), none have resulted in a contiguous high-quality reference genome.

Here, we report near-chromosomal *de novo* reference genome assemblies of a male and a female Bengal tiger from the wild. The female animal is a famous Bengal tigress named Machali from the Ranthambore National Park, India, crowned the Queen Mother of Tigers and is the most widely photographed tiger in the world (https://en.wikipedia.org/wiki/Machali_(tigress)). Machali, a reproductively successful founder, gave birth to seven females and four males between 1999 to 2006.

We have performed a comprehensive annotation of our high-quality genomes and identified several genes that contribute to key traits that make the tiger an apex predator. This included genes involved in skin patterning, tooth development, vision, endurance and olfaction. We also show that the contiguous genomes generated here enable identification of longer stretches of runs of homozygosity (ROH) than previously possible with fragmented genomes. Further, we used the reference quality gene annotation models to analyze signals of positive selection and evolution of big cat-specific traits.

## Results

### Genome sequencing and chromosome-level assembly

We generated high-quality genome assembly using long-read and short-read sequencing data, chromatin conformation (Hi-C) sequence and/or optical mapping (Bionano) data (**Supplementary Table 1a-b**, materials and methods) (Dudchenko et al., 2017; Lieberman-Aiden et al., 2009). The Machali (MC) draft assembly generated using long-read sequence data contained 2,766 contigs, spanning 2.4 Gb and had a contig N50 of 3.26 Mb (**Table 1a;** see materials and methods). Following error correction using PacBio and Illumina data, we obtained the PanTigT.MC.v1 assembly that served as input for Hi-C data scaffolding (Putnam et al., 2016). This resulted in a near-chromosomal assembly, PanTigT.MC.v2 containing 1,054 scaffolds and a scaffold N50 of 145.22 Mb. This represented a ~50x improvement in genome contiguity (see materials and methods). The BUSCO completeness score for this assembly for detection of conserved mammalian single-copy orthologs (n= 4,104 genes; **Table 2**) was 95.4% (Simao et al., 2015; Waterhouse et al., 2018).

Similarly, the initial draft assembly of the Southern Indian (SI) animal obtained using long-read data consisted of 845 contigs with a contig N50 of 29 Mb (**Table 1b**, **Supplementary Table 1a-b;** see materials and methods). After polishing with Illumina short-read sequencing, we used the resulting assembly, PanTigT.SI.v1, as input for scaffolding with Chicago chromatin interaction mapping data and ordered and oriented the contigs, corrected mis-joined and merged overlaps. Integration with Chicago data led to a resolved assembly N50 of 2.3 Mb (PanTigT.SI.v2; **Table 1b**). This is consistent with the finding that Chicago data captures short range chromatin interactions and corrects for misjoins in assemblies obtained using long-read sequencing data alone (Putnam et al., 2016). Integration of this assembly with *de novo* assembled Bionano optical map data (PanTigT.SI_BNG) resulted in an assembly, PanTigT.SI.v3, containing 3,811 scaffolds and a scaffold N50 of 141.0 Mb. The BUSCO completeness score for this assembly was 92.6% (**Table 1b and 2**). Interestingly, further incorporation of Hi-C data did not result in an improvement of the SI genome contiguity (see materials and methods).

To assess for any ordering/orientation errors in the PanTigT.MC.v2 assembly after Hi-C scaffolding, we used the chromosome-level reference genome of domestic cat (*Felis catus*; Genbank: GCA_013340865.1, *Felis catus 9.0*) (Buckley et al., 2020) as well as the PanTigT.SI.v3 assembly to delineate the errors from actual structural changes (see materials and methods). Given the highly conserved karyotype among Felidae species (Wurster-Hill and

132 Gray, 1973), we used the domestic cat reference assembly to merge scaffolds into
133 chromosomes and assign chromosomes based on synteny giving rise to the final near-
134 chromosomal assemblies, PanTigT.MC.v3 and PanTigT.SI.v3 (**Figure 1a-b**). The largest
135 scaffold from PanTigT.MC.v3 and PanTigT.SI.v3 spanned over 230 Mb. Further, using the
136 male PanTigT.SI.v3 assembly, we were also able to identify 3 Y-chromosome-linked scaffolds
137 that spanned ~3 Mb (**Figure 1c**; see materials and methods). Overall, both genome assemblies
138 were highly contiguous, with >93% of the genome contained in the 19 near-chromosome-level
139 scaffolds and were syntenic with each other (**Supplementary Figure 1**). Also, consistent with
140 the evolutionary history, both genomes were highly collinear with the domestic cat reference
141 genome (**Figure 1d, Table 2**).
142 Comparison of the MC and SI genome assemblies to other published Felidae genomes
143 showed that the overall contiguity was better than the domestic cat reference genome. For
144 comparison, the scaffold N50 of the Bengal tiger genomes were ~1.7× (83.8 versus 145-147
145 Mb) longer than that of the domestic cat genome (Felis_catus_9.0, Genbank:
146 GCA_013340865.1), a gold standard feline genome. Unlike the tiger genome assembled *de*
147 *novo* in this study, the domestic cat genome was assembled using physical mapping and
148 sequencing. Also, it was improved iteratively since its publication in 2007 (Pontius et al., 2007).
149 A comparison of our genome assemblies to that of the recently published Malayan tiger genome
150 (*Panthera tigris jacksoni*) showed that they were ~7x (21.2 versus 145-147 Mb) better as
151 assessed by scaffold N50 (Armstrong et al., 2021; Cho et al., 2013)(MalTig1.0, Genbank:
152 GCA_019348595.1) (**Table 3**). Also, a comparison of our genome assemblies to that of the
153 Amur tiger genome showed that they were ~16× more contiguous (8.8 versus 145-147 Mb
154 scaffold N50, respectively) (Armstrong et al., 2021; Cho et al., 2013) (PanTig1.0, Genbank:
155 GCA_000464555.1) (**Table 3**). Although, genomes of other big cats such as jaguar and leopard
156 from the *Panthera* genus have been published (Figueiro et al., 2017; Kim et al., 2016), none are
157 near chromosomal reference assemblies (**Table 3**). Given our assembly included a male Bengal
158 tiger, we were able to identify Y chromosome-associated scaffolds that provides an important a
159 reference for Y-haplotyping of tiger populations.
160
161 **Runs of homozygosity analysis**
162 High-quality near-chromosomal genome assemblies can greatly help in several types of
163 population genetic analysis, especially of threatened and endangered species. Measuring
164 inbreeding due to shared ancestry of parental gametes is one of them. Typically, this is
165 measured by identifying runs of homozygosity (ROH) (Ceballos et al., 2018; Kardos et al., 2016;

166 Pemberton et al., 2012). We aligned short-read Illumina data from four zoo-bred Individuals
167 with known pedigree inbreeding coefficients ranging from 0.21 to 0.28 (Sagar et al., 2021) to
168 MC, SI and the published Malayan tiger genomes (Armstrong et al. 2021) to assess the
169 effectiveness of these assemblies in detecting ROH (**Materials and methods**). We observed on
170 an average 2,895, 2,583 and 2,176 ROH regions that were >100 kb in length in the zoo-bred
171 individuals when using the MC, SI and Malayan tiger genome assemblies, respectively
172 (**Supplementary Tables 1c-e**). Further, we estimated cumulative inbreeding in individuals due
173 to common ancestry dating back to 455, 45 and 4 generations ago (**Materials and methods**).
174 We also observed that all three genomes performed comparably when estimating shorter ROH
175 that represent older ancestry (Khan et al. 2021) (**Figure 1e**) while the MC and SI assemblies
176 performed better at detection of ROH longer than 1 Mb that represents recent inbreeding,
177 compared to the Malayan tiger genome (n= 177 for MC, n= 106 for SI vs n= 94 for Malayan tiger
178 on average) (**Figure 1f**). All estimates of recent inbreeding from ROH longer than 1Mb
179 correlated with the pedigree inbreeding coefficients of the individuals, thus demonstrating the
180 need for near-chromosomal genome assemblies in conservation genomics.
181
182 **Genome Annotation**
183 Our genome assemblies, PanTigT.MC.v3 and PanTigT.SI.v3, both contained 19
184 scaffolds >10 Mb corresponding to the 19 chromosomes (**Figure. 1a-b**). These scaffolds
185 accounted for ~93% of the genome. The average DNA base (GC) content of the assemblies
186 was about 41%. Analysis of the repeat content revealed that ~36% of the genome was
187 repetitive (~870 Mb; **Figure 1a-b** and **Supplementary Table 1f** and **Supplementary Figure 2**)
188 with long interspersed nuclear elements (LINEs) being the dominant family of repeats
189 (**Supplementary Table 1f**).
190 We also annotated the genomes for noncoding RNA elements and created a database
191 of micro RNAs (miRNAs; n=299 in MC and 301 in SI), small nucleolar RNAs (SnoRNA; n=365 in
192 MC and 371 in SI), tRNAs (n=548 and 632 in MC and SI) and other noncoding RNA elements
193 (**Supplementary Table 1g**). We next searched for predicted miRNAs targets using miRanda
194 (Enright et al., 2003). Several miRNAs were predicted to target genes involved in critical
195 biological processes including angiogenesis (n=132 target genes; GO:0001525 and
196 GO:0045765-0045766), brain (n=78; GO:0007420, 1990403), bone (n=73; GO:0060348,
197 0048539, 0030282, 0060349, 0030500) and eye (n=63; GO:0002088, 0043010, 0001654,
198 0048593) and heart development (n = 43; GO:0035904, 0060347, 0007507, 0007512, 0060914,
199 0060973) (**Supplementary Tables 1h-i**). Functional validation studies involving the predicted

200 miRNAs (Ambros, 2004; Lee et al., 1993) shall lead to a better understanding of the post-
201 transcriptional gene expression regulation in tigers and big cats during pre- and post-
202 development.
203 Next, we used the MAKER pipeline (Campbell et al., 2014; Cantarel et al., 2008) to
204 annotate the genomes using protein homology and RNA-seq expression data. We predicted
205 19,931 genes that correspond to 22,718 transcripts in the PanTigT.MC.v3 genome assembly.
206 In the male SI individual PanTigT.SI.v3, we detected 21,126 protein-coding genes that mapped
207 to 24,074 transcripts (**Figure 2a and Supplementary Table 1j**). Overall, the annotation yielded
208 26,068 unique protein-coding between the MC and SI genomes. About 95% of all annotated
209 genes were located on the 19 largest scaffolds corresponding to the numbered chromosomes in
210 the genomes. Using a previously developed annotation pipeline (Suryamohan et al., 2020), we
211 functionally classified the protein coding genes (**Supplementary Table 1j**). A total of 19,640
212 (86%) and 20,412 (85%) predicted proteins contained a canonical start and stop codon in the
213 PanTigT.MC.v3 and PanTigT.SI.v3 assemblies. We found that ~99% of these genes had a
214 corresponding ortholog in either the Human Gene Nomenclature Committee database, NCBI's
215 non-redundant database or the TrEMBL (https://www.ebi.ac.uk/uniprot) database
216 (**Supplementary Tables 1k-n**). High-level gene function classification of the identified proteins
217 revealed 179 solute carriers as the most abundant type of proteins followed by 78 zinc finger
218 proteins (**Supplementary Figure 3** and **Supplementary Tables 1l and 1n**).
219 A tiger typically consumes at least one deer-sized animal each week for survival. Its
220 metabolism, sensory functions and other adaptations are key to it evolutionary success as an
221 apex predator. Using our high-contiguity genomes, we annotated several full-length amino-acid
222 transport (n=7; GO: 0006865) and protein, cholesterol and fatty acid metabolism-related genes
223 (n=21; GO: 0008203, 0019538, 0006631) that likely are crucial for the successful adaptation to
224 its carnivorous diet (**Supplementary Tables 1kand 1m**) (Kim et al., 2016). Further, we
225 identified multiple genes involved in G-protein coupled signaling (n=270) and olfactory receptor
226 activity (n=107) that likely have an important role in sensory functions such as smell, vision and
227 hearing, nervous system development, mate selection and hunting (Barnett et al., 2020).
228 The evolutionary versatility of teeth is an important factor contributing to the success of
229 carnivores. Diet-dependent changes in complexity of dental patterns, tooth morphology,
230 number, function and diet are interlinked (Aristotle et al., 1937). One unique morphological
231 feature that has evolved independently numerous times within the theriodont lineage that
232 includes mammalia, is the saber-tooth morphology of permanent upper canines (Simpson,
233 1941). Tigers have the largest upper canines of all the big cats and is key for its apex predatory

234 status. We identified homologs of gene families known to be involved in tooth development
235 including the *Wnt* family members and homeobox genes (*Hox*) (Sharpe, 1995) *Msx-1*, *Msx-*
236 *2*,(Vastardis et al., 1996), *Dlx-1*, *Dlx-2* and *Barx-1.* We also identified *Pax9* known to be
237 important for molar development. Additionally, our analysis identified several *Fibroblast growth*
238 *factor* (*Fgf*)*,* and *LIM homeodomain transcription factor* (*Lhx*) family members (Zhang et al.,
239 2005) involved in regulating enamel and dentin levels (**Figure 2b**; **Supplementary Tables 1k**
240 **and 1m**). Mouse genetic analysis showed that the *fibroblast growth factor 10* (*Fgf10*) to be a
241 critical factor in determining the length of the teeth (Yokohama-Tamaki et al., 2006). In addition
242 to *Fgf genes*, we also identified *Matrix metalloproteinase-20* (*Mmp-20*) that has been shown to
243 be involved in tooth development (**Figure 2b**; **Supplementary Tables 1k and 1m**). (Thomas
244 and Sharpe, 1998; Thomas et al., 1997).
245 The retractable sharp tiger claws are an important tiger adaptation critical for predation.
246 Claws are a variation in the distal limb integumentary appendages of certain mammals,
247 including most carnivorous animals. Several epithelial-mesenchymal signaling molecules
248 involved in patterning ectodermal derivatives such as teeth, hair, and feathers are also involved
249 in patterning distal epidermal appendages such as claws and nails (Vickaryous and Sire, 2009).
250 Mutations in several *Hox family member* genes have previously been shown to affect claw/nail
251 and limb development in mice and lead to disorders such as brachydactyly in humans (Davis
252 and Capecchi, 1996; Hamrick, 2001; Johnson et al., 2003; Luo et al., 2019; Montague et al.,
253 2014). In the tiger genome besides the several homeobox genes including *Homeobox* family
254 genes (*Hox*), *Msx-1* and *Msx-2,* we identified bone morphogenic protein (*Bmp*) family members
255 known to be involved in early distal limb development (e.g., *Bmp4*, posterior *Hox* genes) and
256 induction of the claw growth, and/or proliferation (**Figure 2b**; **Supplementary Tables 1k and**
257 **1m**).
258 It is estimated that about two-thirds of all extant carnivorous mammals are mainly
259 nocturnal (Roots, 2006). Nocturnality is characterized by an expansion in the number of genes
260 responsible for dim/low-light vision. The retina is a light-sensitive layer of eye tissue and
261 consists of six types of neurons which include light-sensitive photoreceptor cells (rods and
262 cones). Rod cells are responsible for discerning shapes and are highly sensitive to low light
263 while cone cells confer bright light sensitivity as well as color vision. Tigers have more rods in
264 their eyes than cones (Ahnelt et al., 2006). We identified several genes critical for low-light
265 vision, including Neural retina leucine zipper (*Nrl*), S-antigen visual arrestin (*Sag*), G Protein
266 Subunit Alpha Transducin 1 (*Gnat1*), G Protein Subunit Beta 3 (*Gnb3*), Phosphodiesterase
267 genes *Pde6a, Pde6b and Pd36g,* and Cyclic Nucleotide Gated Channel Subunit Alpha 1

268 (*Cnga1*) and Beta 1 (*Cngb1*) (**Figure 2b**; **Supplementary Tables 1k and 1m**). Knockout
269 mutation of *Nrl* leads to the loss of rod cells (Ohsawa and Kageyama, 2008) while mutations in
270 *Gnb3*, *Cnga1* and *Cngb1* have been shown to lead to night blindness in humans and retinal
271 degeneration in chickens (Bareil et al., 2001; Katagiri et al., 2014; Tummala et al., 2006; Vincent
272 et al., 2016).
273 While the tiger's sense of smell, though not critical for hunting, is used for
274 communication between animals particularly, for marking territory, courtship and reproduction.
275 The vomeronasal organ, also called the Jacobson's organ, is primarily responsible for
276 pheromone detection in tigers (Poddar-Sarkar and Brahmachary, 2014). We identified twenty
277 one vomeronasal type-1 receptor genes (V1R), of which twelve were incomplete with truncated
278 sequences, indicating potential pseudogenization, consistent with findings in other carnivorous
279 species (Young et al., 2010). In contrast to the ~400 olfactory receptor genes found in humans
280 (Hasin-Brumshtein et al., 2009), we identified 69 olfactory receptor (OR) genes based on
281 homology to other OR genes (**Figure 2b**; **Supplementary Tables 1k and 1m**) (Montague et al.,
282 2014).
283 Mammalian taste is mainly mediated by receptor cells organized in taste buds on the
284 tongue (Bachmanov et al., 2014). It has been established that the tiger can taste salt, bitter and
285 acidic flavors and to a lesser degree sweetness (Jiang et al., 2012; Li et al., 2005). We
286 identified a fully intact copy of *Tas1r1* (type 1 taste receptor), responsible for sweet taste
287 perception, and two Tas2r (type 2 taste receptor) genes involved in bitter taste perception
288 (Chandrashekar et al., 2000) (**Figure 2b**; **Supplementary Tables 1k and 1m**). Notably, we did
289 not identify a full-length homolog of *Tas1r2*, a sweet taste receptor gene, consistent with
290 observations in other carnivorous mammals (Jiang et al., 2012; Li et al., 2009b).
291 Tigers are the only striped cat of the genus *Panthera* and are most recognizable for their
292 pattern of black vertical stripes on reddish-orange fur. Our high-quality assembly and
293 annotation identified several genes associated with morphometric variation in domestic cats
294 such as pigmentation, coat patterns and other phenotypes. Included among these were
295 homologs of the atrial natriuretic peptide-converting enzyme (*Corin*), membrane-associated
296 transporter protein (*Slc45a2)*, *Agouti-signaling protein* (*Asip*), *Tyrosinase-related protein 1*
297 (*Tyrp1*), *Tyrosinase-related protein 2* (*Tyrp2*), *Tyrosinase* (*Tyr*), *Melanophilin* (*Mlph*), *Tyrosine-*
298 *Protein Kinase Kit* (*Kit*), *Keratin 71* (*Krt71*) and *Aminopeptidase N* (*Laeverin*; *Anpep*). Mutations
299 in these genes can lead to a variety of phenotypes including variation in coat and feet color,
300 stripe patterns and the color of stripes, tail shapes (**Figure 2c-d**) (Gandolfi et al., 2013; Ishida et
301 al., 2006; Kaelin et al., 2012; Luo et al., 2019; Lyons et al., 2005; Xu et al., 2017). Of note, our

302 high-quality genome enabled accurate annotation of long genes with their complete intron-exon
303 structure including that of *Slc45a2,* a ~33kb gene consisting of 7 exons. Using a combination of
304 transcriptome (RNA-seq) and genome data, we were also able to annotate a ~249 kb long full-
305 length copy of *Corin* made of 22 exons. These fully annotated genes will serve as a high-quality
306 resource for understanding coat patterning genes in tigers and other animals.
307
308 **Phylogeny and positive selection analysis**
309 To compare the annotations of the genomes from this study to other big cat species
310 including the Amur tiger, we constructed a phylogeny using annotated protein sequences from
311 13 selected species, namely, dog, Amur tiger, Bengal tigers (this study), lion, leopard, jaguar,
312 clouded leopard, domestic cat, Canada lynx, cougar, horse, and rabbit. The proteins from these
313 species clustered into 26,169 orthogroups (representing 97.2% of the input protein sequences)
314 of which 4,576 were single copy orthologs. The remaining protein sequences did not cluster
315 (854 proteins (3.5%) in SI and 489 genes (2.2%) in MC, respectively). Next, we used the 4,576
316 single-copy orthologs to test for signals of positive selection in big cats using site models in
317 PAML (Yang, 2007). Using Likelihood Ratio Tests (LRT) for models M1-M2 and M7-M8, we
318 identified 1,484 single copy orthologs to be positively selected (p-value <0.05) (**Supplementary**
319 **Tables 1o-q**). Included in this list of positively selected genes, were genes involved in muscle
320 development *Atp1b4*, angiogenesis (*Srpx2*), fatty acid and cholesterol metabolism (*Slc4a9,*
321 *Apod*), mitochondrial respiration (*Cq10a*) and heart development/endurance-related genes (e.g.,
322 *Nkx25*), indicating that these genes are evolving at a faster rate, consistent with the
323 physiological needs of big cats (**Figure 3b** and **Supplementary Tables 1o-q**) (Barnett et al.,
324 2020; Cho et al., 2013; Figueiro et al., 2017; Kim et al., 2016). Further, pathway enrichment
325 analysis of the positively selected genes confirmed enrichment of several metabolism-related
326 pathways including cholesterol metabolism (GO:0008203) and fatty acid metabolism
327 (GO:0006631) that are critical obligate carnivores (**Figure 3c**).
328 **Discussion**
329 We have de novo assembled and annotated genomes of a female and a male Bengal
330 tiger. The resulting assemblies were 17x more contiguous than the published Amur tiger
331 genome and 1.7x more contiguous than the domestic cat genome with a scaffold N50 of over
332 140 Mb, making it the most contiguous near-chromosomal wild felid genomes assembled to
333 date (**Table 3**).
334 We provide here a comprehensive annotation of 26,068 protein-coding genes from the
335 tiger genomes. Additionally, we identified over 3,000 non-coding genes including for the first

336 time a genome-wide analysis of micro-RNAs and their putative target genes in a tiger genome.
337 Functional assignment identified genes and signaling pathways involved in endurance,
338 neurotransmission, teeth, vision, claw and skin patterning, all of which are important for the
339 tiger's adaption as an apex predator. We have performed extensive annotation of genes
340 involved in variations in coat color, stripe patterns and other morphometric variations.
341 Our highly contiguous genome along with the annotations is an important resource for
342 studies on the genetics of Felidae and in general carnivoran development and evolution. For
343 example, it is now feasible to establish homology between tooth components between different
344 big cat species by using the annotation generated in this study. Our annotated gene sets for the
345 Bengal tiger genomes provides a high-quality reference that can be used to support
346 resequencing efforts aimed at studying tiger populations. The genome assemblies presented
347 here demonstrate their utility in population and conservation genetic studies of tigers. They
348 increase the accuracy of the estimates of important population genetic parameters like
349 inbreeding for assessing the threats to a population. Our reference genome along with
350 annotation-guided quantitative studies on wild and inbred, zoo-bred individuals will help in
351 understanding the possible phenotypic effects of inbreeding depression through assessment of
352 ROH. The high-quality genomes present in this study will be important resources that will
353 enable analysis of sequencing data obtained from DNA present in non-invasive scat samples
354 obtained from the wild. Such analysis will be important for understanding and monitoring tiger
355 population genetic diversity in the wild and support informed conservation efforts.

356 **Methods**

357 **Samples and nucleic acid preps**

358 Bengal tiger blood samples used in the study are reported in **Supplementary Table 1a**).
359 Genomic DNA from MC and SI was extracted from whole blood using the MagAttract HMW
360 DNA extraction kit (Qiagen). Sample quality was checked using Qubit 3.0, Nanodrop 8000,
361 Agilent TapeStation 4200 and pulse field gel electrophoresis (Sage Science). Total RNA was
362 isolated from the blood using Trizol (ThermoFisher). A total of 500ng of RNA was used for
363 library preparation.

364

365 **Sequencing**

366 Sequencing libraries for PacBio, Oxford Nanopore (ONT) and Illumina sequencing were
367 constructed as per manufacturer's instructions. Considering the genome size of Panthera Tigris
368 at ~2.4 Gb, we generated ~79.6 Gb of Pacbio long-read data (~33x; N50 = 10.3 Kb) and ~18.28
369 Gb of ONT long-read data (~7.5x, N50 = 8.14 kb) for the female tiger (MC). In addition, short-
370 read Illumina sequencing data 117 Gb, (50x; 250 bp library), 35 Gb (15x; 300 bp library 35 Gb
371 (15x; 500 bp library), and 35 Gb (15x; 800 bp library) was also generated. For the male tiger
372 (SI), a total of 160.3 Gb (64x) ONT data, and ~60 Gb (25x) of Illumina short-read data was
373 generated (**Supplementary Table 1b**). For both animals, the long-read data was filtered using
374 BLAST against the 16s rRNA database (https://ncbiinsights.ncbi.nlm.nih.gov/2020/02/21/rrna-
375 databases/) to identify and remove any potential contaminant reads prior to genome assembly.
376 In addition, blood from an offspring of Machali, ST3 was also obtained.

377 RNA-seq short read (Illumina) data was obtained by sequencing PolyA RNA-sequencing
378 (RNA-seq) libraries prepared from ST3 blood using the Illumina TruSeq stranded messenger
379 RNA kit (**Supplementary Table 1b**).

380

381 **Chicago library preparation and sequencing**

382 Chicago library was prepared as described previously (Putnam et al., 2016). Briefly, for
383 each library, ~500ng of HMW gDNA (mean fragment length of 75 kb) was reconstituted into
384 chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5'
385 overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After
386 ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated
387 to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350
388 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes
389 and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin

390 beads before PCR enrichment of each library. The libraries were sequenced on an Illumina
391 platform. A total of 155 million read pairs (2 x 150 bp) was generated for this library providing
392 ~20x physical coverage of the genome (1-50 kb pairs) (**Supplementary Table 1b**).
393
394 **Bionano optical mapping data generation**
395 High-molecular weight genomic DNA extracted from whole blood from SI was embedded
396 in a thin agarose layer and was labeled and counterstained following the Direct Label and Stain
397 (DLS) Reagent Kit protocol (Bionano Genomics). The labeled sample was then loaded onto a
398 Saphyr chip and run on the Saphyr imaging instrument (BioNano Genomics). A total of 473.7
399 Gb (~190x) optical map data was generated. *De novo* genome assembly using the Bionano
400 Access software for the DLS optical map data produced a 2.63 Gb assembly (PanTigT_BNG)
401 consisting of 182 scaffolds (N50 =145.93 Mb) (**Supplementary Table 1b**).
402
403 **Hi-C library preparation and sequencing**
404 Hi-C library (Dovetail Genomics), was prepared for the two samples (SI and ST3) as
405 described previously (Lieberman-Aiden et al., 2009). Briefly, chromatin was fixed in place with
406 formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the
407 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After
408 ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated
409 to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350
410 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes
411 and Illumina-compatible adapters and sequenced (**Supplementary Table 1b**).
412
413 **Genome assembly**
414 We used Canu (v1.7.1; *corMinCoverage* = 0, *corMhapSensitivity = high and*
415 *correctedErrorRate*=0.12 (MC)/0.15 (SI)) to generate an initial draft assembly for both the study
416 animals. The primary MC draft assembly (PanTigT.MC.v1) consisted of 2.4 Gb spanning 2,766
417 contigs (contig N50 = 3.26 Mb) (**Table 1**) and the SI draft assembly consisted of 845 contigs
418 totaling 2.38 Gb (contig N50 = 29.03 Mb). Next, both assemblies were polished to correct for
419 base call errors using raw PacBio data (MC only) and Illumina data (MC and SI). Briefly,
420 PacBio reads were aligned to the PanTigT.MC.v1 using *pbalign* (pbalign version: 0.3.1)
421 resulting alignment bam file was used as input for polishing using *Arrow*
422 *(*https://github.com/pacificbiosciences/genomicconsensus/*)*. A total of two rounds of polishing
423 were done using *Arrow.* Illumina short-read libraries generated for MC and SI (**Supplementary**

424 **Table 1b**) were used for error correction and conflict resolution using Pilon (v1.22) with the
425 following parameters (*--changes --diploid --fix all*) to produce polished assemblies
426 PanTig.T.MC.v1 and PanTigT.SI.v1 (**Table 1**).
427 Next, PanTigT.MC.v1 was used as the input for scaffolding with Hi-C sequence data
428 using HiRise resulting in an assembly PanTigT.MC.v2 containing 1,054 scaffolds (scaffold N50=
429 145.7 Mb). Similarly, PanTigT.SI.v1 was used as the input for scaffolding with Chicago
430 sequence data using HiRise. This resulted in an assembly PanTigT.SI.v2 containing 6,889
431 scaffolds (scaffold N50 = 2.3 Mb). PanTigT.SI.v2 was then scaffolded with Bionano optical map
432 data to yield PanTigT.SI.v3 assembly comprising of 3,886 scaffolds (scaffold N50 = 141.3 Mb).
433 Interestingly, incorporation of Hi-C data did not improve the contiguity of PanTigT.SI.v3 and did
434 not result in any breaks or joins to the assembly. We also assessed if the order of integration of
435 the different sequencing data types would affect assembly quality. While scaffolding
436 PanTigT.SI.v1 with Hi-C data alone (no Bionano and Chicago) resulted in an assembly with
437 fewer scaffolds (n=692) but lower N50 (125.3 Mb), scaffolding of PanTigT.SI.v1 with only
438 Chicago and Hi-C data resulted in an assembly with 6,461 scaffolds and lower scaffold N50 of
439 48.2 Mb. Thus, we retained the Chicago-scaffolded assembly (PanTigT.SI.v2) for downstream
440 analysis including chromosome assignment. Using the domestic cat reference assembly
441 (Felis_catus_9.0), we merged scaffolds into chromosomes, corrected orientation errors and
442 assigned chromosomes based on synteny using RaGOO (https://github.com/malonge/RaGOO)
443 resulting in the final SI assembly PanTigT.SI.v3 and PanTigT.MC.v3.
444 Next, error-corrected reads obtained from long-read data were used in conjunction with
445 RAILS (v1.4.1) (https://github.com/bcgsc/RAILS) to fill gaps in PanTigT.MC.v3 and
446 PanTigT.SI.v3 using Cobbler with the following parameters: -d 1000 -i 0.95. A total of 21 out of
447 1700 gaps (1.24%) were filled in PanTigT.MC.v3 with the largest gap closed being ~10,025 bp
448 while no significant improvement was observed for PanTigT.SI.v3. A final round of polishing with
449 Racon (v1.3.3) was done on PanTigT.MC.v3.
450
451 **Y-chromosome-linked scaffold identification**
452 To identify the male Y-linked scaffolds, we separately aligned Illumina reads obtained
453 from the female (MC) and the male animal (SI) to PanTigT.SI.v3 assembly using BWA (Li and
454 Durbin, 2009) allowing for two mismatches and one indel. Scaffolds with less than 80%
455 alignment coverage were excluded from further analysis. Then, single-base depths were
456 calculated using SAMtools (Li et al., 2009a) following which coverage and mean depth for each
457 scaffold was calculated. Using the average coverage across a scaffold using either the male or

458 female reads, we identified 3 Y-linked scaffolds Sc4gRnz_35;HRSCAF=35_pilon,
459 Sc4gRnz_53;HRSCAF=53_pilon and Sc4gRnz_8;HRSCAF=8_pilon.
460
461 **Identification of mitochondrial contig**
462 In order to identify the mitochondrial contig, whole mitochondrial sequence
463 (NC_010642.1) for *Panthera tigris tigris* was downloaded from NCBI and all contigs from MC
464 and SI were mapped to it using minimap2. Two distinct mitochondrial contigs (MT) were
465 identified and retained in the primary assemblies.
466
467 **Chromosome painting**
468 Chromosome painting of SI or MC chromosomes with the domestic cat chromosomes
469 was performed using the SatsumaSynteny2 script with default parameters (Grabherr et al.,
470 2010).
471
472 **Runs of Homozygosity (ROH) analysis**
473 ROH were identified as described in Armstrong et al. (2021). We aligned reads from short-
474 read sequencing (**Supplementary Table 1b**) from four zoo-bred tigers with known inbreeding
475 coefficients (Sagar et al., 2021) to MC, SI and Malayan tiger genome assembly (Armstrong et
476 al. 2021). After sequence alignment, variant calling and filtering we used BCFtools/ROH at
477 default settings to determine the allele frequencies. We categorized ROH into three categories
478 of stretches longer than 100Kb, 1Mb and 10Mb.
479
480 **Repeat element identification**
481 We identified the repetitive elements in the genome by combining both homology-based
482 and de novo predictions. Briefly, we used RepeatModeler (v.1.0.11)
483 (http://www.repeatmasker.org/RepeatModeler.html) to construct the species-specific repeat
484 sequence libraries for both SI and MC genome assemblies, and then used these as a query to
485 identify repetitive elements using RepeatMasker.
486
487 **Noncoding RNA Annotation**
488 tRNA-encoding genes were predicted by tRNAscan-SE (version 1.3.1) (Lowe and Eddy,
489 1997) with default parameters. Ribosomal RNA (rRNA) fragments were predicted by aligning to
490 mammalian rRNA sequences by using BLASTN (version 2.2.26) with an E-value cutoff of 1e-10.
491 MicroRNA (miRNA) and small nuclear RNA (snRNA) genes were annotated using INFERNAL

492 (version 1.1.3) (Nawrocki and Eddy, 2013) by searching against the Rfam database (release
493 13.0) (Kalvari et al., 2018). Whole-genome target gene prediction for miRNAs was performed
494 using miRanda (version 3.3a) (Enright et al., 2003) using default parameters. Gene ontology
495 terms for the target genes were obtained from the functional annotation for the genomes.
496
497 **Genome Annotation**
498 Gene prediction was performed on the assemblies PanTigT.SI.v3 and PanTigT.MC.v3
499 using the genome annotation tool, MAKER2 (v2.31.10) (Cantarel et al., 2008) in an iterative
500 process. First, *ab initio* gene prediction was performed by the programs SNAP (v2006- 07-28)
501 (Korf, 2004) and Augustus (v3.2.3) (Stanke and Morgenstern, 2005) using publicly available
502 transcriptome data from the Siberian tiger (SRA091968), a de novo assembled transcriptome
503 from ST3 blood RNA-seq data using Trinity (Grabherr et al., 2011) and a combined protein
504 database consisting of proteins from the UniProt/Swiss-Prot and NCBI non-redundant
505 databases of reviewed proteins. A total of three iterative runs of MAKER was used to refine the
506 gene models and produce the final gene set with an annotation edit distance (AED) cutoff of 0.6.
507 Genome annotation quality was assessed by BUSCO analysis using the conserved core set of
508 mammalian genes (79.3% and 76.9% BUSCO completeness score for MC and SI assemblies,
509 respectively). Next, using the set of complete predicted protein sequences, we performed
510 functional annotation as described previously (Suryamohan et al., 2020) (**Supplementary Table**
511 **1g**).
512 To identify gene families specific to tooth development, claw, taste, neurotransmitter,
513 olfaction and coat color/pattern, we performed a literature-based search (Figueiro et al., 2017;
514 Hamrick, 2001; Luo et al., 2019; Montague et al., 2014) to identify genes potentially involved in
515 these developmental pathways. Completeness of gene models for each trait-specific gene
516 where present (predicted by MAKER) was confirmed by aligning each gene against its homolog
517 using NCBI BLAST. Genes that were partial or incompletely annotated were manually curated.
518 Briefly, for each incomplete gene, the corresponding homolog obtained from UniProt was
519 aligned to the genome (SI/MC) using the 'est2genome' parameter in the Exonerate program
520 with default parameters. The aligned genome coordinates of trait-specific genes were then
521 checked against the annotation file to validate the gene prediction.
522
523 **Phylogeny and positive selection analysis**
524 Proteomes from 12 selected species, namely, dog, Bengal tiger (this study), Amur tiger,
525 lion, leopard, jaguar, clouded leopard, domestic cat, Canada lynx, cougar, horse, and rabbit

526 were obtained from Ensembl Biomart (www.ensembl.org). Next, Orthofinder (v2.5.2) was used
527 to cluster the proteins into orthogroups. Of note, the snow leopard was not included in this
528 analysis due to limited annotation models and lack of a reference genome. Next, single copy
529 orthologs were extracted for evolutionary divergence estimation and positive selection analyses.
530 The single copy orthologs were aligned with muscle (v3.8.31) and alignments trimmed and
531 converted to codon level alignments using trimal (v1.4rev15). For each of these alignments
532 PAML/codeml (v.4.10) (Yang, 2007) was run pairwise to obtain four-fold degenerate bases
533 which are concatenated into a super-matrix based on gene partitions. The dN/dS ratios are also
534 obtained in this process. The resulting supermatrix was used to generate a species tree using
535 iqtree2 (v2.0.3) with GTR+F+I+G4 models along with ultrabootstrap (1000 iterations). Further,
536 PAML/mcmctree was used to estimate evolutionary divergence among the species considered
537 with 95% confidence intervals with a burn-in of 2,000 and sampling 20,000. The evolutionary
538 divergence calibrations between the following species pairs were extracted from timetree.org:
539 *Felis catus* and *Lynx canadensis* (7.9 - 13.1 MYA), *Panthera concolor* and *Felis catus* (8.8 -
540 13.9 MYA), *Panthera leo* and *Panthera pardus* (2.6 - 4.74 MYA), *Neofelis nebulosa* and
541 *Panthera onca* (7.4 - 14 MYA).
542
543 **Site model tests**
544 The single copy orthologs were fit to the following codon substitution models (M0, M1, M2, M7,
545 M8) of PAML/codeml (Yang, 2007). Likelihood Ratio Test (LRT) between M1 vs M2 and M8 vs
546 M7 pairs were used to determine positive selection at Chi-square pvalue <= 0.05. Multiple
547 testing using Benjamini-Hodgberg (BH) method was then performed to determine FDR (False
548 Dicovery Rate) for the p-values obtained above in order to determine significantly positively
549 selected genes among the species. Bayesian Empirical Bayes (BEB) (Yang et al., 2005) results
550 were used to identify the sites of positive selection at >=95% and >=99% confidence intervals.
551
552
553
554 **Software URLs**
555 Canu - https://github.com/marbl/canu
556 RAILS - https://github.com/bcgsc/RAILS
557 RaGOO - https://github.com/malonge/RaGOO
558 PAML – http://abacus.gene.ucl.ac.uk/software/paml.html
559 Orthofinder - https://github.com/davidemms/OrthoFinder

Trimal – http://trimal.cgenomics.org/
Iqtree - http://www.iqtree.org/
Muscle - https://www.ebi.ac.uk/Tools/msa/muscle/
Vcftools - http://vcftools.sourceforge.net/
VEP – https://www.ensembl.org/vep
MAKER - http://www.yandell-lab.org/software/maker.html
BUSCO - https://busco.ezlab.org/
Exonerate - https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
Trinity - https://github.com/trinityrnaseq/trinityrnaseq
Augustus - http://bioinf.uni-greifswald.de/augustus/
SNAP - https://github.com/KorfLab/SNAP
Repeatmasker - https://www.repeatmasker.org/
RepeatModeler - https://www.repeatmasker.org/
D-Genies - http://dgenies.toulouse.inra.fr/
Satsuma - https://github.com/bioinfologics/satsuma2
Pilon - https://github.com/broadinstitute/pilon
miRanda - http://cbio.mskcc.org/miRNA2003/miranda.html
INFERNAL - http://eddylab.org/infernal
PLINK - http://zzz.bwh.harvard.edu/plink/
BWA - https://github.com/lh3/bwa
Samtools - http://www.htslib.org/
Sentieon - https://github.com/Sentieon
Racon - https://github.com/isovic/racon

**Acknowledgements**

This work was partially funded by Senior Fellowship DBT Wellcome Trust India Alliance grant to UR (IA/S/16/2/502714); NCBS data cluster used is supported under project no. 12-R&D-TFR-5.04-0900, Department of Atomic Energy, Government of India). All samples used here from Rajasthan were collected for research presented in Khan et al., 2021, and permissions are detailed there. We thank the Rajasthan Forest Department for their support.

**Data Accessibility**

593 Raw sequencing data (DNA and RNA) as well as both genome assemblies can be accessed at
594 NCBI under BioProject accession nos. PRJNA732096 and PRJNA796358.
595
596

597 **Table 1: Hierarchical genome assembly statistics for (a) MC and (b) SI genomes**

598 **a)**

| Assembly | Size (Gb) | # of contigs | # of scaffolds | Contig/scaffold N50 (Mb) | Longest scaffold (Mb) |
|---|---|---|---|---|---|
| **PanTigT.MC.v1 (LR + SR)** | 2.41 | 2,766 | - | 3.26 | 21.20 |
| **PanTigT.MC.v2 (v1 + Hi-C)** | 2.41 | - | 1,054 | 145.47 | 237.64 |

599

600 **b)**

| Assembly | Size (Gb) | # of contigs | # of scaffolds | Contig/scaffold N50 (Mb) | Longest scaffold (Mb) |
|---|---|---|---|---|---|
| **PanTigT.SI.v1 (LR + SR)** | 2.38 | 845 | - | 29 | 131.97 |
| **PanTigT.SI.v2 (v1 + Chicago)** | 2.53 | - | 6,889 | 2.10 | 29.57 |
| **PanTigT.SI_BNG (BNG)** | 2.63 | - | 182 | 145.93 | - |
| **PanTigT.SI.v3 (v2 + BNG)** | 2.53 | - | 3,811 | 141.30 | 238.76 |
| **ONT+Chicago+Bionano+Hi-C** | 2.53 | - | 3,811 | 141.30 | 238.76 |

601

602

603 **Table 2: Genome assembly statistics**

| Assembly | Machali | SI |
|---|---|---|
| Total sequence length (Gb) | 2.41 | 2.53 |
| Contig N50 (Mb) | 3.27 | 29.00 |
| Number of scaffolds | 1054 | 3,811 |
| Number of scaffolds > 10 Mb (% of assembly) | 19 (94.1) | 19 (93.3) |
| Largest scaffold (Mb) | 237.6 | 238.8 |
| Scaffold N50 (Mb) | 145.47 | 147.0 |
| Number of gaps | 1700 | 5708 |
| BUSCO (%) | 95.4 | 92.6 |

604

605 **Table 3: Comparison of MC and SI assemblies to other Felid genomes**

606

| **Assembly** | **Domestic cat** | **Leopard** | **Jaguar** | **Lion** | **Amur Tiger** | **Malayan Tiger** | **Bengal Tiger ♀ (PanTigT.MC.v3)** | **Bengal Tiger ♂ (PanTigT.SI.v3)** |
|---|---|---|---|---|---|---|---|---|
| Total sequence length (Gb) | 2.52 | 2.57 | 2.50 | 2.50 | 2.42 | 2.42 | 2.41 | 2.53 |
| Contig N50 (Mb) | 41.91 | 0.02 | 0.06 | 0.31 | 0.03 | 0.102 | 3.27 | 29.00 |
| #of scaffolds | 4,525 | 50,377 | 438,792 | 8,061 | 1,479 | 10,077 | 1054 | 3,811 |
| #of scaffolds > 10 Mb (% of assembly) | 19 (96.4) | 84 (75.8) | 0 | 18 (99.2) | 65 (40.9) | 77 (80.0) | 19 (94.1) | 19 (93.3) |
| Largest scaffold (Mb) | 242.1 | 84.1 | 1.4. | 238.9 | 41.6 | 79.7 | 237.6 | 238.8 |
| Scaffold N50 (Mb) | 83.8 | 21.7 | 116.0 | 136.0 | 8.8 | 21.3 | 145.5 | 147.0 |
| #of chromosomes (n) | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 19 |
| Reference | (Buckley et al., 2020) | (Kim et al., 2016) | (Figueiro et al., 2017) | (Armstrong et al., 2020) | (Cho et al., 2013) | (Armstrong et al., 2021) | This study | This study |

607

608 **Figure legends**

609

610 **Figure 1: Bengal Tiger Genome assembly.** (**a-b**) Circos plot representing the near-
611 chromosome-level assembly of female (MC) **(a)** and male (SI) (**b**) Bengal tiger genomes. The
612 two outermost tracks represent the chromosome length (Mb) and ID. Repeat content per 100
613 kb window per chromosome (red line represents the mean genome-wide mean repeat content
614 ~36 %) is shown in black. Gene density (green bars indicate 1-5 genes per 100 kb window per
615 chromosome, orange 5-10, and red > 10 genes) is displayed next. The innermost track
616 represents GC% per 100 kb window per chromosome (blue line represents mean GC% -
617 ~41%). (**c**) Y-linked scaffolds from the male tiger genome assembly. Circos plot track as in (a-
618 b). (**d**) Chromosome painting showing synteny between domestic cat genome (FelCat9.0;
619 Genbank: GCA_000181335.4) and the female Bengal tiger (MC) scaffolds. (**e**) Genomic
620 inbreeding coefficients ($F_{ROH}$) derived from runs of homozygosity (ROH) of different lengths
621 (>100 Kb, >1 Mb and >10 Mb) and, (**f**) box plot of binned distribution of ROH lengths in the four
622 zoo-bred tigers as derived when using the MC, SI and Malayan tiger genomes as the reference
623 genome.

624

625 **Figure 2: Genome Annotation and functional genes.** (**a**), Bar plot of the number of predicted
626 genes and corresponding transcripts observed in MC genome assembly and putative Y-linked
627 scaffolds in SI genome assembly. (**b**) Putative genes involved in various key biological
628 pathways important for apex predatory traints. (**c**) Complete *Slc45a* gene structure including the
629 predicted transmembrane (TM) domains. Multiple sequence alignment of Scl45a c-terminal
630 region showing the A477>V mutation observed in white tiger (**d**) Schematic diagram of full-
631 length *Corin* gene. Multi-species alignment of the Corin LDL receptor domain depicting the
632 known H>Y mutation associated with golden tabby pelage pattern (*Cell Res. 27(7):954–57*).

633

634 **Figure 3: Phylogeny and positive selection analysis.** (**a**) Phylogeny tree constructed using
635 single copy orthologs from selected Panthera and outgroup species. **(b)** Pathway analysis of
636 single copy orthologs within the Panthera lineage showing gene network interaction models. **(c)**
637 Scatter plot of evolutionary rates of single copy key pattern and pelage-related gene orthologs
638 identified in the Bengal tiger genome. cfa: *Canis familiaris* (dog), pts: south Indian Bengal tiger
639 (male), ptm: Machali Bengal tigress (female), pta: *Panthera tigris altaica* (Amur tiger), ple:
640 *Panthera leo* (lion), ppa: *Panthera pardus* (leopard), pon: *Panthera onca* (jaguar), nne: *Neofelis*
641 *nebulosa* (clouded leopard), fca: *Felis catus* (domestic cat), lca: *Lynx canadensis* (Canada lynx),
642 pco: *Puma concolor* (cougar), eca: *Equus caballus* (domesticated horse), ocu: *Oryctolagus*
643 *cuniculus* (rabbit)

## References


Ahnelt, P. K., Schubert, C., Kubber-Heiss, A., Schiviz, A., and Anger, E. (2006). Independent variation of retinal S and M cone photoreceptor topographies: A survey of four families of mammals. Vis Neurosci *23*, 429-435.

Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350-355.

Aristotle, Forster, E. S., and Peck, A. L. (1937). Parts of animals Movement of animals ; Progression of animals / Aristotle ; with an English translation by A.L. Peck and E.S. Forster, (Cambridge, MA: Harvard University Press).

Armstrong, E. E., Khan, A., Taylor, R. W., Gouy, A., Greenbaum, G., Thiery, A., Kang, J. T., Redondo, S. A., Prost, S., Barsh, G.*, et al.* (2021). Recent Evolutionary History of Tigers Highlights Contrasting Roles of Genetic Drift and Selection. Mol Biol Evol *38*, 2366-2379.

Bachmanov, A. A., Bosak, N. P., Lin, C., Matsumoto, I., Ohmoto, M., Reed, D. R., and Nelson, T. M. (2014). Genetics of taste receptors. Curr Pharm Des *20*, 2669-2683.

Bareil, C., Hamel, C. P., Delague, V., Arnaud, B., Demaille, J., and Claustres, M. (2001). Segregation of a mutation in CNGB1 encoding the beta-subunit of the rod cGMP-gated channel in a family with autosomal recessive retinitis pigmentosa. Hum Genet *108*, 328-334.

Barnett, R., Westbury, M. V., Sandoval-Velasco, M., Vieira, F. G., Jeon, S., Zazula, G., Martin, M. D., Ho, S. Y. W., Mather, N., Gopalakrishnan, S.*, et al.* (2020). Genomic Adaptations and Evolutionary History of the Extinct Scimitar-Toothed Cat, Homotherium latidens. Curr Biol *30*, 5018-5025 e5015.

Buckley, R. M., Davis, B. W., Brashear, W. A., Farias, F. H. G., Kuroki, K., Graves, T., Hillier, L. W., Kremitzki, M., Li, G., Middleton, R. P.*, et al.* (2020). A new domestic cat genome assembly based on long sequence reads empowers feline genomic medicine and identifies a novel gene for dwarfism. PLoS Genet *16*, e1008926.

Cabanettes, F., and Klopp, C. (2018). D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ *6*, e4958.

Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. Curr Protoc Bioinformatics *48*, 4 11 11-39.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res *18*, 188-196.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., and Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet *19*, 220-234.

Chandrashekar, J., Mueller, K. L., Hoon, M. A., Adler, E., Feng, L., Guo, W., Zuker, C. S., and Ryba, N. J. (2000). T2Rs function as bitter taste receptors. Cell *100*, 703-711.

Cho, Y. S., Hu, L., Hou, H., Lee, H., Xu, J., Kwon, S., Oh, S., Kim, H. M., Jho, S., Kim, S.*, et al.* (2013). The tiger genome and comparative analysis with lion and snow leopard genomes. Nat Commun *4*, 2433.

Davis, A. P., and Capecchi, M. R. (1996). A mutational analysis of the 5' HoxD genes: dissection of genetic interactions during limb development in the mouse. Development *122*, 1175-1185.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., and Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92-95.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.

Fan, H., Wu, Q., Wei, F., Yang, F., Ng, B. L., and Hu, Y. (2019). Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. Genome Biol *20*, 267.

Figueiro, H. V., Li, G., Trindade, F. J., Assis, J., Pais, F., Fernandes, G., Santos, S. H. D., Hughes, G. M., Komissarov, A., Antunes, A.*, et al.* (2017). Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci Adv *3*, e1700299.

Gandolfi, B., Alhaddad, H., Joslin, S. E., Khan, R., Filler, S., Brem, G., and Lyons, L. A. (2013). A splice variant in KRT71 is associated with curly coat phenotype of Selkirk Rex cats. Sci Rep *3*, 2000.

Ge, H., Lin, K., Shen, M., Wu, S., Wang, Y., Zhang, Z., Wang, Z., Zhang, Y., Huang, Z., Zhou, C.*, et al.* (2019). De novo assembly of a chromosome-level reference genome of red-spotted grouper (Epinephelus akaara) using nanopore sequencing and Hi-C. Mol Ecol Resour *19*, 1461-1469.
Giani, A. M., Gallo, G. R., Gianfranceschi, L., and Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. Comput Struct Biotechnol J *18*, 9-19.
Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.*, et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644-652.
Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alfoldi, J., Di Palma, F., and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics *26*, 1145-1151.
Hamrick, M. W. (2001). Development and evolution of the mammalian limb: adaptive diversification of nails, hooves, and claws. Evol Dev *3*, 355-363.
Hasin-Brumshtein, Y., Lancet, D., and Olender, T. (2009). Human olfaction: from genomic variation to phenotypic diversity. Trends Genet *25*, 178-184.
Ishida, Y., David, V. A., Eizirik, E., Schaffer, A. A., Neelam, B. A., Roelke, M. E., Hannah, S. S., O'Brien, S. J., and Menotti-Raymond, M. (2006). A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. Genomics *88*, 698-705.
Jhala, Y. V., Qureshi, Q., and Nayak, A. K. (2020). Status of tigers, copredators and prey in India, 2018. . In, (National Tiger Conservation Authority, Government of India, New Delhi, and Wildlife Institute of India, Dehradun.).
Jiang, P., Josue, J., Li, X., Glaser, D., Li, W., Brand, J. G., Margolskee, R. F., Reed, D. R., and Beauchamp, G. K. (2012). Major taste loss in carnivorous mammals. Proc Natl Acad Sci U S A *109*, 4956-4961.
Johnson, D., Kan, S. H., Oldridge, M., Trembath, R. C., Roche, P., Esnouf, R. M., Giele, H., and Wilkie, A. O. (2003). Missense mutations in the homeodomain of HOXD13 are associated with brachydactyly types D and E. Am J Hum Genet *72*, 984-997.
Kaelin, C. B., Xu, X., Hong, L. Z., David, V. A., McGowan, K. A., Schmidt-Kuntzel, A., Roelke, M. E., Pino, J., Pontius, J., Cooper, G. M.*, et al.* (2012). Specifying and sustaining pigmentation patterns in domestic and wild cats. Science *337*, 1536-1541.
Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., and Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res *46*, D335-D342.
Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., and Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. Evol Appl *9*, 1205-1218.
Katagiri, S., Akahori, M., Sergeev, Y., Yoshitake, K., Ikeo, K., Furuno, M., Hayashi, T., Kondo, M., Ueno, S., Tsunoda, K.*, et al.* (2014). Whole exome analysis identifies frequent CNGA1 mutations in Japanese population with autosomal recessive retinitis pigmentosa. PLoS One *9*, e108721.
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? Ecol Evol *10*, 4583-4594.
Khan, A., Patel, K., Shukla, S., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y., Kardos, M., and Ramakrishnan, U. (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. bioRxiv.
Kim, S., Cho, Y. S., Kim, H. M., Chung, O., Kim, H., Jho, S., Seomun, H., Kim, J., Bang, W. Y., Kim, C.*, et al.* (2016). Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. Genome Biol *17*, 211.
Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics *5*, 59.
Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009a). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Li, X., Glaser, D., Li, W., Johnson, W. E., O'Brien, S. J., Beauchamp, G. K., and Brand, J. G. (2009b). Analyses of sweet receptor gene (Tas1r2) and preference for sweet stimuli in species of Carnivora. J Hered *100 Suppl 1*, S90-100.
Li, X., Li, W., Wang, H., Cao, J., Maehashi, K., Huang, L., Bachmanov, A. A., Reed, D. R., Legrand-Defretin, V., Beauchamp, G. K., and Brand, J. G. (2005). Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. PLoS Genet *1*, 27-35.
Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.
Liu, Y. C., Sun, X., Driscoll, C., Miquelle, D. G., Xu, X., Martelli, P., Uphyrkina, O., Smith, J. L. D., O'Brien, S. J., and Luo, S. J. (2018). Genome-Wide Evolutionary Analysis of Natural History and Adaptation in the World's Tigers. Curr Biol *28*, 3840-3849 e3846.
Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res *25*, 955-964.
Luo, S. J., Liu, Y. C., and Xu, X. (2019). Tigers of the World: Genomics and Conservation. Annu Rev Anim Biosci *7*, 521-548.
Lyons, L. A., Foe, I. T., Rah, H. C., and Grahn, R. A. (2005). Chocolate coated cats: TYRP1 mutations for brown color in domestic cats. Mamm Genome *16*, 356-366.
Mondol, S., Karanth, K. U., and Ramakrishnan, U. (2009). Why the Indian subcontinent holds the key to global tiger recovery. PLoS Genet *5*, e1000585.
Montague, M. J., Li, G., Gandolfi, B., Khan, R., Aken, B. L., Searle, S. M., Minx, P., Hillier, L. W., Koboldt, D. C., Davis, B. W.*, et al.* (2014). Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. Proc Natl Acad Sci U S A *111*, 17230-17235.
Murchison, E. P., Schulz-Trieglaff, O. B., Ning, Z., Alexandrov, L. B., Bauer, M. J., Fu, B., Hims, M., Ding, Z., Ivakhno, S., Stewart, C.*, et al.* (2012). Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell *148*, 780-791.
Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933-2935.
Ohsawa, R., and Kageyama, R. (2008). Regulation of retinal cell fate specification by multiple transcription factors. Brain Res *1192*, 90-98.
Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., and Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet *91*, 275-292.
Phillippy, A. M. (2017). New advances in sequence assembly. Genome Res *27*, xi-xiii.
Poddar-Sarkar, M., and Brahmachary, R. L. (2014). Pheromones of Tiger and Other Big Cats. In Neurobiology of Chemical Communication, C. Mucignat-Caretta, ed. (Boca Raton (FL)).
Pontius, J. U., Mullikin, J. C., Smith, D. R., Agencourt Sequencing, T., Lindblad-Toh, K., Gnerre, S., Clamp, M., Chang, J., Stephens, R., Neelam, B.*, et al.* (2007). Initial sequence and comparative analysis of the cat genome. Genome Res *17*, 1675-1689.
Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W.*, et al.* (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res *26*, 342-350.
Ripple, W. J., Estes, J. A., Beschta, R. L., Wilmers, C. C., Ritchie, E. G., Hebblewhite, M., Berger, J., Elmhagen, B., Letnic, M., Nelson, M. P.*, et al.* (2014). Status and ecological effects of the world's largest carnivores. Science *343*, 1241484.
Roots, C. (2006). Nocturnal animals, (Westport, Conn.: Greenwood Press).

Sagar, V., Kaelin, C. B., Natesh, M., Reddy, P. A., Mohapatra, R. K., Chhattani, H., Thatte, P., Vaidyanathan, S., Biswas, S., Bhatt, S.*, et al.* (2021). High frequency of an otherwise rare phenotype in a small and isolated tiger population. Proc Natl Acad Sci U S A *118*.

Sharpe, P. T. (1995). Homeobox genes and orofacial development. Connect Tissue Res *32*, 17-25.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210-3212.

Simpson, G. G. (1941). The function of saber-like canines in carnivorous mammals, Vol 1130, (New York City: The American Museum of Natural History).

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res *33*, W465-467.

Suryamohan, K., Krishnankutty, S. P., Guillory, J., Jevit, M., Schroder, M. S., Wu, M., Kuriakose, B., Mathew, O. K., Perumal, R. C., Koludarov, I.*, et al.* (2020). The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat Genet *52*, 106-117.

Sutton, J. T., Helmkampf, M., Steiner, C. C., Bellinger, M. R., Korlach, J., Hall, R., Baybayan, P., Muehling, J., Gu, J., Kingan, S.*, et al.* (2018). A High-Quality, Long-Read De Novo Genome Assembly to Aid Conservation of Hawaii's Last Remaining Crow Species. Genes (Basel) *9*.

Task Force, I. B. f. W. L. (1972). Project Tiger: A Proposal for Preservation of the Tiger (Panthera tigris tigris Linn.) in India. In, (New Delhi, India: Ministry of Agriculture, Government of India), p. p4.

Thomas, B. L., and Sharpe, P. T. (1998). Patterning of the murine dentition by homeobox genes. Eur J Oral Sci *106 Suppl 1*, 48-54.

Thomas, B. L., Tucker, A. S., Qui, M., Ferguson, C. A., Hardcastle, Z., Rubenstein, J. L., and Sharpe, P. T. (1997). Role of Dlx-1 and Dlx-2 genes in patterning of the murine dentition. Development *124*, 4811-4818.

Tummala, H., Ali, M., Getty, P., Hocking, P. M., Burt, D. W., Inglehearn, C. F., and Lester, D. H. (2006). Mutation in the guanine nucleotide-binding protein beta-3 causes retinal degeneration and embryonic mortality in chickens. Invest Ophthalmol Vis Sci *47*, 4714-4718.

Vastardis, H., Karimbux, N., Guthua, S. W., Seidman, J. G., and Seidman, C. E. (1996). A human MSX1 homeodomain missense mutation causes selective tooth agenesis. Nat Genet *13*, 417-421.

Vickaryous, M. K., and Sire, J. Y. (2009). The integumentary skeleton of tetrapods: origin, evolution, and development. J Anat *214*, 441-464.

Vincent, A., Audo, I., Tavares, E., Maynes, J. T., Tumber, A., Wright, T., Li, S., Michiels, C., Consortium, G. N. B., Condroyer, C.*, et al.* (2016). Biallelic Mutations in GNB3 Cause a Unique Form of Autosomal-Recessive Congenital Stationary Night Blindness. Am J Hum Genet *98*, 1011-1019.

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol Biol Evol *35*, 543-548.

Wurster-Hill, D. H., and Gray, C. W. (1973). Giemsa banding patterns in the chromosomes of twelve species of cats (Felidae). Cytogenet Cell Genet *12*, 388-397.

Xu, X., Dong, G. X., Schmidt-Kuntzel, A., Zhang, X. L., Zhuang, Y., Fang, R., Sun, X., Hu, X. S., Zhang, T. Y., Yang, H. D.*, et al.* (2017). The genetics of tiger pelage color variations. Cell Res *27*, 954-957.

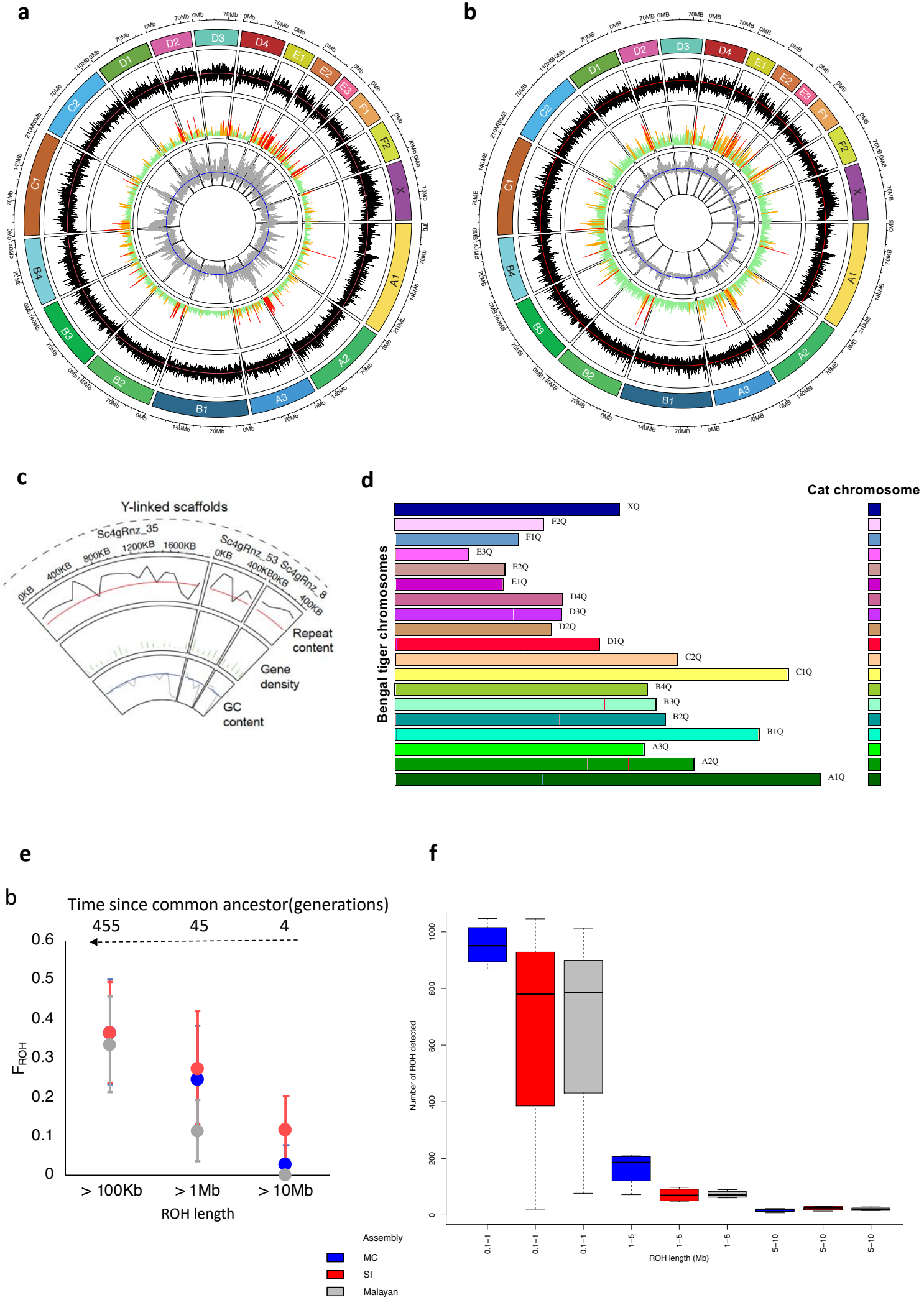
Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*, 1586-1591.

Yang, Z., Wong, W. S., and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol *22*, 1107-1118.

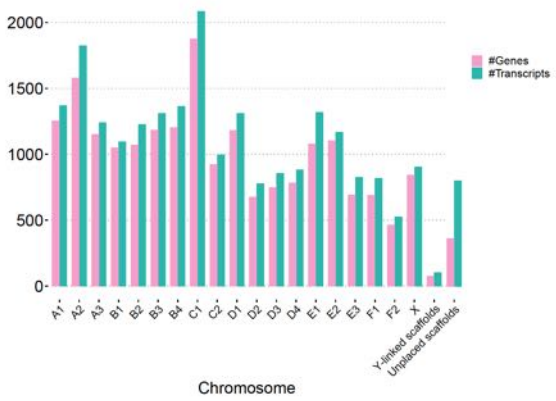
Yokohama-Tamaki, T., Ohshima, H., Fujiwara, N., Takada, Y., Ichimori, Y., Wakisaka, S., Ohuchi, H., and Harada, H. (2006). Cessation of Fgf10 signaling, resulting in a defective dental epithelial stem cell compartment, leads to the transition from crown to root formation. Development *133*, 1359-1366.

Young, J. M., Massa, H. F., Hsu, L., and Trask, B. J. (2010). Extreme variability among mammalian V1R gene families. Genome Res *20*, 10-18.

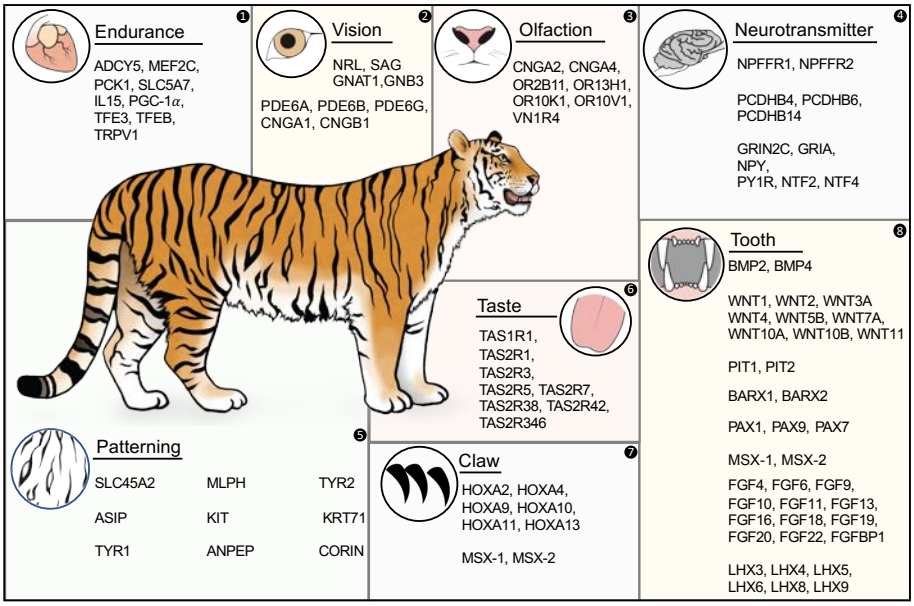
852 Zhang, Y. D., Chen, Z., Song, Y. Q., Liu, C., and Chen, Y. P. (2005). Making a tooth: growth factors,
853 transcription factors, and stem cells. Cell Res *15*, 301-316.
854


a
b
c
Y-linked scaffolds
Sc4gRnz_35
Sc4gRnz_53
Sc4gRnz_8
Repeat content
Gene density
GC content
d
Cat chromosome
Bengal tiger chromosomes
XQ
F2Q
F1Q
E3Q
E2Q
E1Q
D4Q
D3Q
D2Q
D1Q
C2Q
C1Q
B4Q
B3Q
B2Q
B1Q
A3Q
A2Q
A1Q
e
b
Time since common ancestor(generations)
455
45
4
$F_{ROH}$
> 100Kb
> 1Mb
> 10Mb
ROH length
Assembly
MC
SI
Malayan
f
Number of ROH detected
ROH length (Mb)

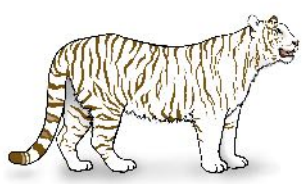
**a**



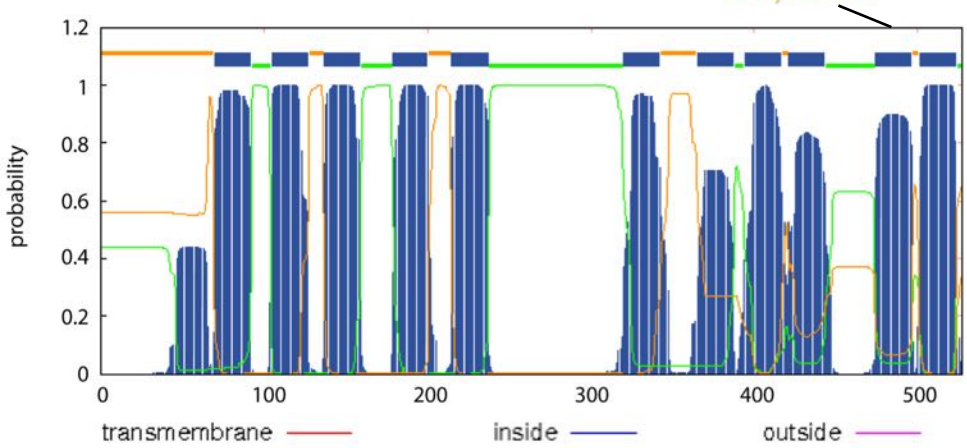

**b**



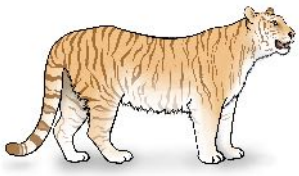

**c**



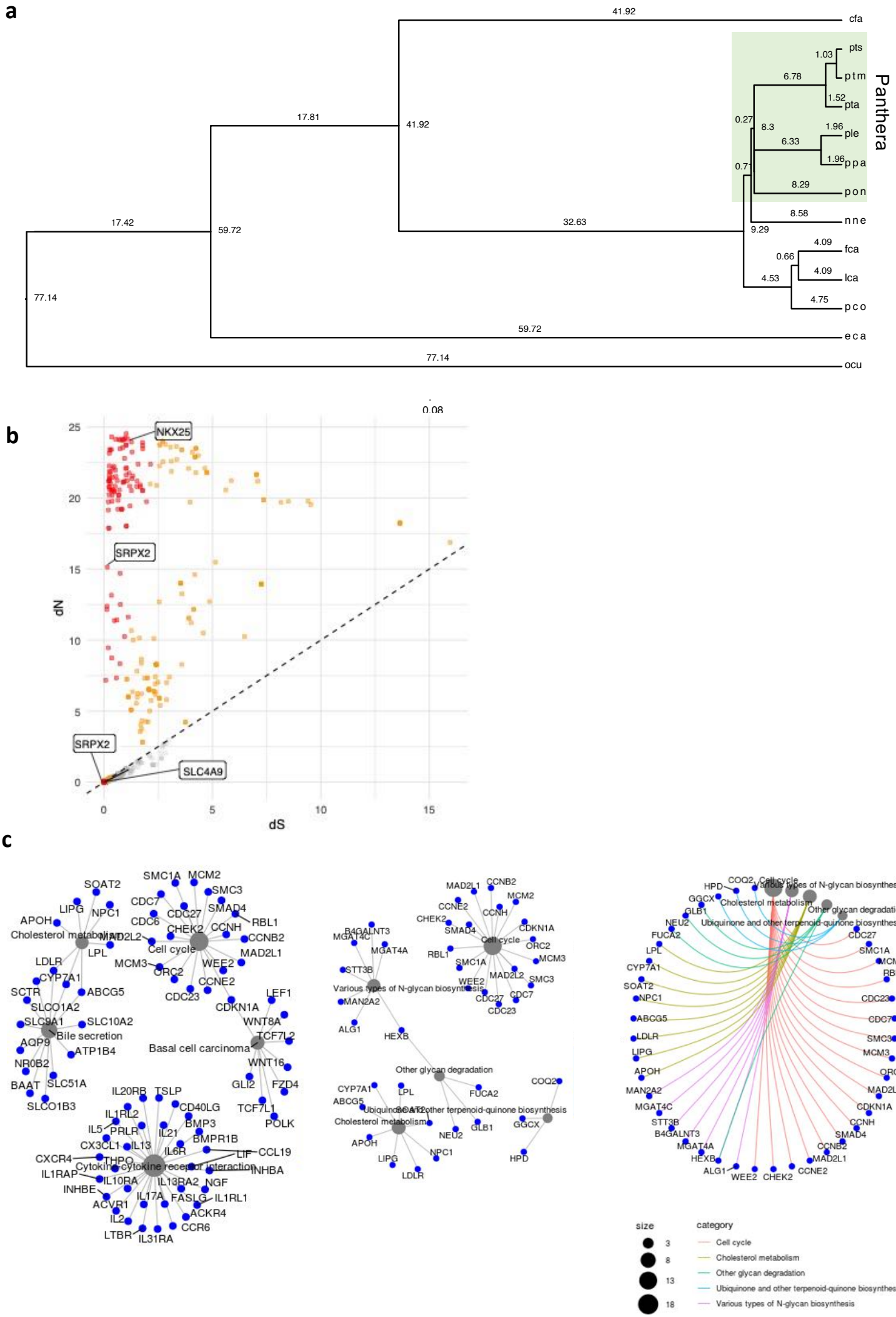
```
DOG/1-599          DCAVLTCMVQ
MOUSE/1-530        DCAALTCMVQ
HUMAN/1-530        DCATLTCMVQ
CAT/1-530          DCAVLTCMVQ
TigerMachli/1-527  DCAVLTCMVQ
TigerSI/1-527      DCAVLTCMVQ
AMUR/1-530         DCAVLTCMVQ
Conservation
```



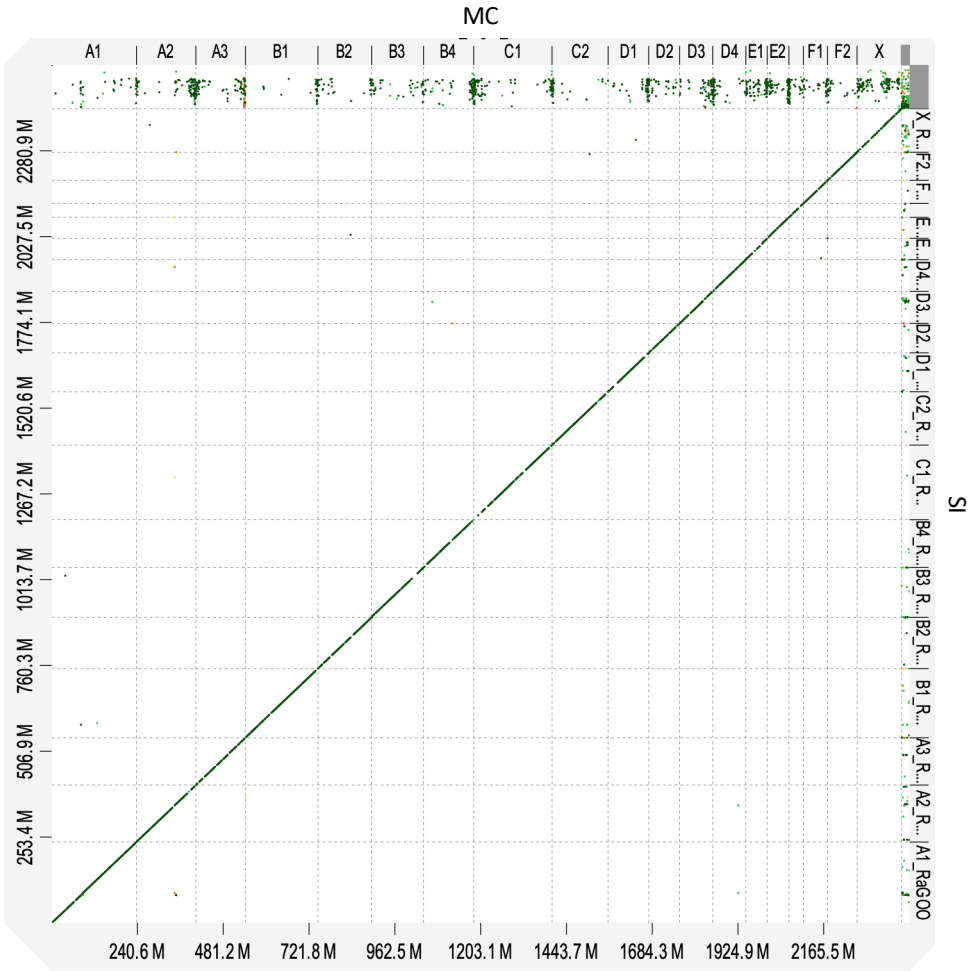

**d**



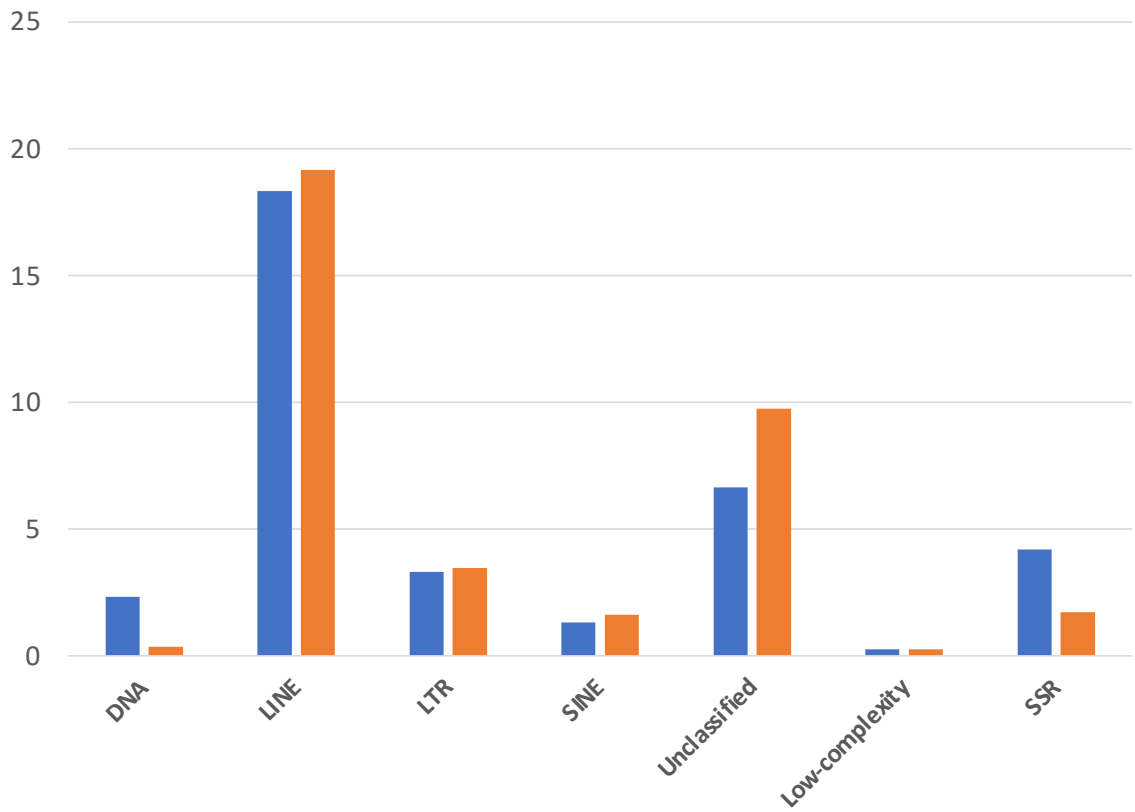
```
                   640       650       660       670       680
MOUSE/1-1113       ESSDNQTCLLPNEDVEECSPSHFKCRSGRCVLGSRRCDGQADCDDDSDEENCGCKERALW
DOG/1-1007         ENSENQTCLMPDEGVEECSPSHFKCGSGRCVLASRRCDGQADCDDDSDEENCGCKERDLW
HUMAN/1-1042       ENSDNQTCLMPDEYVEECSPSHFKCRSGQCVLASRRCDGQADCDDDSDEENCGCKERDLW
TigerMachli/1-1045 ENSDNQTCLMPNEDVEECSPSHFKCGSGRCVLASRRCDGQADCDDDSDEENCGCKERDLW
TigerSI/1-1045     ENSDNQTCLMPNEDVEECSPSHFKCGSGRCVLASRRCDGQADCDDDSDEENCGCKERDLW
AMUR/1-1045        ENSDNQTCLMPNEDVEECSPSHFKCGSGRCVLASRRCDGQADCDDDSDEENCGCKERDLW
CAT/1-1045         ENSDNQTCLMPDEDVEECSPSHFKCGSGRCVLASRRCDGQADCDDDSDEENCGCKERDLW
Conservation
```


a
41.92
cfa
1.03
pts
6.78
ptm
1.52
pta
Panthera
17.81
41.92
0.27
8.3
1.96
ple
6.33
1.96
ppa
0.7
8.29
pon
8.58
nne
17.42
59.72
32.63
9.29
4.09
fca
0.66
4.09
lca
4.53
4.75
pco
77.14
59.72
eca
77.14
ocu
b
0.08
NKX25
SRPX2
SRPX2
SLC4A9
dN
dS
c
Cholesterol metabolism
Cell cycle
Bile secretion
Basal cell carcinoma
Cytokine-cytokine receptor interaction
Various types of N-glycan biosynthesis
Other glycan degradation
Ubiquinone and other terpenoid-quinone biosynthesis
size
category
Cell cycle
Cholesterol metabolism
Other glycan degradation
Ubiquinone and other terpenoid-quinone biosynthesis
Various types of N-glycan biosynthesis

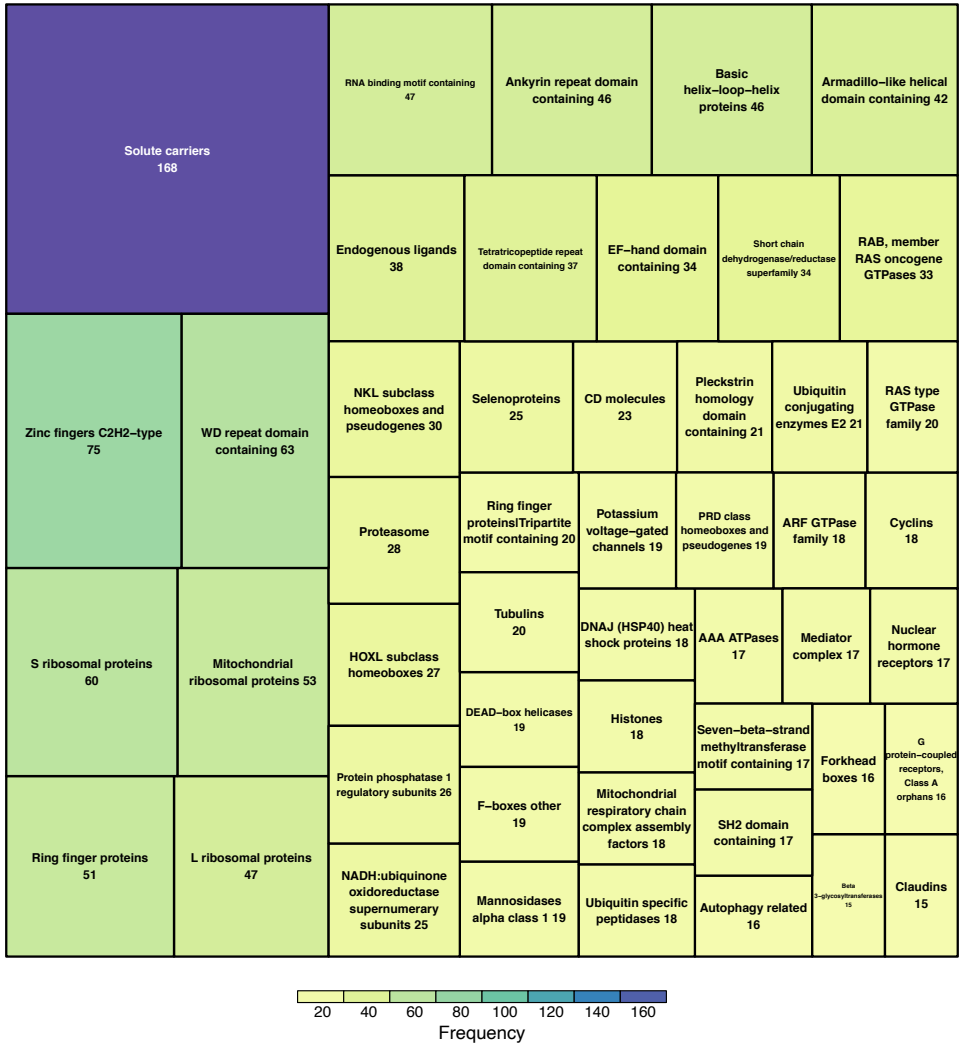
# Supplementary Figures




**Supplementary Figure 1:** Synteny between MC and SI tiger genome assemblies




**Supplementary Figure 2.** Repeat element distribution in MC (blue bars) and SI (orange bars) genomes. LINE: long interspersed nuclear element; LTR: long terminal repeat; SINE: short interspersed nuclear element, SSR: simple short repeat




**Supplementary Figure 3.** Treemap plot of gene families identified in MC genome