Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas

SI Appendix

Include:

Materials and Methods

References

Supplementary Figure S1-S12

Supplementary Table S1-S29

Materials and Methods

Estimation of Red Panda Genome Size. We estimated the genome size of the red panda using K-mer frequency analysis with a K-mer size of 61, following the method in ref. 1. The red panda was estimated to have a genome size of ~2.38 Gb (Fig. S1).

Genome Sequencing. The red panda genome was sequenced using a whole-genome shotgun approach. Genomic DNA was extracted from whole blood sample of a wild male red panda, and libraries of insert size 180 bp to 8 Kb were constructed with TruSeq DNA Library Prep Kit and Mate Pair Library Prep Kit (Illumina), which were then sequenced on the HiSeq 2000 platform. To make the assembly quality of the giant panda genome comparable to that of the red panda genome, we improved the giant panda genome assembly with new data. We generated Illumina sequence data for the same female giant panda (Jingjing) used in the reference genome release (version ailMel1). Since read length of the ailMel1 assembly was rather short (52 bp on average), about 82 × paired-end sequence data of 100 bp were generated for assembly improvement, in combination with published mate-pair sequence data (under GenBank accessions SRX007019 to SRX007029) (2).

Genome Assembly for Contigs and Scaffolds. The red and giant panda genomes were assembled by a number of steps with an aim to obtain the longest continuity while maintaining high base accuracy. Currently there are a number of assembly tools for selection. Some of them perform better in contig/scaffold length, while some of them excel in terms of base quality or scaffold accuracy. To achieve the best assembly, we need to combine the available assemblers, making use of the strength on one assembler and reducing the disadvantage of the same assembler. Multiple assemblies obtained are finally merged together. Fig. S2 shows the flowchart of our assembly working process. Raw sequencing reads were processed to filter out low quality data by Unikalow (*http://sourceforge.net/projects/phusion2/files/unikalow/*), which is part of Phusion2 assembly pipeline (3). Cleaned short-insert paired-end reads were assembled using a string graph based assembler Fermi (4). The assembly tool SOAPdenovo (5) was used to process the cleaned short-insert reads and mate-pair reads. The initial SOAP assembly

was improved using SSPACE (6) and scaffolds were further enhanced by the use of cross_genome, a tool which maps genome synteny to merge scaffolds, (*http://sourceforge.net/projects/phusion2/files/unikalow/cross_genome/*). The synteny-assisted scaffolding method has been successfully applied in the Tasmanian devil genome project (1) using opossum genome assembly and more recently in the grass carp genome assembly (7) using zebrafish genome. For both pandas' genome assembly, the dog genome (CamFam 3.1) was used as reference. Caution was taken to minimize the possibility of scaffold mis-joining based on cross-species genome synteny. Specifically, when a shared chromosomal rearrangement was observed between dog and panda, human genome was used independently for confirmation. After this step, we obtained the longest possible scaffolds: the SOAP-synteny assembly. In our experience, SOAPdenovo produces reliable scaffolds, but lacks accuracy at the contig level. Notably, a certain degree of mismatches, particularly short indels exist in the assembly. For applications in SNP/indel detection and downstream population analysis, an assembly with high base accuracy is needed. This could be achieved using the Fermi assembly to replace contig sequences in the SOAP-synteny assembly, while long scaffold structures from previous steps are kept unchanged. During the merging process, Fermi contigs were aligned firstly to the SOAP-synteny scaffolds and breakpoints were identified. For all aligned regions, sequences in the SOAP-synteny assembly were replaced by the contigs from the Fermi assembly. Gaps were closed when a piece of Fermi sequence bridged two neighboring SOAP contigs (*http://sourceforge.net/projects/phusion2/files/replace/*).

Assessment of Genome Assembly Quality. To assess sequencing and assembly accuracy of the red panda genome, a bacterial artificial chromosome (BAC) library was constructed, sequenced and assembled. Briefly, the pHZAUBAC1 vector (8) was used to construct the BAC library from kidney tissue dissected from a male red panda as used in the RNA-seq transcriptome sequencing. The library consists of 119,808 BAC clones with an average insert size of 137 Kb, giving approximately 7-fold genome equivalents. Five clones (135K16, 136M11, 138J8, 139H13, and 142G16) were selected for sequencing on the ABI 3730xl platform to approximately 6-fold coverage, following the standard Sanger shotgun approach. These capillary reads were then assembled by GAP4 of the

Staden Package (9), and gaps were filled by PCR to obtain complete sequence for each clone. The genome scaffolds were then aligned with BACs using BLAST (E-value = 1e-20, 98% identity). The completeness of the red panda genome assembly was further assessed using the CEGMA and BUSCO methods.

Genome GC Content. We calculated the GC content and sequencing depth of the assembled red panda and giant panda genomes in 10 Kb non-overlapping sliding window. Since red panda genome sequencing was done for a male, two clustered islands were observed (Fig. S3), where the high coverage island corresponded to diploid scaffolds, and the low coverage region to haploid scaffolds. For the giant panda assembly, only one island was observed, since the sequenced individual "Jingjing" is female (Fig. S4). The GC distribution patterns of five Eutherian genomes, including red panda, giant panda, dog, human, and mouse were also compared (Fig. S5).

Transcriptome Sequencing. To facilitate genome annotation, RNA-seq was performed for eight tissues (pituitary, testis, pancreas, liver, stomach, duodenum, small intestine and spleen) from a recently deceased male red panda that was different from the one used for genome sequencing, and for 12 tissues from two recently deceased giant pandas (liver, stomach, small intestine, colon, pallium and testis from one male adult; pituitary, skeletal muscle, ovary, tongue and two skin tissues from one female adult). Paired-end libraries of insert size 500 bp were constructed with TruSeq RNA Library Prep Kit (Illumina), which were then sequenced on Illumina GAIIx and HiSeq 2000 platforms.

Gene Prediction. Protein sequences of human genes (GRCh 37.71) were aligned to the red and giant panda genomes by tBLASTn with an E-value of 1e-10, and regions with no human homolog match were further searched with dog proteins (CanFam 3.1.71). RNA-seq data were mapped to the scaffolds by the aligner SMALT 0.7.4 (*http://www.sanger.ac.uk/resources/software/smalt/)*, and the uniquely matched reads were selected for gene prediction. Based on the above two steps, potential coding regions were determined by extending 3 Kb to either end of the aligned regions. The

corresponding sequences were used to build preliminary gene models by Augustus 2.5.5 (10) and Fgenesh (11) with parameters trained by human genomes. The constructed preliminary gene models were aligned to the human and dog gene model set by BLASTp at E-value of 1e-20. When a red panda (or giant panda) gene hits two or more adjacent human/dog genes, it was manually split according to alignment information. If two or more adjacent genes were aligned to a single human/dog homolog, they were merged together by GeneWise (12), referring to human/dog homolog structure. The longest protein was chosen to represent the merged gene. To remove the genes encoding transposable elements, the gene models were aligned to the Repbase library and the species' repeat element library, using BLASTn with an E-value of 1e-5. For potential contamination from degenerated mobile elements, the predicted protein sequences were also aligned to the repeat protein library, and hits with an E-value of 1e-5 and 1/3 alignable length were filtered out. Finally, predicted proteins less than 30 amino acids were discarded.

Detection of Non-Coding RNAs. Non-coding RNAs of red panda were identified by searching databases using the genome sequence (Table S7). tRNAscan-SE (version 1.3.1) was used with default parameters to search for tRNAs (Table S8), and false-positive predictions from tRNA-derived SINEs were filtered out by two criteria: (1) a final bit score of less than 55 bits (50 bits for tRNA-SeC-TCA), and (2) are identified by only one of the two pre-filter scanners (trnascan 1.3 and EufindtRNA) (13). We searched the red panda rRNA by BLAST using human and dog full-length rRNA as queries (Identity > 90% and E-value < 1e-10). For snoRNA, snRNA, and miRNA, a two-step approach was applied: the genome sequence was first aligned onto the Rfam database (version 11.0), and genomic regions matching those known RNAs were retrieved by extending 3 Kb towards either end, which were then used for RNA prediction. snoRNA were analyzed by Snoscan (version 0.9 beta) for C/D box snoRNA, and snoGPS (version 0.2 beta) for H/ACA snoRNA with default parameters. snRNA and miRNA were identified by Inferal (version 1.1rc2) with E-value < 1e-10 and quality > 30.

Repeat Element Annotation. Repeat elements in the red panda genome were annotated using the following pipeline: (1) The genome was searched against the nucleotide repetitive database of Repbase (version 16.08) using RepeatMasker (version 4.0.2); (2) The genome was searched against the protein repetitive database using RepeatProteinMask; (3) Tandem repeats were identified by TRF (Tandem Repeat Finder, version 4.04) using default parameters; (4) RepeatModeler was used to construct a *de novo* repeat library of the red panda, which was then used as the database to find and classify repeat elements using RepeatMasker. A similar repeat element annotation strategy was applied to the improved giant panda genome, and the published dog and human genomes (Fig. S8; Tables S9-S11).

Identification of Gene Family Expansion and Contraction. We used the TreeFam methodology (14) to build gene families. The red panda and giant panda reference gene sets from this study were used, and protein-coding genes for six Eutherian species, including human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis lupus familiaris*), ferret (*Mustela putorius furo*), tiger (*Panthera tigris*), and polar bear (*Ursus maritimus*) were downloaded from Ensembl release 71 (*http://www.ensembl.org*). While genes shorter than 30 amino acids were filtered out, the longest protein from alternative transcripts was chosen to represent each unique gene, and protein sequences were kept consistent with corresponding CDS. BLASTp was used to search against all of the protein sequences under an E-value of 1e-7, and fragmental alignments of each gene pair were concatenated by SOLAR (*http://sourceforge.net/p/treesoft/code/HEAD/tree/branches/dev/*). A connection (edge) between two nodes (genes) was assigned if more than 1/3 of the region aligned to both genes, and a Hscore that ranged from 0 to 100 was used to weigh the similarity (edge). For two genes G1 and G2, the Hscore was defined as score(G1G2) / max(score(G1G1), score(G2G2)), the score here was the BLAST raw bit score. Finally, gene families were extracted using the clustering algorithm Hcluster_sg (*http://sourceforge.net/p/treesoft/code/HEAD/tree/branches/dev/*), with the following requirements: Hscore > 5, and edge density (total number of edges / theoretical number

of edges) > 1/3. Gene family data from TreeFam were introduced into CAFE 3.0 (15) to evaluate the expansion and contraction of gene families.

Estimate of *TAS1R1* Pseudogenization Time. We estimated the waiting time until inactivation of the *TAS1R1* gene (T_W) using the method in ref. 16 with minor modification. The principle is to utilize a dramatic change in the degree of functional constraint (f) at the nonsynonymous sites accompanying the loss of function. The f value is the ratio of nonsynonymous / synonymous substitution rates prior to functional relaxation, assumed to be constant among carnivores but became one at the time of inactivation (T_P) (Ma, million years ago) in the red panda. Thus, we have

$$T = T_P + T_W$$

and

$$dN = dN_W + dN_P$$

where, T = 30 MYA is the time since divergence between the red panda and ferret (17); dN is the total number of nonsynonymous substitutions per site for the red panda branch; and dN_W and dN_P are the total number of nonsynonymous substitutions per site for the red panda branch during T_W and T_P time, respectively.

If we denote the synonymous substitution rate per site by μ per unit time, we have

$$dN = f\mu T_W + \mu T_P$$

Thus,

$$T_P = (dN - f\mu T)/(\mu(1-f))$$

Under the assumption of $\mu = 10^{-9}$ per site per year, we can estimate f for the red panda and other carnivore branches using PAML. By bootstrapping the codons of *TAS1R1* 10,000 times, we estimated the frequency distribution of T_P. We considered a t value to be valid when it is between 0 and 19 and subsequently derived a posterior distribution of T_P.

References

1. Murchison EP, et al. (2012) Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* 148(4):780-791.
2. Li R, et al. (2010) The sequence and *de novo* assembly of the giant panda genome. *Nature* 463(7279):311-317.
3. Mullikin JC, Ning Z (2003) The Phusion Assembler. *Genome Res* 13(1):81-90.
4. Li H (2012) Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* 28(14):1838-1844.
5. Li R, et al. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20(2):265-272.
6. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578-579.
7. Wang Y, et al. (2015) The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat Genet* 47(6):625-631.
8. Shi X, Zeng H, Xue Y, Luo M (2011) A pair of new BAC and BIBAC vectors that facilitate BAC/BIBAC library construction and intact large genomic DNA insert exchange. *Plant Methods* 7:33.
9. Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. *Methods Mol Biol* 132:115-130.
10. Stanke M, et al. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34(suppl 2):W435-W439.
11. Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 7(suppl 1):S10.
12. Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. *Genome Res*

14(5):988-995.

13. Chan PP, Lowe TM (2009) GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37: D93-D97.
14. Li H, et al. (2006) TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 34(suppl 1):D572-D580.
15. De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22(10):1269-1271.
16. Oda M, Satta Y, Takenaka O, Takahata N (2002) Loss of urate oxidase activity in hominoids and its evolutionary implications. *Mol Biol Evol* 19(5):640-653.
17. Sato JJ, et al. (2009) Deciphering and dating the red panda's ancestry and early adaptive radiation of Musteloidea. *Mol Phylogenet Evol* 53(3):907-922.
18. Zhao S, et al. (2013) Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet* 45(1):67-71.
19. Liu S, et al. (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157(4):785-794.
20. Liu X, Li M, Yu J, Huang X (2005) Study on the amino acids of bamboo ingested by captive giant panda. *J Econom Anim* 9(1):30-34.
21. Liu B, Fan J, Hu T, Zhang F (2008) Determination and nutritional evaluation of amino acids in staple food bamboo for giant panda in Qinling Mountains. *J Anhui Agri Sci* 36(21):9024-9026.
22. Loor JJ, Herbein JH, Polan CE (2002) Trans 18:1 and 18:2 isomers in blood plasma and milk fat of grazing cows fed a grain supplement containing solvent-extracted or mechanically extracted soybean meal. *J Dairy Sci* 85(5):1197-1207.
23. Rule DC, Broughton KS, Shellito SM, Maiorano G (2002) Comparison of muscle fatty acid profiles and cholesterol concentrations of bison, beef cattle, elk, and chicken. *J Anim Sci* 80(5):1202-1211.
24. Li Y, Hu G, Li X (1993) Measure of vitamins and minerals in daily diet of giant pandas. *Foreign Forestry* 23(1):25-27.

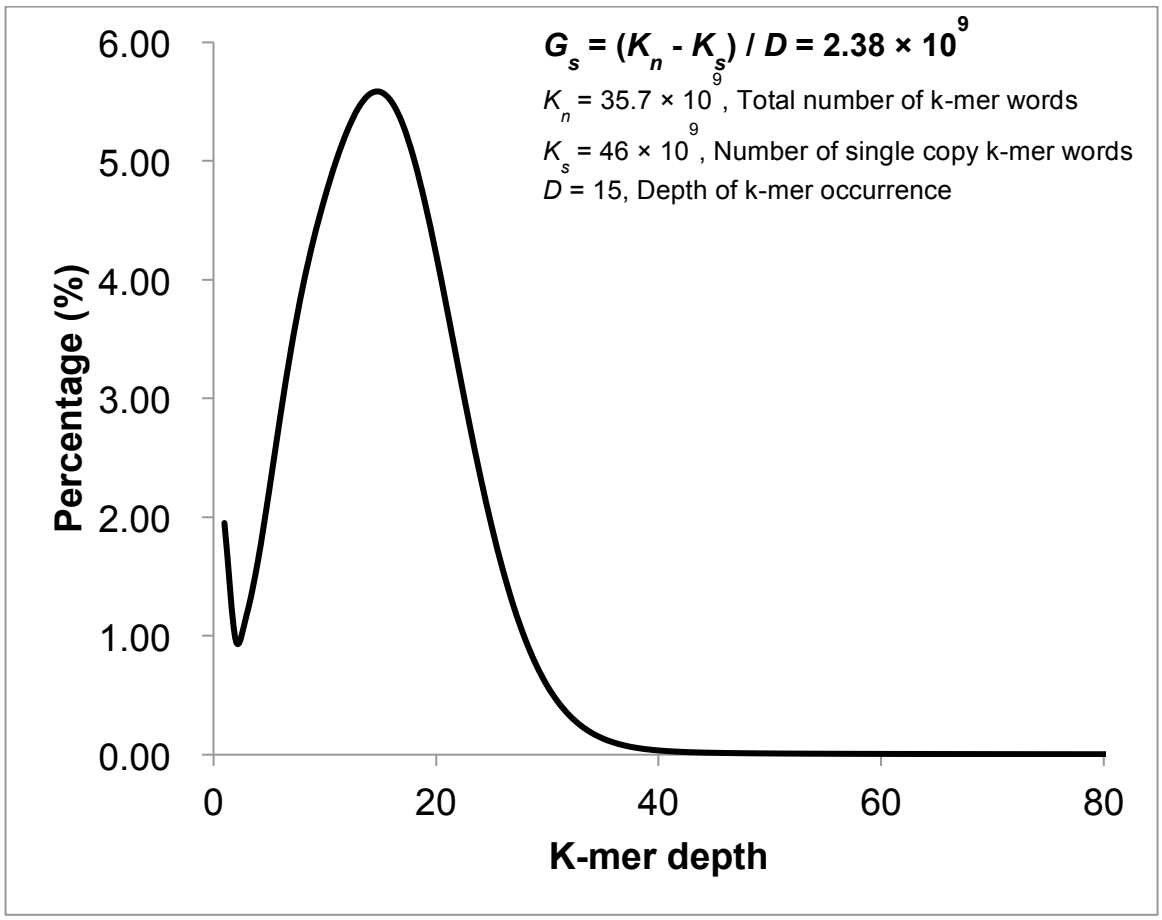


Fig. S1. Genome size estimation of the red panda using K-mer frequency distribution. The x-axis indicates the depth of each unique 61-mer in the red panda genome, and the y-axis denotes the percentage of genome for occurrence of unique 61-mer within the sequence dataset. For this analysis, only paired-end reads were used after contamination filtering and base error correction.

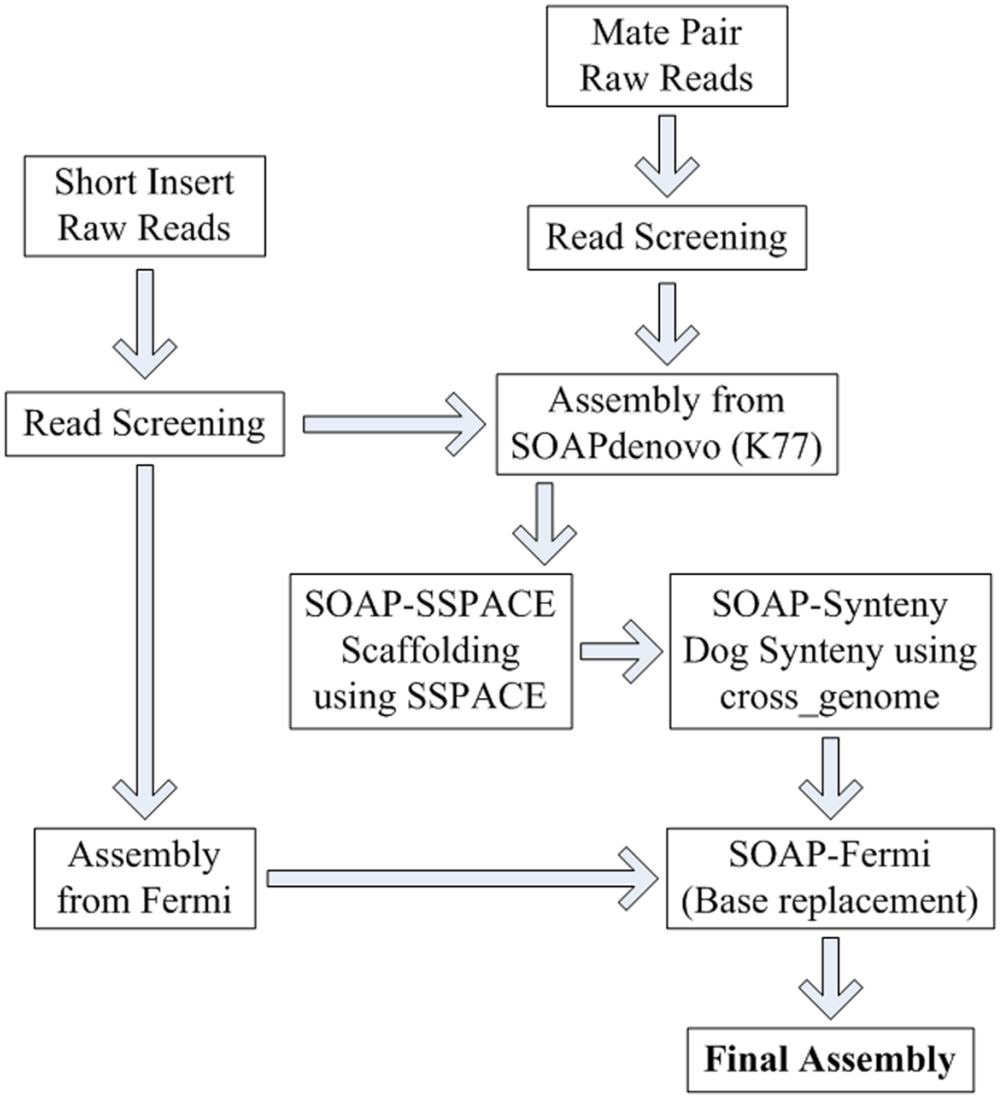


Fig. S2. *De novo* genome assembly flowchart.

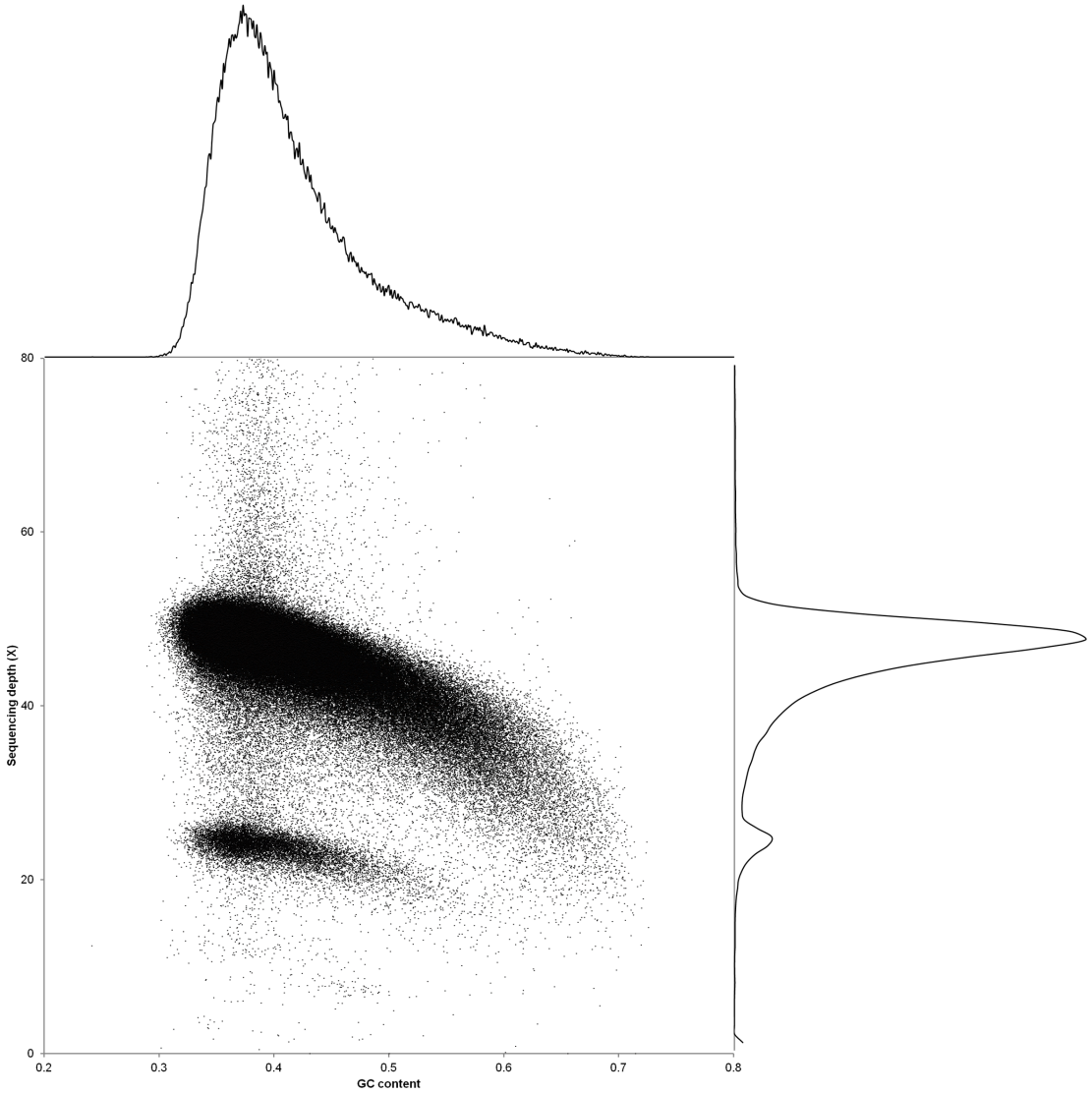


Fig. S3. Distribution of GC content and sequencing depth of the red panda genome. The x-axis represents GC content, and the y-axis indicates average sequencing depth. The filtered reads were aligned onto the assembled genome using SOAPaligner with default parameters, and sequencing depth was then calculated for each base of the genome. Data was calculated per 10 Kb non-overlapping sliding window.

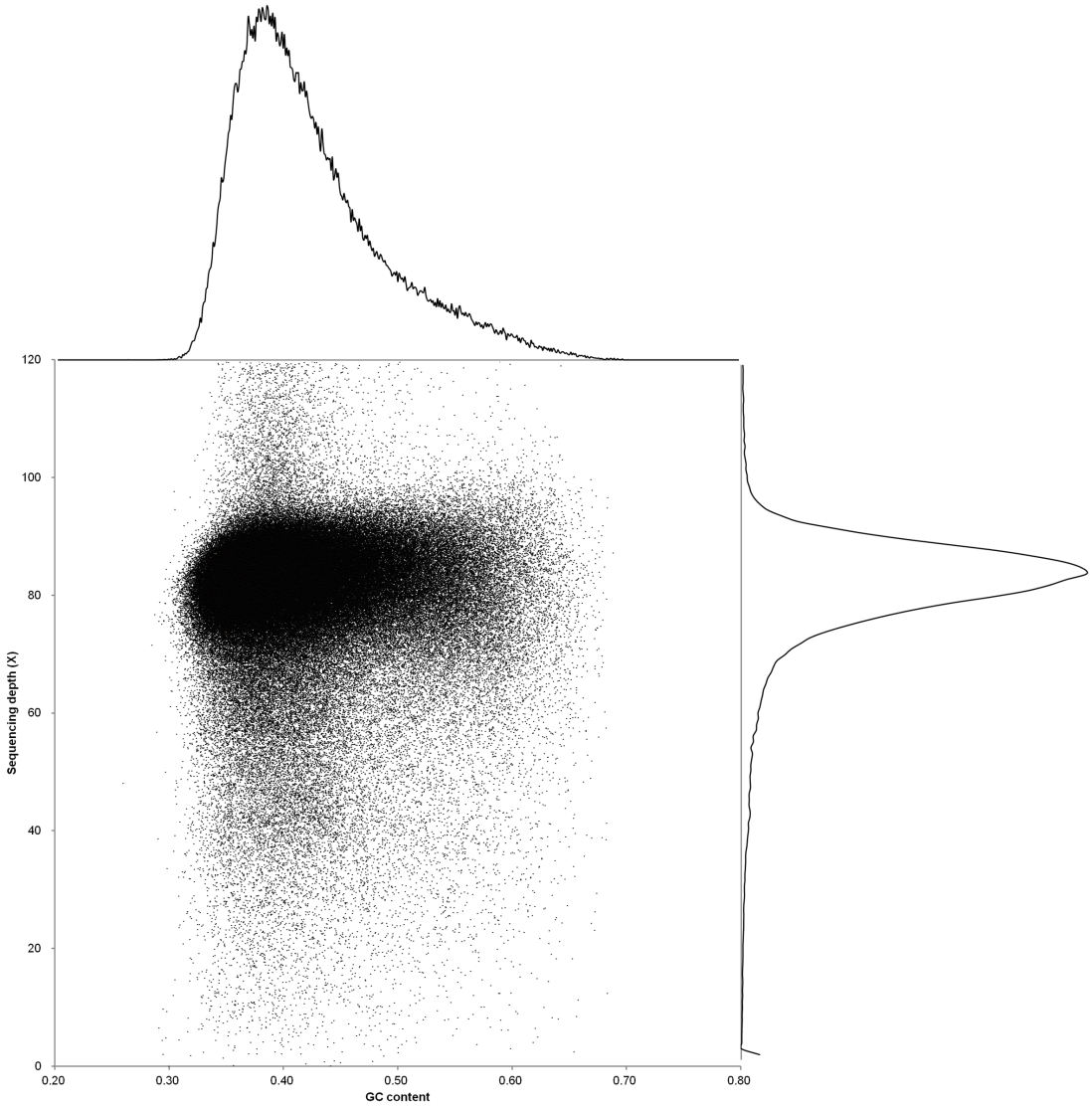


Fig. S4. Distribution of GC content and sequencing depth of the giant panda genome. The x-axis represents GC content, and the y-axis indicates average sequencing depth. The filtered reads were aligned onto the assembled genome using SOAPaligner with default parameters, and sequencing depth was then calculated for each base of the genome. Data was calculated per 10 Kb non-overlapping sliding window.

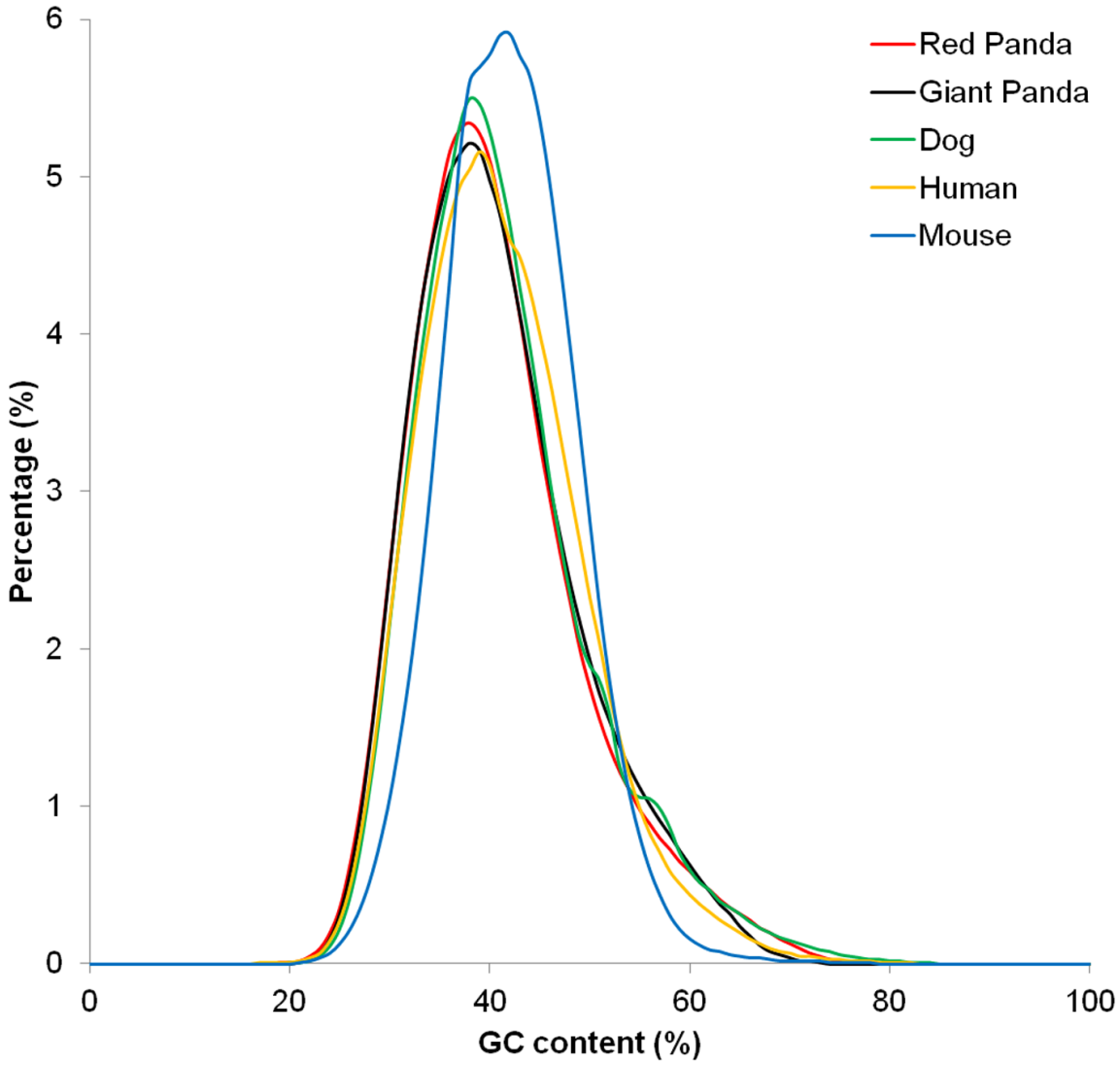
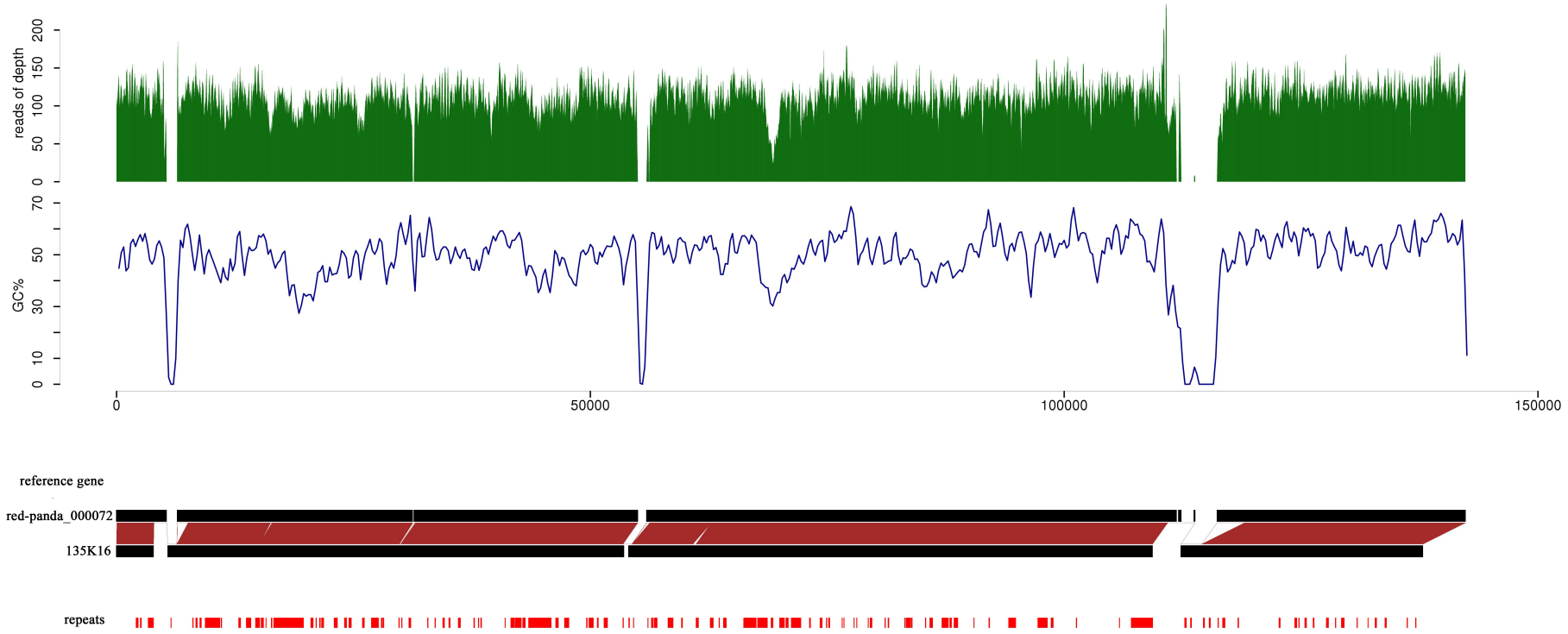
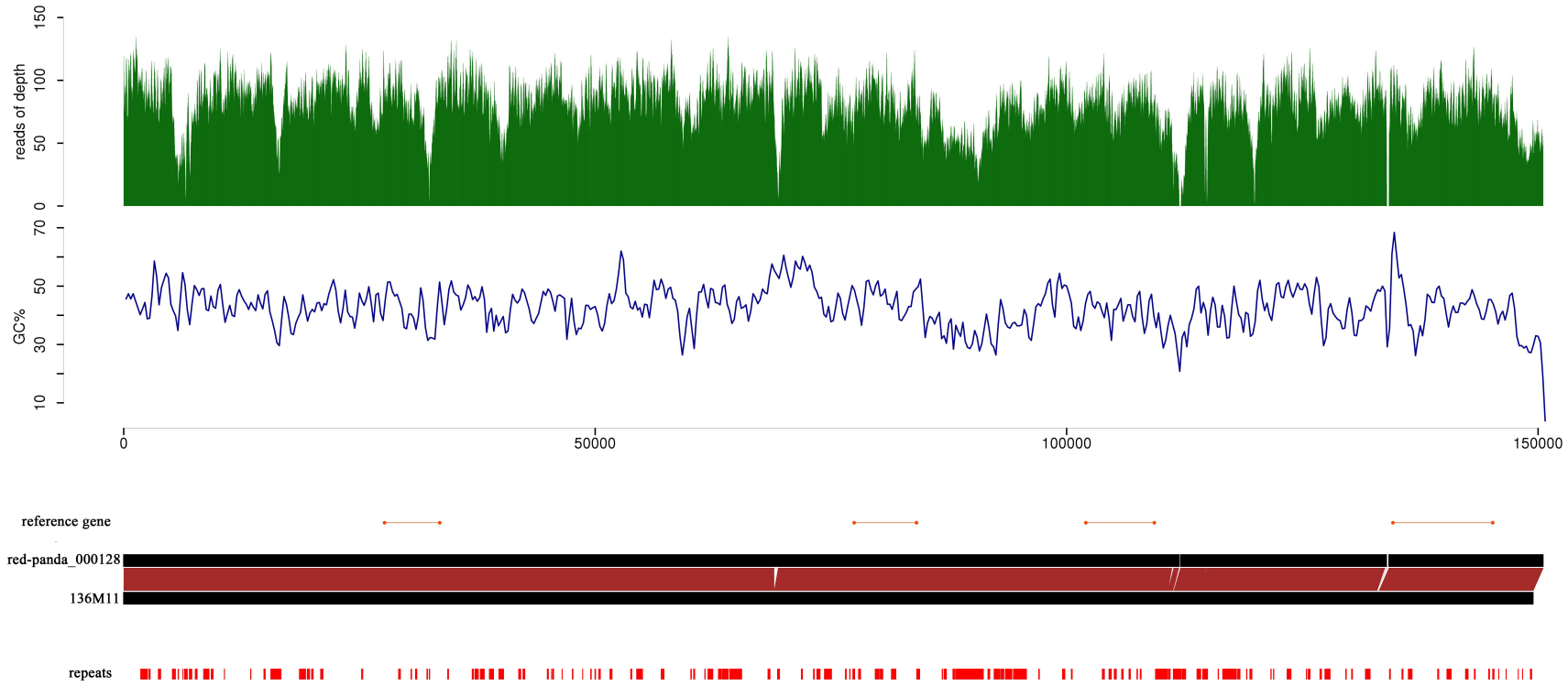


Fig. S5. Comparison of the GC content distribution of five Eutherian genomes in 500 bp non-overlapping sliding windows.

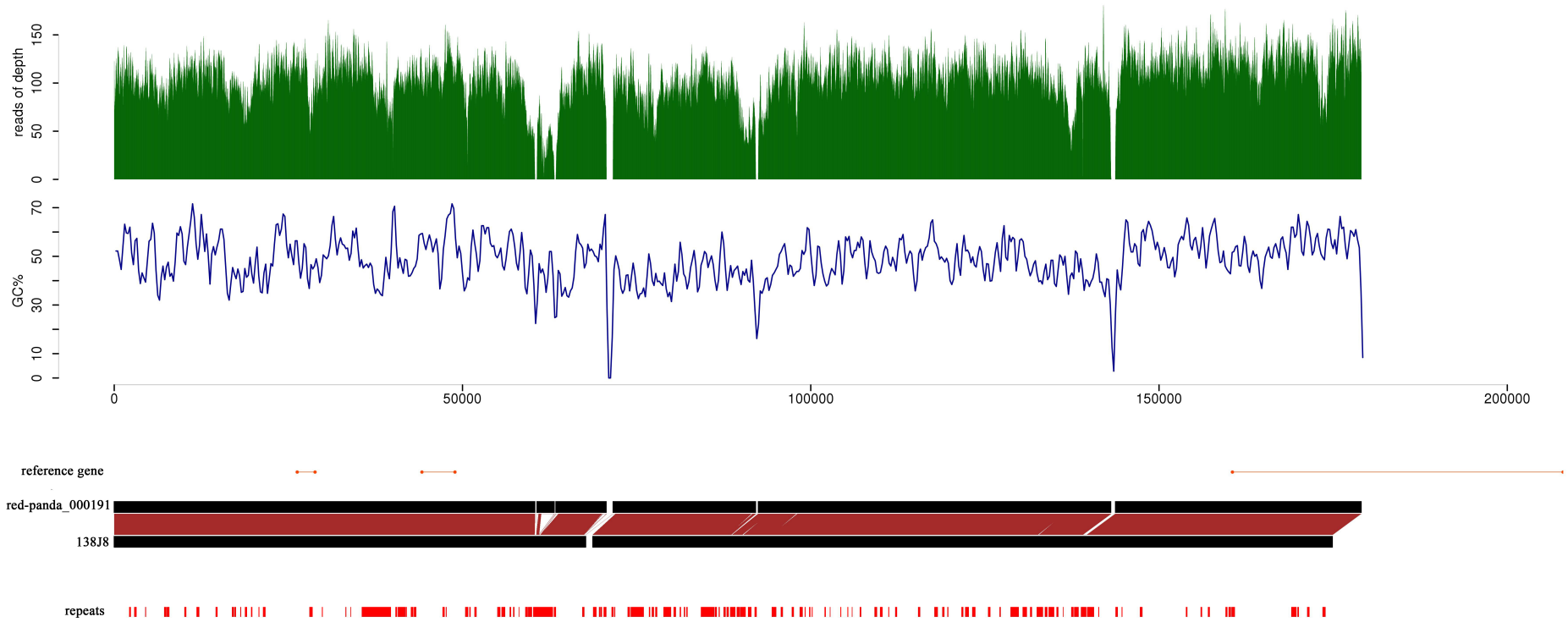
(A) 135K16


reads of depth
0
50
100
150
200
GC%
0
10
30
50
70
0
50000
100000
150000
reference gene
red-panda_000072
135K16
repeats

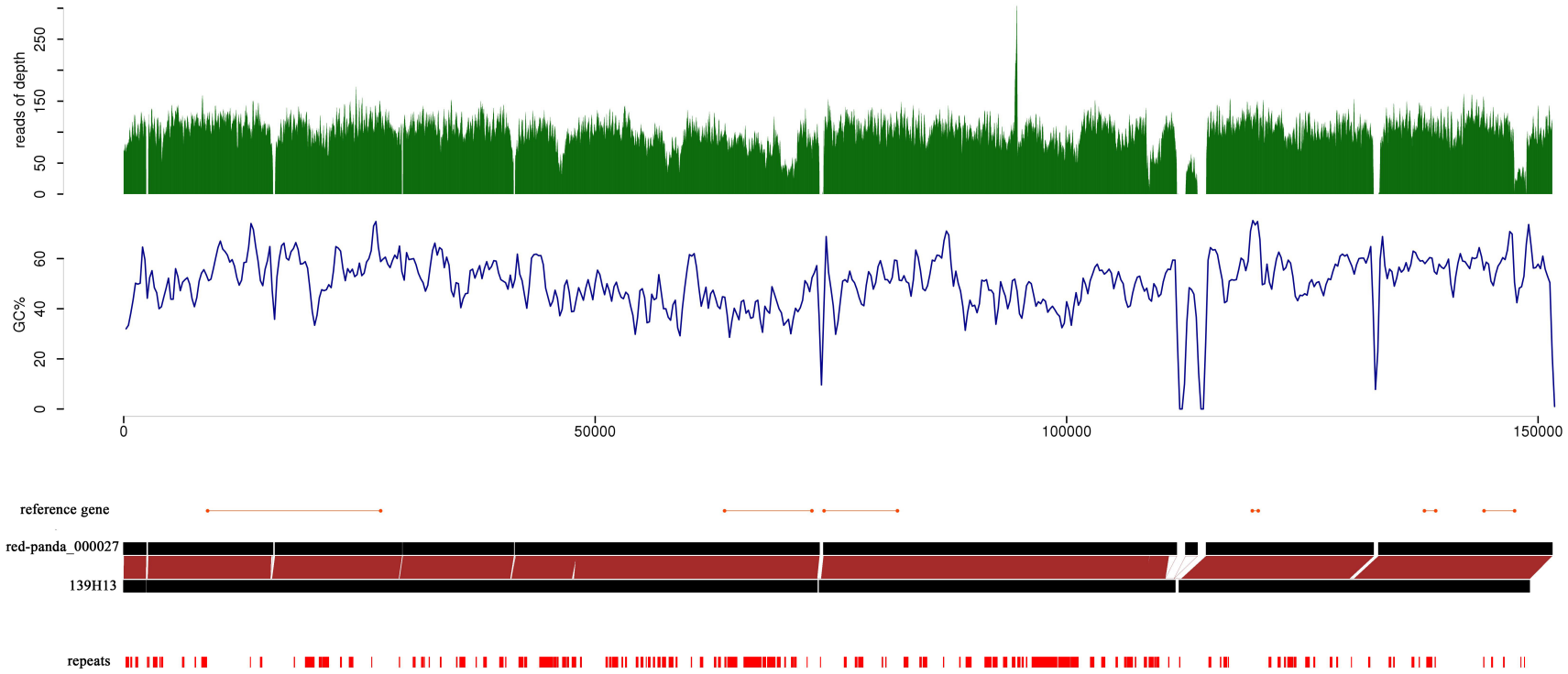
(B) 136M11


reads of depth
150
100
50
0
GC%
70
50
30
10
0
50000
100000
150000
reference gene
red-panda_000128
136M11
repeats

(C) 138J8


reads of depth
0
50
100
150
GC%
0
10
30
50
70
0
50000
100000
150000
200000
reference gene
red-panda_000191
138J8
repeats

(D) 139H13


reads of depth
0
50
150
250
GC%
0
20
40
60
0
50000
100000
150000
reference gene
red-panda_000027
139H13
repeats

(E) 142G16

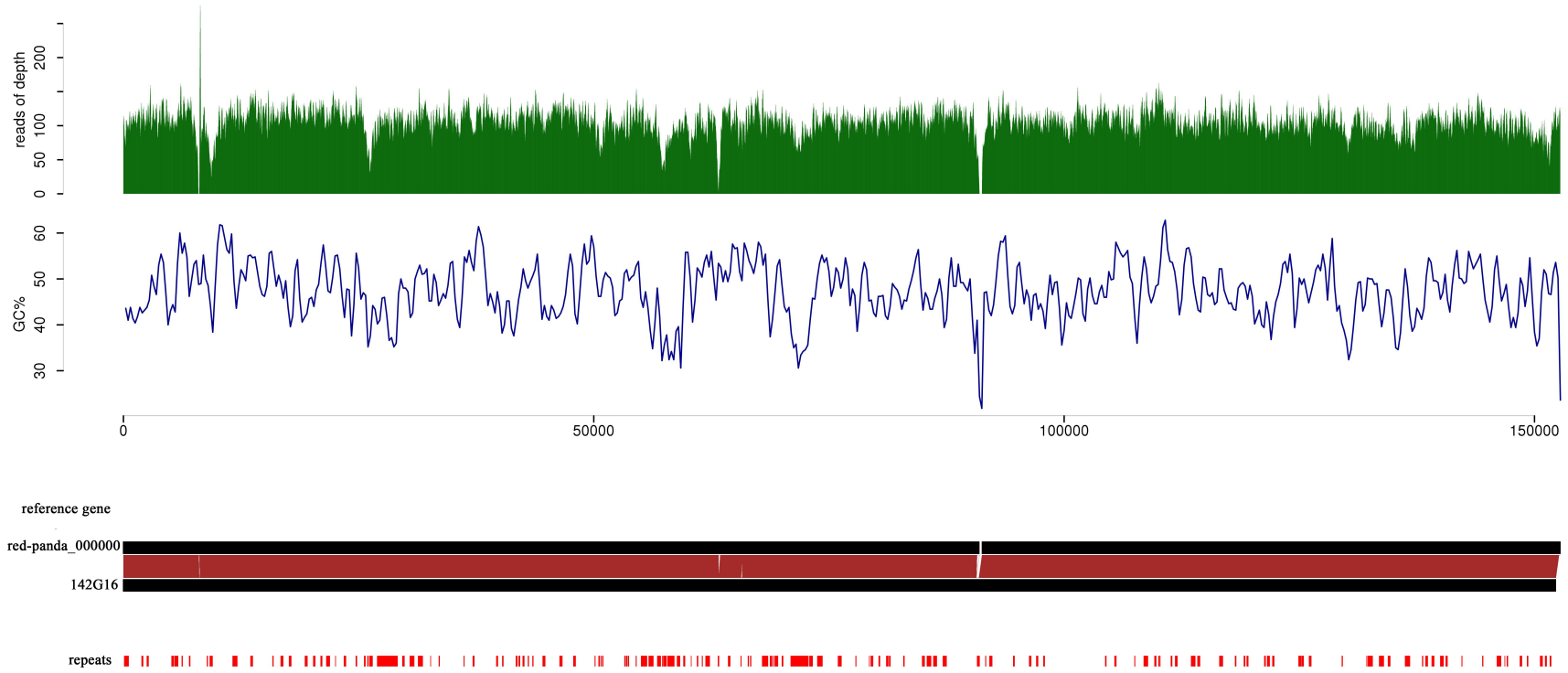


Fig. S6. Comparison of the assembled red panda scaffolds with five BAC clone sequences.

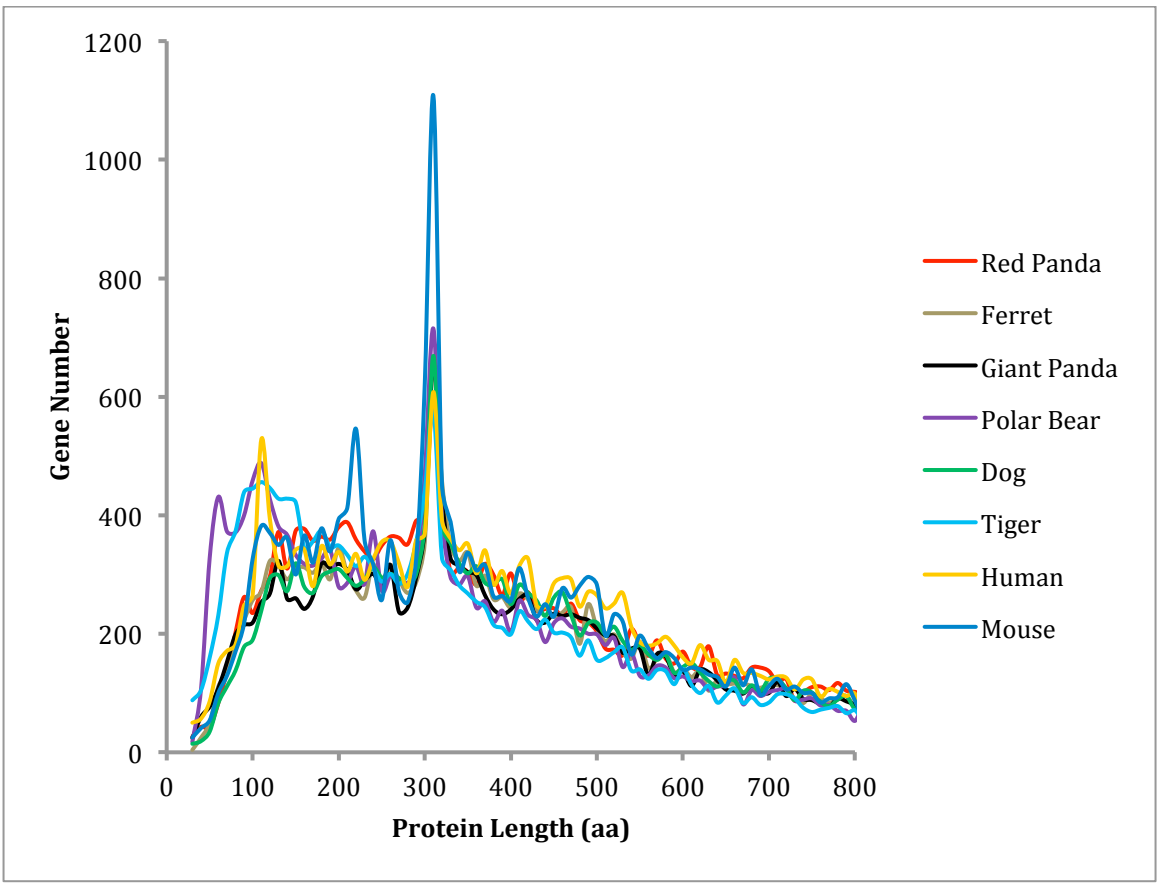


Fig. S7. Distribution of gene size in eight Eutherian species. The x-axis indicates gene size in amino acid length, and the y-axis represents gene number. Note that the peak around 300 amino acids (aa) corresponded to highly duplicated olfactory receptor genes in these mammalian genomes.

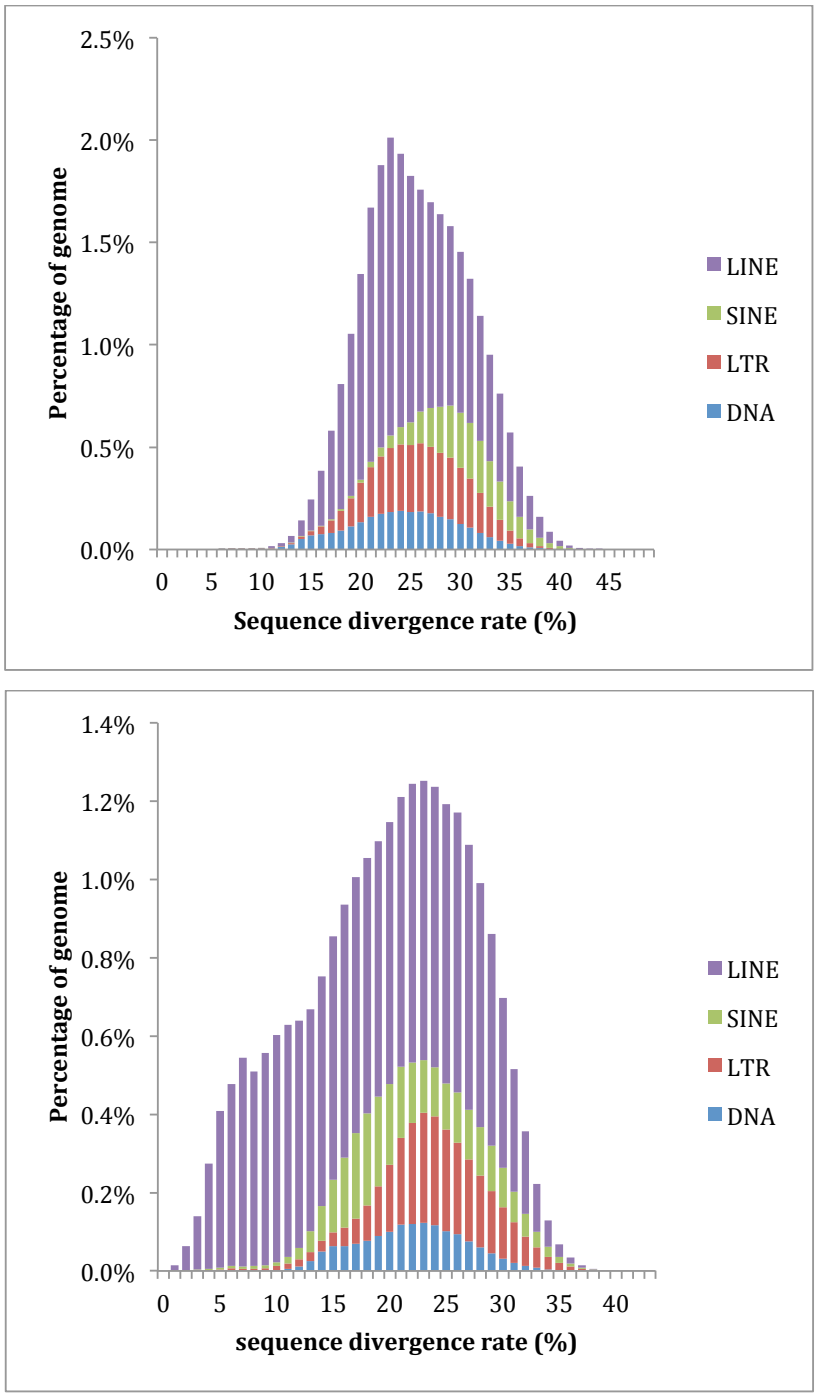


Fig. S8. Divergence distribution of each type of repeat elements in the red panda genome. The top panel shows repeat elements identified by RepeatMasker, and the bottom panel are repeat elements *de novo* identified by RepeatModeler.

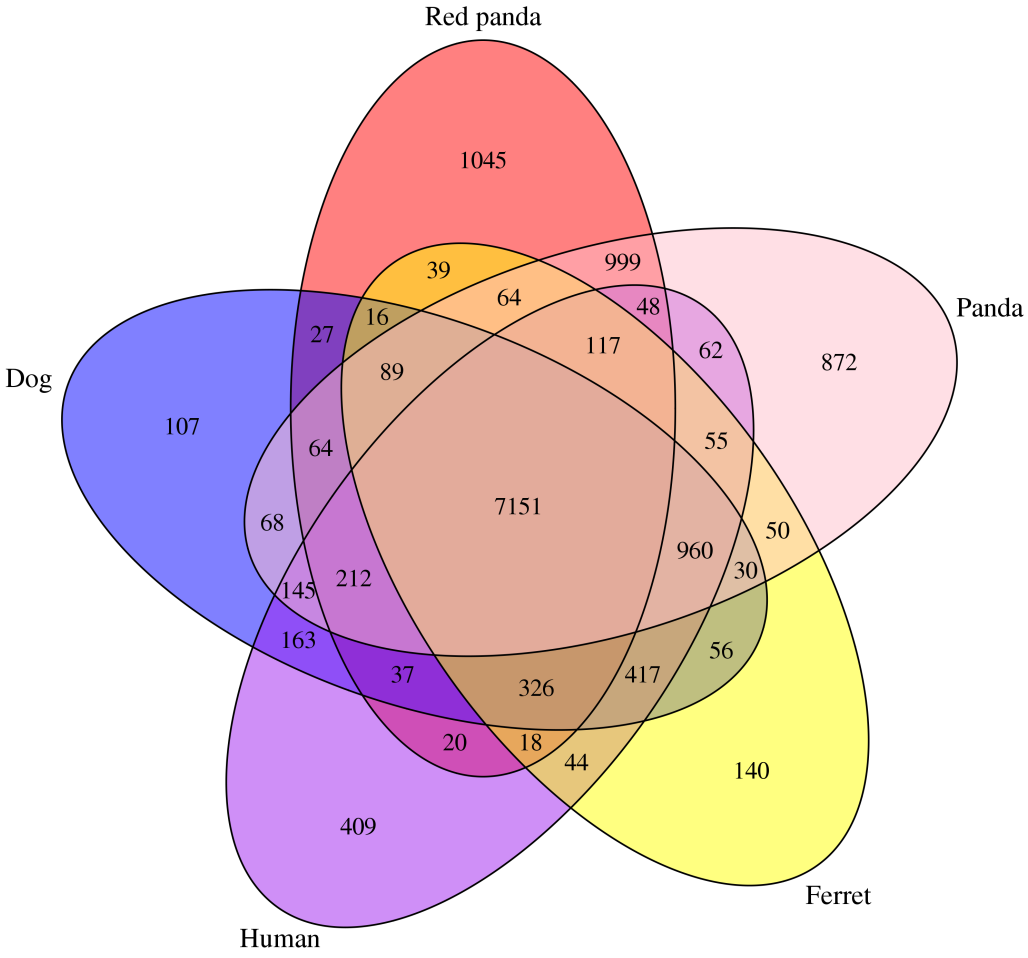


Fig. S9. Venn diagram of gene families shared among five Eutherian species.

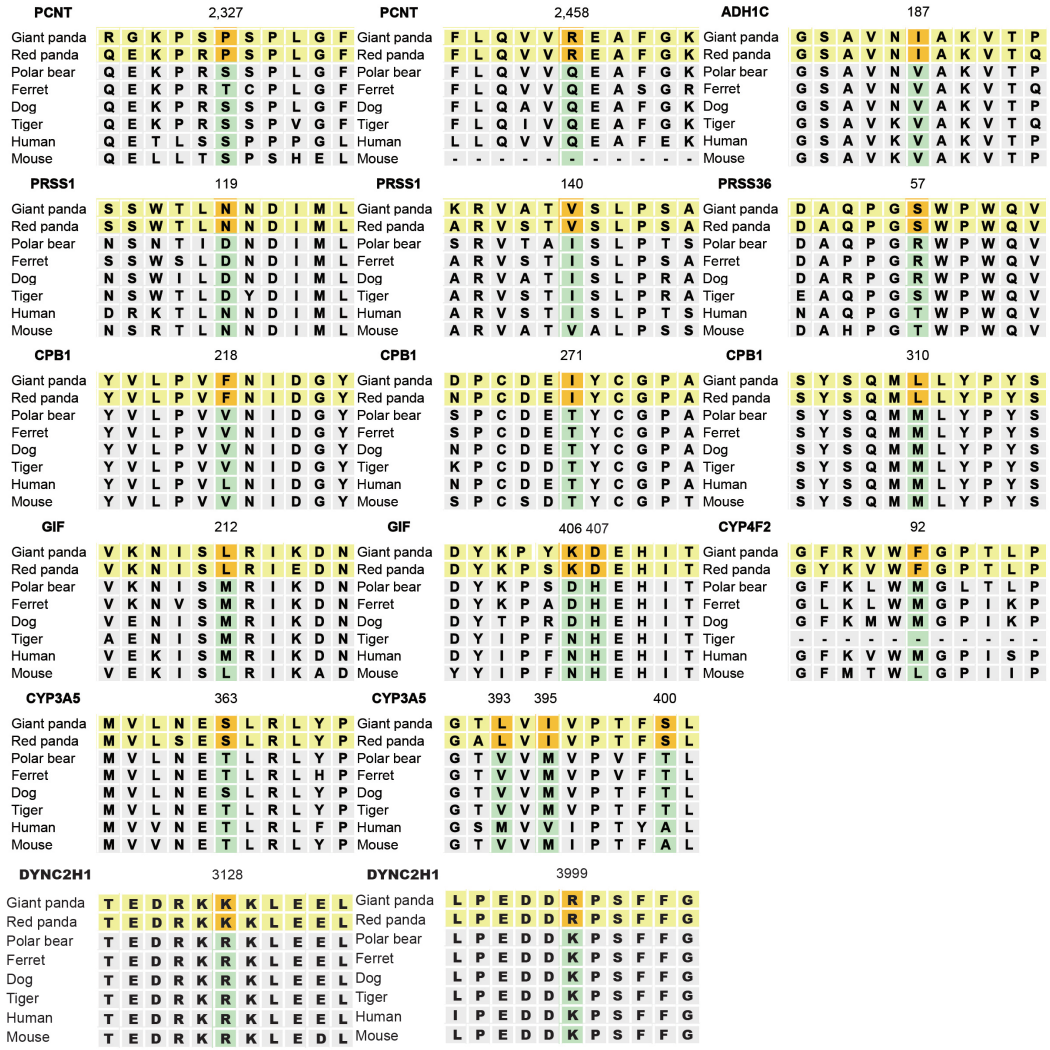


Fig. S10. Convergent amino acid substitutions of nine genes under adaptive convergence closely related with limb development and essential nutrient utilization. The number above the convergent site denotes the amino acid position.

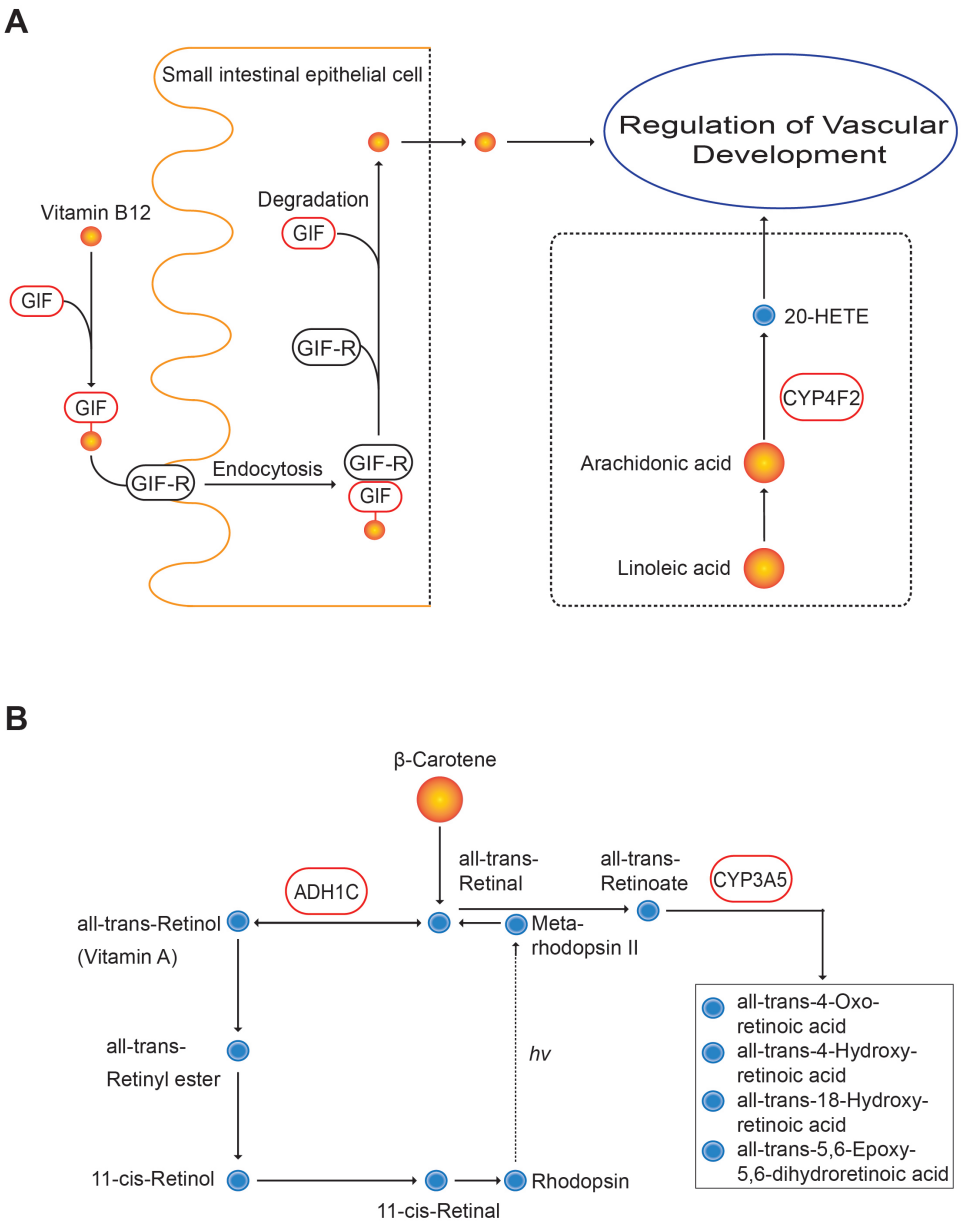


Fig. S11. Adaptively convergent genes involved in the metabolic pathways of essential nutrient utilization. (A) Vitamin B12 and arachidonic acid metabolism. Two adaptively convergent genes (*GIF* and *CYP4F2*) are highlighted by the red box. 20-HETE, 20-hydroxyeicosatetraenoic acid. (B) β-carotene and vitamin A metabolism. Two adaptively convergent genes (*ADH1C* and *CYP3A5*) are highlighted by the red box.

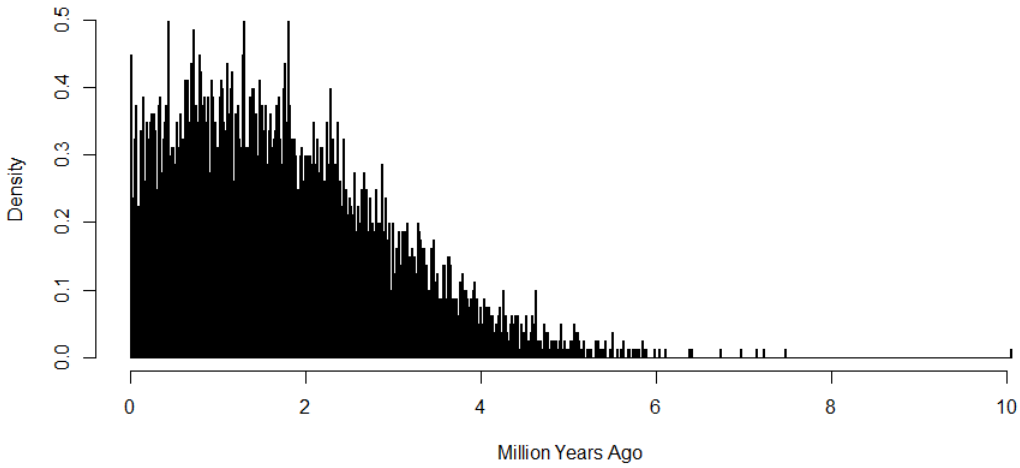


Fig. S12. Posterior probability distribution of the time (T_P) at which the *TAS1R1* gene inactivated.

Table S1. Library statistics of red panda genome sequencing.

Insert size	Libraries	Total paired reads (M)	Read length	Sequence depth (X)
180 bp	1	107.1	100 bp	8.9
500 bp	2	340.4	100 bp	28.4
750 bp	2	306.9	100 bp	25.6
1.5 kb	1	65.4	100 bp	5.5
2 kb	2	153.3	100 bp	12.8
3 kb	1	69.7	100 bp	5.8
5 kb	3	293.4	100 bp	24.4
8 kb	2	123.6	100 bp	10.3
Total	14	1,459.8	-	121.7

Table S2. Summary statistics of the red panda genome assembly.

Red panda	Contig		Scaffold	
	Size (bp)	**Number**	**Size (bp)**	**Number**
N90	24,057	24,512	456,041	990
N80	42,081	17,377	929,949	639
N70	59,958	12,805	1,470,549	437
N60	78,473	9,445	2,203,262	306
N50	98,981	6,826	2,983,736	215
Longest	905,050	---	17,255,363	---
Total Size	2,309,121,644	69,991	2,343,003,074	11,614
≥500 bp	2,306,642,241	60,358	2,343,003,074	11,614
≥2 kb	2,291,277,602	45,096	2,336,940,221	4,739

Table S3. Comparison of the assembled scaffolds with the Sanger-sequenced BAC clone sequences. Note that the BAC library was constructed with a male red panda different from that used in genome sequencing.

BAC ID	Length (bp)	Repeat content (%)	Coverage by initial contigs (%)	Coverage by scaffolds (%)	Number of single-base difference	Number of insertion and deletion	Length of insertion and deletion (bp)	Average read depth on scaffold	Rate of single-base difference (%)
135K16	132,859	38.41	96.41	97.04	228	11	14	112.91	0.172
136M11	149,488	34.38	99.3	99.85	147	5	5	81.69	0.098
138J8	173,910	34.98	98.13	98.87	126	5	5	106.9	0.072
139H13	148,420	32.99	96.84	98.09	79	4	6	102.81	0.053
142G16	152,267	30.13	99.35	99.81	223	9	11	106.53	0.146
Average	151,389	34.18	98.01	98.73	160	7	8	102	0.106

Table S4. Statistics of the completeness of the red panda genome assembly based on 248 core eukaryotic genes (CEGs).

	Number of CEGs	Completeness (%)	Total	Average	Orthology (%)
Complete	242	97.58	517	2.14	72.31
Group 1	63	95.45	114	1.81	53.97
Group 2	55	98.21	117	2.13	76.36
Group 3	59	96.72	122	2.07	76.27
Group 4	65	100	164	2.52	83.08
Partial	248	100	647	2.61	82.66
Group 1	66	100	152	2.3	72.73
Group 2	56	100	134	2.39	82.14
Group 3	61	100	158	2.59	86.89
Group 4	65	100	203	3.12	89.23

Note: Number of CEGs, the number of ultra-conserved CEGs identified in the genome; Completeness, the percentage of CEGs presence; Total, the total number of CEGs present including putative orthologs; Average, the average number of orthologs per CEG; and Orthology, the percentage of detected CEGs that have more than one ortholog.

Table S5. Statistics of BUSCO assessment of the red panda genome assembly and gene set prediction.

Red panda	No. of total BUSCOs	BUSCO mode: genome assembly		BUSCO mode: gene set prediction	
		No. of genes	Percentage	No. of genes	Percentage
Complete BUSCOs	3023	2632	87.07%	2533	83.79%
Complete and single-copy BUSCOs	3023	2597	85.91%	2499	82.67%
Complete and duplicated BUSCOs	3023	35	1.16%	34	1.12%
Fragmented BUSCOs	3023	329	10.88%	389	12.87%
Missing BUSCOs	3023	62	2.05%	101	3.34%

Table S6. General statistics of predicted protein-coding genes in the red panda genome.

Red panda gene		Number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	AUGUSTUS	25,526	26,296	1,262	6.84	184	4,286
	Fgenesh	54,461	12,883	1,271	7.85	162	1,696
Homolog	*Homo sapiens*	21,822	16,403	1,263	6.54	182	2,828
	Canis lupus familiaris	22,364	16,211	1,242	6.49	181	2,836
Final		21,940	19,780	1,672	9.86	170	2,043

Table S7. Predicted non-coding RNAs in the red panda genome.

ncRNA Type	Loci	Average length (bp)	Total length (bp)
tRNA	515	75	38,739
rRNA	961	89	85,941
C/D box	301	111	33,282
H/ACA	117	203	23,783
snRNA	1,515	121	183,766
miRNA	215	98	21,087

Table S8. Comparison of tRNA genes in eight Eutherian genomes.

Summary of tRNA genes	Red panda	Ferret	Giant panda	Polar bear	Dog	Tiger	Mouse	Human
tRNAs decoding Standard 20 AA	512	561	398	620	404	505	378	545
Selenocysteine tRNAs (TCA)	3	0	0	0	1	15	0	1
Possible suppressor tRNAs (CTA,TTA)	0	1	1	1	1	2	0	2
tRNAs with undetermined/unknown isotypes	0	1	1	0	0	0	0	2
Predicted pseudogenes	0	0	0	0	0	0	0	0
Total tRNAs	515	563	400	621	406	522	378	550

Table S9. Detailed composition of repeat elements in the red panda genome.

	Number	Length (bp)	Percent (%)
Class I elements (Retroelements)	**2,495,672**	**686,543,916**	**29.73**
LTR Retrotransposon	**368,917**	**114,519,427**	**4.96**
LTR/Copia	11	879	0.00
LTR/Delta	3	321	0.00
LTR/ERV1	64,286	20,369,905	0.88
LTR/ERVK	1,851	780,731	0.03
LTR/ERVL	110,057	39,226,191	1.70
LTR/ERVL-MaLR	171,712	49,628,108	2.15
LTR/Foamy	2	183	0.00
LTR/Gypsy	14,511	3,030,554	0.13
LTR/Lenti	1	27	0.00
LTR/Pao	2	305	0.00
unclassified LTR	6,481	1,482,223	0.06
non-LTR Retrotransposon	**2,126,755**	**572,024,489**	**24.77**
LINE	**1,502,377**	**485,036,122**	**21.01**
LINE/CR1	46,511	8,468,785	0.37
LINE/Dong-R4	348	77,451	0.00
LINE/Jockey	17	1,746	0.00
LINE/L1	1,104,020	397,311,977	17.21
LINE/L1-Tx1	75	7,415	0.00
LINE/L2	336,654	76,099,498	3.30
LINE/Penelope	613	62,672	0.00
LINE/R2	13	780	0.00
LINE/RTE-BovB	4,274	683,455	0.03
LINE/RTE-X	9,812	2,319,071	0.10
other LINE	40	3,272	0.00
SINE	**624,378**	**86,988,367**	**3.77**
SINE/5S-Deu-L2	1,749	205,906	0.01
SINE/Alu	3	325	0.00
SINE/MIR	453,808	66,415,564	2.88
SINE/tRNA	164,650	19,824,586	0.86
SINE/tRNA-Deu	392	50,023	0.00
SINE/tRNA-RTE	3,743	489,222	0.02
unclassified SINE	33	2,741	0.00
Class II elements (DNA Transposons)	**323,889**	**65,643,081**	**2.84**
DNA Transposon	**321,953**	**65,278,493**	**2.83**
DNA/CMC-EnSpm	4,146	312,598	0.01
DNA/MULE-MuDR	133	18,075	0.00
DNA/PIF-Harbinger	20	2,040	0.00

DNA/PiggyBac	602	205,966	0.01
DNA/TcMar	357	67,036	0.00
DNA/TcMar-Mariner	4,349	584,599	0.03
DNA/TcMar-Pogo	31	3,702	0.00
DNA/TcMar-Tc2	6,309	1,412,390	0.06
DNA/TcMar-Tigger	52,091	15,253,280	0.66
DNA/hAT	7,981	1,230,074	0.05
DNA/hAT-Ac	1,781	209,624	0.01
DNA/hAT-Blackjack	15,412	2,899,673	0.13
DNA/hAT-Charlie	189,583	35,228,987	1.53
DNA/hAT-Tag1	131	17,062	0.00
DNA/hAT-Tip100	34,448	7,196,685	0.31
other DNA Transposon	31	6,049	0.00
unclassified DNA Transposon	4,548	630,653	0.03
RC/Helitron	**1,936**	**364,588**	**0.02**
Satellite	**1,362**	**458,276**	**0.02**
unclassified Satellite	1,352	457,203	0.02
Satellite/acro	1	270	0.00
Satellite/centr	5	275	0.00
Satellite/telo	4	528	0.00
Low_complexity	**366,791**	**70,158,576**	**3.04**
Simple_repeat	**918,727**	**99,777,534**	**4.32**
TRF	**97,368**	**8,309,904**	**0.36**
Other repeats	**77,087**	**642,045**	**0.03**
RNA	611	126,511	0.01
Retroposon/SVA	2	121	0.00
rRNA	1,492	149,335	0.01
scRNA	60	4,740	0.00
snRNA	4,096	309,867	0.01
srpRNA	57	12,732	0.00
tRNA	515	38,739	0.00
Unknown repeats	**160,417**	**20,586,452**	**0.89**
Total repeat elements	**4,371,059**	**952,119,784**	**41.23**

Table S10. Detailed composition of repeat elements in the giant panda genome.

	Number	Length (bp)	Percent (%)
Class I elements (Retroelements)	**2,730,984**	**758,875,315**	**31.86**
LTR Retrotransposon	**395,461**	**124,969,510**	**5.25**
LTR/Copia	6	579	0.00
LTR/ERV	79	13,232	0.00
LTR/ERV1	67,493	21,680,311	0.91
LTR/ERVK	338	138,027	0.01
LTR/ERVL	118,579	43,112,696	1.81
LTR/ERVL-MaLR	185,329	54,863,493	2.30
LTR/Gypsy	16,573	3,554,259	0.15
LTR/Lenti	2	448	0.00
LTR/Pao	3	838	0.00
unclassified LTR	7,059	1,605,627	0.07
non-LTR Retrotransposon	**2,335,523**	**633,905,805**	**26.62**
LINE	**1,562,713**	**516,958,103**	**21.71**
LINE/CR1	51,065	9,345,302	0.39
LINE/Dong-R4	375	86,990	0.00
LINE/I	37	3,677	0.00
LINE/Jockey	48	3,990	0.00
LINE/L1	1,124,247	418,764,418	17.58
LINE/L1-Tx1	174	15,405	0.00
LINE/L2	368,073	84,558,670	3.55
LINE/Penelope	2,640	753,476	0.03
LINE/R1	15	1,112	0.00
LINE/R2	25	1,740	0.00
LINE/RTE	49	3,265	0.00
LINE/RTE-BovB	4,694	758,010	0.03
LINE/RTE-X	11,177	2,653,966	0.11
other LINE	94	8,082	0.00
SINE	**772,810**	**116,947,702**	**4.91**
SINE/5S-Deu-L2	1,758	198,072	0.01
SINE/Alu	4	295	0.00
SINE/Lys	270,737	43,423,846	1.82
SINE/MIR	494,275	72,550,701	3.05
SINE/tRNA	1,458	193,364	0.01
SINE/tRNA-Deu	433	51,716	0.00
SINE/tRNA-RTE	4,117	527,450	0.02
unclassified SINE	28	2,258	0.00
Class II elements (DNA Transposons)	**346,645**	**71,777,977**	**3.01**
DNA Transposon	**344,932**	**71,395,420**	**3.00**

DNA/Chapaev	608	55,708	0.00
DNA/MULE-MuDR	164	15,579	0.00
DNA/PIF-Harbinger	19	2,667	0.00
DNA/PiggyBac	626	221,391	0.01
DNA/TcMar	479	87,965	0.00
DNA/TcMar-Mariner	4,810	655,764	0.03
DNA/TcMar-Pogo	33	3,912	0.00
DNA/TcMar-Tc2	6,714	1,531,326	0.06
DNA/TcMar-Tigger	54,930	16,589,787	0.70
DNA/Transib	627	48,536	0.00
DNA/hAT	8,928	1,387,706	0.06
DNA/hAT-Ac	1,772	218,907	0.01
DNA/hAT-Blackjack	16,972	3,157,905	0.13
DNA/hAT-Charlie	205,719	38,764,862	1.63
DNA/hAT-Tag1	132	16,811	0.00
DNA/hAT-Tip100	37,475	7,947,993	0.33
other DNA	51	9,939	0.00
unclassified DNA Transposon	4,873	678,662	0.03
RC/Helitron	**1,713**	**382,557**	**0.02**
Satellite	**286**	**55,452**	**0.00**
unclassified Satellite	272	51,346	0.00
Satellite/centr	9	2,964	0.00
Satellite/telo	5	1,142	0.00
Low_complexity	**282,569**	**48,218,373**	**2.02**
Simple_repeat	**823,991**	**85,609,513**	**3.59**
TRF	**97,775**	**10,626,786**	**0.45**
Other repeats	**76,677**	**493，465**	**0.02**
RNA	626	125,422	0.01
Retroposon/SVA	6	16,508	0.00
rRNA	883	86,752	0.00
scRNA	60	5,272	0.00
snRNA	2,450	209,515	0.01
srpRNA	44	8,949	0.00
tRNA	400	41,047	0.00
Unknown repeats	**22,412**	**7,615,908**	**0.32**
Total repeat elements	**4,309,131**	**983,272,789**	**41.29**

Table S11. Comparison of repeat elements in the red panda, giant panda, dog, and human genomes.

	Red panda		Giant panda		Dog		Human	
	Length (bp)	**%**	**Length (bp)**	**%**	**Length (bp)**	**%**	**Length (bp)**	**%**
DNA	65,643,081	2.84	71,777,977	3.01	64,889,929	2.71	103,876,654	3.59
LINE	485,036,122	21.01	516,958,103	21.71	456,223,872	19.07	640,619,088	22.11
LTR	114,519,427	4.96	124,969,510	5.25	117,146,888	4.9	271,241,060	9.36
SINE	86,988,367	3.77	116,947,702	4.91	113,239,262	4.73	391,625,775	13.52
Other	179,346,335	7.77	145,003,589	6.09	249,825,363	10.44	84,570,903	2.92
Unknown	20,586,452	0.89	7,615,908	0.32	2,495,596	0.1	1,355,877	0.05
Total	952,119,784	41.23	983,272,789	41.29	1,003,820,910	41.95	1,493,289,357	51.54

Table S12. Sequencing data used for the giant panda genome assembly.

Insert size	Libraries	Total paired reads (M)	Read length	Sequence depth (X)
500 bp*	2	438.6	100 bp	36.6
750 bp*	2	543	100 bp	45.3
1.7-2.8 kb+	5	198	71 bp	11.7
3.7-7.5 kb+	4	245	38 bp	7.8
9.2-12.3 kb+	2	124	35 bp	3.6
Total	15	1,548.6	-	105

* Library prepared in this project; + Mate-pair dataset downloaded from ref. 2.

Table S13. Summary statistics of the giant panda genome assembly.

Giant panda	Contig		Scaffold	
	Size (bp)	**Number**	**Size (bp)**	**Number**
N90	23,145	21,170	1,993,038	255
N80	48,601	14,264	4,389,424	179
N70	72,966	10,297	6,777,671	134
N60	98,258	7,494	8,172,382	102
N50	126,709	5,361	9,947,519	75
Longest	1,471,649	---	32,438,596	---
Total Size	2,381,719,198	107,028	2,428,763,181	57,465
≥500 bp	2,377,554,755	89,491	2,428,763,181	57,465
≥2 kb	2,333,013,924	46,804	2,385,084,718	15,123

Table S14. General statistics of predicted protein-coding genes in the giant panda genome.

Giant panda gene		Number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
De novo	AUGUSTUS	25,046	24,792	1,251	6.85	183	4,024
	Fgenesh	49,374	12,002	1,153	6.73	171	1,892
Homolog	*Homo sapiens*	21,697	16,047	1,236	6.47	180	2,817
	Canis lupus familiaris	22,296	16,548	1,211	6.40	179	2,830
Final		23,371	17,705	1,453	8.32	175	2,219

Table S15. Genome version information and statistics of protein-coding genes for eight Eutherian species used in this study including giant and red pandas.

Species	Latin name	Version	Number of unique gene	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
Human	*Homo sapiens*	GRCh37.71	23,145	45,165	1,619	9	174	5,233
Mouse	*Mus musculus*	GRCm38.71	22,796	35,595	1,557	9	180	4,437
Dog	*Canis lupus familiaris*	CanFam3.1.71	19,856	38,418	1,641	10	172	4,313
Ferret	*Mustela putorius furo*	MusPutFur1.0.72	19,910	38,722	1,602	10	169	4,369
Tiger	*Panthera tigris*	PanTig1.0	20,226	35,544	1,377	8	170	4,818
Polar bear	*Ursus maritimus*	UrsMar1.0	21,142	31,451	1,457	9	169	3,930
Giant panda	*Ailuropoda melanoleuca*	This study	23,371	17,705	1,453	8	175	2,219
Red panda	*Ailurus fulgens*	This study	21,940	19,780	1,672	10	170	2,043

Table S16. Different test strategies and the identified gene number in positive selection analysis.

Strategy No.	Species of interest to be selected as foreground branch	All the species used in positive selection analysis	Number of positively selected gene
1	Giant panda	Giant panda & polar bear	155
2	Red panda	Red panda & ferret	79
3	Giant panda & red panda	Red panda, giant panda, polar bear & ferret	128
4	Giant panda	Giant panda, polar bear, ferret, tiger & dog	55
5	Red panda	Red panda, polar bear, ferret, tiger & dog	143
6	Giant panda & red panda	Red panda, giant panda, polar bear, ferret, tiger & dog	102
Total	---	---	416

Table S17. Significantly enriched KEGG pathways for 416 positively selected genes in giant and red pandas, using the 14,254 genes as the reference. The KEGG pathways associated with the utilization of essential nutrients are in red italics.

KEGG ID	KEGG Pathway	Observed number of genes	Expected number of genes	*P-value*
ko04974	*Protein digestion and absorption*	*14*	*2.49418*	*1.59745E-07*
ko04514	Cell adhesion molecules (CAMs)	8	3.0517	0.0114151
ko05150	Staphylococcus aureus infection	4	1.0857	0.0224824
ko04512	ECM-receptor interaction	6	2.20074	0.0226371
ko05146	Amoebiasis	7	2.84629	0.0239012
ko00740	*Riboflavin metabolism*	*2*	*0.264089*	*0.0269741*
ko05416	Viral myocarditis	4	1.34979	0.0453198
ko00604	Glycosphingolipid biosynthesis - ganglio series	2	0.352119	0.0466749
ko04380	Osteoclast differentiation	7	3.28644	0.0467302
ko04640	Hematopoietic cell lineage	5	1.99534	0.0491144

Table S18. Significantly enriched gene ontology (GO) terms at Biological Process level for 416 positively selected genes in giant and red pandas, using the 14,254 genes as the reference. The GO terms associated with the limb development and utilization of essential nutrients are in red italics.

GO ID	GO Term	Observed number of genes	Expected number of genes	*P-value*
GO:0030574	collagen catabolic process	14	1.99534	8.15841E-09
GO:0044243	multicellular organismal catabolic process	14	2.20074	3.07856E-08
GO:0032963	collagen metabolic process	15	2.81695	1.18739E-07
GO:0044259	multicellular organismal macromolecule metabolic process	15	2.93432	2.06616E-07
GO:0022617	extracellular matrix disassembly	16	3.34513	2.08958E-07
GO:0044236	multicellular organismal metabolic process	15	3.22776	7.33937E-07
GO:0050854	regulation of antigen receptor mediated signaling pathway	8	1.11504	1.15794E-05
GO:0050856	regulation of T cell receptor signaling pathway	5	0.792267	0.00100482
GO:0050858	negative regulation of antigen receptor mediated signaling pathway	4	0.498835	0.00128314
GO:0000042	protein targeting to Golgi	4	0.528178	0.00161188
GO:0071230	cellular response to amino acid stimulus	6	1.2911	0.00168916
GO:1900746	regulation of vascular endothelial growth factor signaling pathway	3	0.264089	0.00184631
GO:0072600	establishment of protein localization to Golgi	4	0.586865	0.00243648
GO:0086069	bundle of His cell to Purkinje myocyte communication	3	0.293432	0.00258046
GO:1902547	regulation of cellular response to vascular endothelial growth factor stimulus	3	0.293432	0.00258046
GO:0010310	regulation of hydrogen peroxide metabolic process	3	0.322776	0.00347141
GO:1903792	negative regulation of anion transport	4	0.645551	0.00351224
GO:0000301	retrograde transport vesicle recycling within Golgi	4	0.645551	0.00351224

GO:0061337	cardiac conduction	6	1.55519	0.00440483
GO:0002710	negative regulation of T cell mediated immunity	3	0.352119	0.0045286
GO:0050855	regulation of B cell receptor signaling pathway	3	0.352119	0.0045286
GO:0014063	negative regulation of serotonin secretion	2	0.117373	0.00495524
GO:0002027	regulation of heart rate	7	2.11271	0.00508309
GO:0010640	regulation of platelet derived growth factor receptor signaling pathway	3	0.381462	0.00576024
GO:0086065	cell communication involved in cardiac conduction	5	1.17373	0.00599719
GO:0050860	negative regulation of T cell receptor signaling pathway	3	0.410805	0.00717339
GO:0050851	antigen receptor mediated signaling pathway	12	5.34047	0.00748964
GO:0009313	oligosaccharide catabolic process	2	0.146716	0.00809873
GO:2001204	regulation of osteoclast development	2	0.146716	0.00809873
GO:0006891	intra Golgi vesicle mediated transport	5	1.26176	0.00816823
GO:0050857	positive regulation of antigen receptor mediated signaling pathway	3	0.440149	0.008774
GO:0006837	serotonin transport	3	0.469492	0.010567
GO:0061515	myeloid cell development	5	1.34979	0.0108314
GO:0009311	oligosaccharide metabolic process	5	1.34979	0.0108314
GO:0034638	phosphatidylcholine catabolic process	2	0.176059	0.0119133
GO:0014062	regulation of serotonin secretion	2	0.176059	0.0119133
GO:0030048	actin filament based movement	8	3.16907	0.0141148
GO:0070252	actin mediated cell contraction	7	2.5822	0.0147383
GO:0034067	protein localization to Golgi apparatus	4	0.968327	0.015259
GO:2001135	regulation of endocytic recycling	2	0.205403	0.0163571
GO:0007229	integrin mediated signaling pathway	7	2.75826	0.0205003
GO:0008016	regulation of heart contraction	9	4.07871	0.0213458
GO:0048739	cardiac muscle fiber development	2	0.234746	0.0213901
GO:0060973	cell migration involved in heart development	2	0.234746	0.0213901

GO:0046643	regulation of gamma delta T cell activation	2	0.234746	0.0213901
GO:0038084	vascular endothelial growth factor signaling pathway	3	0.616208	0.0225282
GO:0042036	negative regulation of cytokine biosynthetic process	3	0.645551	0.0255308
GO:0002707	negative regulation of lymphocyte mediated immunity	3	0.645551	0.0255308
GO:0042269	regulation of natural killer cell mediated cytotoxicity	3	0.645551	0.0255308
GO:0002715	regulation of natural killer cell mediated immunity	3	0.645551	0.0255308
GO:0086005	ventricular cardiac muscle cell action potential	3	0.645551	0.0255308
GO:0050853	B cell receptor signaling pathway	4	1.14439	0.0267559
GO:0030049	muscle filament sliding&actin myosin filament sliding	4	1.14439	0.0267559
GO:0007528	neuromuscular junction development	4	1.14439	0.0267559
GO:0097202	activation of cysteine type endopeptidase activity	6	2.28877	0.0268624
GO:0090161	Golgi ribbon formation	2	0.264089	0.0269741
GO:0001911	negative regulation of leukocyte mediated cytotoxicity	2	0.264089	0.0269741
GO:0010642	negative regulation of platelet derived growth factor receptor signaling pathway	2	0.264089	0.0269741
GO:0045060	negative thymic T cell selection	2	0.264089	0.0269741
GO:1900040	regulation of interleukin 2 secretion	2	0.264089	0.0269741
GO:0060048	cardiac muscle contraction	7	2.93432	0.0276716
GO:0002823	negative regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	3	0.674894	0.0287372
GO:0032731	positive regulation of interleukin 1 beta production	3	0.674894	0.0287372
GO:0045577	regulation of B cell differentiation	3	0.674894	0.0287372
GO:0051180	*vitamin transport*	*3*	*0.674894*	*0.0287372*
GO:0098751	*bone cell development*	*3*	*0.704238*	*0.0321465*
GO:0086091	regulation of heart rate by cardiac conduction	3	0.704238	0.0321465
GO:0071872	cellular response to epinephrine stimulus	2	0.293432	0.0330725
GO:0046629	gamma delta T cell activation	2	0.293432	0.0330725
GO:0043383	negative T cell selection	2	0.293432	0.0330725

GO:0031342	negative regulation of cell killing	2	0.293432	0.0330725
GO:0036035	osteoclast development	2	0.293432	0.0330725
GO:0001956	positive regulation of neurotransmitter secretion	2	0.293432	0.0330725
GO:0033235	positive regulation of protein sumoylation	2	0.293432	0.0330725
GO:2000810	regulation of bicellular tight junction assembly	2	0.293432	0.0330725
GO:0050852	T cell receptor signaling pathway	9	4.43083	0.033996
GO:0017144	drug metabolic process	4	1.23242	0.0340192
GO:0031638	zymogen activation	7	3.08104	0.0348176
GO:0071870	cellular response to catecholamine stimulus	3	0.733581	0.0357573
GO:0071868	cellular response to monoamine stimulus	3	0.733581	0.0357573
GO:0014047	glutamate secretion	4	1.26176	0.0366705
GO:0007030	Golgi organization	6	2.46483	0.0368073
GO:0002820	negative regulation of adaptive immune response	3	0.762924	0.0395676
GO:0035637	multicellular organismal signaling	7	3.16907	0.0396439
GO:0035584	calcium mediated signaling using intracellular calcium source	2	0.322776	0.0396505
GO:0009109	coenzyme catabolic process	2	0.322776	0.0396505
GO:0046475	glycerophospholipid catabolic process	2	0.322776	0.0396505
GO:0070970	interleukin 2 secretion	2	0.322776	0.0396505
GO:0031953	negative regulation of protein autophosphorylation	2	0.322776	0.0396505
GO:1900409	positive regulation of cellular response to oxidative stress&positive regulation of response to oxidative stress	2	0.322776	0.0396505
GO:0036010	protein localization to endosome	2	0.322776	0.0396505
GO:1903010	regulation of bone development	2	0.322776	0.0396505
GO:0071871	response to epinephrine	2	0.322776	0.0396505
GO:0001820	serotonin secretion	2	0.322776	0.0396505
GO:0006941	striated muscle contraction	8	3.90265	0.0422066
GO:0031333	negative regulation of protein complex assembly	6	2.55286	0.0425588

GO:0032732	positive regulation of interleukin 1 production	3	0.792267	0.0435749
GO:0071869	response to catecholamine	3	0.792267	0.0435749
GO:0071732	cellular response to nitric oxide	2	0.352119	0.0466749
GO:0007253	cytoplasmic sequestering of NF kappaB	2	0.352119	0.0466749
GO:0098909	regulation of cardiac muscle cell action potential involved in regulation of contraction	2	0.352119	0.0466749
GO:0043271	negative regulation of ion transport	6	2.61155	0.0466899
GO:0071867	response to monoamine	3	0.821611	0.0477765
GO:0050885	neuromuscular process controlling balance	4	1.37913	0.0484345
GO:0051283	negative regulation of sequestering of calcium ion&release of sequestered calcium ion into cytosol	5	1.99534	0.0491144

Table S19. 70 positively selected genes with nonrandom convergent amino acid substitution between red and giant pandas.

Gene symbol	*P-value*	*FDR*	Position of convergent amino acid substitution	Ancestral amino acid	Descendant amino acid
A2M	9.94E-06	1.74E-03	301	S	T
ABCA6	8.43E-05	9.80E-03	1430	N	D
ACOT11	1.75E-14	9.80E-12	528	L	I
ADH1C	1.75E-04	2.00E-02	187	V	I
BAHCC1	3.31E-05	4.24E-03	1356	S	G
C11orf86	3.64E-04	4.94E-02	95	R	S
C17orf104	1.10E-03	9.30E-02	378	N	Y
C3orf17	4.74E-04	6.24E-02	343	H	C
C7orf49	2.32E-04	3.55E-02	29	V	A
			23	E	K
			147	M	V
CCDC30	1.27E-04	2.79E-02	543	Q	R
CCDC73	7.02E-09	3.62E-06	723	K	L
CD163L1	1.42E-11	2.05E-08		A	T
CDHR2	9.54E-06	3.36E-03	1064	M	T
			921	M	V
CEACAM1	5.27E-05	1.42E-02	89	T	A
			358	R	G
COL16A1	9.50E-07	1.89E-04	58	N	S
COL4A6	9.23E-07	5.65E-04	1095	M	V
			681	R	H
CPB1	3.13E-06	4.73E-04	218	V	F
			310	M	L
			271	T	I
CYP3A5	2.58E-09	3.03E-06	363	T	S
			393	V	L
			400	T	S
			395	M	I
CYP4F2	9.99E-06	2.19E-03	92	M	F
DDX54	5.16E-05	2.16E-02	660	P	L
DENND5B	7.14E-04	7.78E-02	892	S	G
DYNC2H1	1.06E-04	3.72E-02	3128	R	K
			3999	K	R
F7	4.11E-05	1.20E-02	346	Q	R
FAM170A	7.48E-04	7.01E-02	16	P	L
FCRL3	8.09E-04	7.35E-02	384	I	V
GIF	2.42E-04	3.63E-02	406	D	K

			212	M	L
			407	H	D
GPR56	2.70E-04	6.87E-02	248	D	N
			252	H	R
HERC6	8.46E-05	9.80E-03	836	V	A
HEXB	9.19E-04	7.97E-02	522	L	I
IFI16	5.18E-05	6.33E-03	60	A	L
			57	P	R
IRX5	6.48E-04	7.12E-02	126	A	S
ITGAM	1.44E-06	7.77E-04	549	N	S
KIF27	2.79E-04	4.04E-02	664	H	R
			662	R	G
KRT75	1.11E-16	2.40E-13	213	S	G
			202	T	A
			287	L	F
LAG3	9.02E-05	2.20E-02	456	L	P
LETM1	6.47E-07	1.33E-04	683	I	V
LRRC36	1.45E-07	3.13E-05	477	R	Q
MDC1	5.83E-04	7.45E-02	18	E	K
MUC16	2.00E-15	1.58E-12	12412	K	R
			13413	W	Q
			12075	P	L
			13437	V	I
			12631	Q	R
			12088	E	T
			13432	R	P
			13424	Q	R
			13341	Q	S
			12072	F	S
			13491	R	H
			13261	Q	R
			12092	R	G
			13417	Q, H	R
			13339	D	G
MYH1	2.56E-05	1.28E-02	9	I	V
MYH3	3.60E-05	5.18E-03	812	I	V
MYH6	3.70E-04	4.97E-02	1025	S	A
			417	S	N
			1733	E	D
			1734	S	A
MYT1L	3.57E-09	1.00E-06	297	M	V
NFAM1	8.21E-04	9.63E-02	44	S	L
NLRP12	1.24E-09	3.09E-07	607	Q	R

			171	W	R
			285	V	A
NXPE4	5.32E-05	1.42E-02		N	K
OR4D10	1.49E-04	1.63E-02		R	K
PARP9	1.45E-06	7.77E-04	373	E	D
PCNT	3.18E-05	9.68E-03	2327	S	P
			2458	Q	R
PRSS1	3.47E-04	3.62E-02	119	D	N
			140	I	V
PRSS36	7.73E-12	3.45E-09	57	R	S
PTPRC	2.81E-09	6.74E-07	58	S	L
RIBC1	3.70E-04	3.83E-02	141	C	Y
			128	G	S
RTN3	1.58E-04	3.16E-02		T	A
SCRN2	4.14E-04	7.59E-02	388	R	Q
SLC4A1	1.52E-04	3.13E-02	95	G	S
SLC6A13	4.81E-04	4.79E-02	335	V	I
			429	F	L
SORL1	9.45E-10	2.98E-07	1207	H	R
SP140	2.26E-04	4.32E-02	746	R	G
			600	M	I
			589	I	E
			597	V	A
			586	H	L
			575	R	E
			604	K	E
			588	Q	P
			577	R	P
SPTBN5	8.73E-12	3.45E-09	3244	P	L
			3315	Q	R
			855	S	L
			75	I	V
			1002	V	M
			1724	Q	R
TAF7L	2.76E-06	4.22E-04	419	I	L
			411	K	R
TESPA1	1.80E-04	2.87E-02	480	Q	W
TMEM89	8.33E-07	1.40E-04	107	S	C
TRIB3	6.23E-04	6.90E-02	2	R	Q
TRIM22	3.71E-09	8.75E-07	12	E	D
			98	E	K
UGT3A2	4.83E-04	8.62E-02	426	Q	E
XDH	1.15E-05	1.98E-03		Q	R

XKR6	4.82E-05	8.14E-03	360	H	Q
ZNF284	1.03E-03	8.87E-02	53	K	R
ZNF746	3.35E-04	4.65E-02	95	I	V
			103	K	S

Table S20. Gene function, metabolic pathway involved and mutant phenotypes in humans and mice of 70 adaptively convergent genes. Disease/disorder/phenotype in humans are from OMIM (*www.omim.org*) and GeneCards databases (*www.genecards.org*), and mutant phenotypes in mice are from MGI database (*informatics.jax.org*).

Gene symbol	Full gene name	Gene function	Pathways involved	Disease/disorder/phenotype in humans	Mutant phenotype in mice
A2M	alpha-2-macroglobulin	Protease inhibitor	Signaling by GPCR; Rho GTPase cycle	Alpha-2-macroglobulin deficiency; Alzheimer disease	---
ABCA6	ATP-binding cassette, sub-family A, member 6	Energy-dependent transport of a wide variety of substrates across membranes, and may play a role in macrophage lipid homeostasis	Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds; ABC-family proteins mediated transport	---	---
ACOT11	acyl-CoA thioesterase 11	Catalyse the conversion of activated fatty acids to the corresponding non-esterified fatty acid and coenzyme A	---	Obesity	Mice homozygous for a null mutation display resistance to high fat diet induced obesity, inflammation and hepatic steatosis, increased energy expenditure, increased brown adipose tissue amount, and increased food intake
ADH1C	alcohol dehydrogenase 1C (class I), gamma polypeptide	Metabolize a wide variety of substrates, including ethanol, retinol, other aliphatic alcohols, hydroxysteroids, and lipid peroxidation products	Metabolism	Alcohol dependence; Liver cirrhosis; Parkinson disease	Homozygotes for targeted null mutations exhibit impaired metabolism of (and sensitivity to) ethanol and retinol

BAHCC1	BAH domain and coiled-coil containing 1	---	---	---	Homozygous mutation of this gene results in perinatal lethality for the majority of mutants; those that survive exhibit hind leg motor dysfunction
C11orf86	chromosome 11 open reading frame 86	---	---	---	---
C17orf104	chromosome 17 open reading frame 104	---	---	---	---
C3orf17	chromosome 3 open reading frame 17	---	Notch signaling pathway	---	Mice homozygous for a knock-out allele exhibit lethality prior to implantation with a failure to form a blastocyst
C7orf49	chromosome 7 open reading frame 49	Regulator of proteasome	---	Post-traumatic stress disorder and colorectal cancer	---
CCDC30	coiled-coil domain containing 30	---	---	---	---
CCDC73	coiled-coil domain containing 73	---	---	---	---
CD163L1	CD163 molecule-like 1	---	---	---	---
CDHR2	cadherin-related family member 2	Contact inhibition at the lateral surface of epithelial cells	---	---	---
CEACAM 1	carcinoembryonic antigen-related cell adhesion molecule 1	Mediate cell adhesion via homophilic as well as heterophilic binding to other proteins of the subgroup	Adhesion; EGFR1 signaling pathway	Gonorrhea; Adenomatoid tumor	Mice lacking appreciable levels of the two isoforms containing 4 Ig domains and having increased levels of the two isoforms containing

					2 Ig domains are viable and fertile. They are significantly more resistant to mouse hepatitis virus than wild-type mice.
COL16A1	collagen, type XVI, alpha 1	Mediate cell attachment and induce integrin-mediated cellular reactions, such as cell spreading and alterations in cell morphology	ERK signaling	---	---
COL4A6	collagen, type IV, alpha 6	Major structural component of glomerular basement membranes	PI3K-Akt signaling pathway; Pathways in cancer	Deafness, X-linked 6; X-linked non-syndromic sensorineural deafness type dfn	Male mice hemizygous for a knock-out allele are viable, fertile and healthy with no apparent defects in the topology of neuromuscular junctions in the diaphragm muscle
CPB1	carboxypeptidase B1	Preferential release of a C-terminal lysine or arginine amino acid	Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide, LLO) and transfer to a nascent protein; Pancreatic secretion	Trachea leiomyoma; Tricuspid valve insufficiency	---
CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	Oxidize a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics	Metabolism	Tacrolimus dose selection and engraftment syndrome; Salt-sensitive essential,	---

				susceptibility to hypertension	
CYP4F2	cytochrome P450, family 4, subfamily F, polypeptide 2	Oxidize a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics	Metabolism	Vitamin K antagonists toxicity or dose selection	---
DDX54	DEAD (Asp-Glu-Ala-Asp) box polypeptide 54	Has RNA-dependent ATPase activity; Represses the transcriptional activity of nuclear receptors	---	---	---
DENND5B	DENN/MADD domain containing 5B	Convert inactive GDP-bound Rab proteins into their active GTP-bound form	---	---	Mice homozygous for an ENU-induced allele exhibit normal blood lymphocyte populations
DYNC2H1	dynein, cytoplasmic 2, heavy chain 1	Function as a motor for intraflagellar retrograde transport	Class I MHC mediated antigen processing and presentation; Salmonella infection	Short-rib thoracic dysplasia 3 with or without polydactyly; Short-rib thoracic dysplasia 6 with or without polydactyly	Homozygotes for a gene trap allele show complete embryonic lethality with altered heart looping and brain morphology; Chemically induced mutants show randomized heart looping and polydactyly; Holoprosencephaly or exencephaly, micrognathia, and cardiac, renal, airway and eye defects may be observed
F7	coagulation factor VII	Initiate the extrinsic	Hemostasis and	Factor VII deficiency	Mice homozygous for a

		pathway of blood coagulation	formation of fibrin clot (Clotting Cascade)	and proconvertin deficiency, congenital; Decreased susceptibility to myocardial infarction	targeted null mutation developed normally through embryogenesis, and exhibited no vascular defects; however, 70% of homozygous neonates suffered fatal intra-abdominal haemorrhaging and died within 24 hours after birth
FAM170A	family with sequence similarity 170, member A	Act as a nuclear transcription factor that positively regulates the expression of heat shock genes	---	---	---
FCRL3	Fc receptor-like 3	May play a role in regulation of the immune system	---	Autoimmune pancreatitis; Juvenile rheumatoid arthritis	---
GIF	gastric intrinsic factor	Promote absorption of the essential vitamin cobalamin (Cbl) in the ileum	Disease; Metabolism	Intrinsic factor deficiency	---
GPR56	G protein-coupled receptor 56	Regulate the migration of neural precursor cells	Other GPCRs	Bilateral frontoparietal polymicrogyria; Bilateral perisylvian polymicrogyria	Mice homozygous for a null allele exhibit neuronal ectopias in the frontoparietal cortex due to disruptions in the pial basement membrane
HERC6	HECT and RLD domain containing E3 ubiquitin protein ligase	Catalyze the formation of a thioester with ubiquitin before transferring it to a	---	---	---

	family member 6	substrate			
HEXB	hexosaminidase B	Responsible for the degradation of GM2 gangliosides, and a variety of other molecules containing terminal N-acetyl hexosamines	Disease; Sphingolipid metabolism	Sandhoff disease, infantile, juvenile, and adult forms	Homozygous mutants exhibit spasticity, muscle weakness, rigidity, tremors, and ataxia beginning around 4 months of age and resulting in death about 6 weeks later; Mutants accumulate GM2 ganglioside and glycolipid GA2 in brain
IFI16	interferon, gamma-inducible protein 16	May function as a transcriptional repressor	Cytosolic sensors of pathogen-associated DNA; NF-kappaB signaling	---	---
IRX5	iroquois homeobox 5	Establishe the cardiac repolarization gradient by its repressive actions on the KCND2 potassium-channel gene; Required for retinal cone bipolar cell differentiation	---	Hamamy syndrome; Craniofacial dysplasia-osteopenia syndrome	Mice homozygous for a knock-out allele exhibit reduced body size, narrow eye opening, and impaired retinal cone bipolar cell development
ITGAM	integrin, alpha M	Implicated in various adhesive interactions of monocytes, macrophages and granulocytes as well as in mediating the uptake of complement-coated particles	Ras signaling pathway; Regulation of actin cytoskeleton	Systemic lupus erythematous, association with 6; Granulocytopenia	Homozygous null mice exhibit reduced staphylococcal enterotoxin-induced T cell proliferation, reduced neutrophil adhesion to fibrinogen, and defective homotypic aggregation and reduced degranulation of neutrophils

KIF27	kinesin family member 27	Play an essential role in motile ciliogenesis	---	---	Homozygous mice are small and die by 8 weeks and exhibit hydrocephalus, rhinitis and otitis media
KRT75	keratin 75, type II	Play a central role in hair and nail formation	---	Pseudofolliculitis barbae and loose anagen hair syndrome; Ectodermal dysplasia 4, hair/nail type	Mice homozygous for a knock-in mutation that results in the deletion of the highly conserved asparagine residue (N159) in the helix initiation peptide of this gene develop hair shaft and nail abnormalities resembling pachyonychia congenita
LAG3	lymphocyte-activation gene 3	Involved in lymphocyte activation; Bind to HLA class-II antigens	---	---	Mice homozygous for disruptions in this gene have a generally normal phenotype but do display reduced natural killer cell activity and increased T cell response to infection
LETM1	leucine zipper-EF-hand containing transmembrane protein 1	Crucial for the maintenance of mitochondrial tubular networks and for the assembly of the supercomplexes of the respiratory chain	---	Wolf-Hirschhorn syndrome	---
LRRC36	leucine rich repeat containing 36	---	---	---	---
MDC1	mediator Of	Required for checkpoint	Homologous	---	Homozygous mutant mice

	DNA-damage checkpoint 1	mediated cell cycle arrest in response to DNA damage within both the S phase and G2/M phases of the cell cycle	recombination repair		are smaller and display increased susceptibility to ionizing radiation, male infertility, T and B cell abnormalities, and increased genomic instability
MUC16	mucin 16, cell surface associated	Thought to provide a protective, lubricating barrier against particles and infectious agents at mucosal surfaces	Biosynthesis of the N-glycan precursor (dolichol lipid-linked oligosaccharide, LLO) and transfer to a nascent protein	Mucinous ovarian cystadenoma; Pleural cancer	Homozygous null mice are viable and fertile with no gross histological abnormalities; Homozygous male mice father larger litters when crossed to wild-type females
MYH1	myosin, heavy chain 1, skeletal muscle, adult	Muscle contraction	RhoGDI pathway; ERK signaling	Familial gastric cancer	Homozygotes for a targeted null mutation exhibit reduced growth, muscular weakness, kyphosis, and abnormal kinetics of muscle contraction and relaxation
MYH3	myosin, heavy chain 3, skeletal muscle, embryonic	Muscle contraction	RhoGDI pathway; ERK signaling	Arthrogryposis, distal, type 2b; Arthrogryposis, distal, type 2a; Arthrogryposis, distal, type 8	---
MYH6	myosin, heavy chain 6, cardiac muscle, alpha	Muscle contraction	RhoGDI pathway; ERK signaling	Cardiomyopathy, familial hypertrophic, 14 and atrial septal defect 3; Cardiomyopathy, dilated, 1EE	Mice homozygous for a knock-out allele exhibit embryonic lethality associated with heart defects while heterozygotes show cardiac myofibrillar

					disarray, cardiac dysfunction and fibrosis. Mice heterozygous for different knock-in alleles may develop hypertrophic or dilated forms of cardiomyopathy
MYT1L	myelin transcription factor 1-like	May function as a panneural transcription factor associated with neuronal differentiation	---	Mental retardation, autosomal dominant 39	---
NFAM1	NFAT activating protein with ITAM motif 1	May function in immune system as a receptor which activates via the calcineurin/NFAT-signaling pathway the downstream cytokine gene promoters; May be involved in the regulation of B-cell development	---	---	---
NLRP12	NLR family, pyrin domain containing 12	May mediate activation of CASP1 via ASC and promote activation of NF-kappa-B via IKK	NOD-like receptor signaling pathways; Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways	Familial cold autoinflammatory syndrome 2; Nlrp12-associated hereditary periodic fever syndrome	Mice homozygous for a null allele have defects in dendritic and myeloid cell migration and a decreased susceptibility to type IV hypersensitivity reactions; Mice homozygous for a second null allele display increased susceptibility to induced colitis and to chemically-induced tumors

NXPE4	neurexophilin and PC-esterase domain family, member 4	---	---	---	---
OR4D10	olfactory receptor, family 4, subfamily D, member 10	Odorant receptor	Signaling by GPCR	---	---
PARP9	poly (ADP-Ribose) polymerase family, member 9	Play a role in PARP1-dependent DNA damage repair	Apoptotic pathways in synovial fibroblasts and UVA-induced MAPK signaling	---	---
PCNT	pericentrin	Integral component of the filamentous matrix of the centrosome involved in the initial establishment of organized microtubule arrays in both mitosis and meiosis	Mitotic cell cycle	Microcephalic osteodysplastic primordial dwarfism, type II; Seckel syndrome 4	Mice homozygous for ENU mutations exhibit perinatal lethality, polydactyly, and abnormal interneuron migration; Heterozygotes exhibit sporadic seizures
PRSS1	protease, serine, 1 (trypsin 1)	Act on peptide linkages involving the carboxyl group of lysine or arginine	Disease; Degradation of the extracellular matrix	Trypsinogen deficiency; *PRSS1*-related hereditary pancreatitis	---
PRSS36	protease, serine, 36	Has a preference for substrates with an Arg or a Lys residue	---	---	---
PTPRC	protein tyrosine phosphatase, receptor type, C	Required for T-cell activation through the antigen receptor	Class I MHC mediated antigen processing and presentation and L1CAM interactions	T-b+ severe combined immunodeficiency due to CD45 deficiency	---
RIBC1	RIB43A domain with coiled-coils 1	---	---	---	---

RTN3	reticulon 3	Inhibits BACE1 activity and amyloid precursor protein processing	---	---	---
SCRN2	secernin 2	Dipeptidase	---	---	---
SLC4A1	solute carrier family 4 (anion exchanger), member 1 (Diego blood group)	Functions both as a transporter that mediates electroneutral anion exchange across the cell membrane and as a structural protein	Metabolism; Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds	Spherocytosis, type 4; Renal tubular acidosis, distal, AD; Renal tubular acidosis, distal, AR; Ovalocytosis	Homozygotes for null mutations exhibit retarded growth, severe spherocytosis, hemolytic anemia, lack of erythrocyte glycophorin A, mitotic defects, and high postnatal mortality
SLC6A13	solute carrier family 6 (neurotransmitter transporter), member 13	Sodium-dependent GABA and taurine transporter	Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds; Transmission across chemical synapses	Schizophrenia	Mice homozygous for a knock-out allele exhibit reduced taurine levels in the liver and increased taurine levels in the brain
SORL1	sortilin-related receptor, L(DLR Class) A repeats containing	May be implicated in the uptake of lipoproteins and of proteases	---	Alzheimer disease; Early-onset autosomal dominant alzheimer disease	Homozygous mutation of this gene results in decreased femoral artery intimal thickness after cuff placement and abolished angiotensin II stimulated vascular smooth muscle migration and attachment. Two other alleles show an increase in beta-amyloid deposits or peptide in the brain

SP140	SP140 nuclear body protein	May be involved in the pathogenesis of acute promyelocytic leukemia and viral infection	---	---	---
SPTBN5	spectrin, beta, non-erythrocytic 5	---	L1CAM interactions	Macular holes	---
TAF7L	TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated	Probably function as a spermatogensis-specific component of the DNA-binding general transcription factor complex TFIID	---	GPCR pathway; Apoptotic pathways in Synovial fibroblasts	Mice homozygous for a null allele display reduced male fertility, oligozoospermia, teratozoospermia, athenozoospermia, abnormal seminiferous tubules, and reduced testes weight on congenic but not on mixed strain backgrounds
TESPA1	thymocyte expressed, positive selection associated 1	Required for the development and maturation of T-cells	---	---	Mice homozygous for a knock-out allele exhibit impaired late thymocyte development
TMEM89	transmembrane protein 89	---	---	---	---
TRIB3	tribbles pseudokinase 3	Disrupt insulin signaling by binding directly to Akt kinases and blocking their activation	PI-3K cascade	Anoxia	Homozygous null mutants may display hypoactivity, decreased blood pressure and abnormal digit morphology or abnormal mast cell physiology
TRIM22	tripartite motif containing 22	Interferon-induced antiviral protein involved in cell innate immunity	---	MHC class II deficiency; HIV-1	---

UGT3A2	UDP glycosyltransferase 3 family, polypeptide A2	Play an important role in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds	---	Gilbert syndrome	---
XDH	xanthine dehydrogenase	Oxidative metabolism of purines	Metabolism	Xanthinuria, type I; Hereditary xanthinuria.	Homozygotes for a null allele are small and die prematurely while heterozygous females show a lactation defect; most homozygotes for another null allele die within the first month of renal failure associated with uric acid depletion, renal tubular damage, inflammation, fibrosis and oxidative stress
XKR6	XK, Kell blood group complex subunit-related family, member 6	---	---	Keratolytic winter erythema and hydrocele	---
ZNF284	zinc finger protein 284	May be involved in transcriptional regulation	---	---	---
ZNF746	zinc finger protein 746	Play a role in regulation of neuron death	Gene expression	---	---

Table S21. Significantly enriched KEGG pathways for 70 adaptively convergent genes between giant and red pandas, using the 14,254 genes as the reference. The KEGG pathways associated with the utilization of essential nutrients are in red italics.

KEGG ID	KEGG Pathway	Observed number of genes	Expected number of genes	*P-value*
ko04974	*Protein digestion and absorption*	*3*	*0.420853*	*0.00858522*
ko04530	Tight junction	3	0.559487	0.018456
ko00830	*Retinol metabolism*	*2*	*0.227756*	*0.021732*
ko00982	Drug metabolism - cytochrome P450	2	0.24261	0.0244598
ko00980	Metabolism of xenobiotics by cytochrome P450	2	0.257463	0.0273199
ko05204	Chemical carcinogenesis	2	0.282219	0.0323689
ko04610	Complement and coagulation cascades	2	0.282219	0.0323689

Table S22. Significantly enriched gene ontology (GO) terms at Biological Process level for 70 adaptively convergent genes between giant and red pandas, using the 14,254 genes as the reference. The GO terms associated with the limb development and utilization of essential nutrients are in red italics.

GO ID	GO Term	Observed number of genes	Expected number of genes	*P-value*
GO:0050857	positive regulation of antigen receptor mediated signaling pathway	3	0.0742682	5.06204E-05
GO:0050854	regulation of antigen receptor mediated signaling pathway	3	0.188146	0.000865113
GO:0031953	negative regulation of protein autophosphorylation	2	0.0544634	0.00129079
GO:0050855	regulation of B cell receptor signaling pathway	2	0.0594146	0.00154401
GO:0022617	extracellular matrix disassembly	4	0.564439	0.00243705
GO:0009235	*cobalamin metabolic process*	*2*	*0.0891219*	*0.00351132*
GO:0072378	blood coagulation fibrin clot formation	2	0.108927	0.0052342
GO:0042036	negative regulation of cytokine biosynthetic process	2	0.108927	0.0052342
GO:0032731	positive regulation of interleukin 1 beta production	2	0.113878	0.00571445
GO:0045577	regulation of B cell differentiation	2	0.113878	0.00571445
GO:0010927	*cellular component assembly involved in morphogenesis*	*5*	*1.15363*	*0.00589438*
GO:0097202	activation of cysteine type endopeptidase activity	3	0.386195	0.00677684
GO:0032732	positive regulation of interleukin 1 production	2	0.133683	0.00782762
GO:0031952	regulation of protein autophosphorylation	2	0.143585	0.00899676
GO:0045824	negative regulation of innate immune response	2	0.148536	0.0096088
GO:0050777	negative regulation of immune response	3	0.440658	0.0097339
GO:0045214	sarcomere organization	2	0.158439	0.0108868
GO:0007229	integrin mediated signaling pathway	3	0.465414	0.0112902
GO:0032651	regulation of interleukin 1 beta production	2	0.183195	0.0143881
GO:0031348	negative regulation of defense response	3	0.519878	0.0151944
GO:0031638	zymogen activation	3	0.519878	0.0151944

GO:2001056	positive regulation of cysteine type endopeptidase activity	3	0.524829	0.0155825
GO:0050853	B cell receptor signaling pathway	2	0.193097	0.0159073
GO:0030049	muscle filament sliding&actin myosin filament sliding	2	0.193097	0.0159073
GO:0017144	drug metabolic process	2	0.207951	0.0183086
GO:0002920	regulation of humoral immune response	2	0.207951	0.0183086
GO:0002244	hematopoietic progenitor cell differentiation	3	0.559487	0.018456
GO:0032611	interleukin 1 beta production	2	0.212902	0.0191411
GO:0010950	positive regulation of endopeptidase activity	3	0.56939	0.0193275
GO:0071230	cellular response to amino acid stimulus	2	0.217854	0.0199892
GO:0032652	regulation of interleukin 1 production	2	0.217854	0.0199892
GO:0035107	*appendage morphogenesis&limb morphogenesis*	*3*	*0.609*	*0.023039*
GO:0010952	positive regulation of peptidase activity	3	0.623853	0.024524
GO:0032612	interleukin 1 production	2	0.252512	0.0263521
GO:0033013	*tetrapyrrole metabolic process*	*2*	*0.257463*	*0.0273199*
GO:0030239	myofibril assembly	2	0.262414	0.0283019
GO:0007041	lysosomal transport	2	0.267366	0.0292979
GO:0048736	*appendage development&limb development*	*3*	*0.69317*	*0.0321244*
GO:0045582	positive regulation of T cell differentiation	2	0.282219	0.0323689
GO:0042384	*cilium assembly*	*3*	*0.737731*	*0.0375893*
GO:0030574	collagen catabolic process	2	0.336683	0.044632
GO:0045621	positive regulation of lymphocyte differentiation	2	0.341634	0.0458205
GO:0006919	activation of cysteine type endopeptidase activity involved in apoptotic process	2	0.351536	0.0482324
GO:0002698	negative regulation of immune effector process	2	0.351536	0.0482324
GO:0044782	*cilium organization*	*3*	*0.821902*	*0.0491226*
GO:0002027	regulation of heart rate	2	0.356488	0.0494554

Table S23. PCR sequencing results for two convergent amino acid sites of DYNC2H1 in an Asiatic black bear, a striped skunk and a raccoon.

Target gene	Target site	Primer name	Primer sequences	Targeted species	Nucleotide codon confirmed	Amino acid confirmed
DYNC2H1	R3128K	DYNC2H1-U1F	TCCCTTGGTTTCATTTTGGTA	Asiatic black bear / striped skunk / raccoon	AGG	3128R
		DYNC2H1-U1R	AAGATTTTCAAAACTGAAACATT			
	K3999R	DYNC2H1-U2F	ATAGACTGAAAGTTAGATTTAC	Asiatic black bear	AAA	3999K
		DYNC2H1-U2R	GCTTGTGGGGTCTGAWGTAT			
		DYNC2H1-U2F2	GGGTCCCTTTGAATGTGGTGA	striped skunk / raccoon		
		DYNC2H1-U2R	GCTTGTGGGGTCTGAWGTAT			

Table S24. PCR validation results for one frame-shift mutation of *TAS1R1* gene and 19 convergent amino acid sites of nine adaptively convergent genes. The Sanger sequencing results of three giant pandas, three red pandas, a polar bear and a ferret confirmed the existence of the frame-shift mutation and these convergent amino acid sites.

Target gene	Target site	Primer name	Primer sequences	Targeted species	Nucleotide codon confirmed	Amino acid confirmed
TAS1R1	1751 del C (exon 6)	TAS1R1-RP-6F TAS1R1-RP-6R	AGCCTTCTGCCTTTCCTTCC GCCTCGTGTAGTCCTGGATG	red panda	1751 del C (exon 6)	——
DYNC2H1	R3128K	DYNC2H1-GP-1F DYNC2H1-GP-1R	CCTGTTGTCCTCAGTTGCT CATTCTGTCCCTCTTGCTC	giant panda	AAG	3128K
		DYNC2H1-RP-1F DYNC2H1-RP-1R	TTATGCAGAGCACAGAGGC AGGAACAGAGCAGTGACAG	red panda	AAG	3128K
		DYNC2H1-PB-1F DYNC2H1-PB-1R	TGGAGGTGTTAAGAGGAT CACGCCACAGGTAAGGAC	polar bear	AGG	3128R
		DYNC2H1-FE-1F DYNC2H1-FE-1R	GGCTGTTTCTGCTGTTAT GCAATCTTGCATGAGGGA	ferret	AGG	3128R
	K3999R	DYNC2H1-GP-2F DYNC2H1-GP-2R	GGACTGAAGACTCGCTGTG TTAAAGAATGGCTTGTGGG	giant panda	AGA	3999R
		DYNC2H1-RP-2F DYNC2H1-RP-2R	CCTTGGAATAAGCCGAACT ACCTGTCCCGATGAGCACC	red panda	AGA	3999R
		DYNC2H1-PB-2F DYNC2H1-PB-2R	GGGGTTCTGTAAGTTTGG TAAAGAATGGCTTGTGGG	polar bear	AAA	3999K
		DYNC2H1-FE-2F DYNC2H1-FE-2R	CAGATGAGTCTCGGGTAG TTTGGAGGAAGTGATAGTG	ferret	AAA	3999K
PCNT	S2327P	PCNT-GP-1F PCNT-GP-1R	CACTCTTGTTTGGGTTCTGT CTGTGCTGTTCTGCTCCTT	giant panda	CCC	2327P
		PCNT-RP-1F PCNT-RP-1R	TGTGCTCCCAAGCGTGTAAT CCAGTGCCAAAGGGTCCAA	red panda	CCC	2327P
		PCNT-PB-1F	CGTGGCTGGGTTTGTTTA	polar bear	TCC	2327S

	Q2458R	PCNT-PB-1R	TCCGTGCTGTTCTGTTCC			
		PCNT-FE-1F	AAGTTTCCATCTGGCTAGTTC	ferret	ACC	2327T
		PCNT-FE-1R	ACAAGGAGGTCTGGGTGC			
		PCNT-GP-2F	TGCTGTCGGGTCGTTTCTG	giant panda	CGG	2458R
		PCNT-GP-2R	TGCCACGTTCCAAGGTCAA			
		PCNT-RP-2F	TCCGTGAAGGGACCTACCAAA	red panda	CGC	2458R
		PCNT-RP-2R	AGGCCAGTTACTTCCAGAGCACA			
		PCNT-PB-2F	TCTGCTGTTGGGTCGTTT	polar bear	CAG	2458Q
		PCNT-PB-2R	CATGTGCTCACCTGCTCC			
		PCNT-FE-2F	TTCGGTAGTGTCTCCTCTGTCA	ferret	CAG	2458Q
		PCNT-FE-2R	TGCTCCTGGACCACCTTCT			
PRSS1	D119N	PRSS1-F	AGGTGATCTTGCCACACTCC	giant panda / red panda	AAC	119N
		PRSS1-R	CACTCACTGCCAGAGCTCA	polar bear / ferret	GAC	119D
	I140V	PRSS1-F	AGGTGATCTTGCCACACTCC	giant panda	GTC	140V
				red panda	GTA	140V
		PRSS1-R	CACTCACTGCCAGAGCTCA	polar bear	ATC	140I
				ferret	ATA	140I
PRSS36	R57S	PRSS36-F	GAACCTGAWGATCTAGGTA	giant panda / red panda	AGC	57S
		PRSS36-R	GGKGGAGTGGTGGGGCGAGT	polar bear	CGC	57R
		PRSS36-2F	CCWTCAGCCCCATCCCAGGAG	ferret	CGC	57R
		PRSS36-2R	CACGTCASGAARCAGTGAGC			
CPB1	V218F	CPB1-1F	AGTCTATGTAAATTTTAGTAG	giant panda / red panda	TTC	218F
		CPB1-1R	TGCCAATGCAGGTAGTTCC	polar bear / ferret	GTC	218V
	T271I,	CPB1-3F	ACTATATTGATCTCATTGT	giant panda	ATT, CTG	271I, 310L

	M310L			/ red panda		
		CPB1-R	TTAGCTAATTTATCTAGCC	ferret	ACT, ATG	271T, 310M
		CPB1-3F	ACTATATTGATCTCATTGT	polar bear	ACT, ATG	271T, 310M
		CPB1-2R	TACCATACTTACTACCATG			
GIF	M212L	GIF1-F	GGCACTGCTACTTACCAGC	giant panda	TTG	212L
		GIF1-R	AATAACCTGATATGTTTCTG	/ red panda polar bear / ferret	ATG	212M
	D406K, H407D	GIF2-F	CCTCATGTTCTTCTCTGCT	giant panda	AAG, GAT	406K, 407D
		GIF2-R	TCATTTATTCTACATAGTGG	red panda	AAA, GAT	406K, 407D
				ferret	GAC, CAT	406D, 407H
		GIF2-2F	CGGGTACTATATTAGCATCAA	polar bear	GAC, CAT	406D, 407H
		GIF2-2R	CAGTCATTGCCTTGAGAAGC			
CYP4F2	M92F	CYP4F2-F	CTTGGGACRTGGAGAGGCC	giant panda	TTT	92F
		CYP4F2-R	AGAGGARGATGGCTGAGAG	/ red panda polar bear / ferret	ATG	92M
CYP3A5	T363S, V393L, M395I, T400S	CYP3A5-3F	YATAAATGCCTGTTAMCA	giant panda	AGT, TTA, ATA, AGT	363S, 393L, 395I, 400S
		CYP3A5-3R	GACAACAATGATCATTTTGT			
		CYP3A5-6F	CTACTGCATATAAGTCAGTG	red panda	AGT, TTG, ATA, TCT	363S, 393L, 395I, 400S
		CYP3A5-6R	TCCACAAGGACATATGTATC			
		CYP3A5-7F	CCGACAGTCTCCCTTCGTCA	polar bear	ACT, GTG, ATG, ACT	363T,393V, 395M, 400T
		CYP3A5-7R	ATAATTGTACAACCACGTGC			
		CYP3A5-4F	AGTGACTTATGTCCCTACTCT	ferret	ACT, GTG, ATG, ACT	363T,393V, 395M, 400T
		CYP3A5-4R	GAATAATTGTACAGCCACAG			
ADH1C	V187I	ADH1C-F	CTGCAGAGGAAAGCCCATCC	giant panda	ATT	187I
		ADH1C-R	AATGCTCYTCAATTCCTTC	/ red panda polar bear / ferret	GTT	187V

Table S25. Population-level genome dataset or database survey for convergent amino acid substitutions of DYNC2H1 and PCNT in giant and red pandas.

Species	Variant 1 of *DYNC2H1*	Variant 2 of *DYNC2H1*	Variant 1 of *PCNT*	Variant 2 of *PCNT*	Dataset or database used	References
giant panda	AAG	AGA	CCC	CGG	Population resequencing data from 49 giant pandas	Ref. 18
polar bear	AGG	AAA	TCC	CAG	Population resequencing data from 37 polar bears	Ref. 19
dog	AGG	AAA	TCC	CAG	Dog Genome SNP Database with 77 samples	http://dogsd.big.ac.cn/snp/pages/download/snp.jsp
mouse	AGG	AAG	TCC	CAA	dbSNP Build 146	ftp://ftp.ncbi.nlm.nih.gov/snp/
human	AGG	AAA	TCA	CAA	1000 Genomes Project Phase 3 variants, dbSNP Build 146, Exome Variant Server	http://www.internationalgenome.org/data, ftp://ftp.ncbi.nlm.nih.gov/snp/, http://evs.gs.washington.edu/EVS/

Table S26. Comparison of Lys and Arg content in different animals and plants including bamboo. DM, dry matter; USDA database, USDA National Nutrient Database for Standard Reference Release 27, *http://ndb.nal.usda.gov/ndb/nutrients/index*.

Species	Measure part	Lys (g/100g DM)	Arg (g/100g DM)	Data source
Deer (*Odocoileus spp.*)	meat	7.590	6.254	USDA database
Rabbit (*Sylvilagus spp.*)	meat	7.485	5.281	USDA database
Squirrel (Sciuridae)	meat	5.869	4.230	USDA database
Muskrat (*Ondatra zibethica*)	meat	5.312	3.250	USDA database
Amaranth (*Amaranthus spp.*)	leaf	1.528	1.456	USDA database
Chicory (*Cichorium intybus*)	leaf	0.838	1.550	USDA database
Almonds (*Prunus dulcis*)	nuts	0.849	2.579	USDA database
Acorns (*Quercus spp.*)	nuts	0.533	0.656	USDA database
Chestnuts (*Castanea mollissima*)	nuts	0.407	0.767	USDA database
Bamboo (*Fargesia robusta*)	leaf	0.65	0.56	Ref. 20
Bamboo (*Fargesia robusta*)	stem	0.14	0.09	Ref. 20
Bamboo (*Fargesia qinlingensis*)	leaf	0.505	0.441	Ref. 21
Bamboo (*Fargesia qinlingensis*)	stem	0.065	0.039	Ref. 21
Bamboo (*Bashania fargesii*)	leaf	0.361	0.276	Ref. 21
Bamboo (*Bashania fargesii*)	stem	0.081	0.041	Ref. 21

Table S27. Comparison of linoleic acid (18:2) content in different animals and plants including bamboo. DM, dry matter; USDA database, USDA National Nutrient Database for Standard Reference Release 27, *http://ndb.nal.usda.gov/ndb/nutrients/index*.

Species	18:2 (g/100g DM)	Data source
Sunflower (*Helianthus annuus*)	24.19	USDA database
Pecan (*Carya illinoinensis*)	21.38	USDA database
Almonds (*Prunus dulcis*)	12.89	USDA database
Bamboo *(Phyllostachys spp.)*	1.22	USDA database
Clover (*Trifolium repens* and *T. pratense*)	0.8	Ref. 22
Grass (*Dactylis glomerata*)	0.4	Ref. 22
Broccoli raab (*Brassica ruvo*)	0.27	USDA database
Pork *(Sus scrofa)*	2.01	USDA database
Bison (*Bison bison*)	0.87	Ref. 23
Elk (*Cervus elaphus*)	0.81	Ref. 23
Beef cattle (*Bos taurus*)	0.44	Ref. 23

Table S28. Comparison of β-carotene content in different plants including bamboo. DM, dry matter; USDA database, USDA National Nutrient Database for Standard Reference Release 27, *http://ndb.nal.usda.gov/ndb/nutrients/index*.

Species	Measure part	β-carotene (μg/100g DM)	Data source
Carrot (*Daucus carota*)		70,751.49	USDA database
Spinach (*Spinacia oleracea*)		65,418.60	USDA database
Chicory (*Cichorium intybus*)	leaf	42,875.00	USDA database
Dandelion (*Taraxacum officinale*)	leaf	40,652.78	USDA database
Broccoli raab (*Brassica ruvo*)		21,114.09	USDA database
Bamboo (*Fargesia spatgacens*)	leaf	6,620	Ref. 24
Bamboo (*Fargesia spatgacens*)	branch	775	Ref. 24
Bamboo (*Phyllostachys spp.*)	shoot	133.33	USDA database
Pecans (*Carya illinoinensis*)	nuts	30.06	USDA database
Almonds (*Prunus dulcis*)	nuts	1.05	USDA database

Table S29. Ten shared pseudogenes between red and giant pandas that result from nonsense or frame-shift mutations of protein-coding sequences.

Gene symbol	Species	Variation	No. of exon with variation	Length of exon with variation (bp)
OR4K1	red panda	indel	1	933
	giant panda	SNP	1	933
KCNRG	red panda	SNP	1	578
	giant panda	indel	2	238
TAS1R1	red panda	indel	6	929
	giant panda	indel	3, 6	762, 929
OR2AG2	red panda	indel	1	948
	giant panda	indel	1	948
SPEF2	red panda	indel	4	171
	giant panda	SNP	37	87
SYNE2	red panda	indel	24	91
	giant panda	indel	46	266
PSAT1	red panda	indel	8	138
	giant panda	SNP	8	138
SMPD4	red panda	indel	2	84
	giant panda	indel	8	152
SRSF10	red panda	indel	1	65
	red panda	SNP	4	163
	giant panda	indel	6	295
ANKDD1B	red panda	indel	12	202
	giant panda	indel	13	132