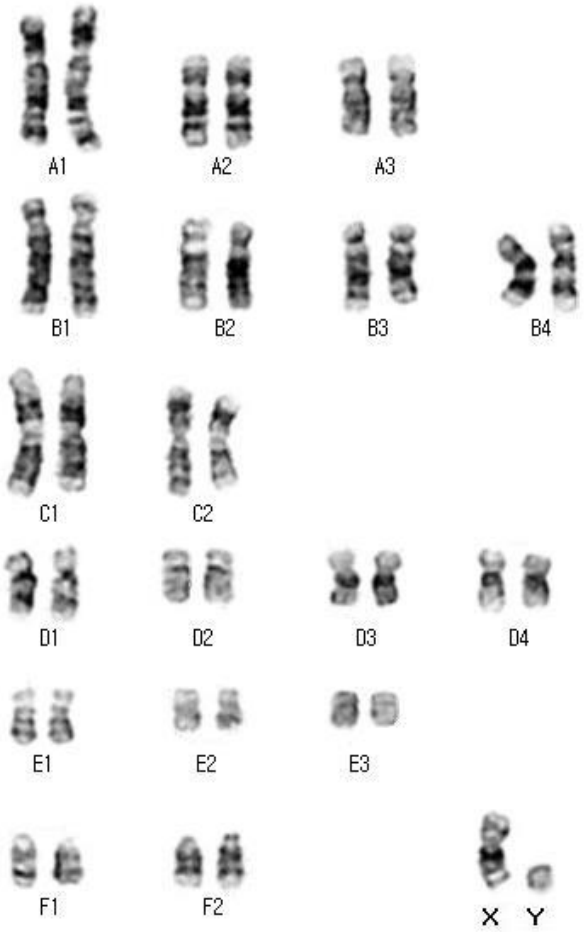
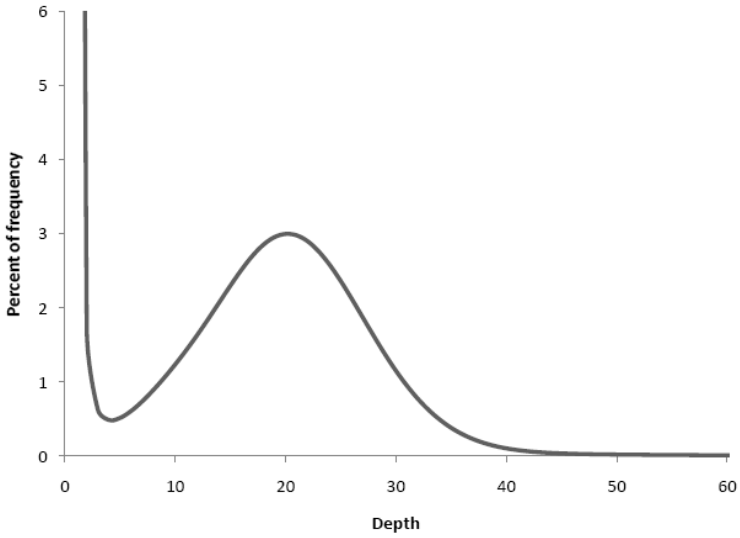
Supplementary Information

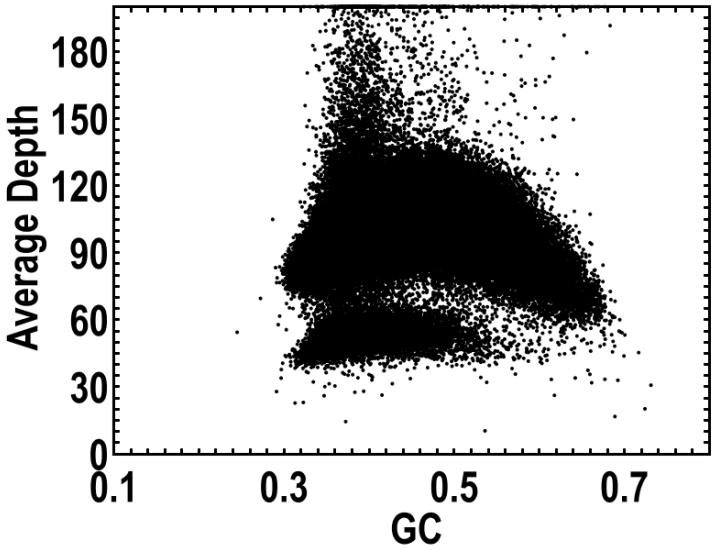
Supplementary Figures



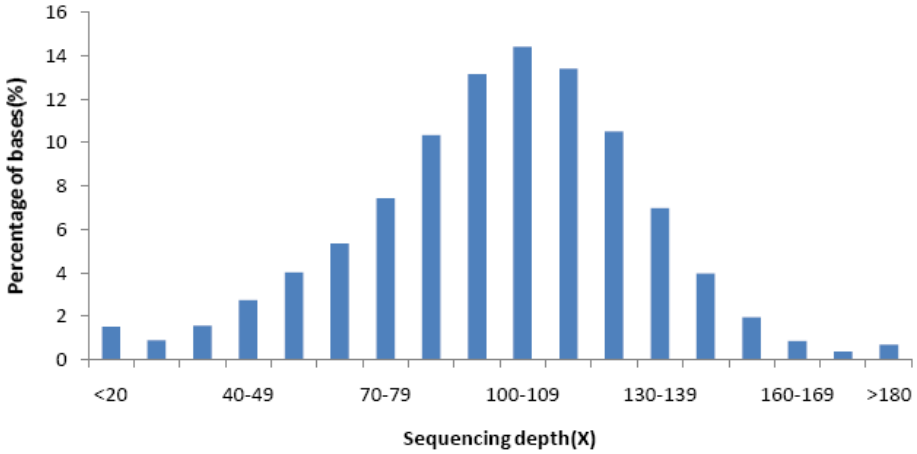
Supplementary Figure S1 | G-banded karyotype of the Amur tiger (TaeGeuk). There were no abnormalities in the chromosomes (2n=38).



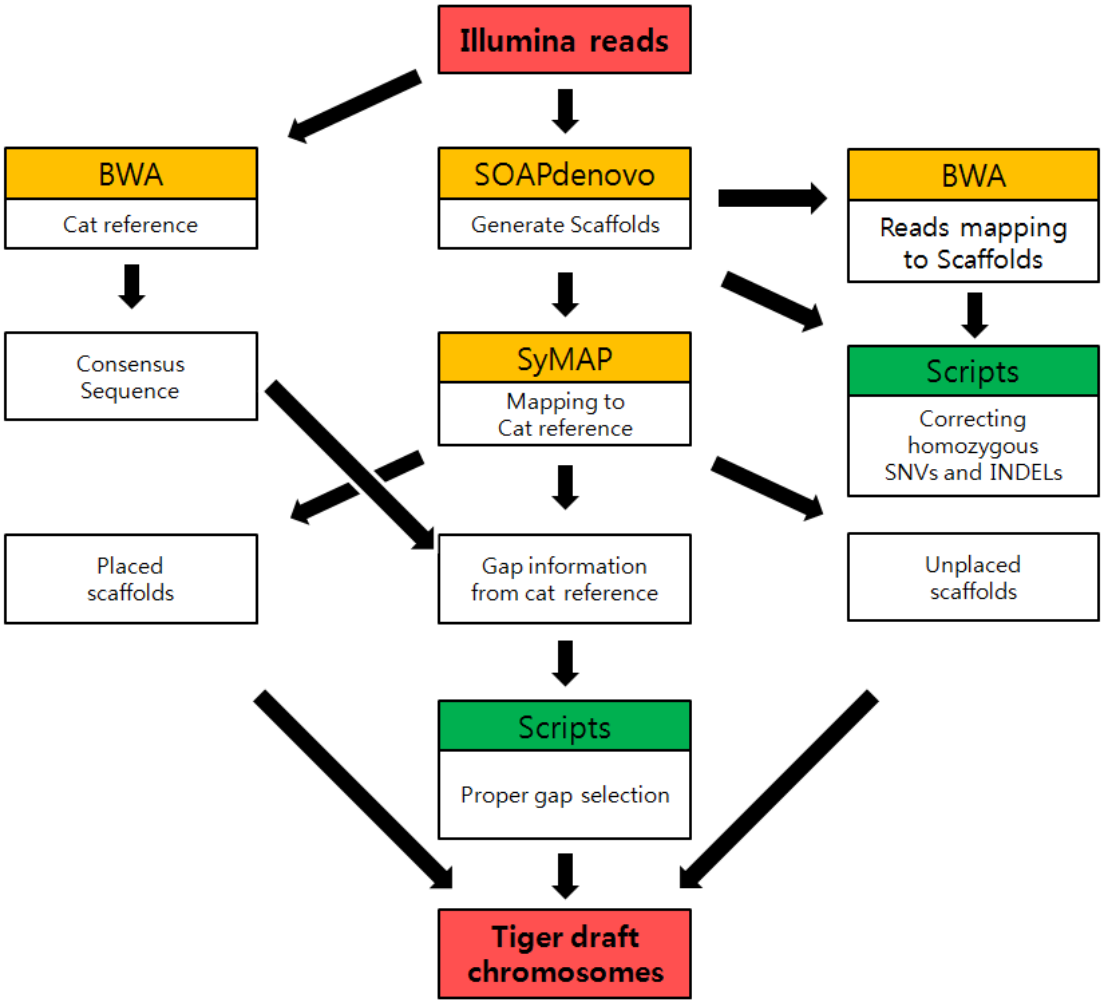
Supplementary Figure S2 | Estimation of genome size using 23-mers. The x-axis denotes the depth coverage of each unique 23-mer in the genome, and the y-axis denotes the occurrence of unique 23-mers within the sequence dataset. In other words, the x-axis represents depth, and the y-axis represents proportion, as calculated by the frequency at that depth divided by the total frequency at all depths.



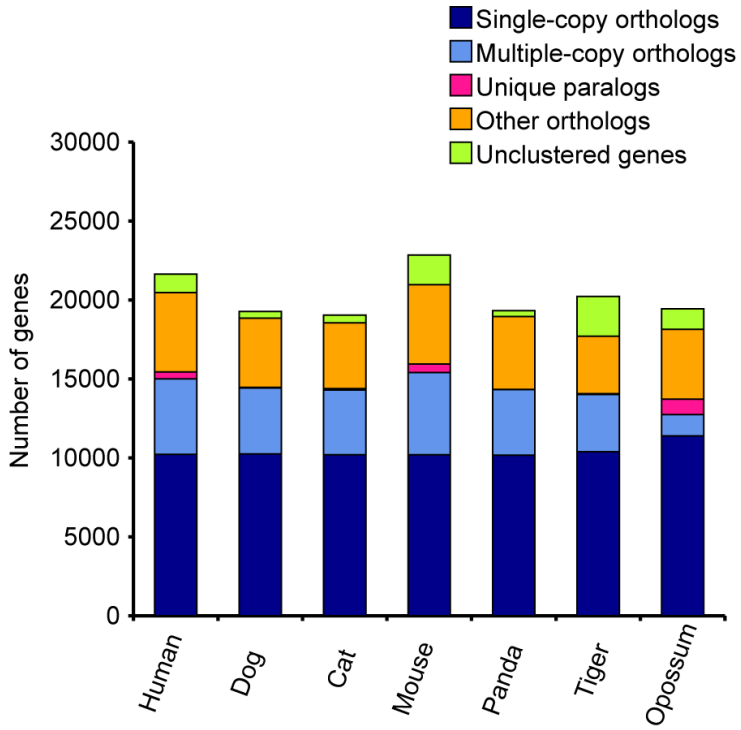
Supplementary Figure S3 | GC contents and sequencing depths. We used 10Kb non-overlapping sliding windows and calculated the GC content and average depth among these windows. To check whether the small block in the bottom of Supplementary Figure 4 originated from contamination from other species, some parts of the scaffolds were chosen from the small block and used as queries to perform BLAST with the nt library. The scaffolds were most closely related to chromosome X of *Felis catus*, which indicates that the small block represents chromosome X of the Amur tiger.



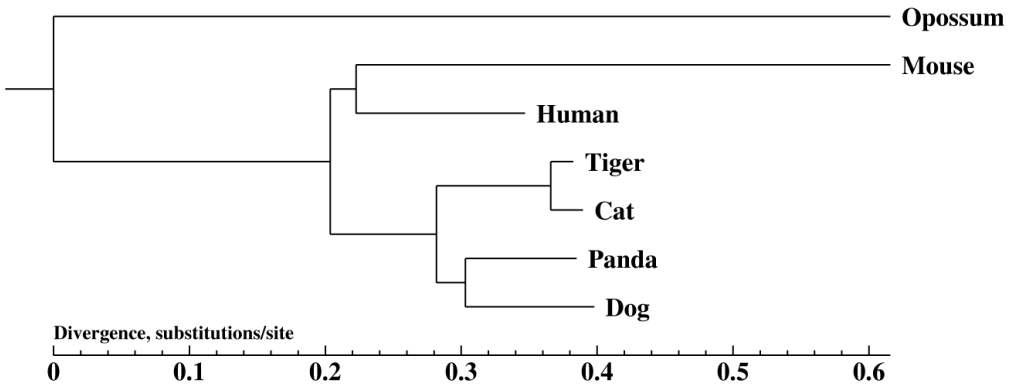
Supplementary Figure S4 | Sequencing depth distribution. Reads were aligned onto the assembled genome sequence using SOAP, allowing three mismatches for each read, and the frequency of each of the covered genome bases was calculated.



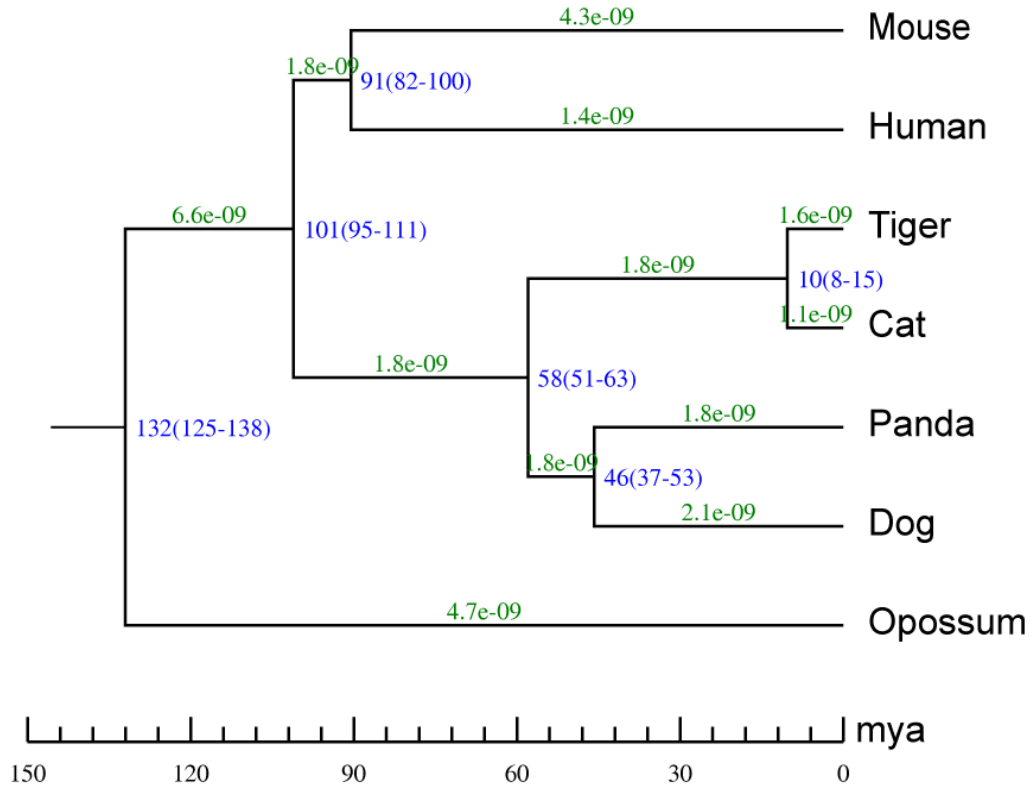
Supplementary Figure S5 | Amur tiger draft genome assembly pipeline showing the software used and the flow of sequence. As the Amur tiger genome is very similar to the cat reference genome, the chromosomal location and ordering information of the tiger scaffolds could be derived from the cat genome. The software and in-house scripts are shown in yellow and green, respectively. Consensus sequencce indicates continuous mapped regions, when the tiger short reads were aligned to the cat genome.



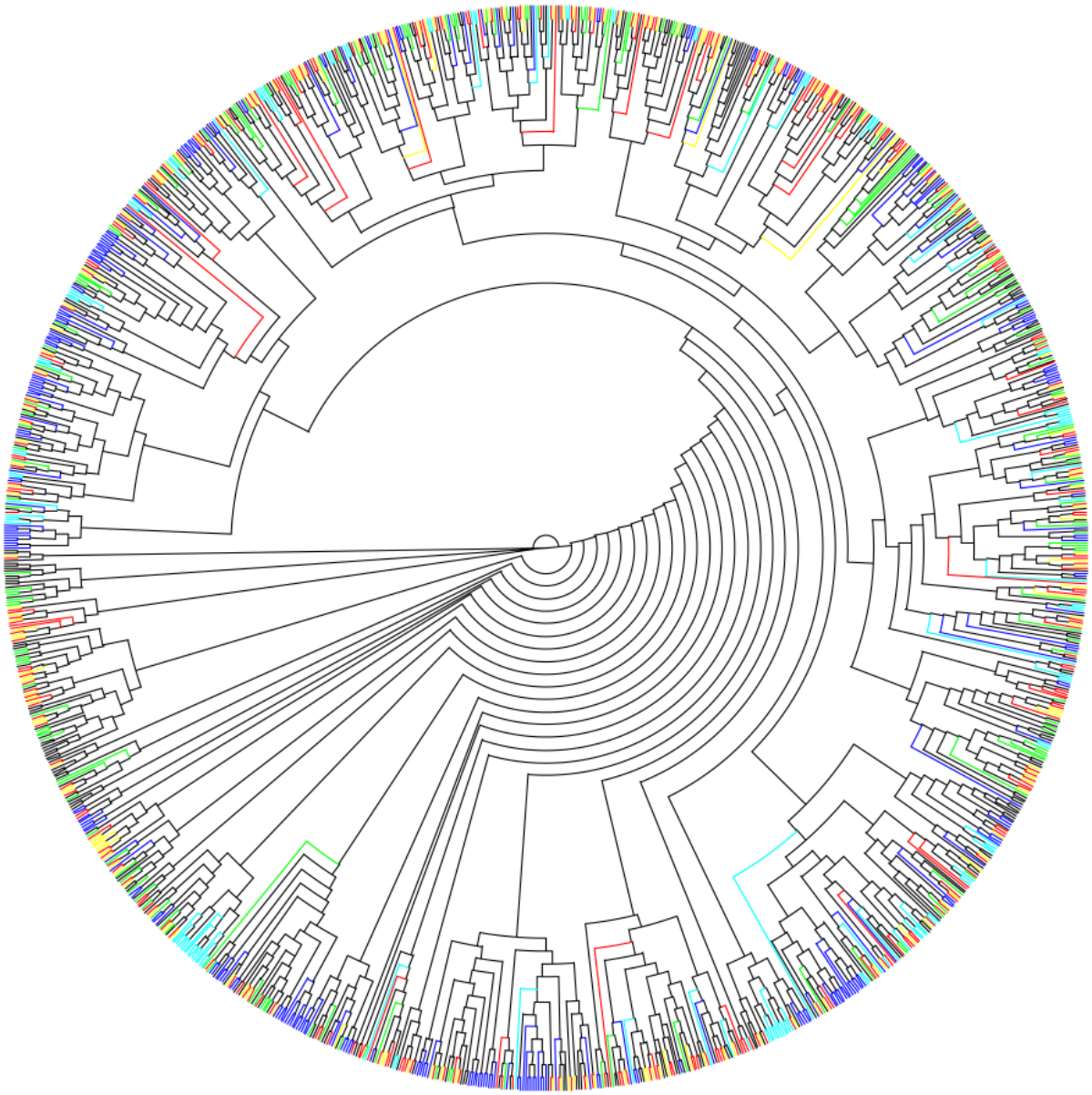
Supplementary Figure S6 | The composition of mammalian orthologous genes. A comparative account of ortholog and paralog genes in the Amur tiger genome is shown in comparison to the genomes of other seven animal species.



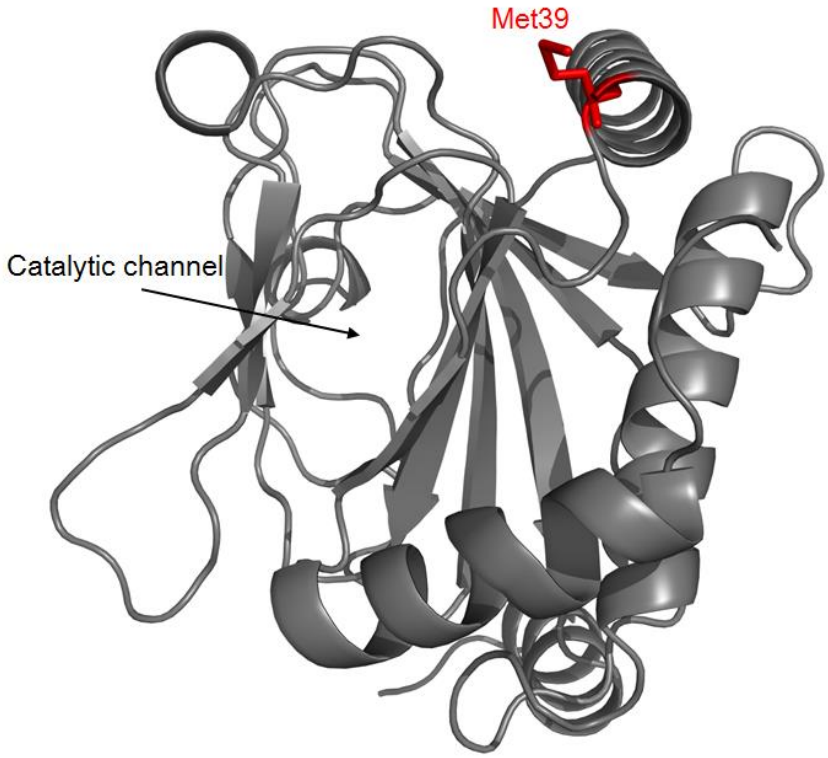
Supplementary Figure S7 | Phylogenetic tree constructed with orthologous genes on 4-fold degenerate sites. The branch length represents the neutral divergence rate.



Supplementary Figure S8 | Estimation of divergence time and substitution rate. The green numbers on the branches are the estimated substitution rate (substitutions per site per year). The blue numbers on the nodes are the divergence time from present (million years ago, Mya). The calibration times of (98.2 Mya) human-dog divergence and (57.5 Mya) cat-dog divergence were derived from the TimeTree database (http://www.timetree.org).

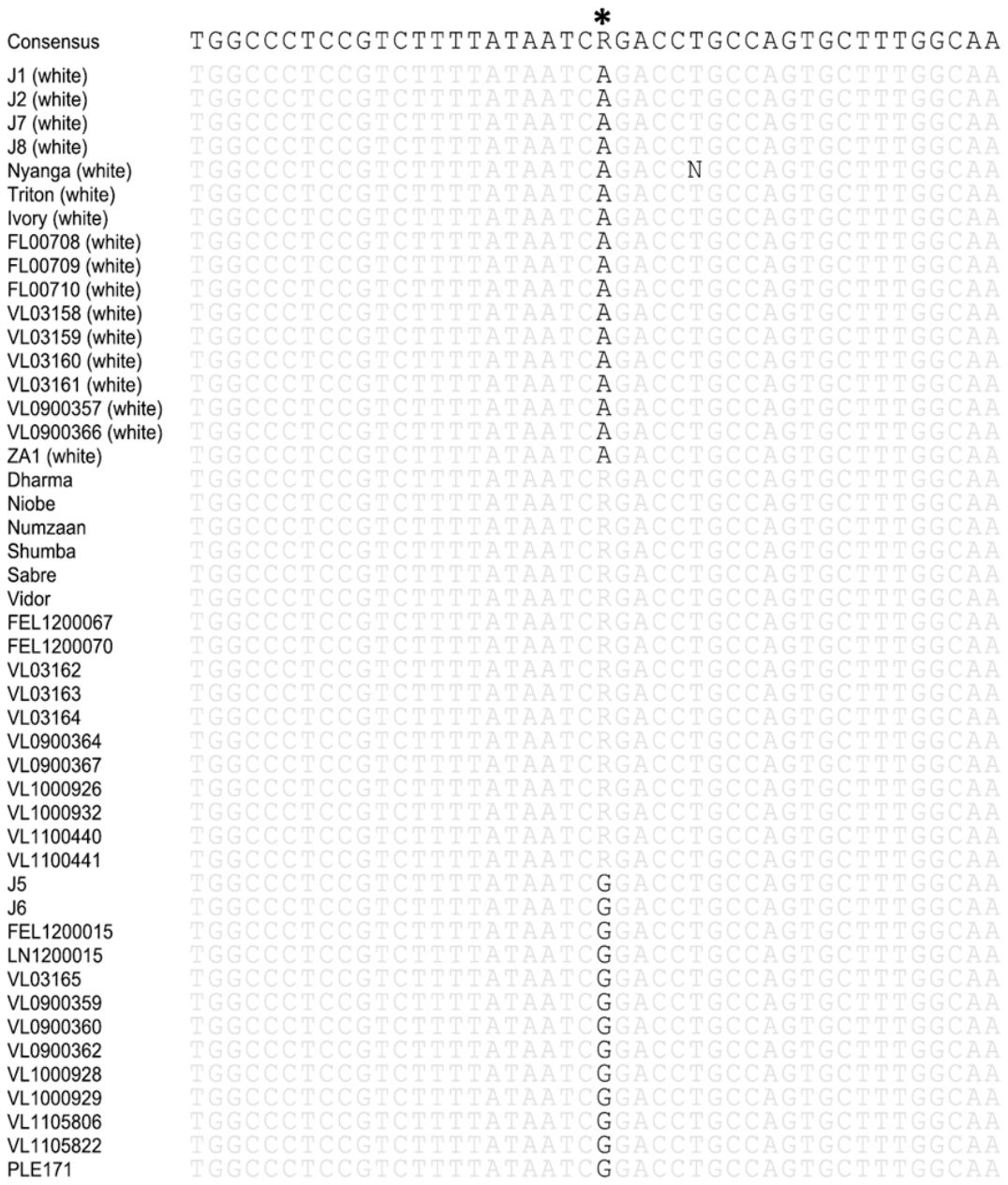


Supplementary Figure S9 | Olfactory receptor gene families expanded in tiger compared to domestic cat. Multiple sequence alignments and clustering were conducted using ClustalW2. Red, yellow, green, black, blue, magenta, and cyan edges in the outside are olfactory receptor genes of tiger, cat, dog, panda, mouse, human, and opossum, respectively.

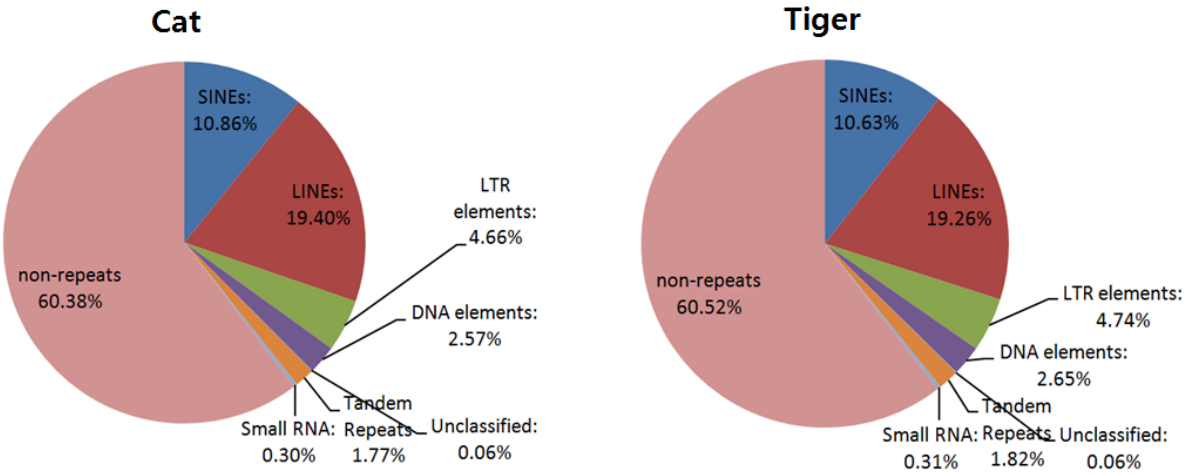


Supplementary Figure S10 | 3D position of snow leopard-specific amino acid change in the *EGLN1* gene. The red stick representation indicates the structural position for Met39>Lys39 amino acid change. The 3D structure was acquired from the Protein Data Bank (PDB) database (Entry: 3HQR).

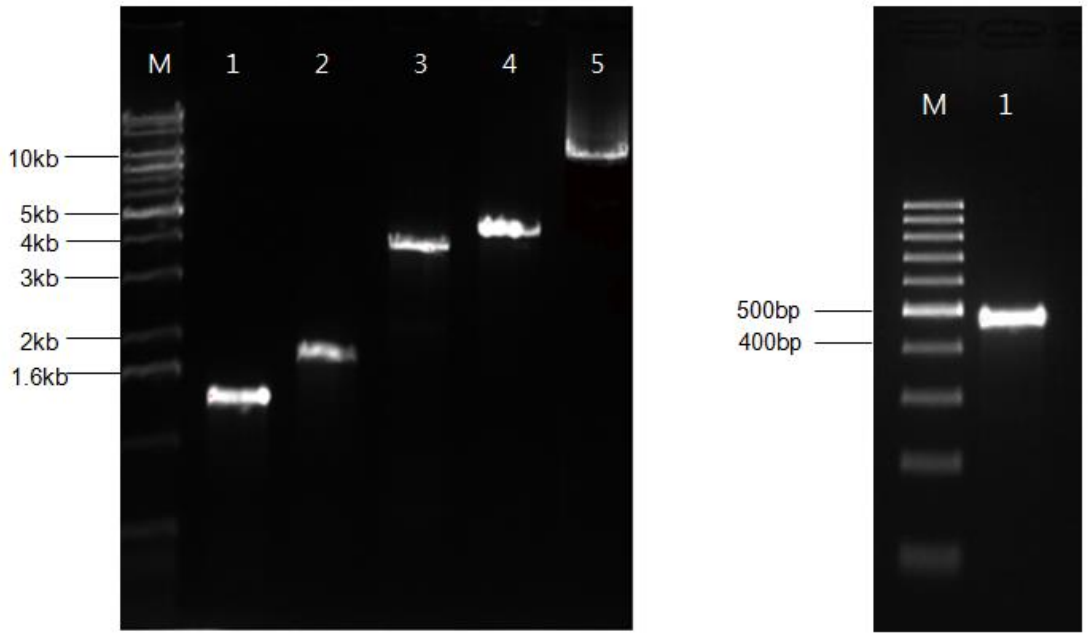
Supplementary Figure S11 | *EPAS1* mutations related to hypoxia in snow leopard. Non conserved amino acids in the big cats are shown in blue and red. The snow leopard-specific amino acids are Ile663 and Arg794.



Supplementary Figure S12 | Alignment of *TYR* exon 1 sequences from 47 lion samples. The *TYR*260G>A polymorphism (*) corresponded to the white lion phenotype. The figure represents a 44-nucleotide sequence of *TYR* exon 1 (base pairs 238-281). 'R' indicates the heterozygous nucleotide A/G.



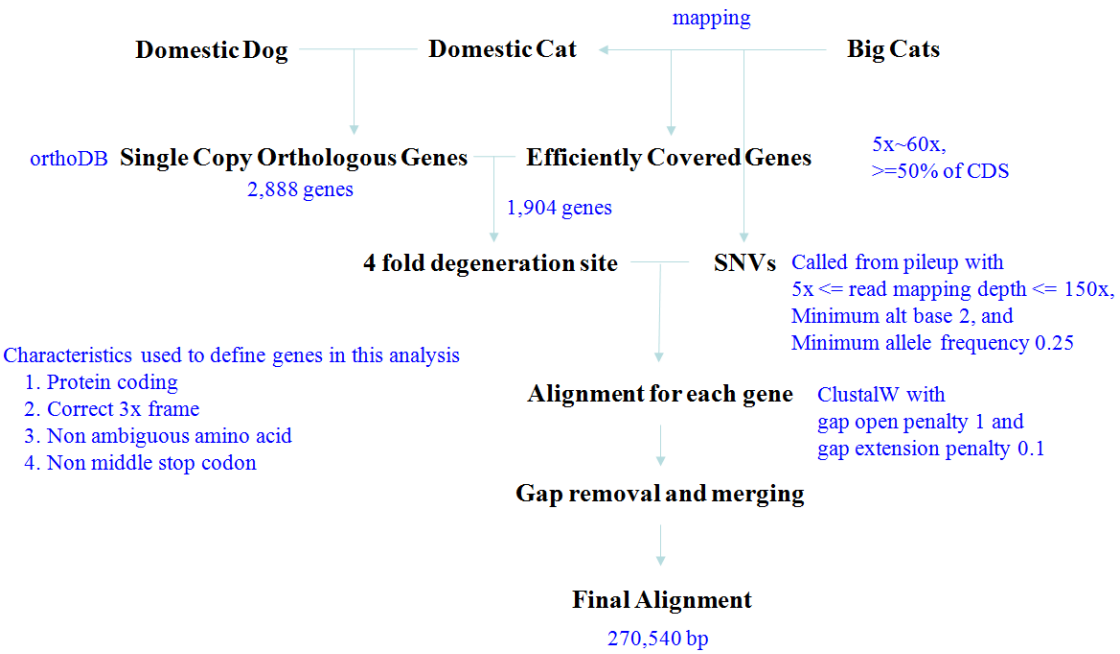
Supplementary Figure S13 | Comparison of repeats between Amur tiger and domestic cat (Felis_catus-6.2). The tiger and cat genomes showed very similar repeat compositions and ratios of repeat components. SINEs are short interspersed elements. LINEs are long interspersed elements. LTR is long terminal repeat.



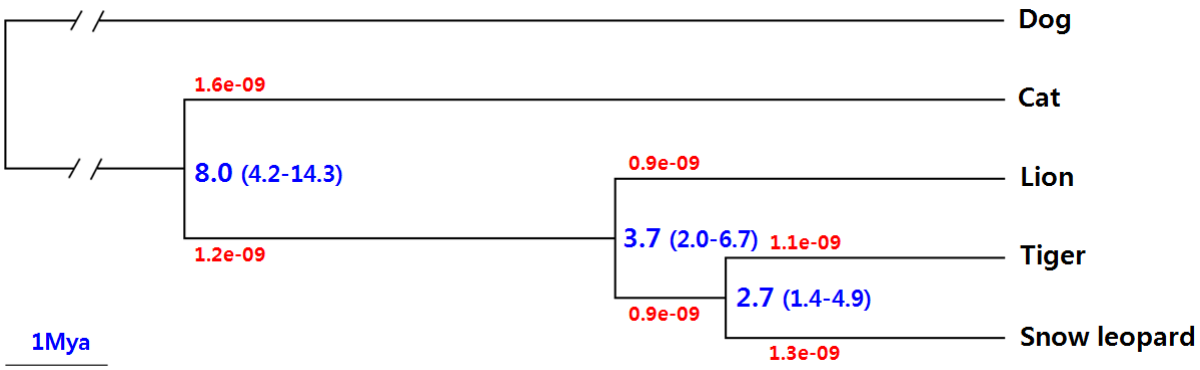
M : 1kb DNA extension ladder
1 : scaffold1490
2 : scaffold1545
3 : scaffold150
4 : scaffold402
5 : scaffold424

M : 100bp ladder
1 : scaffold309

Supplementary Figure S14 | Long range PCR results for tiger scaffold integrity validation. Six scaffolds having chromosomal rearrangements between the Amur tiger and domestic cat were validated by long range PCR experiments followed by the Sanger sequencing method. M Lanes are 1 Kb and 100bp DNA extension ladders. Other lanes are the target regions on each scaffold, supporting the tiger scaffold integrity and the six putative chromosomal rearrangements.

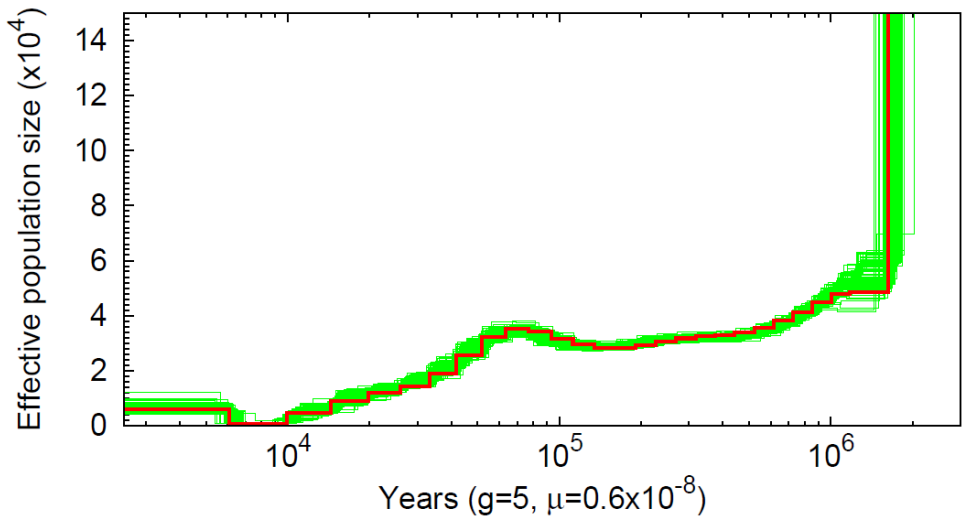


Supplementary Figure S15 | Data set selection for phylogenetic tree analysis among big cats. A total of 1,904 efficiently covered genes, which were extracted from orthologous genes between the domestic cat and dog and mapping results of big cat reads to the domestic cat genome (CAT.66), were used to construct the phylogenetic tree and to calculate the substitution rate of the big cats.

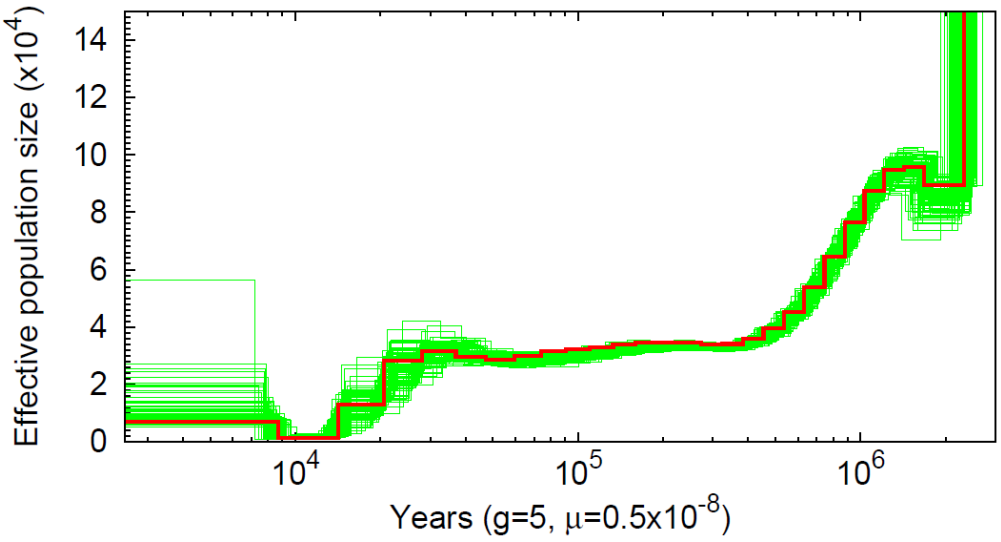


Supplementary Figure S16 | Divergence among domestic and big cats. Divergence times and substitution rates are shown in blue and red, respectively. The 95% confidence intervals of divergence time are shown in parentheses.

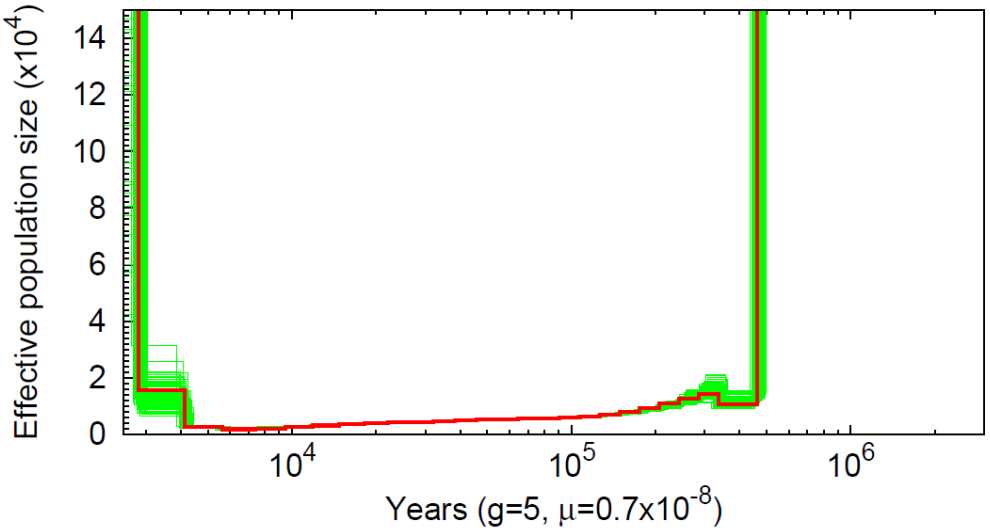
a. TG, μ = 1.1e-09 per site per year



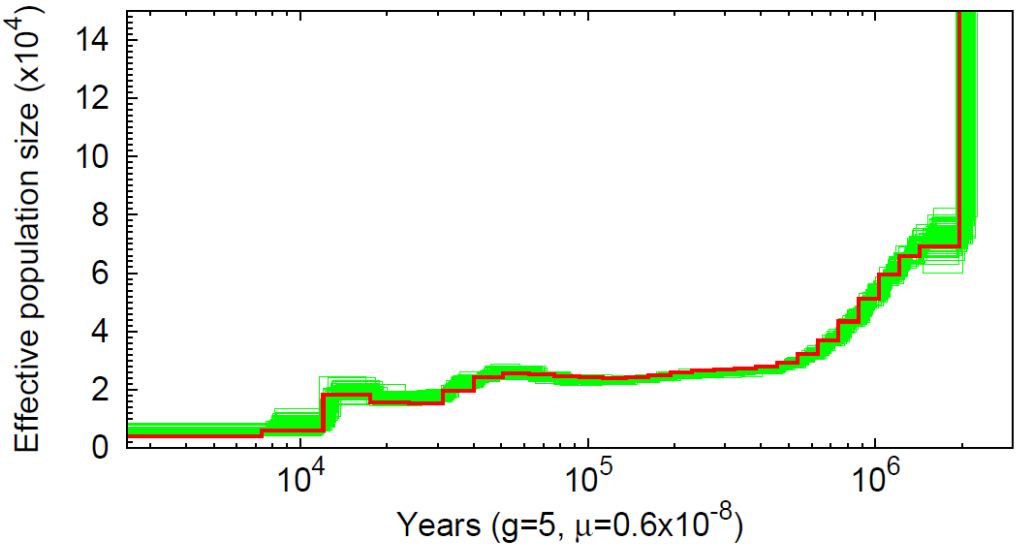
b. LN, μ = 0.9e-09 per site per year



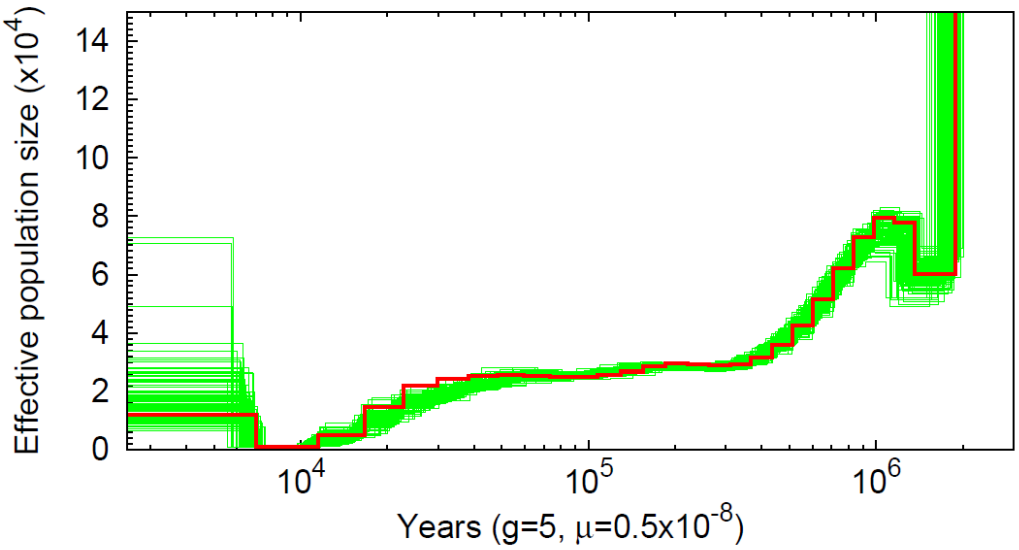
c. SL, μ = 1.3e-09 per site per year



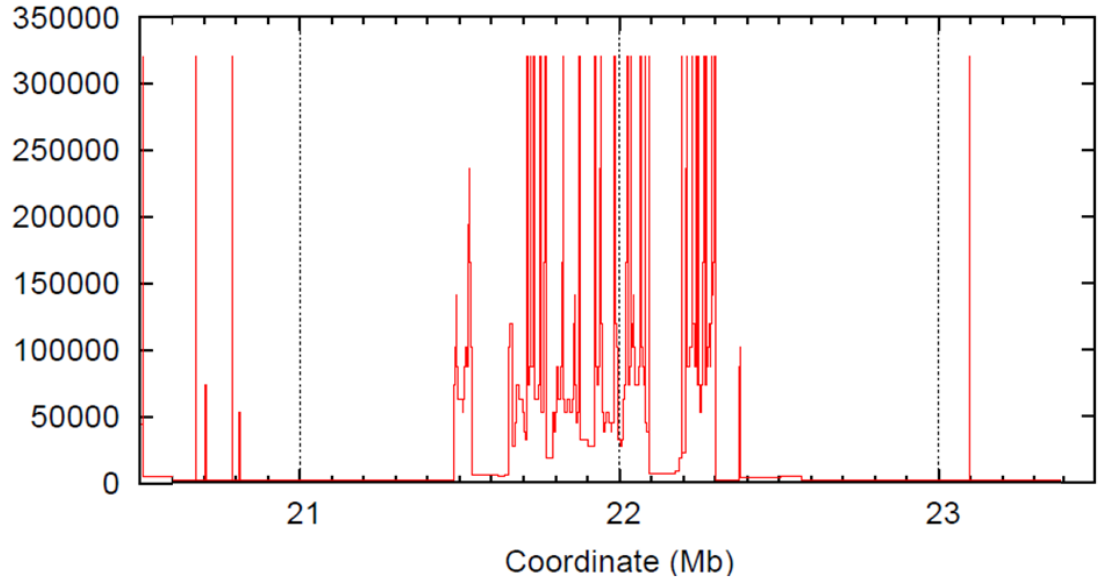
d. WTG, μ = 1.1e-09 per site per year



e. WLN, μ = 0.9e-09 per site per year



Supplementary Figure S17 | Population size history plots. μ of x-axis is mutation rate per generation which equals mutation rate per year multiplied by generation time (g, 5 years in this analysis). Red lines are population size history plots. The green lines indicate 100 bootstrapping results. TG, Amur tiger; LN, African lion; SL, snow leopard; WTG, white tiger; WLN, white African lion.



Supplementary Figure S18 | Sample plot of the most recent common ancestor distribution. X-axis represents chromosome coordinates. Y-axis shows the most recent common ancestor (TMRCA) of the corresponding region of the genome.

Supplementary Tables

Supplementary Table S1 | Sample characteristics

Sample	Name	Birth Date	Sex	Origin
Amur tiger	TaeGeuk	2003. 04. 01	Male	Russian Far East
White tiger	HwaRang	1999. 06. 18	Female	Bengal
African lion	SunDol	2007. 06. 22	Male	Sub-Sahara
White lion	SnowGirl	2005. 01. 18	Female	Sub-Sahara
Snow leopard	N/A	N/A	Female	West Mongolia

Supplementary Table S2 | Information regarding the seven libraries used for the tiger genome assembly

Paired-end libraries	Insert size	Total data (M)	Read length (bp)	Sequence coverage (X)	Physical coverage (X)
Illumina Reads	170bp	79418.44	100	32.55	27.67
	500bp	73543.91	100	30.14	75.35
	800bp	50757.32	100	20.80	83.21
	2Kb	31366.88	49	12.86	262.35
	5Kb	17524.51	49	7.18	366.44
	10Kb	20057.00	49	8.22	838.78
	20Kb	15530.77	49	6.37	1298.99
Total	-----	288198.82	-----	118.11	2952.80

Supplementary Table S3 | Filtered sequence information

Paired-end libraries	Insert size	Total data (Mb)	Average read length	Sequence coverage (X)	Physical coverage (X)
Illumina Reads	170bp	74,486.12	100	30.53	25.95
	500bp	67,494.03	100	27.66	69.15
	800bp	43,861.62	95	17.98	75.83
	2Kb	23,994.11	49	9.83	200.69
	5Kb	14,336.55	49	5.88	299.78
	10Kb	12,570.11	49	5.15	525.68
	20Kb	5,316.72	49	2.18	444.69
Total	-----	242,059.26	-----	99.20	1641.77

Supplementary Table S4 | 23-mer statistics information

K-mer	K-mer_num	K_depth of peak	Genome_size	Used_base	Used_read	Depth
23	48,766,682,683	20	2,438,334,134	89,972,618,150	1,799,452,363	35.71

Supplementary Table S5 | Statistics regarding assembled sequence length

	Contig		Scaffold	
	Size (bp)	Number	Size (bp)	Number
N90	7,434	82,408	2,209,899	291
N80	12,817	58,729	3,731,863	208
N70	18,057	43,354	5,621,845	155
N60	23,646	31,985	7,021,237	118
N50	29,811	23,130	8,840,225	88
Longest	287,365	-----	41,607,841	-----
Total Size	2,350,575,370	-----	2,408,774,396	-----
Total Number (>100bp)	-----	261,299	-----	105,019
Total Number (>2kb)	-----	125,210	-----	1,041

Supplementary Table S6 | Statistics regarding the tiger blood transcriptome

The number of total reads	Total nucleotides (nt)	>Q20 (%)	N base (%)	GC contents (%)
99,886,906	8,989,821,540	98.71	0	52.56

Supplementary Table S7 | Assessment of gene coverage by assembled tiger transcripts

Dataset	Number	Total Length (bp)	Covered by Assembly	With >90% Sequence in one Scaffold		With >50% Sequence in one Scaffold	
				Number	Percent (%)	Number	Percent (%)
All	190,492	136,621,518	96.47	169,027	88.73	186,516	97.91
>200bp	154,550	130,398,346	96.44	136,206	88.13	151,254	97.87
>500bp	61,336	102,810,006	96.32	50,263	81.95	60,112	98.00
>1000bp	35,887	85,032,004	96.36	28,377	79.07	35,212	98.12

Supplementary Table S8 | Assessment of gene coverage by EST sequences

Total EST	Mapped EST	Mapped Percent (%)	ESTs covered by assembly (%)	>90% mapped in one scaffold		>50% mapped in one scaffold		>20% mapped in one scaffold	
				Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
919	909	98.91	90.77	782	85.09	875	95.21	896	97.5

Supplementary Table S9 | Heterozygous SNVs validation by Sanger sequencing

scaffold	position	Ref. Allele	Heterozygous Allele	Forward Primer	Reverse Primer	Status
scaffold177	6288	G	A	CAAGGCAATCAGACACTCAAATG	TAATACCCCTTTTCGTTTCTCCC	Valid
scaffold1514	5425541	C	G	TGTAAGCTCAGGGATGGACAAAT	TTCCTTGTGGTTTGTGTTCCTTT	Valid
scaffold394	1515395	C	T	TGGTGTTCAAGTTCAGGCCTATT	CCAAGCTCCGAACAAACTCTAAA	Valid
scaffold231	2991120	A	T	TTTCCCTAGCAACTGAAGATTCG	CTCTTTTAAGAGATTCCGCCCAT	Valid
scaffold60	17144319	A	C	CATGGCAGATCCATACAACAGAG	CCCCTCATCTCTCCATTCTTTCT	Valid
scaffold259	4852	G	A	CTGGCTCTGAATTTGAGTCCTTC	TCAGCTCAGGGCAAGTATAAAGC	Valid
scaffold60	16891279	T	C	TCTTGCTACTAAAGCCCTTTCCC	CATACAACATTGTGTAACCCCGA	Valid
scaffold1542	7227565	A	G	TACCCCGCACATGTTTTATCTCT	CAAGAATGCATGACAACGAGTTC	Valid
scaffold60	9069100	T	C	GATTTCGCTCTCAATGTAACGCT	CAGAGTGACGCATCTTCAAGCTA	Valid
scaffold231	8046726	T	C	TCAGCAAGGATGTGGAACATCTA	CCAAGAGTAGGACGCTAAGTGGA	Valid
scaffold1542	7509965	A	G	CAACCCCATCTCTATTTGCTGAG	GCAAATCTTCTCTCACCCTGAAA	Valid
scaffold79	9978782	C	G	ATATAGTCCCTGCTCCTTCTGCC	TGCTGGAAATAGTCATCCTGTCA	Valid
scaffold1489	6662547	T	G	ATACGGATAGTGACAGCCAGAGC	GGTACAAATGTGTGGGAGCAAAC	Valid
scaffold60	11360443	A	G	AACCTAGAGAAGGACCCTGGTTG	TAACAGGCCTACTGCCATCCTAA	Valid
scaffold1543	730777	G	C	CCTCTGTGAAATATAAGTGCCCG	AAGGGGTAAGGAGTTTTGCCTAA	Valid
scaffold98	5534196	A	G	CTCAAGGTCTGTTCCCAACAAAT	TAGCAAAAGGCAGACACAGATTG	Valid
scaffold719	2639	A	G	CCCTCCAGTGAGATCACAAGACT	CCAAACAGAGCAGTTGTTAACCC	Valid
scaffold25	102667	C	T	AGCCAGGCATACTGGTTTTATGA	ATAATTATCCAGGTCCCTCACCC	Valid
scaffold1516	2012721	A	G	AGACCTCTCCTTCCTTGCTTCTC	GATGAGTTTTTCCCCAGTCCTCT	Valid
scaffold98	3672260	C	T	GTTCTCAGTTCCCAAAGAGGTGA	CAAGTACCCCCTCCTTCTTTCAT	Valid
scaffold44	11128318	G	A	CAACATCTACCACCTGCACTGTC	TGTGCCAATCTAAATGCTCTTCA	Valid
scaffold17	3173833	T	C	CCTCCTCTCCCTTCTTTTACCTG	CCTGCTACTCATCTTCTCCCAAA	Valid
scaffold44	12259953	A	G	GCAATCGGATAACAACAACACAA	TTGGTTAGGTTTGCTCCTTGGTA	Valid
scaffold1517	434766	G	A	TGGGCATCCGTACTATTTTGTCT	GGCCTTTAATTTGGCTGAGGTAG	Valid
scaffold17	1737076	C	T	CTTAGCATAATTGCATTCCCAGC	AGATGTGGGGTCATCTCACTGTT	Valid
scaffold1544	1526516	A	G	TTTCTCCCCTTTCTTCACCTTTC	CAGGCCTGGGTTCCTACTACTTT	Valid
scaffold44	12983620	A	G	TCATTCATTCATTCATGCAAACC	TTGCAATCTCCACTCATTCTTCA	Valid
scaffold1545	2467666	C	G	AGTGGTTGACCTCTCTGTTGAGG	AATTGTGACATTTGTGCATCCTG	Valid
scaffold140	6858641	G	A	TAGGCAGGAATTTATGTAGCCCA	AGGAAAGAGAGGGAAAGAAAGGG	Valid
scaffold36	7381721	T	C	CAGACTCCACACTGCCAACATAG	GAAGAAATCGCAGGAAGACAAAA	Valid
scaffold450	54872	T	C	CCAGGGTAGGACAACTGCTAATG	AGGAGCAGATATGAATGCTGAGG	Valid
scaffold55	6397965	G	A	GACAACAACTTATGATCTGGGGC	GGTGTGTGTGTGTGTTGTGTTGT	Valid
scaffold154	2065870	T	C	TGCTTTCAGTTCAAAACAGTCCA	GGACTTAGGTCAGCAAGCAAAGA	Valid
scaffold154	15027008	A	G	ACTTGTTGGGTTGCTCCTAATCA	TAGTCCATGCAGAAGAGGGAAAG	Valid
scaffold28	16428997	T	G	TTTCATTGTGCCAAACATGAGTC	TGAGGCACAGATGACTCGAATAG	Valid
scaffold28	14976886	C	T	CCATACATGGGATAAAGTGGCAT	GTGCCCCTGTCATTACTTTTGAT	Valid
scaffold55	5949709	A	G	GACAGAGGAAATTGGCAGAGAGA	CACTGCCCAAATGTCATACTCAG	Valid
scaffold1466	2307023	G	A	AAAGCAGAGAAGCAGCAGAAAAA	CCATCACTGAACACCTGGTATGA	Valid
scaffold209	5230856	A	G	AACTTATTTGACCACAGAGCCCA	ATGAGGTCAGGGACATTTTCAGA	Valid
scaffold291	5928984	G	C	TGCTCTGTACCTTTCCTTGCTTC	ACAAGGACATGCCTCAGAGTAGG	Valid
scaffold66	383710	G	A	AATTTTTGAGGAACCTCCAGAGC	CATACGGGAGGTAAAGGGATTTC	Valid
scaffold66	1173812	T	C	CCATGCCCAGGTTTAACTGTAAG	ACACACAGATGGGTAGGAATGCT	Valid
scaffold589	202985	G	T	TTGACTTTGGATTGTCTTACCCG	AGCAGTGGTCACCTTCAGACAG	Invalid
scaffold108	3043398	C	T	CGTTCAAGAGTGGTTGACATCAC	CTGCTACCATCCTCGTATTCAGC	Valid
scaffold1495	2082842	T	C	ACAGACACAGGCCTCATCATACA	GGGATGAACTTCCTTCTCTTGGT	Valid
scaffold1521	3765007	A	G	CCTCATCAGTGCCTTGTTAGCTT	ACTTGAAGCACTTTGCAGGATGT	Valid
scaffold127	4048573	T	C	TATACCCACCTAGCGACCAAAGA	TGCTGTAGGCTATCACTGCTCTG	Valid
scaffold159	1766548	A	G	TATATGATTATGATCGGTGGGGG	GCCAGTTCTACTCCTGTGCTTGT	Valid
scaffold119	202468	C	A	GTCTACTGGGGAGAACACGAAGA	CAAATGAATGGACAGATGGATCA	Valid
scaffold96	4314581	A	G	AACACATTTGCAAGAAGGTGGAT	GCTTCTCCACATCACTCTTCACA	Valid
scaffold6	452759	A	G	CAGAGCCCACTGGTTCAGTTAAT	GATGCTGTGTTCCTCACAGACAC	Valid
scaffold1470	962388	G	A	TGTCCACAATGTGTGGTCTCTCT	TGGGCTCTGGAACAGAGTAAAAC	Valid
scaffold1470	3964793	G	A	GGATATGTGCCATATTCCTCAGC	TTTTTGCTCCTTTGTCTCCCATA	Valid
scaffold34	10117124	C	A	CCAGGACCCCAAGGTTTTTAAT	TGGGAAATACAGGCTTCCAGTTA	Valid
scaffold296	1502666	G	C	GACATGAAGCATAAGACATACGCA	TAATCCTCCCTGATTTTCCTTCC	Valid
scaffold214	850545	A	G	TGAGGATATGTGTGTGGAGACAG	CAGCAGTGGAGAGAATGCTTTG	Valid
scaffold1499	13935005	C	A	ACTCTGTGCCAGCCACTATCCTA	TTCCAAGTGGAGGATGGATCTAA	Valid
scaffold188	3041173	G	A	CCTTCAATGGCTCTCTTTGTCAT	CTCGAGAAAGTGCACAAAATTCA	Valid
scaffold99	5314488	T	C	AGCCCAACCACAAGATGTCTAAA	CTTACAGAGGACCCATCACCAAC	Valid
scaffold26	5799966	C	T	ATTAAAGTCTGGGATGAGGAGCC	TTTCACCTGCCTAGCAAAAAGAG	Valid
scaffold80	7918908	G	T	AACTAACCATCACTCTCCTGCCA	GTTATCGTTGGGACATTTTTGGA	Invalid
scaffold130	97757	G	A	TCAAGGAAAGACTGTCTGCACTG	GCTGTTGCTTGCTGAGACACTTA	Valid
scaffold103	5404043	C	T	CGATTTCAGAAGGAATAAGCCCT	TTTCTAACTCACCTCCCCATCAA	Valid
scaffold103	1181159	A	G	TTCCCTCCCTACTGTTTGTTTGA	TCATCACAACCTGCAAGAATTTG	Valid
scaffold80	8909665	G	A	AATGCAGCTCTATTGGCAAGAAC	ATGACCACCCACAATCTCTAAGC	Invalid
scaffold406	448629	G	A	CCATAGATCCTTGCAGTGACAGA	CTTTGCCAAAAAGAAAGCACTGT	Valid
scaffold406	206792	C	G	TTGGTGAGGAAATGGAACTCTGT	ATTCCTGTCACACATCCCTTCAT	Valid
scaffold1528	2348013	T	C	CTGACAGTTTCTGCCGATTTCTT	CTGTGATGGTTGTGCTCCTACAG	Valid
scaffold190	234363	C	T	TTCATTCCTTTTCGTTGCTGAAT	GATATGGAGAAATTGGGACCCTC	Valid
scaffold37	8154853	A	T	ACATTGCTTTTTCTCCTGCTGAC	GCCAAATAATAACCATCTCATTCCA	Valid
scaffold29	1880518	C	T	TCCAGACAGATAAGCCACATCAA	GCACCACTCCACCTAAGTCAATC	Valid
scaffold218	1115175	T	C	TTGCTGATCAGTTGACCTTGAAA	CGTTCTCTTAATGGTTCTGTGGG	Valid
scaffold83	10654280	G	A	ATACCCACCATACAGCTCAGGAA	AGCTTGAGTGAGAAGCTCTGGAA	Valid
scaffold1470	6503038	T	C	CAGGAGAGTGGAACGAAGTCTGT	ACAGATAAAGGGTCCCACAAATG	Valid
scaffold15	4603302	A	G	ACGTCGGTCAAGATCACAAACTT	GCCAGACCCAATTGACAGATAAC	Valid
scaffold34	7457992	T	C	GTACGTGGGAAACGGAAGTACAG	TCAAAAGCCTAGAAATGGACAGC	Valid
scaffold63	3590345	C	G	AATTTGTGCAAGTTCTAGCGGAG	CACCACTTTTTCACACTTATGTTTTG	Valid
scaffold29	3774771	A	C	ATGGAGCCAGAAGATACATTCCA	GGAATCTGACCATGATGGAGAAC	Valid

Supplementary Table S10 | Evaluation of completeness of the tiger genome assembly using core mammals gene mapping approach (CEGMA)

Parameter		Number	Percent (%)
Total KOGs		441	
One KOG align one gene		412	93.42
One KOG align one gene	overlap>0.8	321	72.79
	overlap>0.5	394	89.34
One KOG align several genes		3	0.68
One KOG align no gene		26	5.9

Note: KOG is mammalian orthologous gene sequences

Supplementary Table S11 | Statistics regarding mapping of tiger raw reads to the cat genome (Felis_catus-6.2**)**

	Total bases (except N)	no depth filter	≥5 depth	coverage (no depth filter)	coverage (up to 5 depth)
chrA1	233,159,525	229,329,978	224,952,444	0.9836	0.9648
chrA2	163,356,921	160,639,547	156,946,924	0.9834	0.9608
chrA3	137,575,932	135,439,549	132,493,776	0.9845	0.9631
chrB1	198,861,767	195,550,553	191,512,332	0.9833	0.963
chrB2	148,785,170	146,238,354	143,241,571	0.9829	0.9627
chrB3	142,933,721	140,469,301	137,315,865	0.9828	0.9607
chrB4	138,891,385	136,252,630	133,185,093	0.981	0.9589
chrC1	215,112,611	211,894,281	207,644,444	0.985	0.9653
chrC2	152,468,354	150,184,565	147,210,325	0.985	0.9655
chrD1	112,035,993	109,789,647	107,245,456	0.9799	0.9572
chrD2	85,466,620	84,108,810	82,288,361	0.9841	0.9628
chrD3	91,285,072	89,709,860	87,527,445	0.9827	0.9588
chrD4	91,542,936	89,934,189	87,768,719	0.9824	0.9588
chrE1	58,807,627	57,599,556	55,726,247	0.9795	0.9476
chrE2	59,775,008	58,413,217	56,555,516	0.9772	0.9461
chrE3	39,319,762	38,453,200	37,118,737	0.978	0.944
chrF1	67,430,147	66,304,962	64,714,706	0.9833	0.9597
chrF2	81,618,509	80,386,026	78,764,822	0.9849	0.965
chrX	119,642,964	114,115,672	108,427,336	0.9538	0.9063
Avg.	-	-	-	0.9809	0.9564

Supplementary Table S12 | Statistics regarding mapping of tiger raw reads to the tiger scaffolds

Total Reads	Mapped reads	Properly mapped reads	Singletons	Unmapped	Coverage
1,143,615,842	1,085,799,891 (94.94%)	1,070,162,056 (93.58%)	6,900,923 (0.60%)	57,815,951 (5.05%)	44.84

Supplementary Table S13 | Variation statistics regarding mapping of tiger raw reads to the tiger scaffolds

	Homozygous	Heterozygous	Total
SNVs	719,258	1,262,207	1,981,465
INDELs	103,065	289,117	392,182
Total	822,323	1,551,324	2,373,647

Supplementary Table S14 | Statistics regarding consensus sequence generated by mapping tiger raw reads to the cat genome (Felis_catus-6.2**)**

	Avg. length	Max. length
chrA1	5,129	96,817
chrA2	4,804	83,984
chrA3	5,018	107,166
chrB1	4,813	73,953
chrB2	4,861	80,217
chrB3	4,706	84,378
chrB4	4,714	143,963
chrC1	5,100	80,899
chrC2	5,053	84,737
chrD1	4,763	107,739
chrD2	5,067	92,522
chrD3	4,786	81,672
chrD4	4,792	94,269
chrE1	3,982	109,356
chrE2	4,031	82,936
chrE3	4,060	75,689
chrF1	4,744	81,693
chrF2	5,207	62,817
chrX	2,902	87,216

Supplementary Table S15 | Variation statistics regarding mapping of tiger raw reads to the cat genome (Felis_catus-6.2**)**

	Homozygous	Heterozygous	Total
SNVs	43,407,200	1,471,052	44,878,252
INDELs	5,784,116	209,366	5,993,482
Total	49,191,316	1,680,418	50,871,734

Supplementary Table S16 | Statistics regarding the Amur tiger draft chromosomes

	Total bases	# of A,C,G,T	# of N	# of scaffolds	Avg. length of scaffolds	Max. length of scaffold	Min. length of scaffold
chrA1	243,492,181	232,275,858	11,216,323	36	6,560,146	32,696,139	298,198
chrA2	190,495,254	163,297,928	27,197,326	40	4,179,683	13,728,702	57,285
chrA3	144,011,757	134,866,479	9,145,278	22	6,254,790	15,081,062	324,270
chrB1	222,683,385	201,742,861	20,940,524	27	7,616,073	23,356,477	136,373
chrB2	154,295,958	147,227,104	7,068,854	28	5,363,328	25,943,428	144,758
chrB3	150,246,213	142,454,758	7,791,455	17	8,576,661	41,605,936	128,514
chrB4	144,888,701	139,336,917	5,551,784	30	4,744,080	20,060,599	158,178
chrC1	223,586,761	214,870,707	8,716,054	35	6,258,464	17,690,132	162,664
chrC2	160,670,131	154,115,635	6,554,496	24	6,534,066	23,545,505	154,141
chrD1	125,709,129	108,826,731	16,882,398	32	3,488,721	22,133,838	41,474
chrD2	87,703,667	85,139,235	2,564,432	16	5,417,977	21,439,391	29,738
chrD3	103,759,264	90,419,476	13,339,788	28	3,297,702	16,120,292	118,778
chrD4	97,290,273	89,103,429	8,186,844	23	3,972,896	21,005,123	44,215
chrE1	66,408,731	60,756,433	5,652,298	21	3,000,195	10,647,627	159,720
chrE2	64,743,307	58,729,907	6,013,400	24	2,541,384	17,289,880	36,796
chrE3	47,874,673	39,154,156	8,720,517	22	1,848,219	6,319,830	67,679
chrF1	68,695,903	66,668,743	2,027,160	20	3,411,870	18,887,390	80,919
chrF2	91,576,383	85,144,388	6,431,995	23	3,774,408	11,788,466	97,428
chrX	142,585,357	112,864,775	29,720,582	103	1,162,958	9,521,833	60,170
total	2,530,717,028	2,326,995,520	203,721,508	571	-	-	-

Supplementary Table S17 | Genome comparison between Amur tiger and the domestic cat reference sequence (Felis_catus-6.2**)**

	Amur tiger	Domestic cat
# of chromosome	19	19
Total Bases	2,530,717,028	2,428,540,393
# of A	681,551,092	399,279,516
# of T	682,242,750	399,368,055
# of G	481,692,618	289,203,563
# of C	481,509,060	289,055,332
# of N	203,721,508	1,051,633,927
GC Content (%)	41.4	42.0

Supplementary Table S18 | Raw read filtering statistics

Sample	The number of raw read pairs	The number of proper read pairs	The percentage of proper read pairs	Estimated sequencing depth from raw read pairs	Estimated sequencing depth from proper read pairs
White tiger	427,476,607	365,433,486	85.49	32.06	27.41
Snow leopard	539,322,392	445,497,709	82.60	40.45	33.41
African lion	487,495,504	417,666,943	85.68	36.56	31.33
White lion	417,957,994	362,046,532	86.62	31.35	27.15

Supplementary Table S19 | Statistics regarding predicted protein-coding genes

Gene set		Number	Average transcript length (bp)	Average CDS length (bp)	Average no. of exons per gene	Average exon length (bp)	Average intron length (bp)
De novo	AUGUSTUS	20,682	52,995	1,394	8.85	158	6,574
	GENSCAN	45,131	37,076	1,257	7.91	159	5,182
Homolog	*H.sapiens*	19,912	25,350	1,392	8.14	171	3,355
	C.familiaris	24,432	20,165	1,193	7.15	167	3,085
	M.musculus	22,150	21,940	1,263	7.41	170	3,227
	F.catus	21,858	19,190	1,169	6.64	176	3,194
EST		8,214	183,578	560	2.10	267	166,627
cDNA		985	20,988	984	5.48	180	4,470
GLEAN		17,929	48,142	1,628	9.49	171	5,476
Synteny		16,907	38,144	1,598	9.66	165	4,220
Final		20,226	35,544	1,377	8.09	170	4,818

Supplementary Table S20 | Summary of evidence from GLEAN gene models

	≥20% overlap		≥50% overlap		≥80% overlap	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
P (single)	15	0.084	389	2.170	2,240	12.494
P (more)	2,193	12.232	2,388	13.319	2,603	14.518
H (single)	2	0.011	14	0.078	90	0.502
H (more)	4	0.022	39	0.218	367	2.047
C (single)	0	0	1	0.006	2	0.011
C (more)	0	0	0	0	0	0
P+H	15,118	84.321	14,608	81.477	11,895	66.345
P+C	10	0.056	12	0.067	19	0.106
H+C	11	0.061	18	0.100	29	0.162
P+H+C	556	3.101	372	2.075	239	1.333

P, *ab initio* prediction; H, homology-based; C, cDNA/EST expressed genes. According to the number of gene sources providing support, the evidence was further separated into single (one gene source) and multiple (two or more gene sources). The overlap threshold is relative to the CDS region of GLEAN genes.

Supplementary Table S21 | Statistics regarding domestic cat's predicted protein-coding genes

Gene set		Number	Average transcript length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	*AUGUSTUS*	20,563	50464.82	1440.33	8.88	162.17	6220.17
	GENSCAN	41,339	40403.16	1333.8	8.27	161.31	5375.22
Homolog	*H.sapiens*	26,628	32224.44	1566.59	8.8	178.04	3930.94
	C.familiaris	25,693	26239.48	1486	8.48	175.23	3309.23
	B.taurus	24,847	27404.2	1506.94	8.41	179.15	3494.1
Final		19,048	36841.82	1600.06	9.28	172.35	4254.37

The average transcript lengths do not contain UTR. Three approaches were employed in gene prediction: Homolog (*H.sapiens*, *C.familiaris* and *B.taurus*), *De novo* (*GENSCAN*, *AUGUSTUS*).

Supplementary Table S22 | Summary of evidence for the domestic cat's final gene models

	≥20% overlap		≥50% overlap		≥80% overlap	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
P (single)	574	3.01	771	4.05	911	4.78
P (more)	364	1.91	181	0.95	45	0.24
H (single)	485	2.55	605	3.18	1335	7.01
H (more)	727	3.82	1229	6.45	4051	21.26
P+H	16898	88.71	16262	85.37	12706	66.71

P: *ab initio* prediction; H: homology-based; According to number of gene sources support, the evidence was further separated into single (with one gene source) and more (with two or more gene sources). The overlap threshold is relative to the CDS region of final gene sets.

Supplementary Table S23 | The number of genes with homology or functional classification by each method

		Number	Percent (%)
Total		20,226	
Annotated	InterPro	15,554	76.901
	GO	13,004	64.293
	KEGG	13,854	68.496
	SwissProt	18,004	89.014
	TrEMBL	18,143	89.701
Unannotated		1,963	9.705

Supplementary Table S24 | Summary of non-coding RNA in the tiger genome

Type		Copy Number	Average length (bp)	Total length (bp)	% of genome
miRNA		418	91	38,036	0.00158
tRNA		259	76	19,759	0.00082
rRNA	rRNA total	554	85	47,112	0.00196
	18S	22	165	3,627	0.00015
	28S	98	126	12,346	0.00051
	5.8S	1	52	52	0.00000
	5S	433	72	31,087	0.00129
snRNA	snRNA total	1,704	118	201,017	0.00835
	CD-box	313	94	29,365	0.00122
	HACA-box	235	139	32,769	0.00136

Supplementary Table S25 | Statistical analysis of gene families

Species	Total Genes number	Unclustered genes	Family number	Unique families	Average genes per family
Tiger	20,226	2,514	14,954	20	1.18
Human	21,642	1,167	15,911	109	1.29
Dog	19,281	432	15,605	11	1.21
Cat	19,048	497	15,421	25	1.20
Mouse	22,843	1,877	15,743	93	1.33
Panda	19,329	375	15,895	4	1.19
Opossum	19,448	1,304	15,787	210	1.15

Supplementary Table S26 | Genes in Felidae-specific gene family. A total of 129 genes are assigned in 58 different Felidae-specific gene families.

Genes in Felidae-specific gene family	Genes in Felidae-specific gene family
PTIG0005478.1, cat_01603	PTIG0007338.1, cat_01016
PTIG0004080.1, cat_00807	PTIG0013176.1, cat_01225
PTIG0019786.1, cat_05610	PTIG0018813.1, cat_01269
PTIG0014681.1, cat_08399	PTIG0011961.1, cat_16017
PTIG0014101.1, cat_04260	PTIG0020205.1, cat_16461
PTIG0004303.1, cat_00338	PTIG0000566.1, cat_04688
PTIG0004940.1, PTIG0009531.1, cat_01806	PTIG0009197.1, cat_08154
PTIG0015008.1, PTIG0010995.1, PTIG0012338.1, PTIG0003385.1, PTIG0004166.1, PTIG0003566.1, PTIG0004164.1, cat_12851	PTIG0006488.1, cat_10203
PTIG0018602.1, cat_14155	PTIG0004862.1, cat_07017
PTIG0016645.1, cat_09480	PTIG0009495.1, cat_04143
PTIG0007799.1, cat_15068	PTIG0011593.1, cat_04237
PTIG0016141.1, cat_13610	PTIG0015503.1, cat_05586
PTIG0019220.1, cat_18831, cat_06725	PTIG0011646.1, cat_15439
PTIG0008984.1, cat_04413	PTIG0014861.1, cat_03537
PTIG0019060.1, cat_01242	PTIG0015331.1, cat_09489, cat_07161
PTIG0011514.1, cat_01290	PTIG0009505.1, cat_07332
PTIG0006997.1, cat_07950	PTIG0007398.1, cat_07832
PTIG0019120.1, cat_11128	PTIG0010939.1, PTIG0004547.1, cat_17792
PTIG0009682.1, cat_12934	PTIG0001747.1, cat_10703
PTIG0002569.1, PTIG0007587.1, cat_13825	PTIG0003962.1, cat_13561
PTIG0004544.1, cat_15616, cat_08459	PTIG0015324.1, cat_00414
PTIG0014604.1, cat_04496	PTIG0015315.1, cat_09140
PTIG0003766.1, cat_18849	PTIG0009646.1, cat_10121
PTIG0020137.1, cat_18654, cat_13317	PTIG0008016.1, cat_15105
PTIG0011319.1, cat_03126	PTIG0006026.1, cat_09024
PTIG0000704.1, cat_12747	PTIG0016903.1, cat_03839
PTIG0008020.1, cat_11391	PTIG0006398.1, cat_09378
PTIG0005273.1, cat_15562	PTIG0012210.1, cat_10034
PTIG0014045.1, cat_17444	PTIG0018881.1, cat_02261

Supplementary Table S27 | Genes in tiger/cat-specific gene family. A total of 52 genes are assigned in 20 different tiger-specific gene families, and 100 genes are assigned in 25 different cat-specific gene families.

Genes in tiger-specific gene family	Genes in cat-specific gene family
PTIG0007979.1, PTIG0008027.1, PTIG0008031.1	cat_10935, cat_04582
PTIG0018115.1, PTIG0018618.1, PTIG0018632.1	cat_04437, cat_01868
PTIG0014679.1, PTIG0013125.1	cat_18486, cat_14701
PTIG0009547.1, PTIG0015344.1, PTIG0015362.1	cat_13006, cat_09913
PTIG0007343.1, PTIG0011597.1, PTIG0000689.1, PTIG0005374.1, PTIG0010764.1	cat_03160, cat_06448, cat_03956, cat_11634, cat_06466, cat_02522, cat_06190, cat_01557, cat_06351, cat_09112, cat_00979, cat_09905, cat_14303, cat_16288, cat_04532, cat_18794, cat_14610, cat_01305, cat_06329, cat_17981, cat_12417, cat_08720, cat_11304, cat_01389, cat_08951, cat_01772, cat_01185, cat_00548, cat_02995, cat_17460, cat_07960, cat_04058
PTIG0016220.1, PTIG0015481.1, PTIG0000295.1	cat_17526, cat_18361, cat_17226
PTIG0011713.1, PTIG0012584.1	cat_10104, cat_13903, cat_04809
PTIG0009878.1, PTIG0001570.1, PTIG0005044.1, PTIG0005301.1	cat_16374, cat_15880
PTIG0005534.1, PTIG0016635.1, PTIG0016644.1	cat_17229, cat_03791, cat_18456, cat_09731, cat_15515, cat_01473, cat_09815, cat_10824, cat_16082, cat_10165, cat_18354, cat_18957, cat_09925, cat_11495
PTIG0011076.1, PTIG0019744.1, PTIG0019745.1	cat_11181, cat_09719
PTIG0003344.1, PTIG0016123.1	cat_07366, cat_08869, cat_18414, cat_14192
PTIG0015852.1, PTIG0020184.1	cat_16310, cat_10570, cat_14898
PTIG0007855.1, PTIG0010213.1	cat_16069, cat_14310
PTIG0017862.1, PTIG0009264.1	cat_18701, cat_04928
PTIG0017229.1, PTIG0001594.1	cat_05570, cat_15670, cat_11451
PTIG0014914.1, PTIG0003330.1	cat_14710, cat_13984
PTIG0001297.1, PTIG0001298.1	cat_18615, cat_18759, cat_14143
PTIG0020199.1, PTIG0020188.1, PTIG0020195.1	cat_14078, cat_09197
PTIG0012448.1, PTIG0012449.1	cat_03866, cat_15372
PTIG0006789.1, PTIG0006276.1	cat_05363, cat_17719, cat_10826
	cat_02985, cat_12816
	cat_01230, cat_15506
	cat_15407, cat_08875
	cat_00343, cat_05829
	cat_12686, cat_06619

Supplementary Table S28 | Annotated domains of Felidae-specific protein families

InterProScan ID	Description	# of domains in tiger & cat	P-value
IPR007087	Zinc finger, C2H2	42	8.09E-14
IPR015880	Zinc finger, C2H2-like	23	2.07E-09
IPR013087	Zinc finger C2H2-type/integrase DNA-binding domain	17	5.06E-07
IPR013783	Immunoglobulin-like fold	12	4.86E-05
IPR007110	Immunoglobulin-like	12	0.00201
IPR000679	Zinc finger, GATA-type	7	3.23E-03
IPR001487	Bromodomain	5	2.77E-02
IPR002453	Beta tubulin	5	0.02767
IPR003599	Immunoglobulin subtype	7	0.06505
IPR003961	Fibronectin, type III	4	0.08292
IPR007707	Transforming acidic coiled-coil	4	0.08292
IPR012722	T-complex protein 1, zeta subunit	4	0.08292
IPR013106	Immunoglobulin V-set	4	0.08292
IPR020831	Glyceraldehyde/Erythrose phosphate dehydrogenase family	4	0.08292
IPR003598	Immunoglobulin subtype 2	6	0.1013
IPR000217	Tubulin	3	1.85E-01

Supplementary Table S29 | Annotated domains of Amur tiger-specific protein families (Fisher's exact test using a conservative 5% false-discovery-rate criterion)

InterProScan ID	Description	# of domains in Amur tiger	P-value
IPR003596	Immunoglobulin V-set, subgroup	6	3.62E-08
IPR003599	Immunoglobulin subtype	6	3.62E-08
IPR013106	Immunoglobulin V-set	6	3.62E-08
IPR007110	Immunoglobulin-like	6	4.31E-07
IPR003598	Immunoglobulin subtype 2	3	0.000258
IPR003380	Transforming protein Ski	2	0.005102
IPR005818	Histone H1/H5	2	0.005102
IPR005819	Histone H5	2	0.005102
IPR018533	Forkhead box protein, C-terminal	2	0.005102
IPR020810	Enolase, C-terminal	2	0.005102

Supplementary Table S30 | GO enrichment of genes that were expanded in Amur tiger

GO_ID	GO_Term	GO_Class	Pvalue	Adjusted Pvalue	# of genes
GO:0004984	olfactory receptor activity	MF	5.75E-185	3.89E-182	289
GO:0003735	structural constituent of ribosome	MF	4.70E-134	1.59E-131	145
GO:0005840	Ribosome	CC	1.91E-133	4.32E-131	145
GO:0030529	ribonucleoprotein complex	CC	1.72E-124	2.91E-122	148
GO:0007186	G-protein coupled receptor signaling pathway	BP	2.98E-106	4.04E-104	302
GO:0006412	Translation	BP	3.75E-100	4.23E-98	150
GO:0004888	transmembrane signaling receptor activity	MF	1.43E-97	1.39E-95	293
GO:0005198	structural molecule activity	MF	7.80E-97	6.60E-95	169
GO:0007166	cell surface receptor signaling pathway	BP	1.14E-93	8.58E-92	306
GO:0043232	intracellular non-membrane-bounded organelle	CC	9.29E-77	4.84E-75	192
GO:0004871	signal transducer activity	MF	2.25E-74	1.02E-72	295
GO:0034645	cellular macromolecule biosynthetic process	BP	2.08E-70	8.79E-69	280
GO:0010467	gene expression	BP	4.38E-67	1.65E-65	275
GO:0050794	regulation of cellular process	BP	6.45E-58	2.30E-56	450
GO:0065007	biological regulation	BP	7.60E-58	2.57E-56	460
GO:0044249	cellular biosynthetic process	BP	1.92E-53	5.64E-52	291
GO:0009058	biosynthetic process	BP	4.45E-50	1.21E-48	294
GO:0023052	Signaling	BP	1.88E-49	4.89E-48	330
GO:0007165	signal transduction	BP	6.52E-49	1.63E-47	327
GO:0032991	macromolecular complex	CC	8.55E-48	2.03E-46	214
GO:0016021	integral to membrane	CC	8.71E-48	2.03E-46	322
GO:0009987	cellular process	BP	3.47E-46	7.84E-45	752
GO:0007154	cell communication	BP	2.90E-45	6.13E-44	330
GO:0044444	cytoplasmic part	CC	2.23E-40	4.44E-39	155
GO:0044425	membrane part	CC	3.18E-37	6.15E-36	330
GO:0050896	response to stimulus	BP	9.32E-35	1.75E-33	334
GO:0005622	Intracellular	CC	1.29E-27	2.30E-26	429
GO:0044464	cell part	CC	3.54E-23	5.70E-22	440
GO:0005737	Cytoplasm	CC	7.30E-22	1.15E-20	160
GO:0044260	cellular macromolecule metabolic process	BP	2.81E-21	4.33E-20	341
GO:0043229	intracellular organelle	CC	3.78E-19	5.56E-18	247
GO:0043170	macromolecule metabolic process	BP	5.43E-17	7.83E-16	367
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	MF	6.44E-14	8.72E-13	20
GO:0006351	transcription, DNA-dependent	BP	9.04E-14	1.20E-12	122
GO:0044267	cellular protein metabolic process	BP	1.38E-13	1.76E-12	194
GO:0016020	Membrane	CC	4.69E-13	5.88E-12	364
GO:0015934	large ribosomal subunit	CC	7.35E-12	7.78E-11	14
GO:0006355	regulation of transcription, DNA-dependent	BP	1.18E-11	1.19E-10	113
GO:0044424	intracellular part	CC	2.71E-10	2.44E-09	256
GO:0006865	amino acid transport	BP	3.09E-10	2.75E-09	16
GO:0019538	protein metabolic process	BP	5.72E-10	5.03E-09	220
GO:0044237	cellular metabolic process	BP	1.49E-09	1.26E-08	355
GO:0015171	amino acid transmembrane transporter activity	MF	3.37E-09	2.79E-08	16
GO:0008270	zinc ion binding	MF	5.57E-09	4.54E-08	173
GO:0003676	nucleic acid binding	MF	7.41E-09	5.77E-08	202
GO:0000786	Nucleosome	CC	1.22E-08	9.18E-08	23
GO:0034622	cellular macromolecular complex assembly	BP	1.57E-08	1.17E-07	32
GO:0044238	primary metabolic process	BP	5.84E-08	4.16E-07	384
GO:0005525	GTP binding	MF	7.04E-08	4.91E-07	54
GO:0003964	RNA-directed DNA polymerase activity	MF	1.02E-07	6.93E-07	8
GO:0016070	RNA metabolic process	BP	1.41E-07	9.29E-07	125
GO:0006457	protein folding	BP	2.65E-07	1.68E-06	29
GO:0006334	nucleosome assembly	BP	4.65E-07	2.74E-06	23
GO:0006278	RNA-dependent DNA replication	BP	9.64E-07	5.48E-06	8
GO:0046914	transition metal ion binding	MF	1.82E-06	1.02E-05	181
GO:0090304	nucleic acid metabolic process	BP	2.40E-06	1.32E-05	147
GO:0005874	Microtubule	CC	0.00022326	0.001028195	11
GO:0044430	cytoskeletal part	CC	0.00030319	0.0013684	24
GO:0008199	ferric iron binding	MF	0.00040817	0.001760065	8
GO:0006826	iron ion transport	BP	0.00056703	0.002399233	8
GO:0006879	cellular iron ion homeostasis	BP	0.00056703	0.002399233	8
GO:0019725	cellular homeostasis	BP	0.00085292	0.003542512	16
GO:0006352	transcription initiation, DNA-dependent	BP	0.00132693	0.005417668	9
GO:0008152	metabolic process	BP	0.0017736	0.006745654	413
GO:0004594	pantothenate kinase activity	MF	0.00229384	0.008627375	3

Supplementary Table S31 | Statistics regarding mapping to tiger scaffolds

Samples	All reads	Mapped reads	Unmapped reads	Mapped read percentage	Unmapped read percentage	Average mapping depth
Amur tiger	873,618,858	854,642,533	18,976,325	97.83	2.17	32
White tiger	730,866,972	716,487,035	14,379,937	98.03	1.97	25
Snow leopard	890,995,418	782,670,464	108,324,954	87.84	12.16	24
African lion	835,333,886	774,106,594	61,227,292	92.67	7.33	29
White lion	724,093,064	665,693,631	58,399,433	91.93	8.07	24

Supplementary Table S32 | Variant sites in *Panthera* species for unique amino acid changes and genetic diversity

Species	All variant sites	Homozygous SNV sites	Heterozygous SNV sites	Indel sites	homoSNV sites in coding region	heteroSNV sites in coding region	Indel sites in coding region
Amur tiger	2,025,421	524,138	1,171,350	329,933	5,410	8,786	824
White tiger	4,090,852	1,582,058	1,755,504	753,290	7,862	9,240	865
Snow leopard	15,850,394	12,956,696	555,451	2,338,247	87,570	6,418	3,080
African lion	19,159,093	15,071,123	1,406,491	2,681,479	97,239	12,254	3,485
White lion	18,632,694	14,900,837	1,149,507	2,582,350	92,318	10,477	3,118

Supplementary Table S33 | Pathway analysis for big cat-specific genes having functional changes

KEGG pathway	Genes having functional changes	P-value
Propanoate metabolism	ACSS1,SUCLG2,MLYCD,EHHADH,ABAT,ACACB,HADHA,ALDH3A2	0.000032
Inositol phosphate metabolism	CDIPT,PIK3C2G,IMPA1,PIP5KL1,PIK3C2A,INPP5J,PLCG2,INPP4B,PIP5K1A,INPP5B	0.000092
Phosphatidylinositol signaling system	CDIPT,PIK3C2G,IMPA1,PIK3C2A,INPP5J,PLCG2,INPP4B,PIP5K1A,INPP5B,PIK3R2	0.00017
Histidine metabolism	DDC,ABP1,HDC,ALDH1A3,HAL,ALDH3A2	0.000242
Fatty acid metabolism	ACADVL,CPT2,ACADS,EHHADH,ACADL,HADHA,ALDH3A2,ACSBG1	0.000384
beta-Alanine metabolism	MLYCD,EHHADH,ALDH1A3,ABAT,HADHA,ALDH3A2	0.000775
alpha-Linolenic acid metabolism	PLB1,PLA2G4E	0.002141
mTOR signaling pathway	EIF4B,TSC1,STK11,ULK2,EIF4E2,PIK3R2,DDIT4	0.00242
mRNA surveillance pathway	PPP2R1B,SYMPK,UPF2,CSTF2,SMG6,RNPS1,NXF1,CASC3,CPSF1,DAZAP1	0.002569
Glycerophospholipid metabolism	ACHE,CDIPT,PLB1,PPAP2C,AGPAT9,LYPLA1,AGPAT4,PLA2G4E,AGPAT2	0.002654
Butanoate metabolism	ALDH5A1,ACADS,EHHADH,ABAT,HADHA	0.003381
Adipocytokine signaling pathway	IRS4,PRKCQ,STK11,LEPR,PRKAG1,PRKAG2,ACACB,IRS1,AGRP,ACSBG1,TRADD	0.005188
Nicotinate and nicotinamide metabolism	NNT,BST1,NT5C2,NAPRT1	0.005768
Biotin metabolism	HLCS	0.006898
Valine, leucine and isoleucine biosynthesis	LARS2,IARS2	0.00943
Selenocompound metabolism	SEPSECS,TXNRD2,MARS2	0.009687
Base excision repair	LIG1,NEIL2,NEIL1,MBD4,XRCC1,SMUG1	0.009855
Insulin signaling pathway	IRS4,PRKCZ,EXOC7,PRKAG1,PRKAG2,PDE3A,ACACB,IRS1,PYGM,TSC1,SHC1,TRIP10, EIF4E2,PIK3R2	0.011139
Lysine degradation	DLST,EHHADH,TMLHE,MLL3,NSD1,HADHA,ALDH3A2	0.011253
Glycerolipid metabolism	GK2,PPAP2C,AGPAT9,AGPAT4,AGPAT2,ALDH3A2	0.011846
Phenylalanine metabolism	DDC,ALDH1A3,PAH	0.013593
Ether lipid metabolism	PLB1,PPAP2C,PLA2G4E	0.018371
RNA transport	UPF2,RNPS1,NXF1,CASC3,EIF4B,EIF3D,NUP62,POP1,DDX20,TGS1,EIF2B2,EIF2B3,EIF4E2, GEMIN5,EIF2B5,NUP210L	0.021246
Fat digestion and absorption	APOA4,PPAP2C,PLA2G4E,AGPAT2,MTTP	0.024835
SNARE interactions in vesicular transport	SNAP29,STX2,VAMP7,USE1,GOSR1	0.024835
VEGF signaling pathway	PLCG2,SPHK1,NFATC4,MAPKAPK2,PLA2G4E,PIK3R2	0.026286
ABC transporters	ABCB9,ABCB8,ABCB10,ABCC8,ABCB6,ABCG4,ABCA5	0.026381
Aminoacyl-tRNA biosynthesis	SEPSECS,MARS2,EPRS,LARS2,IARS2,AARS2,MTFMT	0.029849
Sulfur relay system	TST,NFS1	0.032811
Linoleic acid metabolism	PLB1,PLA2G4E	0.032811
Valine, leucine and isoleucine degradation	ACADS,EHHADH,BCKDHB,ABAT,HADHA,ALDH3A2	0.034466
Fc gamma R-mediated phagocytosis	DNM1L,PPAP2C,PLCG2,SPHK1,PIP5K1A,PRKCE,PLA2G4E,PIK3R2	0.035663
Thiamine metabolism	NFS1	0.036962
Cysteine and methionine metabolism	TST,DNMT3A,BHMT,AHCYL2	0.037283
Folate biosynthesis	GGH,FPGS	0.044049
Nucleotide excision repair	RPA1,ERCC6,XPC,LIG1,ERCC4,ERCC1	0.04418
PPAR signaling pathway	ACOX2,SLC27A1,CPT2,GK2,EHHADH,ACADL,CYP8B1,ACSBG1	0.048026

Supplementary Table S34 | Pathway analysis for feline-specific genes having functional changes

KEGG pathway	Genes having functional changes	P-value
ECM-receptor interaction	ITGB7,ITGA3,RELN,SV2A,THBS3	0.000742
Inositol phosphate metabolism	PIK3C2B,PLCD1,INPP4B	0.002649
Phosphatidylinositol signaling system	PIK3C2B,PLCD1,INPP4B	0.003373
Focal adhesion	ITGB7,GSK3B,ITGA3,SHC1,RELN,BIRC3,THBS3	0.004653
Primary immunodeficiency	JAK3,ADA,BLNK	0.004695
Endocytosis	RABEP1,AP2A1,PDCD6IP,PSD2,ARAP1,DNM2	0.017982
Endocrine and other factor-regulated calcium reabsorption	AP2A1,DNM2	0.018888
Sphingolipid metabolism	GALC,SMPD3	0.020693
Pentose and glucuronate interconversions	DCXR	0.022482
Pathogenic Escherichia coli infection	TUBA4A,ABL1	0.022589
Amino sugar and nucleotide sugar metabolism	PGM3,UXS1	0.026656
Glycosphingolipid biosynthesis - ganglio series	ST6GALNAC6	0.030216
Hedgehog signaling pathway	GSK3B,GLI3	0.03109
ErbB signaling pathway	GSK3B,SHC1,ABL1	0.031777
Viral myocarditis	MYH11,ABL1	0.033443
Lysine degradation	PLOD3,SETD1A	0.03842
Melanogenesis	EP300,ADCY7,GSK3B	0.038865
Basal cell carcinoma	GSK3B,GLI3	0.041043
Collecting duct acid secretion	SLC12A7	0.043466
Pentose phosphate pathway	H6PD	0.043466
Pathways in cancer	FGF18,EP300,GSK3B,ITGA3,BIRC3,ABL1,GLI3,RALGDS	0.048299

Supplementary Table S35 | Statistical orthologous gene numbers

Species	orthologous gene numbers
Tiger:Cat:Human	7,415
Shared with Mouse	6,155
Shared with Dog	6,365
Shared with Panda	6,062

Note: Genes were required to fall into regions of large scale synteny between genomes, and to have completely aligned coding regions of human annotation, not to have frame-shift indels or altered gene structures, and not to show signs of recent duplication.

Supplementary Table S36 | Pathway analysis using positively selected genes

KEGG pathway	P-value	Positively selected genes
Endocrine and other factor-regulated calcium reabsorption	0.000421	AP2A2,PLCB4,CLTB
Drug metabolism - cytochrome P450	0.001581	FMO1,CYP2D6
Carbohydrate digestion and absorption	0.002447	PIK3CB,HK2
Acute myeloid leukemia	0.002468	PIK3CB,PML,RPS6KB2
Bacterial invasion of epithelial cells	0.002673	CLTB,PIK3CB,ARHGAP10
Nitrogen metabolism	0.005284	CA9,CA5A
Inositol phosphate metabolism	0.006009	PLCB4,PIK3CB
mTOR signaling pathway	0.006009	PIK3CB,RPS6KB2
Phosphatidylinositol signaling system	0.007212	PLCB4,PIK3CB
Type II diabetes mellitus	0.007863	PIK3CB,HK2
Amoebiasis	0.009544	PLCB4,ACTN4,PIK3CB
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	0.010047	ACTN4,DSG2,ITGA2B
Insulin signaling pathway	0.016048	PIK3CB,HK2,RPS6KB2
Systemic lupus erythematosus	0.016285	HIST1H4L,HIST1H4K,ACTN4
Hypertrophic cardiomyopathy (HCM)	0.016749	MYH7,TPM4,ITGA2B
Cytosolic DNA-sensing pathway	0.020658	TREX1,ZBP1
Dilated cardiomyopathy	0.02136	MYH7,TPM4,ITGA2B
Calcium signaling pathway	0.022762	SLC8A3,PLCB4,TNNC2,PTAFR
Olfactory transduction	0.023162	OR2A1,OR7C1,OR10Q1,OR1J4
TGF-beta signaling pathway	0.026948	RPS6KB2,TFDP1
Endocytosis	0.027812	AP2A2,RAB11FIP3,CLTB,PML
Fc gamma R-mediated phagocytosis	0.031161	PIK3CB,RPS6KB2
ErbB signaling pathway	0.040566	PIK3CB,RPS6KB2
Cardiac muscle contraction	0.042259	MYH7,TPM4

Supplementary Table S37 | Identification of rapidly evolving categories

	Higher Ka/Ks				
Threshold	0.05	0.01	0.001	1.00E-04	1.00E-05
Observed[a]	106	86	64	52	41
Expected[b]	129	86	52	33	22
P-value[c]	0.9061	0.5041	0.1732	0.055	0.0266

a: number of significant categories; b: average number of significant categories identified in 10,000 random sets; c: proportion of random sets which have as many or more categories as observed in the dataset.

Supplementary Table S38 | Identification of slowly evolving categories

	Lower Ka/Ks								
Threshold	0.05	0.01	0.001	1.00E-04	1.00E-05	1.00E-06	1.00E-07	1.00E-08	1.00E-09
Observed[a]	67	66	63	55	49	42	31	27	22
Expected[b]	98	70	51	41	36	32	22	17	15
P-value[c]	0.9832	0.619	0.1918	0.15	0.142	0.112	0.103	0.089	0.0456

a: number of significant categories; b: average number of significant categories identified in 10,000 random sets; c: proportion of random sets which have as many or more categories as observed in the data set.

Supplementary Table S39 | Lineage-specific GO analysis between tiger and cat

	Tiger			Cat		
Threshold	0.05	0.01	0.001	0.05	0.01	0.001
Observed[a]	116	53	25	44	24	18
Expected[b]	77	42	10	28	11	6
p-value[c]	0.1253	0.011	0.0088806	0.0988	0.067628	0.032034

a: number of significant categories; b: average number of significant categories identified in 10,000 random sets; c: proportion of random sets which have as many or more categories as observed in the data set.

Supplementary Table S40 | Results of rapidly evolving categories

#GO ID	gene number	GO name	GO Categories	d_N/d_S	Amino Acid divergence	P-value
GO:0004930	223	G-protein coupled receptor activity	molecular_function	0.2896	0.024327	1.49E-152
GO:0004984	135	olfactory receptor activity	molecular_function	0.3056	0.036221	3.16E-129
GO:0016021	1316	integral to membrane	cellular_component	0.208	0.010151	8.25E-95
GO:0005886	1062	plasma membrane	cellular_component	0.183	0.009573	1.43E-47
GO:0005576	457	extracellular region	cellular_component	0.2163	0.010388	9.53E-35
GO:0009897	57	external side of plasma membrane	cellular_component	0.2959	0.015072	9.63E-25
GO:0007275	113	multicellular organismal development	biological_process	0.248	0.01187	1.94E-23
GO:0005615	252	extracellular space	cellular_component	0.2047	0.010882	6.60E-21
GO:0007283	97	spermatogenesis	biological_process	0.2501	0.011805	1.04E-19
GO:0004872	149	receptor activity	molecular_function	0.2235	0.009598	2.97E-17
GO:0006954	95	inflammatory response	biological_process	0.2472	0.012773	4.07E-17
GO:0007218	32	neuropeptide signaling pathway	biological_process	0.3392	0.012868	1.90E-15
GO:0005777	36	peroxisome	cellular_component	0.3353	0.012317	2.37E-14
GO:0006955	105	immune response	biological_process	0.2405	0.011106	3.45E-14
GO:0007186	142	G-protein coupled receptor signaling pathway	biological_process	0.1987	0.010216	3.50E-14
GO:0006805	44	xenobiotic metabolic process	biological_process	0.2893	0.01268	7.60E-14
GO:0004867	27	serine-type endopeptidase inhibitor activity	molecular_function	0.3106	0.016726	3.93E-13
GO:0060333	22	interferon-gamma-mediated signaling pathway	biological_process	0.289	0.013204	4.63E-11
GO:0010951	23	negative regulation of endopeptidase activity	biological_process	0.316	0.015587	9.36E-11
GO:0005887	343	integral to plasma membrane	cellular_component	0.1736	0.00869	1.51E-10
GO:0004888	26	transmembrane signaling receptor activity	molecular_function	0.2634	0.016773	1.62E-10
GO:0020037	36	heme binding	molecular_function	0.2559	0.011981	3.91E-10
GO:0005882	25	intermediate filament	cellular_component	0.2651	0.014729	2.01E-09
GO:0000166	162	nucleotide binding	molecular_function	0.203	0.00913	2.99E-09
GO:0050900	36	leukocyte migration	biological_process	0.2586	0.010026	4.86E-08
GO:0016324	67	apical plasma membrane	cellular_component	0.2343	0.008645	5.61E-08
GO:0009055	65	electron carrier activity	molecular_function	0.2356	0.009831	7.47E-08
GO:0006935	40	chemotaxis	biological_process	0.2377	0.012394	9.28E-08
GO:0005179	28	hormone activity	molecular_function	0.3315	0.014263	1.46E-07
GO:0006281	93	DNA repair	biological_process	0.2179	0.007908	2.25E-07
GO:0003674	199	molecular_function	molecular_function	0.1894	0.009107	3.15E-07
GO:0045087	94	innate immune response	biological_process	0.2057	0.010231	3.39E-07
GO:0016042	23	lipid catabolic process	biological_process	0.2752	0.011518	4.90E-07
GO:0005929	21	cilium	cellular_component	0.2794	0.011625	5.26E-07
GO:0005789	216	endoplasmic reticulum membrane	cellular_component	0.1869	0.008438	6.06E-07
GO:0005792	85	microsome	cellular_component	0.208	0.008901	7.54E-07
GO:0005575	150	cellular_component	cellular_component	0.1968	0.009303	1.03E-06
GO:0005125	52	cytokine activity	molecular_function	0.2345	0.011391	1.43E-06
GO:0008150	171	biological_process	biological_process	0.189	0.008756	1.90E-06
GO:0004252	46	serine-type endopeptidase activity	molecular_function	0.2233	0.010051	2.12E-06
GO:0007204	33	elevation of cytosolic calcium ion concentration	biological_process	0.1874	0.011275	6.28E-06

Supplementary Table S41 | Results of slowly evolving categories

#GO ID	gene number	GO name	GO Categories	d_N/d_S	Amino Acid divergence	P-value
GO:0010467	242	gene expression	biological_process	0.0732	0.002782	2.84E-35
GO:0007268	130	synaptic transmission	biological_process	0.0785	0.003756	2.59E-17
GO:0045202	40	synapse	cellular_component	0.0613	0.0036	2.64E-15
GO:0000398	68	nuclear mRNA splicing, via spliceosome	biological_process	0.0584	0.002194	7.54E-15
GO:0016070	107	RNA metabolic process	biological_process	0.0645	0.002365	1.93E-14
GO:0007399	101	nervous system development	biological_process	0.075	0.003109	2.09E-14
GO:0008286	49	insulin receptor signaling pathway	biological_process	0.0709	0.00573	3.22E-14
GO:0008543	45	fibroblast growth factor receptor signaling pathway	biological_process	0.0576	0.003835	1.05E-13
GO:0016071	99	mRNA metabolic process	biological_process	0.0589	0.002168	4.51E-13
GO:0008380	74	RNA splicing	biological_process	0.0619	0.002072	1.65E-12
GO:0006414	53	translational elongation	biological_process	0.0435	0.00266	1.03E-11
GO:0016044	37	cellular membrane organization	biological_process	0.0528	0.00237	2.17E-11
GO:0043161	26	proteasomal ubiquitin-dependent protein catabolic process	biological_process	0.0345	0.001183	3.02E-11
GO:0000209	47	protein polyubiquitination	biological_process	0.0667	0.002629	3.12E-11
GO:0046777	47	protein autophosphorylation	biological_process	0.0752	0.003911	3.29E-11
GO:0060021	21	palate development	biological_process	0.0553	0.002122	7.58E-11
GO:0005200	23	structural constituent of cytoskeleton	molecular_function	0.0649	0.003025	2.25E-10
GO:0007409	25	axonogenesis	biological_process	0.0517	0.00283	2.48E-10
GO:0004713	26	protein tyrosine kinase activity	molecular_function	0.0594	0.003794	2.57E-10
GO:0014069	31	postsynaptic density	cellular_component	0.058	0.001994	5.75E-10
GO:0016192	58	vesicle-mediated transport	biological_process	0.0699	0.002574	6.15E-10
GO:0018279	37	protein N-linked glycosylation via asparagines	biological_process	0.0652	0.002484	6.91E-10

Supplementary Table S42 | Tiger-specific GOs under rapid evolution

Go ID	Go name	Taxonomy	Gene Number	Tiger Mean *Ka/Ks*	Cat Mean *Ka/Ks*	P-value
GO:0004930	G-protein coupled receptor activity	molecular_function	223	0.211163	0.120473	1.75E-23
GO:0004984	olfactory receptor activity	molecular_function	135	0.210356	0.117013	3.85E-23
GO:0005886	plasma membrane	cellular_component	1062	0.052482	0.050925	1.16E-13
GO:0015629	actin cytoskeleton	cellular_component	72	0.115289	0.109031	1.36E-11
GO:0012505	endomembrane system	cellular_component	24	0.062953	0.029472	1.59E-11
GO:0006112	energy reserve metabolic process	biological_process	41	0.062886	0.00336	5.20E-09
GO:0007275	multicellular organismal development	biological_process	113	0.24135	0.218099	1.05E-08
GO:0008021	synaptic vesicle	cellular_component	22	0.076001	0.040245	1.99E-08
GO:0000166	nucleotide binding	molecular_function	162	0.215402	0.185111	3.12E-08
GO:0055085	transmembrane transport	biological_process	157	0.130818	0.063701	9.61E-08
GO:0006355	regulation of transcription, DNA-dependent	biological_process	304	0.173607	0.119945	2.35E-07
GO:0042470	Melanosome	cellular_component	44	0.072953	0.029109	2.50E-06
GO:0006351	transcription, DNA-dependent	biological_process	353	0.141428	0.108387	8.92E-06
GO:0006879	cellular iron ion homeostasis	biological_process	30	0.094467	0.024621	2.34E-05
GO:0030018	Z disc	cellular_component	27	0.095806	0.044286	3.15E-05
GO:0019717	Synaptosome	cellular_component	37	0.090089	0.041846	6.11E-05
GO:0043154	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	biological_process	25	0.147305	0.086255	6.21E-05
GO:0043066	negative regulation of apoptotic process	biological_process	94	0.126623	0.100882	7.11E-05
GO:0005938	cell cortex	cellular_component	32	0.092195	0.055152	0.000301
GO:0006184	GTP catabolic process	biological_process	73	0.128157	0.04869	0.000319
GO:0045202	Synapse	cellular_component	40	0.075984	0.042005	0.000377
GO:0003924	GTPase activity	molecular_function	103	0.108368	0.045003	0.000377
GO:0005096	GTPase activator activity	molecular_function	31	0.242979	0.06166	0.000682
GO:0016055	Wnt receptor signaling pathway	biological_process	51	0.11763	0.057197	0.000742
GO:0007601	visual perception	biological_process	61	0.152073	0.006502	0.000998
GO:0003729	mRNA binding	molecular_function	20	0.162308	0.019308	0.001081
GO:0031410	cytoplasmic vesicle	cellular_component	41	0.107544	0.065301	0.001301
GO:0016020	Membrane	cellular_component	238	0.169225	0.157351	0.001624
GO:0005856	Cytoskeleton	cellular_component	111	0.103067	0.0828	0.001631
GO:0004871	signal transducer activity	molecular_function	80	0.104259	0.003981	0.00166
GO:0016071	mRNA metabolic process	biological_process	99	0.073283	0.037783	0.002634
GO:0016070	RNA metabolic process	biological_process	107	0.078623	0.043412	0.002813
GO:0000910	Cytokinesis	biological_process	23	0.101798	0.056073	0.002986
GO:0008217	regulation of blood pressure	biological_process	21	0.08874	0.039484	0.003069
GO:0009790	embryo development	biological_process	28	0.209814	0.068383	0.003398
GO:0051082	unfolded protein binding	molecular_function	60	0.116632	0.104256	0.004178
GO:0005200	structural constituent of cytoskeleton	molecular_function	23	0.075316	0.036237	0.004301
GO:0034142	toll-like receptor 4 signaling pathway	biological_process	26	0.063347	0.054847	0.004442
GO:0007399	nervous system development	biological_process	101	0.081574	0.056249	0.005229
GO:0006936	muscle contraction	biological_process	34	0.081264	0.021875	0.005345
GO:0006397	mRNA processing	biological_process	53	0.120032	0.050146	0.005753
GO:0017124	SH3 domain binding	molecular_function	38	0.179972	0.111585	0.005851
GO:0005794	Golgi apparatus	cellular_component	258	0.112461	0.059338	0.00632
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	biological_process	29	0.113454	0.072839	0.006446
GO:0034130	toll-like receptor 1 signaling pathway	biological_process	24	0.058732	0.051953	0.006446
GO:0008286	insulin receptor signaling pathway	biological_process	49	0.068434	0.020738	0.006969
GO:0003682	chromatin binding	molecular_function	68	0.154143	0.097193	0.007268
GO:0003676	nucleic acid binding	molecular_function	121	0.195874	0.162948	0.008097
GO:0043434	response to peptide hormone stimulus	biological_process	29	0.26784	0.095789	0.008599
GO:0035023	regulation of Rho protein signal transduction	biological_process	25	0.140319	0.132507	0.008624
GO:0034641	cellular nitrogen compound metabolic process	biological_process	84	0.128567	0.114524	0.009101
GO:0009408	response to heat	biological_process	21	0.14973	0.073249	0.009132
GO:0030659	cytoplasmic vesicle membrane	cellular_component	26	0.142113	0.074746	0.00948

Supplementary Table S43 | Cat-specific GOs under rapid evolution

Go ID	Go name	Taxonomy	Gene Number	Tiger Mean *Ka/Ks*	Cat Mean *Ka/Ks*	P-value
GO:0003723	RNA binding	molecular_function	198	0.155977	0.175731	5.35E-07
GO:0008624	induction of apoptosis by extracellular signals	biological_process	37	0.110118	0.151265	0.000196
GO:0016021	integral to membrane	cellular_component	1316	0.08882	0.119716	4.89E-15
GO:0005515	protein binding	molecular_function	1836	0.05162	0.116463	0.000186
GO:0005737	cytoplasm	cellular_component	1430	0.046448	0.048348	0.000187
GO:0005789	endoplasmic reticulum membrane	cellular_component	216	0.023564	0.146209	8.20E-06
GO:0008284	positive regulation of cell proliferation	biological_process	120	0.00593	0.087667	2.12E-07
GO:0007411	axon guidance	biological_process	104	0.005821	0.060511	1.45E-07
GO:0005578	proteinaceous extracellular matrix	cellular_component	56	0.005096	0.110785	1.31E-05
GO:0031012	extracellular matrix	cellular_component	45	0.004464	0.102696	3.77E-05
GO:0019904	protein domain specific binding	molecular_function	66	0.00401	0.061782	5.47E-13
GO:0005769	early endosome	cellular_component	50	0.003822	0.082039	1.88E-08
GO:0045121	membrane raft	cellular_component	46	0.003775	0.113407	1.24E-06
GO:0010628	positive regulation of gene expression	biological_process	41	0.003627	0.122742	2.12E-06
GO:0005770	late endosome	cellular_component	39	0.002661	0.120469	2.12E-05
GO:0000902	cell morphogenesis	biological_process	21	0.002241	0.043471	1.99E-13
GO:0009952	anterior/posterior pattern specification	biological_process	29	0.002084	0.089508	2.18E-08
GO:0030182	neuron differentiation	biological_process	29	0.001993	0.047827	3.25E-10

Supplementary Table S44 | Unique amino acid changes validation by Sanger sequencing

Gene	scaffold	position	Allele change	Amino acid change	Forward Primer	Reverse Primer	SNV status
EGLN1 (snow leopard)	scaffold219	34956771	T>A	K39M	CTACTCACTTACTTTCG TCCGGC	AAGCTGGCGCTGGAGTA CAT	Valid
EPAS1 (snow leopard)	scaffold81	86375291	G>A	V663I	AAGTATCAACAGCAGCT GGGAAG	GATTACAGTGGGAGTCT GCTGCT	Valid
EPAS1 (snow leopard)	scaffold81	86377097	T>C	C794R	GGCATTGGTCTTTCAGA TGAGTT	GCCCCTGCTGTAGTAAA AATCTG	Valid
TYR (white lion)	scaffold1503	26676209	G>A	R87Q	AGAGCCTGATGGAGAA GGAATG	CGCTAAAGTGAGGTAG GCAAGAA	Valid

Supplementary Table S45 | Sample information to verify the observed mutation in *EGLN1*

Species	Individual I.D.	Sex	Source of Sample	Location Born	Subspecies
Tiger	Pti33	M	CERI, Dr. Mitch Bush	Captive	
	Pti109	M	Smithsonian National Zoological Park, Dr. Mitch Bush	Captive	P.t. sumatrae
	Pti111	F	Amur Carnivore Project, Dr. Howard Quigley	Wild, Russia	P.t. altaica
	Pti113	F	Amur Carnivore Project, Dr. Howard Quigley	Wild, Russia	P.t. altaica
	Pti117	F	Amur Carnivore Project, Dr. Howard Quigley	Wild, Russia	P.t. altaica
	Pti138	M	Amur Carnivore Project, Dr. Howard Quigley	Wild, Russia	P.t. altaica
	Pti148	M	Minnesota Zoological Gardens, Dr. Frank Wright	Captive, Russia	P.t. altaica
	Pti217	M	Suzhuo Zoo	Wild, China	
	Pti257	M	Singapore Zoological Gardens, Dr. Paolo Martelli	Captive, Indonesia	P.t. corbetti
	Pti289	F	Busch Gardens Zoological Park, Dr. Ray L. Ball	Captive	
Snow leopard	Pun6	M	San Antonio Zoological Gardens	Captive	
	Pun10	F	New York Zoological Park	Captive	
	Pun16	M	Philadelphia Zoological Garden	Captive	
	Pun17	F	Cheyenne Mountain Zoo	Captive	
	Pun19	M	Cheyenne Mountain Zoo	Captive	
	Pun28	M	Cheyenne Mountain Zoo	Captive	
	Pun29	M	Cheyenne Mountain Zoo	Captive	
	Pun30	F	Cheyenne Mountain Zoo	Captive	
	Pun35	M	John Ball Zoological Garden	Captive	
	Pun63	M	PUEBLO ZOO	Captive	
	Pun83	F	Tallinn Zoological Park, Dr. Vladimir Fainstein	Wild, Kyrgyzstan	
	Pun89	F	Urumqi Zoo, Mr. Han Bin	Wild, Xinjian, China	
	Pun91	M	Lanzhou Zoo, Mr. Zhao Guorui	Wild, Gansu, China	
	Pun92	F	Novosibirsk Zoo	Wild, Novosibirsk, Russia	
Leopard	Ppa50	F	Zoologischer Garten Berlin, Dr. Goltenboth	Captive	P.p. melas
	Ppa113	M	National Zoological Gardens Dehiwela, Sri Lanka, Dr. Sriyanie Miththapala	Captive	P.p. delacouri
	Ppa145	M	Tallinn Zoo Park, Dr. Vladimir Fainstein	Captive	P.p. orientalis
	Ppa147	M	Leningrad Zoo Park, Dr. Ivan Korneev	Captive	P.p. saxicolor
	Ppa203	Unk	Welsh Mountain Zoo, Dr. Nick Jackson	Captive	P.p. saxicolor
Clouded leopard	Nne32	M	Smithsonian National Zoological Park, Dr. Mitch Bush	Captive, Unknown	
	Nne52	M	Samut Prakan Crocodile Farm, Dr. JoGayle Howard	Wild, Thailand	N.n. nebulosa
	Nne82	F	Negara Zoo	Wild, Borneo	N.n. diardi
Lion	Ple184	F	Sakkarbaug Zoo, Dr. Rawal	Captive, India	
	Ple185	F	Sakkarbaug Zoo, Dr. Rawal	Wild, India	
	Ple262	F	Messerli Veterinary Disease Project, Dr. Melody Roelke	Wild, Tanzania	
	Ple275	F	Serengeti Lion Project, Dr. Craig Packer	Wild, Tanzania	
	Ple287	M	Serengeti Lion Project, Dr. Craig Packer	Wild, Tanzania	
	Ple295	F	Serengeti Lion Project, Dr. Craig Packer	Wild, Tanzania	
	Ple584	M	Messerli Veterinary Disease Project, Dr. Melody Roelke	Wild, Tanzania	
	Ple614	F	Messerli Veterinary Disease Project, Dr. Melody Roelke	Wild, Tanzania	
	Ple865	Unk	Dr. Kathy Alexander	Wild, Botswana	
	Ple1039	F	Dr. Kathy Alexander	Wild, Botswana	

Supplementary Table S46 | Overview of samples and associated genotypes. Full location detail for the lion populations: Ukutula Lodge & Lion Centre, North West Province, South Africa (Ukutala); Tsau Conservancy, Greater Timbavati, Limpopo, South Africa (Tsau); Johannesburg Zoo, Johannesburg, Gauteng, South Africa (Johannesburg); Tiger's Preserve, Myrtle Beach, North Carolina, United States of America (Myrtle); Ouwehands Dierenpark, Rhenen, Netherlands (Ouwehands); Etosha National Park, Namibia (Etosha); Kgalagadi Transfrontier Park, Northern Cape, South Africa (Kgalagadi); Greater Mapungubwe Transfrontier Conservation Area, Limpopo, South Africa (Mapungubwe); Kruger National Park, South Africa (Kruger). Genotype nomenclature: wildtype (wt); white (w).

Animal ID	Sex	Date of Birth	Phenotype	Population	Sire (if known)	Dam (if known)	Expected genotype	Justification	176C>T	260G>A
									T59M	R87Q
FEL1200067		24-May-11	Tawny	Ukutula	UM013	VL03159 (white)	w/wt	White dam	C T	G A
FEL1200070		24-May-11	Tawny	Ukutula	UM013	VL03159 (white)	w/wt	White dam	C T	G A
VL03158	F	25-Feb-06	White	Ukutula	VL1000932		w/w	White phenotype	C C	A A
VL03159	F	25-Aug-07	White	Ukutula	VL1000932		w/w	White phenotype	C C	A A
VL03160	F	25-Jan-08	White	Ukutula	VL1000932		w/w	White phenotype	C C	A A
VL03161	F	25-Jan-08	White	Ukutula	VL1000932		w/w	White phenotype	C C	A A
VL03162	F	25-Aug-07	Tawny	Ukutula	VL1000932				C C	G A
VL03163	F	25-Jan-08	Tawny	Ukutula	VL1000932				C C	G A
VL03164	M	01-Nov-05	Tawny	Ukutula	VL1000932				C C	G A
VL03165	M	07-May-08	Tawny	Ukutula					T T	G G
VL0900357	F	23-Apr-09	White	Ukutula	VL1000932		w/w	White phenotype	C C	A A
VL0900359	F	25-Nov-08	Tawny	Ukutula					C C	G G
VL0900360	F	16-Feb-09	Tawny	Ukutula	VL1000932				C C	G G
VL0900362	M	16-Feb-09	Tawny	Ukutula	VL1000932				C C	G G
VL0900364	M	24-Mar-09	Tawny	Ukutula	VL1000932				C C	G A
VL0900366	M	24-Mar-09	White	Ukutula	VL1000932		w/w	White phenotype	C C	A A
VL0900367	M	23-Apr-09	Tawny	Ukutula	VL1000932				C C	G A
VL1000926			Tawny	Ukutula					C C	G A
VL1000928	F	<2010	Tawny	Ukutula					C C	G G
VL1000929	F	<2010	Tawny	Ukutula					C C	G G
VL1000932	M	31-Aug-01	Tawny	Ukutula			w/wt	White offspring (6)	C C	G A
VL1100440	F	2011	Tawny	Ukutula			w/wt	White dam	C T	G A
VL1100441	F	2011	Tawny	Ukutula			w/wt	White dam	C T	G A
J1	M	1998	White	Tsau			w/w	White phenotype	C C	A A
J2	F	2003	White	Tsau	Triton (white)		w/w	White phenotype	C C	A A
J5	F	05-Jul	Tawny	Tsau					C T	G G
J6	F	05-Jul	Tawny	Tsau					C T	G G
J7	F	18-Feb-08	White	Tsau	J1 (white)	J2 (white)	w/w	White phenotype	C C	A A
J8	M	18-Feb-08	White	Tsau	J1 (white)	J2 (white)	w/w	White phenotype	C C	A A
Dharma	F	Sep-00	Tawny	Johannesburg			w/wt	White offspring (3)	C C	G A
Niobe	F	15-May-01	Tawny	Johannesburg	Naas	Simone	w/wt	White offspring (3)	C C	G A
Numzaan	M	31-Jul-99	Tawny	Johannesburg	Wotan (white)	Scar	w/wt	White sire, offspring (1)	C C	G A
Nyanga	F	18-Feb-01	White	Johannesburg	Thor (white)	Scarlet2	w/w	White phenotype	C C	A A
Sabre	F	20-Mar-99	Tawny	Johannesburg	Wotan (white)	Scar	w/wt	White sire, offspring (1)	C C	G A
Shumba	F	05-Feb-99	Tawny	Johannesburg	Thor (white)	Vagti	w/wt	White sire, offspring (1)	C C	G A
Triton	M	27-Nov-99	White	Johannesburg	Thor (white)	Vagti	w/w	White phenotype	C C	A A
Vidor	M	27-Nov-99	Tawny	Johannesburg	Thor (white)	Vagti	w/wt	White sire, offspring (1)	C C	G A
Ivory	M	08-Nov-08	White	Myrtle			w/w	White phenotype	C C	A A
ZA1	M	<2008	White	Ouwehands			w/w	White phenotype	C C	A A
FEL1200015	F	2010	Tawny	Etosha			wt/wt	control population	C C	G G
LN1200015	F	<2008	Tawny	Kgalagadi			wt/wt	control population	C T	G G
VL1105822	M	Jan/Feb-08	Tawny	Tuli block			wt/wt	control population	C C	G G
VL1105806	F	07-Sep	Tawny	Kruger			wt/wt	control population	C C	G G
PLE171			Tawny	Kruger			wt/wt	control population	C C	G G
FL00708		<2004	White				w/w	White phenotype	C C	A A
FL00709		<2004	White				w/w	White phenotype	C C	A A
FL00710		<2004	White				w/w	White phenotype	C C	A A

Supplementary Table S47 | Concordance between expected genotype and obtained genotype. For each identified polymorphism (column 1), the obtained genotype (column 2) is compared to the expected genotype (Columns 3-6), and the level of concordance is indicated (Column 7).

Polymorphism	Genotype	Expected Genotype				Total individuals	Concordance
		wt/wt	w/wt	w/w	wt/_		
*TYR*260G>A	GG	5			8	13	100%
*TYR*260G>A	GA		11		6	17	100%
*TYR*260G>A	AA			17		17	100%
*TYR*176C>T	TT				1	1	100%
*TYR*176C>T	TC	1	4		2	7	86%
*TYR*176C>T	CC	4	7	17	11	39	44%
Total individuals:		5	11	17	14	47	

Supplementary Table S48 | Primer sequences for the amplification and sequencing of the white lion candidate mutation (*TYR*260G>A)

Primer name	Primer use	Primer sequence
TYR_ex1_1F	PCR and sequencing	GATCCGTGAAGACGAGGGTA
TYR_ex1_1R	PCR and sequencing	TCGCTTCTCTGTGCAGTTTG
TYR_e1_white lion_F	Semi-nested PCR	CTGTCCAGCGTGGACGGGTG

Supplementary Table S49 | Summary of evolutionarily conserved regions in the tiger genome

Total coverage (bp)	% Genome coverage	All gene coverage (bp)	% Gene coverage
77,102,208	3.2	1,375,111	0.19

Supplementary Table S50 | Summary of Segmental Duplications

Cutoff	Block #	Median size (bp)	Genome coverage (Mb)
>1kb	10,691	1,668	11.15449
>5kb	627	6,110	2.45428
>10kb	32	11,577	0.30372
>50kb	0	0	0

Supplementary Table S51 | Statistics regarding synteny. "SourceIndex" indicates the species in which syntenic gaps were found. "-", "I" and "D" represent gaps, lineage-specific insertions, and lineage-specific deletions, respectively. Yellow boxes highlight species lineage-specific insertions, while light blue boxes indicate deletions. Orange and blue boxes highlight family lineage-specific insertions and deletions, respectively.

a. Lineage-specific insertions and deletions

Phylogenetic information

[-]sourceIndex	Tiger	Cat	Dog	Human	Mouse	Count
0,1,2,3	-	-	-	-	I	4311
0,1,2,4	-	-	-	I	-	1104
0,1,2	-	-	-	I	I	1396
0,1,3,4	-	-	I	-	-	870
0,1,3	-	-	I	-	I	22
0,1,4	-	-	I	I	-	26
0,1	-	-	I	I	I	732
0,2,3,4	-	I	-	-	-	130
0,2,4	-	I	-	I	-	5
0,2	-	I	-	I	I	3
0,3,4	-	I	I	-	-	7
0,3	-	I	I	-	I	2
0,4	-	I	I	I	-	2
0	-	I	I	I	I	122
1,2,3,4	-	D	D	D	D	104
1,2,3	-	D	D	D	-	6
1,2,4	-	D	D	-	D	2
1,2	-	D	D	-	-	4
1,3,4	-	D	-	D	D	10
1,3	-	D	-	D	-	1
1,4		D			D	4
1	-	D	-	-	-	290
2,3,4	-	-	D	D	D	656
2,3			D	D		29
2,4	-	-	D	-	D	17
2	-	-	D	-	-	899
3,4	-	-	-	D	D	898
3	-	-	-	D	-	1952
4	-	-	-	-	D	6833

Note: *P.tig* is the reference genome in the multiZ alignment.

b. Color index

Color Index:	Species specific	Family specific
Insertion	I	I
Deletion	D	D

Supplementary Table S52 | Statistics regarding lineage-specific insertions and deletions (indels). In accordance with the divergence time tree, we obtained the time to most recent common ancestor (TMRCA) for each selected species and tiger. We also obtained TMRCA of the Felidae family and dog. Then, the average mutation rate of indels was calculated for each species.

	Tiger	Cat	Dog	Human	Mouse	*Felidae*_specific
TMRCA (Myr)	10.00	10.00	58.00	103.00	103.00	58.00
Insertions/Myr	10.40	13.00	15.00	10.72	41.85	11.31
Deletions/Myr	12.20	29.00	15.50	18.95	66.34	12.62
InDels/Myr	22.60	42.00	30.50	29.67	108.19	23.93

Supplementary Table S53 | Synteny breaks between tiger and domestic cat genomes

Tiger scaffold	Tiger chromosome	Cat chromosome	Start of tiger scaffold	End of tiger scaffold	Start of cat chromosome	End of cat chromosome	Avg.%ID
scaffold20	chrB4	chrB4	1,488,246	20,061,016	107,012,319	125,668,453	93
		chrX	4,299	1,473,387	73,047,809	74,854,810	89
scaffold26	chrA2	chrA2	857	12,624,254	105,367,957	118,010,880	93
		chrB3	12,649,013	13,727,346	105,270,110	106,329,886	93
scaffold61	chrA1	chrA1	609,708	2,542,326	234,140,802	238,328,006	90
		chrX	628	599,391	79,386,885	80,185,908	91
scaffold65	chrA3	chrA3	570,477	8,927,705	41,881,778	50,215,314	93
		chrE2	2,177	567,493	10,192,520	10,736,574	89
scaffold79	chrB3	chrB3	191	10,377,140	92,373,737	102,799,738	92
		chrX	10,388,968	11,222,387	56,418,539	58,772,022	93
scaffold96	chrC1	chrB4	2,509	634,675	64,379,291	65,235,876	92
		chrC1	645,618	7,495,312	120,875,936	127,782,767	93
scaffold98	chrB3	chrA1	5,239,260	5,320,393	181,084,247	181,183,472	76
		chrB3	657	5,115,306	22,756,637	27,701,226	92
scaffold109	chrE1	chrC1	4,331,848	5,286,603	96,414,816	97,872,873	92
		chrE1	20,155	4,315,233	51,435,123	55,725,894	92
scaffold142	chrF2	chrD1	14,518	2,162,419	10,456,194	12,565,329	92
		chrF2	2,185,511	8,010,532	59,192,979	65,053,178	93
scaffold150	chrB4	chrA3	5,707	1,690,066	71,128,944	72,837,533	93
		chrB4	1,699,654	6,729,120	133,693,003	138,704,750	91
scaffold159	chrF2	chrA1	3,447	1,492,556	175,308,603	176,815,577	92
		chrF2	1,494,622	3,634,129	38,484,786	40,650,446	93
scaffold226	chrE1	chrD2	9,210	143,469	22,167,496	22,294,304	89
		chrE1	167,125	10,647,756	27,163,459	37,623,116	93
scaffold234	chrE2	chrE2	5,788	778,208	13,189,342	14,019,982	91
		chrX	788,018	1,258,410	93,928,288	94,374,940	90
scaffold244	chrB1	chrB1	2,259,197	13,085,878	98,571,018	109,514,462	93
		chrD4	9,509	2,253,940	23,158,891	25,425,318	94
scaffold249	chrX	chrA1	4,434,128	5,305,617	25,799	1,001,237	93
		chrX	6,508	4,419,938	16,444,294	20,883,251	93
scaffold255	chrA2	chrA2	7,445	3,711,360	165,389,623	169,035,055	89
		chrD1	3,724,398	3,819,954	1,008	96,401	93
scaffold316	chrB4	chrB4	547,449	5,181,807	67,423,010	72,063,516	93
		chrX	2,914	546,247	59,753,753	60,311,528	93
scaffold1461	chrE1	chrB4	4,108	1,505,181	64,806,639	66,619,091	93
		chrE1	1,525,183	10,248,697	12,742,949	21,439,819	93
scaffold1545	chrB4	chrB4	57,632	8,822,392	91,376,868	100,211,049	93
		chrD2	3,372	49,800	22,613,719	22,664,610	94
scaffold1486	chrA1	chrA1	1,913,973	16,285,132	184,023,829	194,847,457	93
		chrA1	5,206	1,809,904	178,533,363	180,353,532	93
		chrC2	16,292,822	16,882,400	85,664,281	86,435,248	93
scaffold309	chrA1	chrA1	5,183	702,778	233,427,809	234,129,317	92
		chrA1	926,032	1,722,635	232,378,826	233,179,930	91
scaffold1490	chrA3	chrA3	1,390	1,028,659	13,590,691	14,628,901	84
		chrA3	1,030,821	3,412,178	10,824,083	13,225,570	91
scaffold105	chrA3	chrA3	1,257,120	3,600,886	27,978,900	30,355,569	92
		chrA3	19,615	1,247,638	30,835,948	32,053,879	92
scaffold1454	chrB1	chrB1	7,329	3,503,883	201,360,351	204,776,126	88
		chrB1	3,513,656	6,645,199	191,986,897	195,134,304	91
scaffold402	chrB1	chrB1	117,867	571,931	204,786,020	205,238,072	90
		chrB1	1,332	109,668	199,133,595	199,244,639	90
scaffold424	chrD3	chrD3	1,041	1,278,552	15,827,979	17,101,090	93
		chrD3	1,296,519	1,529,629	26,422,068	26,653,975	91
scaffold191	chrX	chrX	11,887	7,852,486	20,904,805	28,839,740	92
		chrX	7,856,780	9,519,650	6,692,288	8,357,805	92
scaffold258	chrX	chrX	158,804	687,568	53,173,599	53,752,805	92
		chrX	9,237	137,872	49,922,761	50,066,411	86
scaffold268	chrX	chrX	899,215	1,018,438	65,088,932	65,198,446	90
		chrX	1,739	582,367	81,756,200	82,446,661	90

Supplementary Table S54 | Chromosomal rearrangement between tiger and domestic cat

	Intra-chromosomal rearrangement				Inter-chromosomal rearrangement			
Cutoff (Kb)	scaffold	Breakpoint	Tiger changed	Cat changed	scaffold	Breakpoint	Tiger changed	Cat changed
5	77	129	21	75	26	26	8	4
10	58	92	15	56	20	20	8	4
20	34	50	9	29	19	19	8	4
50	19	26	4	15	19	19	8	4
100	16	20	3	12	18	18	7	4
200	9	11	2	7	16	16	7	4
500	6	7	2	4	15	15	6	4

Note: Syntenic segment is defined as continuous regions without any order or orientation change. Different cutoffs of syntenic segment, from 5Kb to 500Kb, were used to find inter-chromosomal and intra-chromosomal breakpoints. The dog genome was used as an outgroup.

Supplementary Table S55 | Integrated chromosomal rearrangement between tiger and domestic cat. The SyMAP and LASTZ approaches were integrated, and genomic positions, strand, and changed species were derived from results of the LASTZ approach.

Type	Tiger scaffold	Cat chromosome	Start of tiger scaffold	End of tiger scaffold	Start of cat chromosome	End of cat chromosome	Cat strand	Species different from common ancestor	Number of supporting unique long mate pairs
Inter-chromosomal rearrangements	scaffold20	chrX	4,255	1,475,684	73,047,765	74,298,532	+	Tiger changed	
		chrB4	1,484,996	20,061,759	107,009,106	125,669,096	+		
	scaffold26	chrB3	29	1,083,891	105,267,370	106,332,703	+	Cat changed	
		chrA2	1,084,009	13,729,574	105,367,523	118,032,905	-		
	scaffold79	chrB3	18	10,383,675	92,373,564	102,806,254	+	Tiger changed	6
		chrX	10,387,784	11,224,900	57,926,544	58,773,198	-		
	scaffold96	chrB4	79	636,682	64,376,861	65,039,188	+	Tiger changed	
		chrC1	637,373	7,499,292	120,867,722	127,786,402	+		
	scaffold109	chrE1	1	4,329,577	51,414,147	55,740,335	+	Tiger changed	1
		chrC1	4,329,581	5,288,321	96,412,954	97,383,070	-		
	scaffold142	chrD1	11,160	2,165,537	10,452,855	12,568,494	+	Cat changed	
		chrF2	2,165,700	8,011,555	59,172,821	65,054,191	+		
	scaffold150	chrA3	1	1,691,025	71,123,232	72,838,560	+	Tiger changed	56
		chrB4	1,693,780	6,731,089	133,687,071	138,706,717	+		
	scaffold234	chrX	61,601	496,389	93,971,024	94,382,289	+	Tiger changed	
		chrE2	496,668	1,272,003	13,238,760	14,022,740	-		
	scaffold244	chrD4	615	2,256,466	23,156,337	25,434,092	-	Cat changed	
		chrB1	2,256,564	13,086,430	98,568,390	109,515,015	+		
	scaffold249	chrX	17,818	4,427,280	16,456,322	20,890,542	+	Cat changed	
		chrA1	4,433,908	5,308,825	22,480	904,041	-		
	scaffold1461	chrE1	19	8,732,391	12,741,537	21,447,335	-	Tiger changed	
		chrB4	8,750,421	10,255,625	65,040,005	66,550,172	-		
	scaffold1545	chrD2	470	55,011	22,608,581	22,667,397	-	Tiger changed	24
		chrB4	56,261	8,822,945	91,375,508	100,211,599	+		
Intra-chromosomal rearrangements	scaffold309	chrA1	3,364	938,577	233,189,587	234,130,994	-	Tiger changed	848
		chrA1	938,578	1,731,245	232,391,123	233,189,587	+		
	scaffold1490	chrA3	1,030,687	3,219,663	10,823,949	13,031,206	+	Tiger changed	1,127
		chrA3	3,219,664	3,347,235	13,031,206	13,158,446	-		
	scaffold1454	chrB1	2,329,233	2,566,080	192,563,359	192,800,276	+	Cat changed	
		chrB1	3,140,392	6,648,794	201,352,213	204,780,505	-		
	scaffold402	chrB1	2,291	460,825	204,781,970	205,238,638	-	Cat changed	455
		chrB1	464,212	574,771	199,132,694	199,245,940	+		
	scaffold424	chrD3	1	1,284,827	15,826,967	17,107,620	+	Cat changed	115
		chrD3	1,293,393	1,547,855	26,403,266	26,657,196	-		
	scaffold191	chrX	8,025	7,854,478	20,901,054	28,841,714	+	Tiger changed	
		chrX	7,856,448	9,522,475	6,689,573	8,358,191	-		

Supplementary Table S56 | Primer information for tiger scaffold integrity validation

scaffold	Forward Primer	Reverse Primer	Tiger scaffold integrity
scaffold150	CGGTGATGCCTCTTGATTGCCTG	ACGTGGGTCAGATGTCTCAGCGG	Valid
scaffold1545	CCTGCCAGTCTGCCCTGTAAGCA	TGTGGTGGAAGTTAGGGACCACCG	Valid
scaffold309	TGCAGGATGGATACCTGGGGAGG	ACAACCGGTCGACACAGAAGCCA	Valid
scaffold1490	GCAAGTGACAAGAGCCGAGGGGT	GAGGGAGAAGGCACCCGAAGTGA	Valid
scaffold402	GGCTCCCGCACGAGAGACAAGAT	GGTACGCACGCAACAACAGAGGC	Valid
scaffold424	ATCCCCCGAAGGTACCGCTGTCT	CAGCCACGTGCACCAGTTCTCAA	Valid

Supplementary Table S57 | Heterozygous SNV rates in several species

Species	Heterozygous SNV rate	The number of heterozygous SNVs	The number of heterozygous SNVs after Repeat masked	Genome size
Gorilla	0.00163	4,943,562	-	3,041,976,159
Orang-Utan (Sumatran)	0.00120	3,700,000	-	3,090,000,000
Giant panda	0.00112	2,682,349	-	2,400,000,000
Chimpanzee	0.00095	-	-	-
White tiger	0.00073	1,755,504	1,052,537	2,408,774,396
Naked mole rat	0.00068	1,870,000	-	2,744,000,000
Human (Korean)	0.00066	2,071,417	1,225,851	3,147,272,210
Orang-Utan (Bornean)	0.00065	2,000,000	-	3,090,000,000
African lion	0.00058	1,406,491	858,664	2,408,774,396
Amur tiger (TaeGeuk)	0.00049	1,171,350	724,381	2,408,774,396
White lion	0.00048	1,149,507	688,034	2,408,774,396
Tasmanian devil	0.00032	1,057,507	-	3,300,000,000
Snow leopard	0.00023	555,451	299,418	2,408,774,396
Domestic cat	0.00012	327,000	-	2,700,000,000

Supplementary Table S58 | Statistics regarding mapping to CAT.66 genome

Species	All reads	Mapped reads	Unmapped reads	Mapped read percentage	Unmapped read percentage	Average mapping depth
Amur tiger	873,618,858	517,734,353	355,884,505	59.26	40.74	27
White tiger	730,866,972	401,456,779	329,410,193	54.93	45.07	20
Snow leopard	890,995,418	452,625,291	438,370,127	50.80	49.20	24
African lion	835,333,886	458,641,686	376,692,200	54.91	45.09	20
White lion	724,093,064	383,655,225	340,437,839	52.98	47.02	19

Supplementary Table S59 | Variant sites in *Panthera* species for phylogenetic analysis

Species	All SNV sites	homoSNV sites	heteroSNV sites
Amur tiger	34,381,497	31,969,041	2,412,456
Snow leopard	34,014,630	31,953,087	2,061,543
African lion	27,621,623	25,447,786	2,173,837

Supplementary Methods

Genome sequencing and assembly

Sample preparation

The male Amur tiger (TaeGeuk), white tiger, African lion, and white African lion samples for genome sequencing were acquired from Everland Zoo of Korea (Supplementary Table S1). TaeGeuk, who has well-documented pedigree information, was confirmed as an Amur tiger by assessment of diagnostic subspecies specific synapomorphic mitochondrial DNA gene markers[1]. The snow leopard sample was acquired from the Conservation Genome Resource Bank for Korean Wildlife, Seoul National University, and the original sample (tissue) was imported from Mongolia with CITES permits (Mongolian export permit no. 0390, South Korean import permit no. ES2012-02507).

Amur tiger genome sequencing

Libraries of different insert sizes were constructed at BGI, Shenzen, China. The insert sizes of the libraries were 170bp, 500bp, 800bp, 2Kb, 5Kb, 10Kb, and 20Kb. Three short insert paired-end libraries (170bp, 500bp, and 800bp), and four long insert mate-pair libraries (2Kb, 5Kb, 10Kb, and 20Kb) were sequenced using HiSeq2000. A total of 288.20 billion base pairs were produced from 27 lanes. The average read length was 76.62bp (Supplementary Table S2). No chromosomal abnormalities were detected, and this lack of abnormality was confirmed by G-banded karyotype (Supplementary Fig. S1).

Raw read filtering for *de novo* assembly

In order to reduce sequencing error effects on the assembly, sequence reads were filtered for high quality[51] (Supplementary Table S3). Filtered, 242Gb sequences (around 100X average depth) were generated with an average read length of 77bp from 15 different (7 insert-size) libraries. The filtering criteria for exclusion were as follows:

1) Reads with ambiguous bases (represented by the letter N) for more than 5% of bases or poly-A structure contents.

2) Reads with 60% low-quality bases (base quality < 5) for the small insert-size libraries (170, 500 and 800bp) and reads with 30% low-quality bases for the large insert-size libraries.

3) Reads with adapter contamination: reads with more than 10bp aligned to the adapter sequence (no more than 3bp mismatch allowed).

4) Small insert-size reads in which read1 and read2 overlapped over 10bp or longer (10%

mismatch allowed).

5) PCR duplications (reads are considered duplications when read1 and read2 of the two paired end reads are identical).

6) Low-quality ends (obvious in terms of GC percentage composition) were trimmed directly.

Panthera species genome sequencing and filtering

Panthera species (white tiger, snow leopard, African lion, and white African lion) were sequenced at Theragen BiO Institute (TBI), Korea, using HiSeq2000 with read and insert lengths of 90bp and ~400bp, respectively.

For the comparative analysis, the sequence reads of the *Panthera* species were also filtered. Finally, on average 71.6 billion base pairs was produced, and the average genome coverage was ~30X (Supplementary Table S18). The filtering criteria for exclusion were as follows:

1) Reads with ambiguous bases (represented by the letter N) exceeds 10%.

2) Average quality of the read is under 15.

3) Nucleotides under quality 15 exceed 10% of a read.

4) For any read which contains an adapter sequence:

   A. More than 10bp of the tail of the first read and the head of the index adapter are identical.

   B. More than 10bp of the tail of the second read and the head of the universal adapter complementary sequence are identical.

5) Any read which contains PhiX or process control sequences.

De novo assembly of the Amur tiger genome

Prior to assembly, the sequencing errors were corrected based on K-mer frequency information[51]. For the Amur tiger genome assembly, we chose K=17bp and corrected sequencing errors for the 17-mers (expected frequency of 17-mers was 68) with a frequency lower than 18. In total, we corrected 0.21% of the bases and trimmed 3.97% of the bases from the filtered reads.

We estimated the Amur tiger genome size using K-mer analysis with K-mer size of 23bp. The genome size was calculated using the following formula: G=K_num/K_depth (K_num is the total number of K-mer, and K_depth is the frequency occurring more often than the others). The size of Amur tiger genome was estimated at 2.44Gb (Supplementary Fig. S2, Supplementary Table S4).

The Amur tiger genome was assembled using SOAPdenovo software[13], which employs the *de Brujin* graph algorithm. Only highly qualified data were used in the genome assembly. SOAPdenovo mainly consists of three key steps. Using this software, a total of 2.40Gb of scaffolds was assembled. The N50 of the scaffolds was 8.8Mb, and the size of the longest scaffold was 41.61Mb (Supplementary Table S5).

1) Contig construction: firstly, data from the short insert size library were used to construct a *de Bruijn* graph. Then, tips, merged bubbles, connections with low coverage, and small repeats were removed.

2) Scaffold construction: all qualified reads were realigned onto the contig sequences. We then calculated the amount of shared paired-end relationships between each pair of contigs, weighted the rate of consistent and conflicting paired-ends, and then constructed the scaffolds step-by-step, from short insert-size paired-ends, to long distance paired-ends.

3) Gap closure: The gaps between the constructed scaffolds were mainly composed of repeats that were masked off during scaffold construction. To close these gaps, the paired-end information was used to retrieve read pairs with one end mapped to a unique contig while the other was located in the gap region. Subsequently, a local assembly for these collected reads was performed.

The mapping depths of the assembled Amur tiger genome showed a normal distribution, indicating that there was no severe bias caused by repetitive sequences (Supplementary Figs. S3, S4). We used 10Kb non-overlapping sliding windows and calculated the GC content and average depth among these windows. To check whether the small block in the bottom of Supplementary Figure S3 originated from contamination from other species, some parts of the scaffolds were chosen from the small block and used as queries to perform BLAST[34] with the nt library. The scaffolds were most closely related to chromosome X of *Felis catus*, which indicates that the small block represents chromosome X of the Amur tiger.

Amur tiger draft chromosomes

Amur tiger draft chromosomes were constructed using the cat reference genome (Felis_catus-6.2). As the Amur tiger genome is very similar to the cat reference genome (Supplementary Table S11), the chromosomal location and ordering information of the tiger scaffolds could be derived from the cat genome (Supplementary Fig. S5). The methods used are detailed below.

1) Homozygous variations (719,258 of SNVs and 103,065 of indels), found by mapping the tiger short reads to the tiger scaffolds, were replaced with the nucleotides of the raw reads (Supplementary Tables S12, S13). We aligned the short reads of the Amur tiger to the scaffolds using BWA-0.5.9[52] with seed length 50 and edit distance 2, and variations were called from the depth of 5X ~ 300X using SAMtools-0.1.18[53].

2) The corrected tiger scaffolds were mapped to the cat genome, Felis_catus-6.2, using

SyMAP[49], and then the chromosomal location and order of the scaffolds were decided upon. Gaps between the located tiger scaffolds were estimated based on gaps in the alignment with the cat genome. When two tiger scaffolds overlapped, 20Kb was used as the size of the gap between the two scaffolds, because the longest insert size was 20Kb.

3) Additionally, we aligned tiger short reads to the cat genome using BWA with the same option described above, and continuous mapped regions (consensus sequences) with mapping depth over 1X were extracted (Supplementary Table S14). A total of 43,407,200 homozygous SNVs were used to generate the consensus sequences (Supplementary Table S15). The gap sequences between the tiger scaffolds were substituted with the consensus sequences if they were fully covered by the consensus sequences. As a result, 18 gaps (45,578 bp in length) in the tiger draft chromosomes were filled with the consensus sequences, which were generated by substituting the cat genome sequences with 956 tiger homozygous SNVs.

Finally, 571 of 674 scaffolds, 99.6% of total scaffold length, were placed in the Amur tiger chromosomes (Supplementary Tables S16, S17).

Assembly quality

In order to assess the assembly quality, the tiger blood transcriptome was sequenced using HiSeq2000 with read and insert lengths of 90bp and ~300bp, respectively. The transcriptome was assembled using Trinity[54]. The assembled transcripts were aligned to the assembly sequences using Blat with default parameters except an identity cutoff of 80%. As a result, >96% of the assembled transcripts were covered in the tiger scaffolds (Supplementary Tables S6, S7).

Published cat EST sequences were mapped to the tiger genome sequence (http://hgdownload.cse.ucsc.edu/goldenPath/felCat4/bigZips/est.fa.gz). The coverage rate was higher than 90%. Most of the unmapped EST segments were on the head or tail of the ESTs, and many of these segments were identified as having vector contamination (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html). The mapping rate of the cat EST sequences was 98.9% (Supplementary Table S8).

Assembly quality was also assessed by mapping the tiger short reads to the tiger scaffold resulting in 95% of the reads mapped (Supplementary Table S12). A total of 719,258 homozygous SNVs and 1,262,207 heterozygous SNVs were found (Supplementary Table S13). A total of 78 heterozygous SNVs were validated by Sanger sequencing method, and 75 of them (96.2%) were true heterozygous SNVs (Supplementary Table S9).

Additionally, analysis of the tiger draft genome assembly for core eukaryotic genes[14] revealed homologs for >93.4% of conserved genes in the assembly (Supplementary Table S10).

Annotation of the assembled Amur tiger genome

Repeat annotation

We searched the genome for tandem repeats using the Tandem Repeats Finder[55] version 4.04. Transposable elements (TEs) were identified in the genome by homology-based approaches. The homology-based approach was used with Repbase[56] version 16.08, the commonly used database of known repeats. We used this database to find repeats using corresponding software such as RepeatMasker[57] version 3.3.0 and RMBlast[58] version 1.2. RepeatMasker was applied for DNA-level identification in combination with Repbase (Supplementary Fig. S13).

Gene prediction

Amur tiger genes were predicted as follows (Supplementary Tables S19, S20).

1) *De novo* prediction: *de novo* prediction was performed on the repeat masked genome based on HMM (hidden Markov model). Programs applied were AUGUSTUS[32] (version 2.5.5) and GENSCAN (version 1.0)[33].

2) Homology-based prediction: Homologous proteins of other species (from the Ensembl 60 release) were mapped to the genome using TblastN (Blast 2.2.23)[34] with an E-value cutoff of 1E-5. The aligned sequences, as well as its query protein, were then filtered and passed to GeneWise[35] (version 2.2.0) to search for accurate spliced alignments.

3) cDNA/EST-based prediction: Cat EST and full length cDNA sequences (from UCSC) were aligned to the genome using BLAT[36] (blat-34, identity ≥ 0.90, coverage ≥ 0.90) to generate spliced alignments. For EST results, spliced alignments were linked according to overlap using PASA[37].

4) Integration evidence: Source evidence generated from the three approaches mentioned above was integrated using GLEAN[38] to produce a consensus gene set.

5) Synteny detection: The Amur tiger genome sequence was aligned to two well assembled and annotated genomes (human and domestic cat) using LASTZ[59] (version 1.02). Then, mapped results yielding information on homologous proteins were filtered by syntenic 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. We used *de novo* and homology-based gene prediction methods, and the gene sets from two methods were merged to form a comprehensive and non-redundant reference gene set (Supplementary Tables S21, S22). We clustered the genes from all the input sets with a cut off of genomic overlap greater than 50 bp for each gene locus; we get larger CDS from human-derived, dog-derived and cow-derived. Finally, if no homologous

gene mapped, the *de novo* prediction was used. We had a much stricter cutoff for *de novo* genes than for homology genes. The one with the larger CDS from Genscan and Augustus was chosen and was required to have more than 30% aligning rate to the SwissProt/TrEMBL database[60,61] and had to have more than 3 exons.

Gene function annotation

Gene functions were assigned according to the best alignment match using Blastp with SwissProt and TrEMBL databases (Uniprot release 2011-01)[60,61]. The motifs and domains of genes were determined by InterProScan[62] (version 4.7) against protein databases, including ProDom[63], PRINTS[64,65], Pfam[66], SMART[67], PANTHER[68], and PROSITE[69]. Gene Ontology[70] IDs for each gene were obtained from the corresponding InterPro entries. All genes were aligned against KEGG[71] (Release 58) proteins, and the pathway in which the gene might be involved (Supplementary Table S23).

Detection of non-coding RNAs

We detected four types of non-coding RNA in the Amur tiger genome by searching databases using the whole genome sequence. tRNAscan-SE[72] (version 1.23) was performed on a SINE premasked genome in order to search for reliable tRNA positions. snRNA and miRNA were sought through a two-step method: after alignment with BLAST, INFERNAL was used to search for putative sequences in the Rfam database[73] (Release 9.1). Though *Panthera tigris* rRNA sequences were not available, we searched the genome using human full-length rRNA as queries for possible rRNA positions by BLAST with some restrictions (Supplementary Table S24).

Orthologous gene clusters

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. There were 17,841 orthologous groups and 8,875 single-copy gene families (Supplementary Fig. S6, Supplementary Table S25). We used the Treefam methodology[39] 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.

BLASTP was used with all of the protein sequences against a database containing a protein dataset of all species under E-value 1E-7, and fragmental alignments were conjoined for each gene pair by Solar[74]. We assigned a connection (edge) between two nodes (genes) if more than 1/3 of the region was aligned to both genes. An H-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 here is the BLAST raw score.

Gene families were extracted by clustering using Hcluster_sg[75]. We used the average distance for the hierarchical clustering algorithm, requiring the H-score to be larger 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 outgroup genes.

Genome evolution

Conserved sequences among genomes

We examined the Amur tiger's genomic elements that may have been evolutionarily conserved across species. PhastCons[76] was adopted to identify conserved elements with conservation scores, given a multiple alignment and a phylo-HMM. The parameter settings were "--target-coverage 0.3 --expected-length 45 --rho 0.31", and the phylogenetic model for non-conserved regions was produced by phyloFit in the PHAST package[77]. A total of 3.2% (77Mb) of the Amur tiger's genome sequence was conserved, which accounts for 0.19% of the genic region (Supplementary Table S49).

Lineage-specific insertions and deletions (indels)

For the synteny analysis, MULTIZ[78] was used to integrate all pairwise alignments together to identify conserved elements among the tiger, cat (Felis_catus-6.2), dog, human, and mouse genomes. For blocks longer than 500bp, species-specific and family-specific indels were counted according to the alignment data; indels located within 50bp of the end of the block, and pairs of indels less than 50bp apart were filtered out.

In accordance with the divergence time tree (see Supplementary Fig. S8), we obtained the time to most recent common ancestor (TMRCA) for each selected species and tiger. We also obtained TMRCA of the Felidae family and dog. Then, the average mutation rate of indels was calculated for each species (Supplementary Tables S51, S52).

Segmental duplication (SD)

Duplications were detected using whole-genome assembly comparison (WGAC). The self-alignment was generated by the LASTZ tool, with most parameters set as defaults. Before aligning, repeat sequences were masked, and the genome assembly was split into several

small sub-files. The maximum simultaneous gap allowed during aligning was 100bp. After aligning, we conducted self-to-self sequence alignment and identified 10,691 recent segmental duplicated fragments (>90% identity, >1Kb length) with a total length of 11.2Mb (0.47%) in the Amur tiger's whole genome assembly. Its recent segmental duplication rate was similar to that of giant panda (10.4Mb, 0.43%)[51] and lower than that of dog (43.8Mb, 1.73%)[79] as detected using the same whole-genome assembly comparison method (Supplementary Table S50). This low segmental duplication rate may be due to an over-collapsing of duplications, a common error seen in assemblies from next generation sequencing technologies[80].

Chromosomal rearrangement

Among the alignment data generated from SyMAP[49], when one scaffold happened to be mapped to several physically distant cat (Felis_catus-6.2) chromosomal locations, they were considered to be inter- or intra-chromosomal rearrangement events of the Amur tiger genome relative to the cat genome (Supplementary Table S53). A total of 30 breaks, i.e., large-size inter- or intra-chromosomal rearrangements, were detected using SyMAP. The species (tiger and domestic cat)-specific genomic rearrangements were also analyzed. We performed the dog vs. tiger and cat vs. tiger whole-genome pair-wise alignments using LASTZ software on the repeat-masked genomes. Using these methods, we identified clusters of unique alignments with well-defined order and orientation. These clusters were defined as syntenic segments. Different cutoffs (5Kb to 500Kb) were tested for the minimum syntenic segment size and various counts of intra-chromosomal and inter-chromosomal breakpoints were calculated (Supplementary Table S54).

There was a total of 18 chromosomal rearrangement (12 inter- and 6 intra-chromosomal rearrangements) overlaps (Supplementary Table S55) when the results from SyMAP and LASTZ were integrated by comparing syntenic break positions. As the tiger assembly was generally fragmented, we carefully validated the 18 syntenic breaks to examine the assembly integrity by aligning long insert mate-pair libraries (2Kb, 5Kb, 10Kb, and 20Kb) to the tiger scaffolds. Finally, we reported six putative chromosomal rearrangements (two inter- and four intra-chromosomal rearrangements) between the tiger and cat. All six rearrangements were validated by long range PCR experiments followed by the Sanger sequencing method (Supplementary Fig. S14, Supplementary Table S56).

Gene evolution

Phylogenetic analysis of Amur tiger and other mammals

A phylogenetic tree of the Amur tiger and the other sequenced genomes was constructed using single-copy gene families. For each species, 4-fold degenerate sites were extracted

from each family and concatenated to one super gene. A substitution model (HKY85+gamma+I) was selected, and PhyML[81] was used to construct the phylogenetic tree (Supplementary Fig. S7).

Estimation of divergence time and substitution rate

We estimated the divergence time for the seven mammalian species (tiger, human, dog, mouse, giant panda, opossum, and domestic cat (Felis_catus-6.2)) using single copy gene families and 4-fold degenerate sites (neutral substitution rate per year). PhyML molecular dating was adopted to estimate the neutral evolutionary rate and species divergence time using the program MCMCTREE[82], which is implemented in the PAML package[83]. The divergence time of tiger and cat was estimated to be 10 (8-15) MYA, and the substitution rate of tiger (1.6E-09) was similar to that of human (1.4E-09) (Supplementary Fig. S8).

Phylogenetic analysis of Amur tiger and other Felidae

We also estimated the divergence time among Felidae specifically, using dog genome as an outgroup. All filtered reads were mapped to the cat reference genome sequence, CAT.66 build from Ensembl (Supplementary Table S58). Mapping was conducted using the BWA program with default options.

Single nucleotide variations (SNVs) were scanned from SAMtools pileup files. Variations were detected from the depth of 5-150X (Supplementary Table S59). Reads of the Amur tiger were filtered using the same criteria used for the other big cat genomes. Then, variations were scanned.

By using 2,888 orthologous genes between the domestic cat and dog from OrthoDB[84] and mapping the big cat reads to the domestic cat, 1,904 efficiently covered genes (depth: 5-60X, ≥50% of CDS coverage) were found. SNVs detected in this way were used to select big cat gene sequences by substituting the variations, and the sequences were multiply-aligned for all of the genes. Gapped regions of the multiply aligned genes were removed, and, finally, 270,540bp on the 4-fold degenerate sites were collected and used to construct the phylogenetic tree (Supplementary Fig. S15). We adopted <16.0 and >1.0 million years as fossil record constraints of divergence time for domestic cat vs. big cats and snow leopard vs. tiger, respectively. The estimated divergence time was similar to that predicted by fossil records and close to the 10.8MYA date developed by a recent comprehensive Felidae phylogeney. Tiger, lion, and snow leopard diverged at about 2-4 MYA (Supplementary Fig. S16), and the snow leopard was closer to the tiger than to the lion, affirming earlier inferences[15,16]. The phylogenetic tree construction was extremely consistent under bootstrapping analysis (100% for all of the branching points).

Gene expansion and contraction

Among 14,425 orthologous gene families, 103 shared orthologous gene families were specific to the tiger and the cat. The Felidae-specific gene families contained 129 genes and 287 InterPro domains, and gene list of the Felidae-specific gene families and tiger/cat-specific gene families are shown in Supplementary Tables S26, S27, respectively (sequence data is available at http://tigergenome.org). The Felidae- and Amur tiger-specific protein domains are shown in Supplementary Tables S28, S29.

Lineage-specific gene family expansion and contraction may be associated with specific functions, phenotypes, and physiology. We determined the expansion and contraction of the orthologous protein families among seven mammalian species (tiger, cat (Felis_catus-6.2), dog, human, mouse, giant panda, and opossum) using CAFÉ 2.2[40] with 0.001080 of lambda option. GO of all tiger genes was annotated by InterPro. ChiSquare test followed by Fisher's exact test (p-value≤0.01) was used to test for over-represented functional categories among expanded genes and 'genome background' genes; Fisher's exact test was used when any expected value of count was below 5, which would have made the ChiSquare test inaccurate[41]. Compared to the feline common ancestor, the tiger genome is enriched in olfactory receptor activity, G-protein coupled receptor signaling pathway, signal transducer activity, amino acid transport, and protein metabolic process (Supplementary Table S30). All olfactory receptor gene families, which were expanded in the tiger compared to the feline common ancestor, are shown in Supplementary Fig. S9.

Positively selected genes (PSG)

Genome-wide scan for positively selected genes (PSGs) in mammals can provide insight into the dynamics of genome evolution, the genetic basis for differences among species, and the functions of individual genes[85]. We used conserved genome synteny methodology[19] to establish a high-confidence orthologous gene set. Briefly, whole-genome multiple alignments were performed between human (hg19) and other species (cat (Felis_catus-6.2), dog (CanFam2.0), mouse (mm9), and panda (ailMel1) genomes) by the Lastz alignment pipeline. We collected all the human protein-coding genes from RefSeq[43], KnownGene[44], and VEGA[45], and mapped them to the other species via the syntenic regions. We then filtered the resulting 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 (Supplementary Table S35), most of which also correspond to genes in the panda, dog, and mouse genomes.

To detect tiger genes evolving under positive selection, we aligned ortholog genes by PRANK[46] and used the optimized branch-site model of PAML (version 4.5) and likelihood ratio tests (LRTs) (P-value ≤ 0.05). A total of 178 PSGs were identified (Supplementary Data 1). A GO annotation download from Ensembl was used to assign GO categories to 7,415 orthologs. ChiSquare test followed by Fisher's exact test (P-value ≤ 0.01) was used to test for over-represented functional categories among PSGs (Supplementary Data 2); Fisher's exact test was used when any expected value of count was below 5, which would have made the

ChiSquare test inaccurate[41]. KEGG pathway analysis results for the Amur tiger's PSGs are shown in Supplementary Table S36.

Rapid evolution

We used an approach based on *Ka*/*Ks*[47,48] to identify GO categories that significantly above or below average in the tiger genome. Firstly, the *Ka* and *Ks* rates are estimated by PAML from all aligned bases with quality score > 20 in orthologs, using the F3x4 codon frequency model and the REV substitution matrix. To explore the evolution function catalog, we download the Gene Ontology (GO) annotation of human gene from the Ensembl database (release-69). We calculated the average Ka and Ks values for all genes that have annotated GO as following equations (S1, S2).

$$k_a = \frac{\sum_{ieT} a_i}{\sum_{ieT} A_i} \qquad \text{(S1)}$$

$$k_s = \frac{\sum_{ieT} s_i}{\sum_{ieT} S_i} \qquad \text{(S2)}$$

Where T is the number of annotated GO genes, a_i and A_i are the numbers of non-synonymous substitutions and sites, and s_i and S_i are the numbers of synonymous substitutions and sites in gene *i*, as estimated by PAML, respectively.

The expected proportion of non-synonymous substitutions p_A in a GO category was then estimated (S3).

$$p_A = \frac{k_a \sum_{ieC} A_i}{k_a \sum_{ieC} A_i + k_s \sum_{ieC} S_i} \qquad \text{(S3)}$$

For a given GO category C, the probability p_c of observing an equal or higher number of non-synonymous substitutions was calculated assuming a binominal distribution (S4).

$$p_c = \sum_{j-a_c}^{a_c+s_c} \binom{a_c+s_c}{j} p_A^j (1-p_A)^{a_c+s_c-j} \qquad \text{(S4)}$$

where a_C and s_C are the total number of non-synonymous and synonymous substitutions in GO category *C*, respectively.

To determine whether the GO categories are evolving under significantly high constraints, we repeated this procedure 10,000 times on the same dataset after randomly permuting the GO annotations to test whether p_C is less than a threshold value. Then, we acquired the GO

categories if the p-value was less than 0.05 (Supplementary Tables S37, S38). Finally, 41 and 22 GO categories were selected to be rapidly and slowly evolving, respectively (Supplementary Tables S40, S41).

We also used a similar approach to the binomial test described above to identify GO categories that have an excess of non-synonymous changes on one lineage. Lineage-specific rates were estimated using a rooted tree including human, tiger, and cat.

For lineages x and y, the average proportion of non-synonymous substitutions were calculated by the following formula (S5).

$$p_x = \frac{x}{x+y} \text{ , which } x = \sum_i x_i \text{ , } y = \sum_i y_i \qquad \text{(S5)}$$

Where $x = \sum_i x_i$, $y = \sum_i y_i$, x is the total number of non-synonymous substitutions in the x lineage, y is the total number of non-synonymous substitutions in the y lineage, and the divergence of the proportion of non-synonymous substitution numbers in different lineages between the observed and expected obeys binomial distribution, the formula is as in the following equation (S6).

$$p_c = \sum_{j-x_c}^{x_c+y_c} \binom{x_c+y_c}{j} p_x^j (1-p_x)^{x_c+y_c-j} \qquad \text{(S6)}$$

As described for the absolute rate tests, we then computed this statistic for every GO category with more than 20 orthologs, as well as for every category in 10,000 randomly permuted data sets (Supplementary Table S39). The tiger- and cat (Felis_catus-6.2)-specific GOs under rapid evolution are shown in Supplementary Tables S42, S43, respectively.

Unique amino acid changes

We investigated *Panthera* lineage-specific amino acid changes by comparison with the known genes from the human, dog, and mouse (from the Ensembl 69 release). We used lion and snow leopard gene sets by mapping reads to the tiger scaffolds and substituting SNVs (Supplementary Tables S31, S32). Artifacts from the multiple sequence alignments (ClustalW2[42]) limitation were removed by filtering option with ≥1/2 of coverage and ≥ of well-matched amino acids (consensus string is '*', ':', or '.'). A total of 3,646 genes had the big cat-specific amino acid changes, and 5,882 genes had unique amino acid changes that are shared in the feline lineage (big cats and domestic cat (Felist_catus-6.2)). Pathway analysis for the genes having functional changes (PolyPhen2[17]) in big cat and feline lineage are shown in Supplementary Tables S33, S34, respectively.

The protein sequences of *EGLN1*, *EPAS1*, and *TYR* were produced by mapping big cat raw

reads to the Amur tiger genome. After variant calling using SAMtools (the depth of 5-150X), both homozygous and heterozygous substitutions were selected for the multiple sequence alignment. The snow leopard-specific amino acid change in *EGLN1* (Met39) was located adjacent to the catalytic channel (Supplementary Fig. S10). We also found snow leopard-specific amino acid changes in the *EPAS1* gene such as Ile663 and Arg794 (Supplementary Fig. S11). Arg794 was found to be possibly damaging causing protein functional changes by computational prediction.

Four nucleotide variants causing amino acid changes in snow leopard (*EGLN1* and *EPAS1*) and white lion (*TYR*) were validated by Sanger sequencing, and they were the snow leopard- and white lion-specific variations, respectively (Supplementary Table S44). Two snow leopard-specific amino acid changes in the *EPAS1* gene were heterozygous variations.

Population genetic analysis of white lion and snow leopard

To verify the proposed candidate mutation in *TYR* gene, a total of 52 samples were included in this study (Supplementary Fig. S12, Supplementary Tables S46, S47), and one DNA sample was made available without extraction history; the following protocols were used for the remaining samples. DNA was extracted from 3 blood and 18 hair samples with the DNeasy Blood & Tissue Kit (Qiagen), following the manufacturer's protocol and a published user-developed protocol, respectively. DNA was extracted from 3 FTA blood samples following protocol 4 in a previously published paper[86]. DNA was obtained from 3 tissue and 24 blood samples (22 EDTA blood samples, 2 heparin blood samples); tissue samples were cut up finely, and red blood cells in blood samples lysed prior to enzymatic digestion (Proteinase K) and phenol:chloroform:isoamyl alcohol extraction[87] with a modified lysis buffer. DNA dilutions were determined based on quantification with a NanoDrop® ND-1000 Spectrophotometer (ThermoScientific) or amplification success. Five samples (2 blood and 3 hair samples) were dropped because of lack of amplification or insufficient sequence quality.

PCR and sequencing: Primers (Supplementary Table S48) to amplify the first part of *TYR* exon 1, including the candidate mutation, were designed with Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)[88] based on the domestic cat genome sequence assembly[8].

Amplification and sequencing was performed following three different protocols.

Veterinary Genetics Laboratory: A 20µl reaction was set up containing 0.2µM forward and reverse primers, 0.0125U Super-Therm Gold DNA Polymerase with 1x Buffer and 1.5mM $MgCl_2$ (Separation Scientific), 250µM dNTPs (Life Technologies), and 0.5µl DNA (5 to 50 ng/µl). PCR was performed on a GeneAmp ® PCR system 9700: 95°C for 5 min; 35 cycles of 95°C for 1 min, 58°C for 30 sec, and 72°C for 30 sec, and a final extension of 72°C for 30 min. The PCR product was cleaned up with MSB® Spin PCRapace (Invitek) following

Protocol 1 with a final elution volume of 20µl. A 10µl sequencing reaction was set up containing a 2µl ABI Prism® BigDye® Terminator v3.1 mix and 1µl buffer (Applied Biosystems), a 0.32µM forward or reverse primer, and a 6µl PCR product. The sequencing reaction was performed on a GeneAmp ® PCR system 9700 according to the Perkin Elmer temperature and time requirements. Sequencing products were purified with an ethanol precipitation and analyzed on a 3130xl Genetic Analyzer (Applied Biosystems).

Institute of Environmental Sciences: A 20µl reaction was set up containing 0.4µM forward and reverse primers, 200µM dNTPs (BioLine), a 0.1U/µl Taq Polymerase with 1x PCR buffer and 1mM $MgCl_2$ (Qiagen), 0.4mg/ml bovine serum albumin (Promega), and 2µl DNA (2-20 ng/µl). PCR was performed on a Bio-Rad S1000 Thermal Cycler PCR system 9700: 94°C for 4 min; 40 cycles of 94°C for 20 sec, 54.5 or 56.6°C for 1 min, and 72°C for 1 min; and a final extension of 72°C for 10 min. PCR products were sent to Macrogen (Amsterdam, The Netherlands) for sequencing.

LifeTechnologies Conservation Genetics Laboratory: A 15µl reaction was set up containing 0.4µM forward and reverse primers, a 0.05U/µl MyTaq HS DNA Polymerase with 1x Buffer and 3mM $MgCl_2$ and 250µM dNTPs (Bioline), and 1.5µl DNA. A touch down PCR was performed on a GeneAmp ® PCR system 9700: 94°C for 10 min; 10 cycles of 94°C for 15 sec, 60-51°C (decreasing 1°C at each cycle) for 30 sec, and 72°C for 45 sec; 30 cycles of 94°C for 15 sec, 50°C for 30 sec, and 72°C for 45 sec; and a final extension of 72°C for 30 min. PCR products were assessed with an agarose electrophoresis and purified with illustra ExoStar (GE Healthcare Life Sciences). If deemed necessary, a second semi-nested PCR reaction was performed with a nested Forward primer following the same amplification and purification protocols as the first PCR. A 10µl sequencing reaction was set up containing a 2µl ABI Prism® BigDye® Terminator v3.1 mix and 1µl buffer (Applied Biosystems), a 0.2µM reverse primer, and a 1-1.5µl PCR product. The sequencing reaction was performed on a GeneAmp ® PCR system 9700 according to the Perkin Elmer temperature and time requirements. Sequencing products were purified with an ethanol precipitation and analyzed on a 310 Genetic Analyzer (Applied Biosystems).

***Analysis*:** All data files were analyzed in Geneious v.6.0.3 (Biomatters; http://www.biomatters.com). Two sequence variants were evaluated. The degree of concordance between expected and observed genotypes were calculated, to test whether the molecular variant that differed from the other felid sequences could be responsible for the white phenotype in lion.

To verify the observed mutation in the *EGLN1* gene, samples from an additional 42 individuals from 5 species (snow leopard, tiger, lion, leopard, and clouded leopard) were sequenced (Supplementary Table S45). Primer information for the *EGLN1* gene is listed in Supplementary Table S44. A 15µl reaction was set up containing 0.4µM forward and reverse primers, 0.05U/µl AmpliTaq DNA Polymerase with 10x Buffer and 3mM MgCl2 and 250µM dNTPs, and approximately 20ng DNA. A touch down PCR was performed on a GeneAmp® PCR system 9700: 94°C for 10 min; 10 cycles of 94°C for 15 sec, 60-51°C (decreasing 1°C at each cycle) for 30 sec, and 72°C for 45 sec; 30 cycles of 94°C for 15 sec, 50°C for 30 sec,

and 72°C for 45 sec; and a final extension of 72°C for 30 min. PCR products were assessed with an agarose. Cycle sequencing reactions consisted of a 0.25U BigDye® Terminator v3.1 Ready Reaction Mix, a 0.075μM primer, 5μL of sequencing buffer (Applied Biosystems), 1.5μL of purified PCR product, and enough water for a 10μL reaction. Cycle sequencing was performed under the following conditions: 94°C for 10 sec, 52°C for 5 sec, and 72°C for 2 min for 45 cycles. Products from cycle sequencing reactions were run on an ABI 3730 DNA Analyzer. Sequence results were visualized and edited in Geneious® (Biomatters; http://www.biomatters.com). All snow leopards had the same amino acid change in *EGLN*1 (Met39), which was not observed in any of the other felid species tested.

Genetic Diversity

Heterozygous SNVs of two tiger genomes, two lion genomes, and one snow leopard genome were calculated by mapping raw reads to the Amur tiger genome (Supplementary Table S32). All genome sizes were assumed to be the same as that of the Amur tiger. Heterozygous SNVs of a recently sequenced human genome (Korean female, from the Korea Personal Genome Project[89], using the same sequencing machine and methods as those used for the present study) were calculated by mapping raw reads to the human reference genome (hg.19). The number of heterozygous SNVs, as well as the genome sizes of the eight species used for comparison, were cited individually from their initial genome sequence reports[8,22,28,30,48,51,90]. Overall, the span of tigers' genetic diversity was similar to that of human, but higher than those of lion and snow leopard. Among the tiger subspecies, the genetic diversity of TaeGeuk (Amur tiger) was relatively lower than that of white tiger (Supplementary Table S57).

Demographic history

The history of population size helps to develop insights into evolution. Based on the pairwise sequentially Markovian coalescent model (PSMC)[31], we inferred detailed population size histories of Amur tiger (TG), African lion (LN), snow leopard (SL), white tiger (WTG), and white lion (WLN).

Using SNV datasets scanned with the all the big cat sequencing reads mapped to Felis_catus-6.2 (BWA-0.5.9 and SAMtools-0.1.18), the consensus sequences of each big cat were constructed and then divided into non-overlapping 100-bp bins marked as homozygous or heterozygous. The resultant bin sequences for their sex chromosomal parts were removed and then they were taken as the input of the PSMC estimation. To test the estimation accuracy, bootstrapping was performed by randomly resampling 100 sequences from the original sequences. Using the neutral mutation rates calculated previously (Supplementary Information section 4.3), the raw PSMC outputs were scaled to time and population sizes (Supplementary Fig. S17). To study the detailed distribution of time to the most recent

common ancestor (TMRCA) and its relationship with evolution, parameters of each bin were decoded with the software and scaled to time (Supplementary Fig. S18). Climate change and migration are two important factors influencing the population size. Thus, we obtained atmospheric surface air temperature (Tsuf) and global relative sea level (RSL) data of the past 3 million years[50] from National Climatic Data Center (NCDC) and combined them together with the 5 big cat demographic data into a single plot).

Since the neutral mutation rate varies with chromosome regions and mutation types, the time and size scaling of our inference could be affected by inaccurate estimation of neutral mutation rates. The inferred history of snow leopard shows a distinct shape, especially in early (0.5 Myr) and recent (3 Kyr) stages. This could be an artifact probably due to the model limitations occurring when there is extremely low genetic diversity.

Supplementary References

51. Li, R. *et al*. The sequence and de novo assembly of the giant panda genome. *Nature* **463**, 311-317 (2010).

52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

53. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

54. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol*. **29**, 644-652 (2011).

55. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. **27**, 573-580 (1999).

56. Repbase Release 16.02 at http://www.girinst.org/repbase/index.html

57. Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker at http://repeatmasker.org

58. RMBlast at http://www.repeatmasker.org/RMBlast.html

59. LASTZ at http://www.bx.psu.edu/miller_lab/dist/README.lastz-1.02.00/README.lastz-1.02.00a.html

60. UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. **40**, D71-75 (2012).

61. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. **28**, 45-48 (2000).

62. Zdobnov, E. M. & Apweiler, R. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848 (2001).

63. Bru, C. *et al*. The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res*. **33**: D212-215 (2005)

64. Attwood, T. K. & Beck, M. E. PRINTS--a protein motif fingerprint database. *Protein Eng*. **7**, 841-848 (1994).

65. Attwood, T. K., Beck, M. E., Bleasby, A. J. & Parry-Smith, D. J. PRINTS--a database of protein motif fingerprints. *Nucleic Acids Res*. **22**, 3590-3596 (1994)

66. Punta, M. *et al*. The Pfam protein families database. *Nucleic Acids Res*. **40**, D290-301 (2012).

67. Ponting, C. P., Schultz, J., Milpetz, F. & Bork, P. SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nucleic Acids Res*. 27, 229-

232 (1999).

68. Mi, H. *et al*. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284-288 (2005).

69. Hulo, N. *et al*. The PROSITE database. *Nucleic Acids Res*. **34**, D227-D230 (2006).

70. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet*. **25**, 25-29 (2000).

71. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. **28**, 27-30 (2000).

72. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955-964 (1997).

73. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*. **33**, D121-124 (2005).

74. Solar at http://treesoft.svn.sourceforge.net/viewrc/treesoft/branches/dev/solar

75. Hcluster_sg at http://pypi.python.org/pypi/hcluster/0.2.0

76. Siepel, A. *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034-1050 (2005).

77. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform*. **12**, 41-51 (2011).

78. Blanchette, M. *et al*. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708-715 (2004).

79. Nicholas, T. J. *et al*. The genomic architecture of segmental duplications and associated copy number variants in dogs. *Genome Res*. **19**, 491–499 (2009).

80. Alkan, C., Sajjadian, S. & Eichler, E. E. Limitations of next-generation genome sequence assembly. *Nat Methods.* **8**, 61-65 (2011).

81. PhyML at http://code.google.com/p/phyml/

82. MCMCTREE at http://nebc.nerc.ac.uk/bioinformatics/docs/mcmctree.html

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

84. Waterhouse, R. M., Zdobnov, E. M., Tegenfeldt, F., Li, J. & Kriventseva, E. V. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. *Nucleic Acids Res*. **39** (Database issue), D283-288 (2011).

85. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, el70 (2005).

86. Smith, L. M. & Burgoyne, L. A. Collecting, archiving and processing DNA from wildlife samples using FTA databasing paper. *BMC Ecol*. **4**, 4-15 (2004).

87. Budowle, B., Smith, J., Moretti, T. & DiZinno, J. *DNA Typing Protocols: Molecular Biology and Forensic Analysis* (Eaton Publishing/BioTechniques Books. Natick, MA USA, 2000)

88. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol*. *Biol*. **132**, 365-386 (2000).

89. KPGP projects at http://opengenome.net

90. Scally, A. *et al*. Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169-175 (2012).