The tiger genome and comparative analysis with lion and snow leopard genomes

Yun Sung Cho1, Li Hu2, Haolong Hou2, Hang Lee3, Jiaohui Xu2, Soowhan Kwon4, Sukhun Oh4, Hak-Min Kim1, Sungwoong Jho1, Sangsoo Kim5, Young-Ah Shin1, Byung Chul Kim1,6, Hyunmin Kim6, Chang-uk Kim1, Shu-Jin Luo7, Warren E. Johnson8, Klaus-Peter Koepfli9, Anne Schmidt-Ku¨ntzel10, Jason A. Turner11, Laurie Marker12, Cindy Harper13, Susan M. Miller13,14, Wilhelm Jacobs15, Laura D. Bertola16, Tae Hyung Kim6, Sunghoon Lee1,6, Qian Zhou2, Hyun-Ju Jung6, Xiao Xu7, Priyvrat Gadhvi1, Pengwei Xu2, Yingqi Xiong2, Yadan Luo2, Shengkai Pan2, Caiyun Gou2, Xiuhui Chu2, Jilin Zhang2, Sanyang Liu2, Jun Ding6, Ying Chen2, Linfeng Yang2, Yulan Yang2, Jiaju He2, Sha Liu2, Junyi Wang2, Chul Hong Kim9, Hwanjong Kwak6, Jong-Soo Kim1, Seungwoo Hwang17, Junsu Ko6, Chang-Bae Kim18, Sangtae Kim19, Damdin Bayarkhagva20, Woon Kee Paek21, Seung-Jin Kim6,22, Stephen J. O’Brien9,23, Jun Wang2,24,25 & Jong Bhak1,6,26,27

Tigers and their close relatives (Panthera) are some of the world’s most endangered species. Here we report the de novo assembly of an Amur tiger whole-genome sequence as well as the genomic sequences of a white Bengal tiger, African lion, white African lion and snow leopard. Through comparative genetic analyses of these genomes, we find genetic signatures that may reflect molecular adaptations consistent with the big cats’ hypercarnivorous diet and muscle strength. We report a snow leopard-specific genetic determinant in EGLN1 (Met39 > Lys39), which is likely to be associated with adaptation to high altitude. We also detect a TYR260G > A mutation likely responsible for the white lion coat colour. Tiger and cat genomes show similar repeat composition and an appreciably conserved synteny. Genomic data from the five big cats provide an invaluable resource for resolving easily identifiable phenotypes evident in very close, but distinct, species.

DOI: 10.1038/ncomms3433

1 Personal Genomics Institute, Genome Research Foundation, Suwon 443-270, Republic of Korea. 2 BGI-Shenzhen, Shenzhen 518083, China. 3 Tiger and Leopard Conservation Fund in Korea and College of Veterinary Medicine, Seoul National University, Seoul 151-742, Republic of Korea. 4 Samsung Everland Zoo, Yongin 449-715, Republic of Korea. 5 School of Systems Biomedical Science, Soongsil University, Seoul 156-743, Korea. 6 Theragen BiO Institute, TheragenEtex, Suwon 443-270, Republic of Korea. 7 Peking-Tsinghua Center for Life Sciences, School of Life Sciences, Peking University, Beijing 100871, China. 8 Smithsonian Conservation Biology Institute, National Zoological Park 1500 Remount Road, Front Royal, Virginia 22630, USA. 9 Theodosius Dobzhansky Center for Genome Bioinformatics, St Petersburg State University, St Petersburg 199034, Russia. 10 Life Technologies Conservation Genetics Laboratory, Cheetah Conservation Fund, Otjiwarongo 9000, Namibia. 11 Global White Lion Protection Trust, Tsau Conservancy, Greater Timbavati Region, PO Box 858, Limpopo 1380, South Africa. 12 Cheetah Conservation Fund, Otjiwarongo 9000, Namibia. 13 Veterinary Genetics Laboratory, Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, South Africa. 14 Department of Nature Conservation, Tshwane University of Technology, Private Bag X680, Pretoria 0001, South Africa. 15 Ukutula Lodge & Lion Centre, NorthWest Province Brits 0250, South Africa. 16 Institute of Environmental Sciences (CML), Leiden University, PO Box 9518, Leiden RA 2300, The Netherlands. 17 Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Republic of Korea. 18 Department of Life, Sangmyung University, Seoul 110-743, Republic of Korea. 19 Department of Biology, Sungshin Women’s University, Seoul 142-732, Republic of Korea. 20 Department of Molecular Biology and Genetics, School of Biology and Biotechnology, National University of Mongolia, Ulaanbaatar 210646A, Mongolia. 21 National Science Museum, Daejeon 305-705, Republic of Korea. 22 CHA Cancer Institute, CHA University, Seoul 135-081, Republic of Korea. 23 Oceanographic Center, Nova Southeastern University, Ft Lauderdale, Florida 33304, USA. 24 Department of Biology, University of Copenhagen DK-2200, Copenhagen, Denmark. 25 King Abdulaziz University 21589 Jeddah, Saudi Arabia. 26 Advanced Institutes of Convergence Technology, Nano Science and Technology, Suwon 443-270, Republic of Korea. 27 Program in Nano Science and Technology, Department of Transdisciplinary Studies, Seoul National University, Suwon 443-270, Republic of Korea. Correspondence and requests for materials should be addressed to S.J.O’B. (email: lgdchief@gmail.com) or to J.W. (email: wangj@genomics.cn) or to J.B. (email: jongbhak@genomics.org).
The tiger (*Panthera tigris*), the largest felid species on Earth and a widely recognized symbol for wildlife conservation, is one of the world’s most endangered species. Tigers are a keystone species and natural indicators of the health of the ecological communities in which they are found. The current estimates of wild tigers range from just 3,050 to 3,950 individuals. It is postulated that without conservation measures tigers will soon become extinct in the wild, thus turning the preservation efforts into a major goal of conservation efforts. Tigers comprise of nine genetically validated subspecies from the existing wild tiger populations into a major goal of conservation. It is postulated that without conservation measures tigers will soon become extinct in the wild, thus turning the preservation efforts into a major goal of conservation efforts. Previous genetic studies using mitochondrial and nuclear loci have helped to elucidate the phylogenetics and population genetics of tigers, and the low coverage genome of the domestic cat (*Felis catus*) has provided insights into felid evolution. However, no whole-genome reference sequence has been reported for the tiger, or any of the existing wild tiger populations into a major goal of conservation efforts. Tigers comprise of nine genetically validated subspecies from the existing wild tiger populations into a major goal of conservation efforts. Previous genetic studies using mitochondrial and nuclear loci have helped to elucidate the phylogenetics and population genetics of tigers, and the low coverage genome of the domestic cat (*Felis catus*) has provided insights into felid evolution. However, no whole-genome reference sequence has been reported for the tiger, or any of the *Panthera* species, thus limiting current understanding of genetic diversity and demography.

We report the first tiger genome sequence assembly and annotation as well as a comparative analysis of the lion (*Panthera leo*) and snow leopard (*Panthera uncia*) genomes. We describe genotypic variation and genotype association analyses with species-specific phenotypes and adaptation. *Panthera* whole-genome sequences provide valuable information on genome organization, evolutionary divergence and overall endemic diversity.

### Results

**The Amur tiger genome.** The DNA of a 9-year-old male Amur tiger from Everland Zoo in Korea was sequenced by Illumina HiSeq2000 (Supplementary Fig. S1, Supplementary Tables S1–S3). Sequence reads were assembled using SOAPdenovo into scaffolds (2.4 Gb in length) having an N50 length of 8.84 Mb (contig N50 length of 29.8 kb; Table 1, Supplementary Figs S2–S4, Supplementary Tables S4 and S5, Methods). Assembly quality was assessed by aligning the assembled tiger blood transcripts and EST sequences onto the tiger scaffolds (>96% coverage and 98.9% mapping rate, respectively), and heterozygous single nucleotide variants (SNVs) were validated by the Sanger method (Supplementary Tables S6–S9, Supplementary Methods). Additionally, analysis of the tiger draft genome assembly for core eukaryotic genes revealed homologues for >93.4% of conserved genes in the assembly (Supplementary Table S10). The tiger genome sequence shows 95.6% similarity to the domestic cat (Supplementary Table S11) from which it diverged approximately 10.8 million years ago (MYA); human and gorilla have 94.8% similarity and diverged around 8.8 MYA (from *Tree*). This high similarity allowed us to improve the assembly of the tiger genome by using the recently completed high coverage (12× coverage) domestic cat genome (Supplementary Fig. S5, Supplementary Tables S12–S17, Supplementary Methods). For comparative genomic analysis of big cats, we also sequenced four other *Panthera* genomes (Table 1, Supplementary Tables S1 and S18): a white Bengal tiger (*Panthera tigris tigris*), an African lion, a white African lion and a snow leopard; their genome sequences were aligned with the domestic cat and tiger genomes.

### Adaptation of the big cats

The assembled Amur tiger genome was predicted to contain 20,226 protein-coding genes (Supplementary Tables S19–S23, Supplementary Methods) and 2,935 non-coding RNAs (Supplementary Table S24, Supplementary Methods). To create a detailed annotation of the tiger proteome, gene clusters were constructed using seven mammalian genomes (tiger, cat, human, dog, mouse, giant panda and opossum). The tiger proteome contained 14,954 orthologous gene families. Among these, 14,425 orthologous gene families were shared by all seven genomes, whereas 103 orthologous gene families were exclusively shared by the tiger and cat (Fig. 1a, Supplementary Fig. S6, Supplementary Table S25). The Felidae-specific gene families contained 287 InterPro domains (Supplementary Tables S26–S29). Based on the comparison of orthologous gene families among seven mammalian species, the Amur tiger genome displays 381 expanded and 1,790 contracted gene families compared with the feline common ancestor (Fig. 1b, Supplementary Figs S7 and S8). The tiger genome is particularly enriched in olfactory receptor activity (GO:0004984, *P* = 5.75 × 10−185, ChiSquare test followed by a Fisher’s exact test, 289 genes), G-protein coupled receptor signalling pathway (GO:0007186, *P* = 2.98 × 10−106, 302 genes), signal transducer activity (GO:0004871, *P* = 2.25 × 10−124, 295 genes), amino-acid transport (GO:0006865, *P* = 3.09 × 10−6, 16 genes) and protein metabolic process (GO:0019538, *P* = 5.72 × 10−122, 220 genes) (Supplementary Fig. S9, Supplementary Table S30). In most cats, smell has an important role in social behaviour such as territory ownership and mating, while vision and hearing are important for hunting.

Additionally, we investigated *Panthera* lineage-specific amino-acid changes by comparison with the known genes from the human, dog and mouse. A total of 3,646 genes had amino-acid changes specific to big cats (Amur tiger, white tiger, snow leopard, African lion and white lion; Supplementary Tables S31 and S32), and 5,882 genes had amino-acid changes unique to the feline lineage (big cats plus domestic cat). Among these, 1,376 genes had big cat-specific amino-acid changes that were found to be protein functional changes according to computational predictions (PolyPhen2). Metabolism pathways associated with protein and fatty acid, which are important sources of energy, were enriched with genes having *Panthera*-specific functional changes; histidine metabolism (*P* = 0.00024, Fisher’s significant).

### Table 1 | Global statistics of the Panthera genomes.

| Sequencing (species) | Insert size | Total data (Gb) | Sequence coverage (×) |
|----------------------|-------------|-----------------|-----------------------|
| Amur tiger           | 170, 500, 800 bp | 203.72          | 83.5                  |
|                      | 2, 5, 10, 20 kb | 84.48           | 34.6                  |
| White tiger          | 400 bp       | 86.35           | 32.1                  |
| Snow leopard         | 400 bp       | 108.94          | 40.5                  |
| African lion         | 400 bp       | 98.47           | 36.6                  |
| White lion           | 400 bp       | 84.43           | 31.4                  |

| Amur tiger assembly  | N50 (kb) | Longest (kb) | Size (Gb) |
|----------------------|----------|--------------|-----------|
| Contig               | 29.8     | 287          | 2.35      |
| Scaffold             | 8,840    | 41,607       | 2.41      |

| Amur tiger annotation| Number | Total length (Mb) | Percentage of genome |
|----------------------|--------|-------------------|----------------------|
| Genes                | 20,226 | 718.9             | 29.5                 |
| Repeats              | —      | 958.9             | 39.3                 |

The statistics were based on Amur tiger genome size (2.44 Gb), estimated by K-mer analysis. Contigs and scaffolds above 100 bp length were included in the statistics.
are evolving under significantly high constraints (Supplementary Table S33, Supplementary Data 2). We also identified GO categories, which were enriched for myosin heavy chain (MYH7) and taurine-conjugated putrescine 4 (TPM4) and S34). These signals of amino-acid metabolism have been found in the human genome, but not in other mammalian species. Although the tiger had unique amino-acid changes in both genes that were positively charged, was found in the snow leopard (Fig. 2a, Supplementary Fig. S10, Supplementary Table S44), indicating a significant change in charge that might alter protein function. This Met39 residue was replicated as a genetically fixed substitution in 14 additional snow leopards, whereas the ancestral Lys39 was monomorphic in a sampling of 28 individuals of Panthera and Neofelis (clouded leopard) (Supplementary Table S45, Supplementary Methods). Naked mole rats have also adapted to hypoxia in a different manner by having unique amino-acid changes in different positions of EGLN1 (Pro15, Arg17 and Arg36). Ile663 and Arg794 in EPAS1 are two additional snow leopard-specific changes (Supplementary Fig. S11), and Arg794 was predicted to confer a functional change on the protein. Taken together, these EGLN1 and possibly EPAS1 variants are provocative candidates that may have contributed to the snow leopard’s acquisition of an alpine, high altitude ecological niche.

Tyrosinase (TYR) mutational variants cause white coat colour in the domestic cat, and TYR mutations are related to human oculocutaneous albinism 1 (ref 24). The genetic basis of the white tiger with white fur and dark stripes is revealed as an amino-acid change (A477V) in the transporter protein SLC45A2 (ref. 25). Therefore, we examined the pigment-associated gene EPAS1 variants are provocative candidates that may have contributed to the snow leopard’s acquisition of an alpine, high altitude ecological niche.

Genetic landscape of the snow leopard and white lion. In addition to the Amur tiger data, we used sequence data from the four other big cats to investigate the genetic basis of several unique physiological or phenotypic traits. Snow leopards generally live in alpine areas, 3,350–6,700 m above sea level, in Central Asia. Recent genome-wide association studies implicated two human loci EGLN1 (Egl nine homologue 1) and EPAS1 (endothelial PAS domain-containing protein 1) as mediating high-altitude adaptation. We examined mutational substitutions in mammalian EGLN1 and EPAS1 genes and found that the snow leopard had unique amino-acid changes in both genes that were not found in other mammalian species. Although EGLN1 is highly conserved in mammals, Met39 (non-polar), instead of Lys39 (positively charged), was found in the snow leopard (Fig. 2a, Supplementary Fig. S10, Supplementary Table S44), indicating a significant change in charge that might alter protein function. This Met39 residue was replicated as a genetically fixed substitution in 14 additional snow leopards, whereas the ancestral Lys39 was monomorphic in a sampling of 28 individuals of Panthera and Neofelis (clouded leopard) (Supplementary Table S45, Supplementary Methods). Naked mole rats have also adapted to hypoxia in a different manner by having unique amino-acid changes in different positions of EGLN1 (Pro15, Arg17 and Arg36). Ile663 and Arg794 in EPAS1 are two additional snow leopard-specific changes (Supplementary Fig. S11), and Arg794 was predicted to confer a functional change on the protein. Taken together, these EGLN1 and possibly EPAS1 variants are provocative candidates that may have contributed to the snow leopard’s acquisition of an alpine, high altitude ecological niche.

Genomic comparison between the tiger and other mammals. Although repeat characteristics can vary considerably among closely related species, the tiger and cat genomes showed very similar repeat compositions (39.3% versus 39.2%, respectively), as well as ratios of repeat components, including tandem repeats.
Snow leopard (15)  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Tiger (12)  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Lion (12)  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Leopard (5)  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Clouded leopard (3)  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Dog  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Human  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 53
Naked mole rat  BR2NGKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 66
Mouse  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 207
Rat  LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH 127

Table 1 | Amino-acid sequences unique to white lion, snow leopard and tigers. The numbers in parentheses are number of individuals. (w) denotes white type and (wt) denotes wild type.

| Species          | Amino-acid Sequence | Number of Individuals |
|------------------|---------------------|----------------------|
| White lion       | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 18 |
| Tawny lion       | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 17 |
| White tiger      | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 14 |
| Snow leopard     | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 12 |
| Dog              | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 12 |
| Human            | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 12 |
| Naked mole rat   | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 66 |
| Mouse            | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 207 |
| Rat              | LRPGNQKTPKLALFLYVPCMHKHHGICVDDFLGKETQQI5GDEVRALH | 127 |

Figure 2 | EGLN1 and TYR mutations related to hypoxia in snow leopard and white fur in white lion. (a) Alignment of mammalian EGLN1 amino-acid sequences. Amino acids unique to the snow leopard (216th residue in human EGLN1), naked mole rat and rodents are shown in red, grey and blue, respectively. The number of individuals genotyped in this study is listed in parentheses. (b) Alignment of mammalian TYR sequences. Amino-acid sequences unique to the white lion (87th residue in human TYR) are shown in red, and tawny lion having heterozygous allele (G/A) are shown in grey; X represents amino acid of R/Q. The numbers in parentheses are number of individuals. ‘w’ denotes white type and ‘wt’ denotes wild type.

and transposable elements (Supplementary Fig. S13), suggesting a similar genome architecture between domestic cats and tigers. By contrast, for the great apes, the ratio of repeat components was considerably different between species, especially between human and orangutan, which diverged about 12 MYA. Additionally, we estimated the evolutionarily conserved sequences (77 Mb, 3.2%), segmental duplication (11.2 Mb, 0.47%) and lineage-specific insertions and deletions in the tiger genome (Supplementary Tables S49–S52, Supplementary Methods).

To detect genome-wide structural differences, we aligned the tiger scaffolds to the cat genome using dog genome as a reference after masking repeats. A total of 571 of the 674 tiger scaffolds (length >20 kb, 99.6% of the total scaffold length) were aligned with the cat genome sequence, and as high as 98.8% of gene-coding regions and 98.3% (2.38 Gb) of conserved synteny blocks were shared by the tiger and cat genomes. We detected a rather high level of genomic synteny, containing six breaks with large-size chromosomal segmental rearrangement between the tiger and cat genomes (Fig. 3, Supplementary Fig. S14, Supplementary Tables S53–S56, Methods). These consisted of two inter- and four intra-chromosomal rearrangements. Divergence in genomic structure among closely related species is considered as a major factor underlying species diversification, as gene flow requires recombination in collinear chromosomes, and the reduction in recombination associated with chromosomal rearrangements results in a partial reproductive barrier. These structural variations may be one of the important factors underlying species diversification among the felines.

The level of within-species genetic diversity, as measured by the rate of heterozygous SNVs, of the tiger (0.00049–0.00073) and lion (0.00048–0.00058) genomes was found to be similar to that of human (0.00066) (Supplementary Table S57, Supplementary Methods). Interestingly, diversity of the snow leopard genome was nearly half that of the other Panthera species and slightly lower than that of the Tasmanian devil, which is purported to display a low level of genetic diversity (Fig. 4a). We also estimated the occurrence of a marked bottleneck around the last glacial maximum 20 ky ago (7–70 ky) using a pairwise sequentially Markovian coalescent (PSMC) model inference of tiger demographic history based on SNV distribution (Fig. 4b, Supplementary Figs S15–S18, Supplementary Tables S58 and S59, Methods). A similar bottleneck was estimated a bit earlier (72–108 kyr) based on mitochondrial DNA coalescence. White lion (0.00048) and domestic cat (0.00012) have both undergone multiple rounds of close inbreeding during breed development and therefore would display lower SNV diversity bias as a consequence. Therefore, we investigated the genetic diversity of Panthera using the rate of heterozygous SNVs and confirmed that the genetic diversity within a single individual coincided with that deduced from mitochondrial sequences of several individuals.
Many whole-reference genome studies used few close species genomes that can be directly compared with the reference genome constructed. Although we did not have the resources to construct lion and leopard reference genomes, and hence were not able to show all the structural variations on the genomes, our ‘close species comparative genomics’ approach, utilizing at least one reference species, heralds a new level of genome studies. It is because those very close Panthera species have distinct species-specific and readily identifiable phenotypes that can be associated quickly to mutations by comparing the homologous genes of interest as shown in the fur colouring (white lion) and high-altitude adaptation (snow leopard). If sufficiently distinct phenotypes are biologically curated, genetic mutations causing species specificity can be systematically detected using next generation sequencing. Once such candidate genetic mutations are confirmed in the set of species genomes, experimental validations can be carried out, as in the additional 47 lion samples here, for targeted genes. This genetic variation comparison using whole genomes among species and subspecies can thus provide valuable insight and information for the whole family’s conservation. Our data from tigers, lions and snow leopard can provide a rich and diverse genome resource that could be used in future studies of conservation and population genomics so that the genetic underpinnings of local adaptation and potential inbreeding and/or outbreeding in wild and captive populations can be illuminated and thereby help ensure the future survival of these majestic species.

**Methods**

**Genome sequence assembly and annotation.** The blood samples used for genome sequencing were acquired from the Everland Zoo of Korea (Amur tiger, white Bengal tiger, African lion and white African lion) following the Everland Zoo (Korea) ethical guidelines and procedures, and a muscle sample was obtained from a Mongolian snow leopard carcass preserved in the Conservation Genome Resource Bank for Korean Wildlife, Seoul National University. No animals were killed or captured as a result of this study. Libraries for the Amur tiger genome were constructed at BGI, Shenzhen, and the insert sizes of the libraries were 170 bp, 500 bp, 800 bp, 2 kb, 5 kb, 10 kb and 20 kb. The libraries were sequenced using HiSeq2000. Other big cat genomes were sequenced at Theragen BiO Institute (TBI), Korea, using HiSeq2000 with read and insert lengths of ~90 bp and ~100 bp, respectively.

The corrected reads were used to complete the genome assembly using SOAPdenovo. First, the short insert size library (170 bp, 500 bp and 800 bp) data were used to construct a de Brujin graph. Second, all reads were realigned with the contig sequences. The amount of shared paired-end relationships between pairs of contigs was calculated and weighted with the rate of consistent and conflicting paired ends, before constructing the scaffolds step by step from the short insert size paired ends to the long distant paired ends. Third, the gaps between the constructed scaffolds were closed using the paired-end information to retrieve read pairs where one end mapped to a unique contig while the other was located in the gap region.

The Amur tiger genome were predicted using three approaches. First, de novo prediction was performed using the repeat-masked genome using AUGUSTUS (version 2.5.5) and GENSCAN (version 1.0). Second, homologous proteins in other species were mapped to the genome using tBLASTn (Blast 2.2.23) with an E-value cutoff of 1E-5. The aligned sequence and its query protein were then filtered to search for accurately spliced alignments. Third, cat EST and full-length cDNA sequences (from UCSC) were aligned to the genome using BLAT to produce a consensus gene set. Then, the Amur tiger genome sequence was aligned to two well-assembled and annotated genomes (human and domestic cat) using LASTZ (version 1.02). Finally, mapped results yielding information on homologous proteins were filtered by syntactic blocks of genome sequences. We also predicted the domestic cat (Felis_catus-6.2) gene set, because the gene set of the cat genome is preliminary.

**Orthologous gene families.** A comparative analysis was used to examine the rate of protein evolution and the conservation of gene repertoires among orthologs in the genomes of the Amur tiger, dog, human, mouse, giant panda, domestic cat (Felis_catus-6.2) and opossum. We used the TreeFam methodology to define a gene family as a group of genes that descended from a single gene in the last
common ancestor of a considered species. We assigned a connection (edge) between two nodes (genomes) if more than 1/3 of the region was aligned to both genes. The score (ES score [minimum edge weight]) that ranged from 0 to 100 was used to weigh the similarity (edge). For two genes, G1 and G2, the H-score was defined as score (G1G2)/max (score (G1G1), score (G2G2)), where the score shown is the BLAST raw score. Gene families were extracted by clustering using Hcluster. We used the average distance for the hierarchical clustering algorithm, requiring the H-score to be greater than five, and the minimum edge density (total number of edges/theoretical number of edges) to be larger than 1/3. The clustering for a gene family would also stop if it already had one or more of the out-group genes.

We determined the expansion and contraction of the orthologous protein family by analyzing syntenic blocks with rigorous conditions to get large-scale synteny of high-alignment quality, and a conservation of exon–intron structure. Finally, we found 7,415 1:1 high-quality ortholog genes to analyze, most of which also correspond to genes in the other species via the syntenic regions. We then filtered the resulting outputs were scaled to time and population sizes. We obtained atmospheric surface temperature and global relative sea level data of the past 5 million years.

**References**

1. Luo, S. J. et al. Phylogeography and genetic ancestry of tigers (Panthera tigris). PLoS Biol. 2, e442 (2004).
2. Smith, J. L. D., Ahern, S. C. & McDougal, C. Landscape analysis of tiger distribution and habitat quality in Nepal. Conserv. Biol. 12, 1338–1346 (1998).
3. Chundawat, R. S. et al. Panthera tigris. In: IUCN 2012. IUCN red list of threatened species. version 2012.1 www.iucnredlist.org (2011).
4. Wiston, J. et al. Bringing the tiger back from the brink—the six percent solution. PLoS Biol. 8, e1000485 (2010).
5. Luo, S. J., Johnson, W. E. & O'Brien, S. J. Applying molecular genetic tools to tiger conservation. Integr. ZooL. 5, 351–362 (2010).
6. Driscoll, C. A. et al. Mitochondrial phylogeography illuminates the origin of the extinct Caspian tiger and its relationship to the Amur tiger. PLoS One 4, e4125 (2009).
7. Kitipipit, T. & Linacre, A. The complete mitochondrial genome analysis of the tiger (Panthera tigris). Mol. Biol. Rep. 39, 5745–5754 (2012).
8. Pontius, J. U. et al. Initial sequence and comparative analysis of the cat genome. Genome Res. 17, 1675–1689 (2007).
9. Mullikin, J. C. et al. Light whole genome sequence for SNP discovery across domestic cat breeds. BMC Genomics 11, 406 (2010).
10. Allenford, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709 (2010).
11. Murchison, E. P. et al. Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. Cell 148, 780–791 (2012).
12. Shao, Q. et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat. Genet. 45, 67–71 (2012).
13. Li, H. et al. SOAP2: an improved ultrafast tool from short read alignment. Bioinformatics 25, 1966–1967 (2009).
14. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23, 1061–1067 (2007).
15. Johnson, W. E. et al. The late Miocene radiation of modern Felidae: a genetic demography solution. Science 311, 73–77 (2006).
16. Corbett, J. The Temple Tiger and More Man-eaters of Kumaon (Oxford Univ. Press, London, 1954).
17. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249 (2010).
18. Iriarri, K. J. et al. Sequencing and comparative genomic analysis of 1227 Felis catus cDNA sequences enriched for developmental, clinical and nutritional phenotypes. BMC Genomics 13, 31 (2012).
19. Qiu, Q. et al. The yak genome and adaptation to life at high altitude. Nat. Genet. 44, 946–949 (2012).
20. Charudatt, M. et al. The role of incentive programs in conserving the snow leopard. Conserv. Biol. 17, 1512–1520 (2003).
21. Xu, S. et al. A genome-wide search for signals of high-altitude adaptation in Tibetans. Mol. Biol. Evol. 28, 1003–1011 (2011).
22. Kim, E. B. et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479, 223–227 (2011).
23. Schönheit, J.-B., Künstler, A., Eizenberg, O., O’Brien, S. J. & Menotti-Raymond, M. Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat colour phenotypes of the albinos and brown loci. J. Hered. 96, 289–301 (2005).
24. King, R. A. et al. Tyrosine gene mutations in oclocutaneous albinism 1 (OCA1): definition of the phenotype. Hum. Genet. 113, 502–513 (2003).
25. Optiz, S., Kämssmann-Kellner, B., Kaufmann, M., Schwingr, E. & Zühlik, C. Detection of 53 novel DNA variations within the tyrosinase gene and accumulation of mutations in 17 patients with albinism. Hum. Mutat. 23, 630–631 (2004).
26. Xu, X. et al. The genetic basis of white tigers. Curr. Biol. 23, 1031–1035 (2013).
27. Mayer, C., Leese, F. & Tödllein, R. Genome-wide analysis of tandem repeats in Daphnia pulex—a comparative approach. BMC Genomics 11, 277 (2010).
28. Locke, D. P. et al. Comparative and demographic analysis of orang-utan genomes. Nature 469, 529–533 (2011).
29. Zhang, J., Wang, X. & Podlaha, O. Testing the chromosomal speciation hypothesis for humans and chimpanzees. Genome Res. 14, 845–851 (2004).
30. Miller, W. et al. Genetic diversity and population structure of the endangered marsupial Sarcoptes harrisii (Tasmanian devil). Proc. Natl Acad. Sci. USA 108, 12348–12353 (2011).
31. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496 (2011).
32. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34, W435–W439 (2006).
33. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94 (1997).
34. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).
Acknowledgements

This work was supported by the Industrial Strategic Technology Development Program, 10040231, Bioinformatics platform development for next generation bioinformation analysis' funded by the Ministry of Knowledge Economy (MKE, Korea). This work was partly supported by the National Research Foundation of Korea (NRF-2011-0019745) and grants from the Ministry of Science, ICT & Future Planning (MSIP), Republic of Korea (2012R1A1A2043851 and NRF-2008-2040707). This research was also supported in part by Russian Ministry of Science Mega grant no.11.G34.31.0068; SJ O'Brien Principal Investigator. Scaffolds (sequence data) and gene sets are available at http://tigergenome.org. We thank Dr Rui-Qiang Li for contributing to the manuscript. We thank Maryana Bhak for editing. Authors thank many people not listed as authors who gave us feedback, samples and encouragement, especially Won-shik Min. This project was initiated with the generous support of TheragenEtex, Seong-Jin Kim and Jin-Up Goh. We thank the following people and organizations for providing valuable samples, data and time for this project: Johannesburg Zoo, in particular Director Dominic Moss for pedigree information and permission to use samples, Janine Fearon and Lucia Muuulu for laboratory work done at the Life Technologies Conservation Genetics Laboratory of the Czechat Conservation Fund, Ukutula Lodge & Lion Centre, Global White Lion Protection Trust and Tsau Conservancy, Bhagavan "Doc" Antle from Tiger's Preserve, Myrtle Beach, and Ouwehands Dierenpark for contributing to this project with their lion population, Dr Carlos Driscoll for making one control sample (PLE171) available to this project, and Life Technologies and Bimatters for generous support with software (Life Technologies and Generous) and equipment (Life Technologies) of the Genetics Laboratory at the Czechat Conservation Fund. Finally, we thank the many conservationists who have helped protect endangered big cats throughout the world.

Author contributions

Y.S.C., L.H., H.H., H.L. and J.X. contributed equally to this work. The tiger genome project was initiated by J.B., B.C.K., H.L., T.H.K., S. Lee., Sangtgae K., C.-B.K., S.-J.K., W.K.P. and Jun W. Library construction, sequencing, bioinformatics data processing and analysing genetic variation data were carried out by L.H., J.X., S.J., Y.-A.S., Q.Z., H.K., C.-U.K., Y.X., Y.L., S.P., C.G., X.C., J.Z., Sanyang L., Jing H., Y.C., L.Y., Y.Y., Jaee H., S.-I.L., Junyi W., J.-S.K., H.-M.K., Y.S.C., T.H.K., Sangsoo K., I.B. and Jun W. Several big cat genome re-sequencing were performed by H.-J.J. and C.H.K. PCR validations were performed by H.-J.J. and Hwanjong K., S. Kwon., S.O., W.K.P., H.L. and D.B. provided samples, advice and associated information. Y.S.C., L.H., H.H., S.-J.L, W.J., K.-P.K.X., P.G., S.H., J.K., C.-B.K., H.L., Sangtgae K., Sangsoo K., S.I.O., Jun W., and J.B. wrote, edited and revised the manuscript. A.S.-K. conducted overall project coordination, data analysis and presentation of lion DNA analysis and overseeing/planning of laboratory work done at CCF (Cheetah Conservation Fund). J.A.T. carried out the lion DNA project initiation for samples of the Tsau Conservancy and the Johannesburg Zoo. L.M. did support of the laboratory work done at CCF. C.H. performed communication and project initiation at VGL (Veterinary Genetics Laboratory) and oversaw the laboratory work done at VGL, sampling and funding. S.M.M. at VGL did planning and execution of the laboratory work, pedige verification and reference samples of other populations. Wilhelm J. did pedigree information processing and sampling at the Ukutula Lodge. L.B. did the sample processing and laboratory work of Ouwehands Dierenpark and PLE171.

Additional information

Accession codes: The Amur tiger whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number ATCQ00000000. The version described in this paper is the first version, ATCQ01000000. Raw DNA and RNA sequencing reads have been submitted to the NCBI Sequence Read Archive database (SRA074975, SRA091968).

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions

How to cite this article: Cho, Y. S. et al. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat. Commun. 4:2433 doi: 10.1038/ncomms3433 (2013).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/