Supplementary Materials for

INSIG1 parallel substitution drives lipid/sterol metabolic plasticity mediating desert adaptation in ungulates

Xinmei Li[1†], Ziyi He [1†], Anguo Liu[1], Fanxin Meng[1], Xiao Zhang[1], Nana Li [1], Huan Liu[1], Yuyi Lu[1], Zhipei Wu[1], Xixi Yan[1], Nange Ma[1], Zhenyu Wei[1], Wei Wang[1], Xixi He[1], Kunyu Ma[1], Yu Jiang[1], Chao Tong[3*], Bo Xia[1*], Yu Wang[1,2*]

*Corresponding author. E-mail: wang_yu@nwsuaf.edu.cn; imed23@nwafu.edu.cn; tongchao1990@gmail.com

This PDF file includes:

Figs. S1 to S12#
Tables S1 to S3 and S17#

Other Supplementary Materials for this manuscript include the following:

Tables S4 to S15#

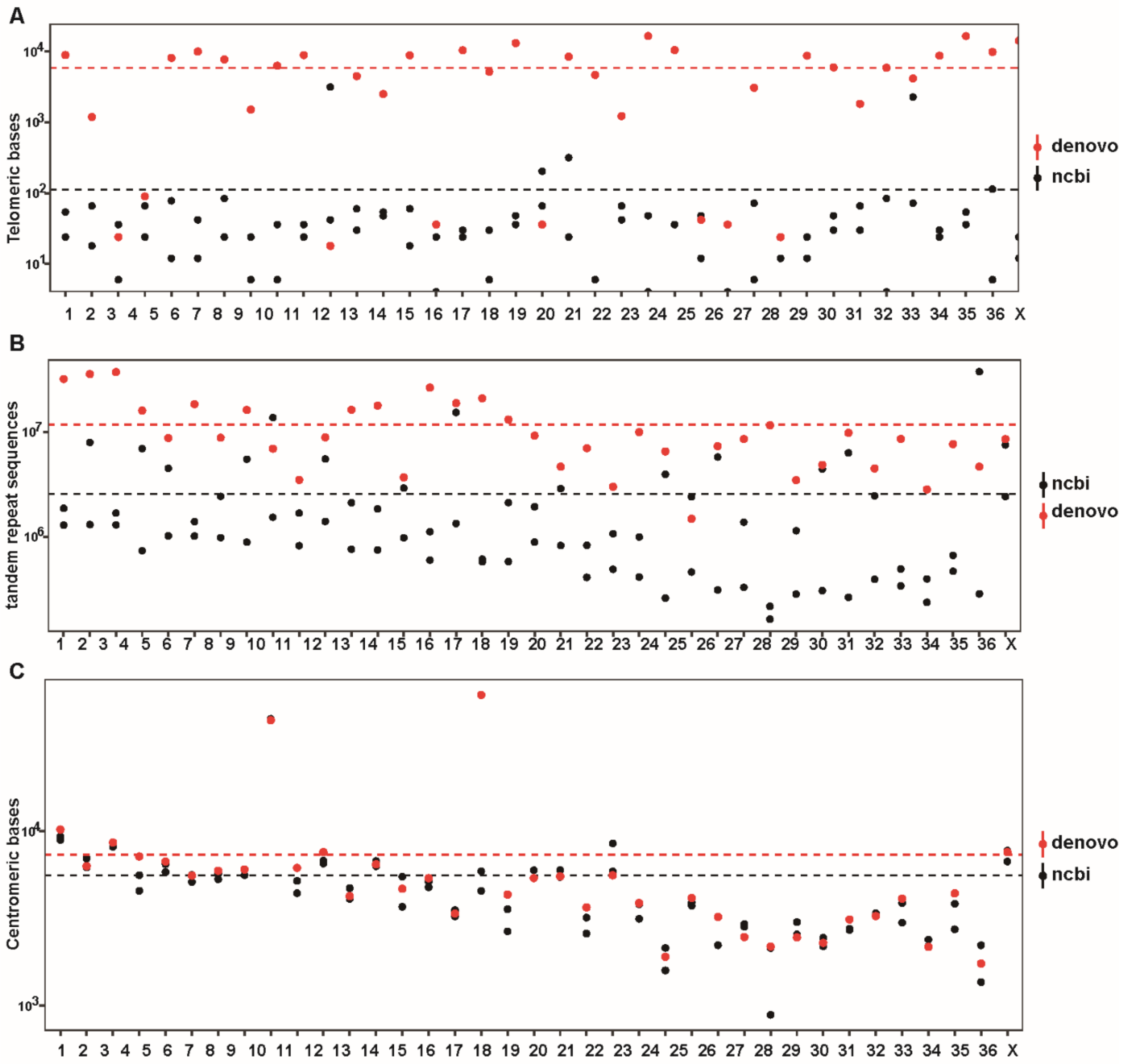


Fig. S1 The completeness of telomere and tandem repeat sequence.

(A) Statistical distribution of centromeric bases on each chromosome of *de novo* assembly (red dot) and ncbi reference genome of dromedary and wild Bactrian camel (black dots). The number of bases in telomeric repeats within 10 Kb of chromosome ends. Dashed lines indicate the autosome-wide mean for the respective color of points.

(B) Tandem repeat sequences on each chromosome were identified by Tandem repeats finder (TRF).

(C) The mean number of bases on each chromosome identified as “Satellite” by RepeatMasker for the *de novo* (red) assembly and ncbi reference genome of dromedary and wild Bactrian camel (black). Dashed lines indicate the autosome-wide mean.

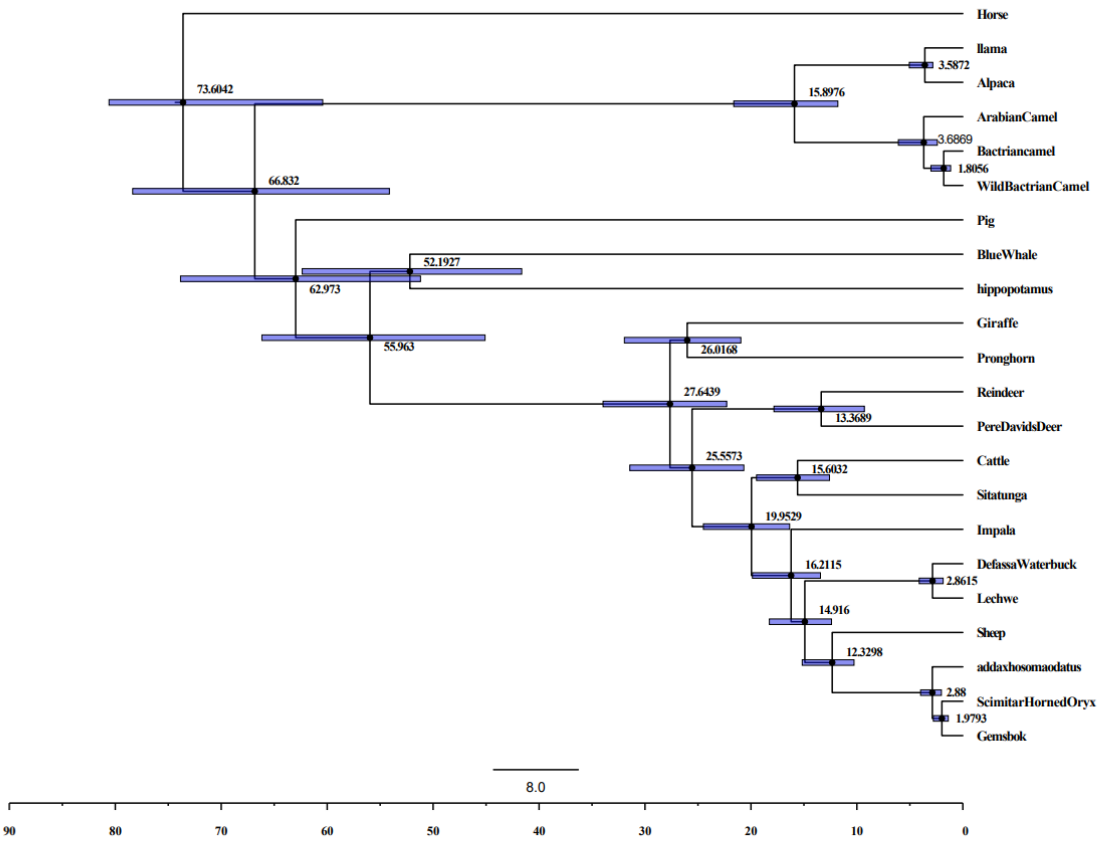
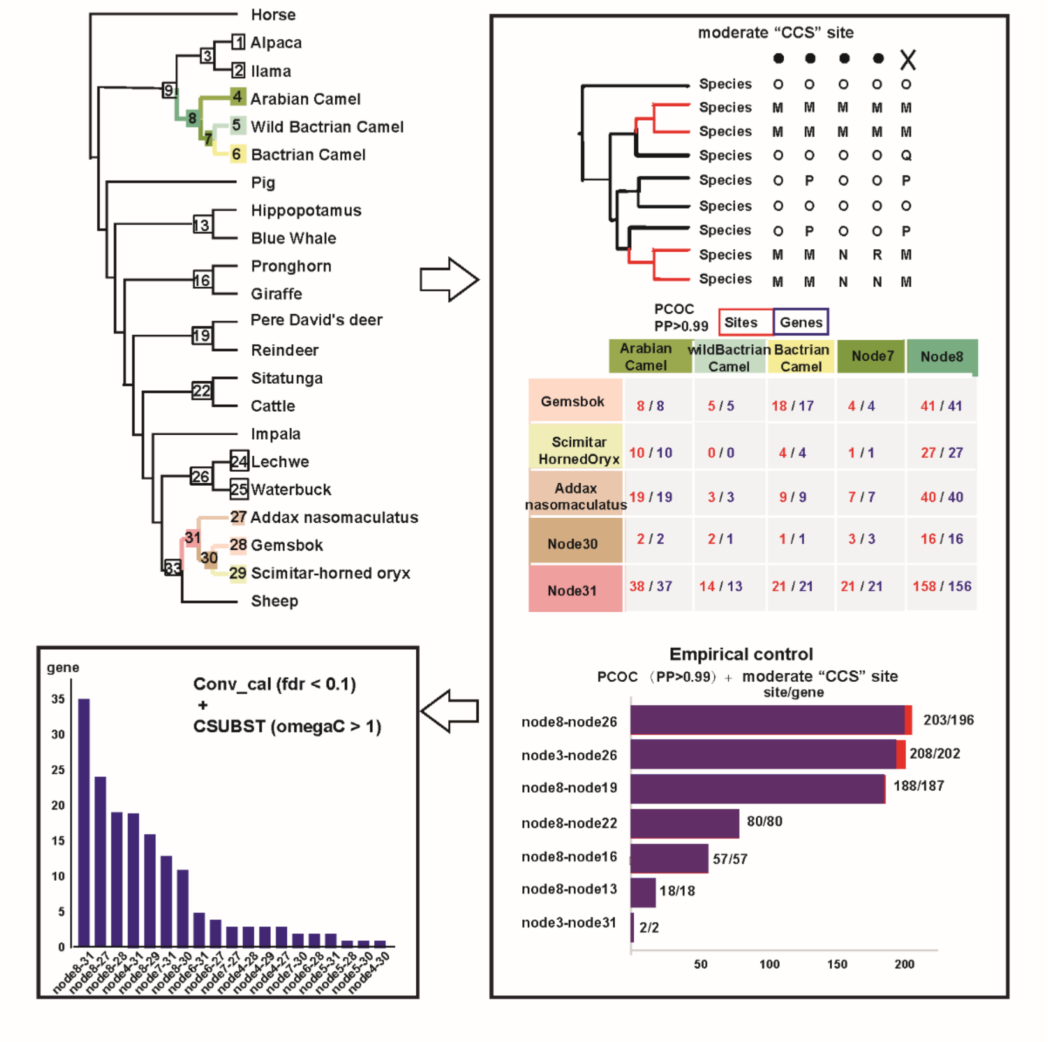


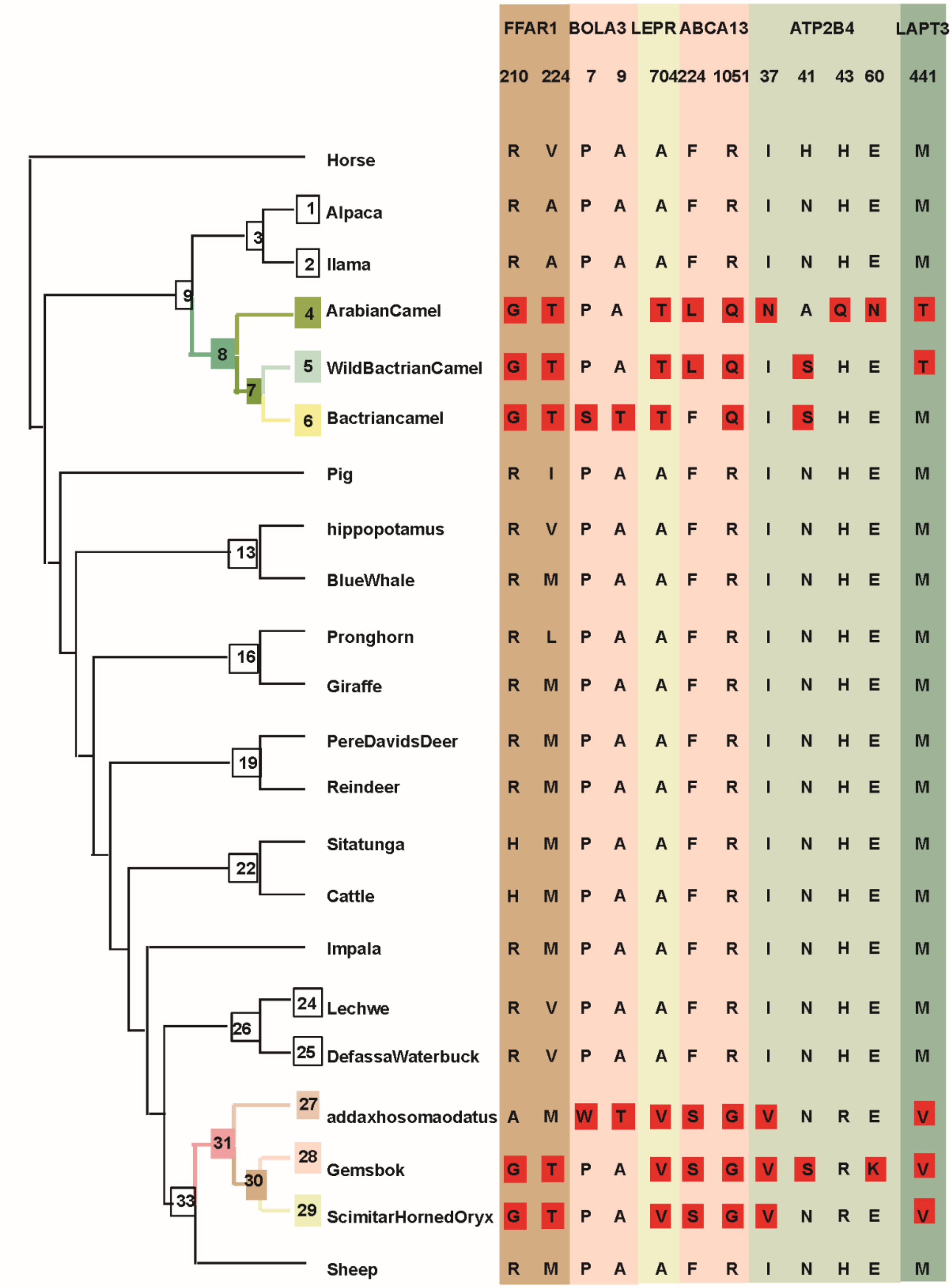
Fig. S2 Estimated divergence times among 22 ungulates.

The blue rectangles at the nodes represent 95% confidence intervals of the corresponding estimated divergence times.



	Arabian Camel	wildBactrian Camel	Bactrian Camel	Node7	Node8
Gemsbok	8 / 8	5 / 5	18 / 17	4 / 4	41 / 41
Scimitar HornedOryx	10 / 10	0 / 0	4 / 4	1 / 1	27 / 27
Addax nasomaculatus	19 / 19	3 / 3	9 / 9	7 / 7	40 / 40
Node30	2 / 2	2 / 1	1 / 1	3 / 3	16 / 16
Node31	38 / 37	14 / 13	21 / 21	21 / 21	158 / 156

Fig. S3 Strategies for screening nonrandom candidate convergent gene sets.



	FFAR1		BOLA3		LEPR	ABCA13		ATP2B4				LAPT3
	210	224	7	9	704	224	1051	37	41	43	60	441
Horse	R	V	P	A	A	F	R	I	H	H	E	M
Alpaca	R	A	P	A	A	F	R	I	N	H	E	M
llama	R	A	P	A	A	F	R	I	N	H	E	M
ArabianCamel	G	T	P	A	T	L	Q	N	A	Q	N	T
WildBactrianCamel	G	T	P	A	T	L	Q	I	S	H	E	T
Bactriancamel	G	T	S	T	T	F	Q	I	S	H	E	M
Pig	R	I	P	A	A	F	R	I	N	H	E	M
hippopotamus	R	V	P	A	A	F	R	I	N	H	E	M
BlueWhale	R	M	P	A	A	F	R	I	N	H	E	M
Pronghorn	R	L	P	A	A	F	R	I	N	H	E	M
Giraffe	R	M	P	A	A	F	R	I	N	H	E	M
PereDavidsDeer	R	M	P	A	A	F	R	I	N	H	E	M
Reindeer	R	M	P	A	A	F	R	I	N	H	E	M
Sitatunga	H	M	P	A	A	F	R	I	N	H	E	M
Cattle	H	M	P	A	A	F	R	I	N	H	E	M
Impala	R	M	P	A	A	F	R	I	N	H	E	M
Lechwe	R	V	P	A	A	F	R	I	N	H	E	M
DefassaWaterbuck	R	V	P	A	A	F	R	I	N	H	E	M
addaxhosomaodatus	A	M	W	T	V	S	G	V	N	R	E	V
Gemsbok	G	T	P	A	V	S	G	V	S	R	K	V
ScimitarHornedOryx	G	T	P	A	V	S	G	V	N	R	E	V
Sheep	R	M	P	A	A	F	R	I	N	H	E	M

Fig. S4 Adaptive convergent genes FFAR1, BOLA3, ABCA13, ATP2B4, LAPT3.

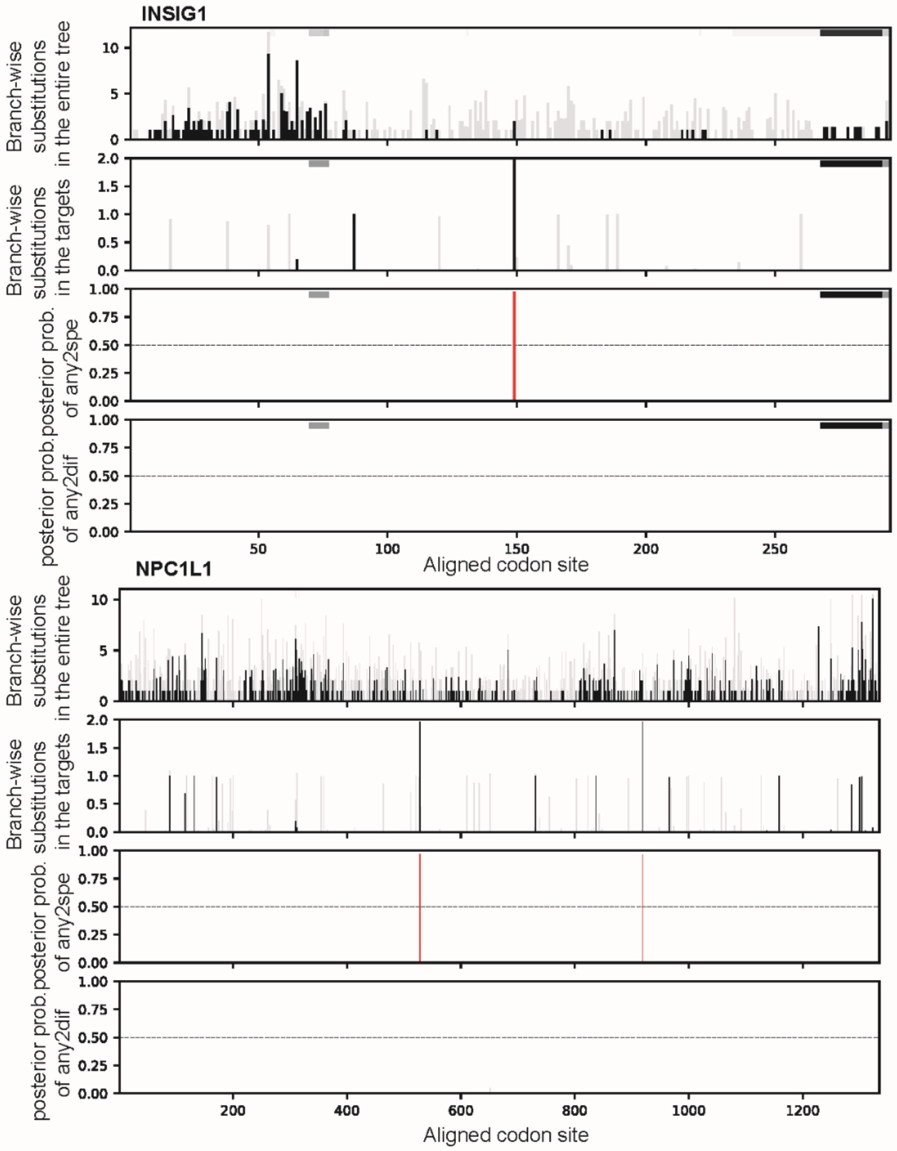


Fig. S5 The probability of substitution for each codon site of INSIG1 and NPC1L1 calculated by CSUBST. Black and gray represents nonsynonymous and synonymous substitutions, respectively. And red/blue is non-synonymous combinatorial substitutions.

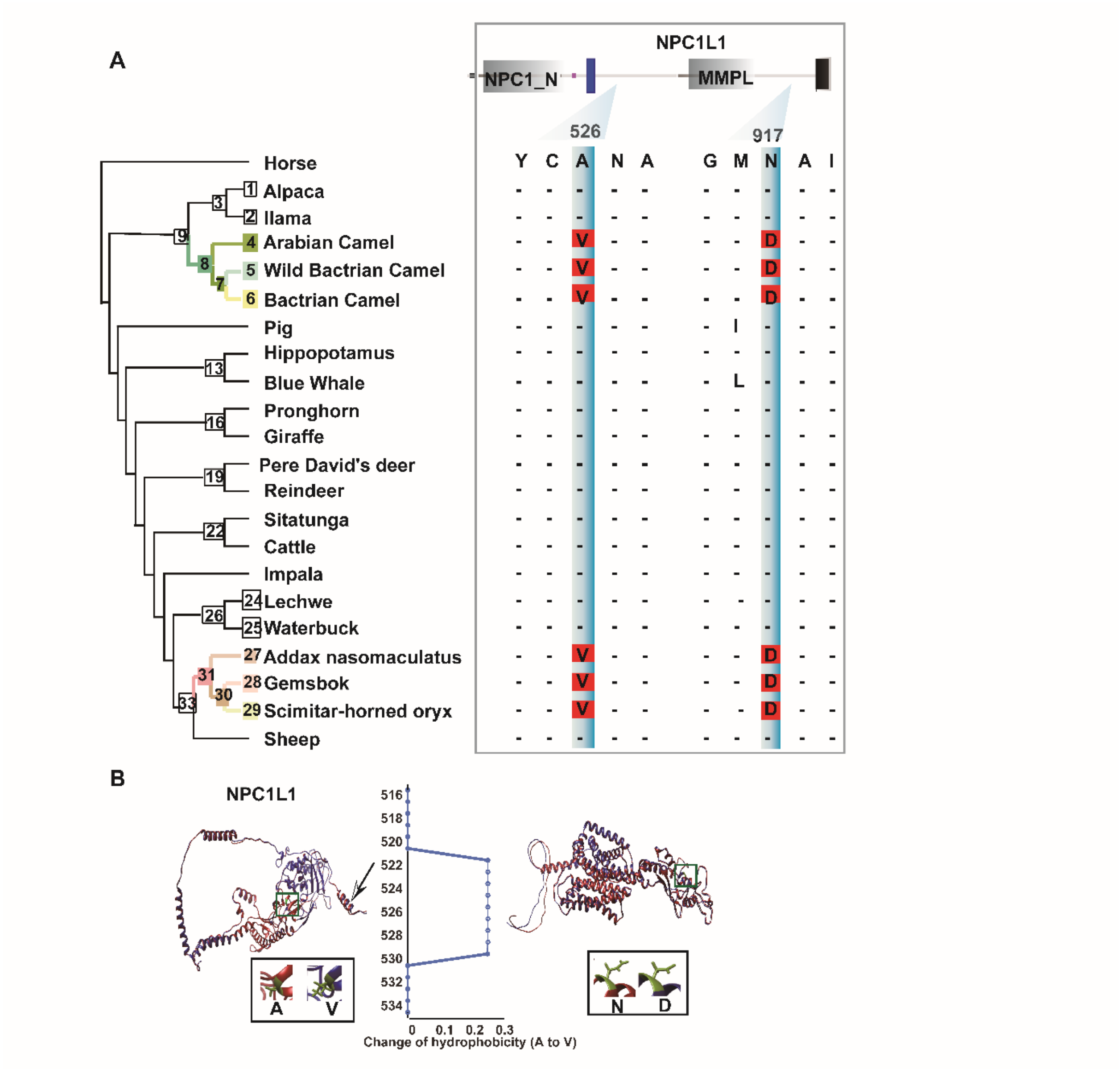


Fig. S6 adaptive convergent mutation in NPC1L1

(A) Parallel substitutions at conserved sites of NPC1L1 protein in *Camelus* and Hippotraginae lineages.

(B) Mutation site in the 3D protein structures of NPC1L1 was indicated in green. The red chains represent the normal protein structure, and the blue chains represent protein structure after parallel mutations. The arrows indicate the locations where structural changes occurred before and after the mutation. The hydrophobicity prediction of NPC1L1. Amino acid substitution of A526V in NPC1L1 increase the amino acids hydrophobicity from 522 to 530.

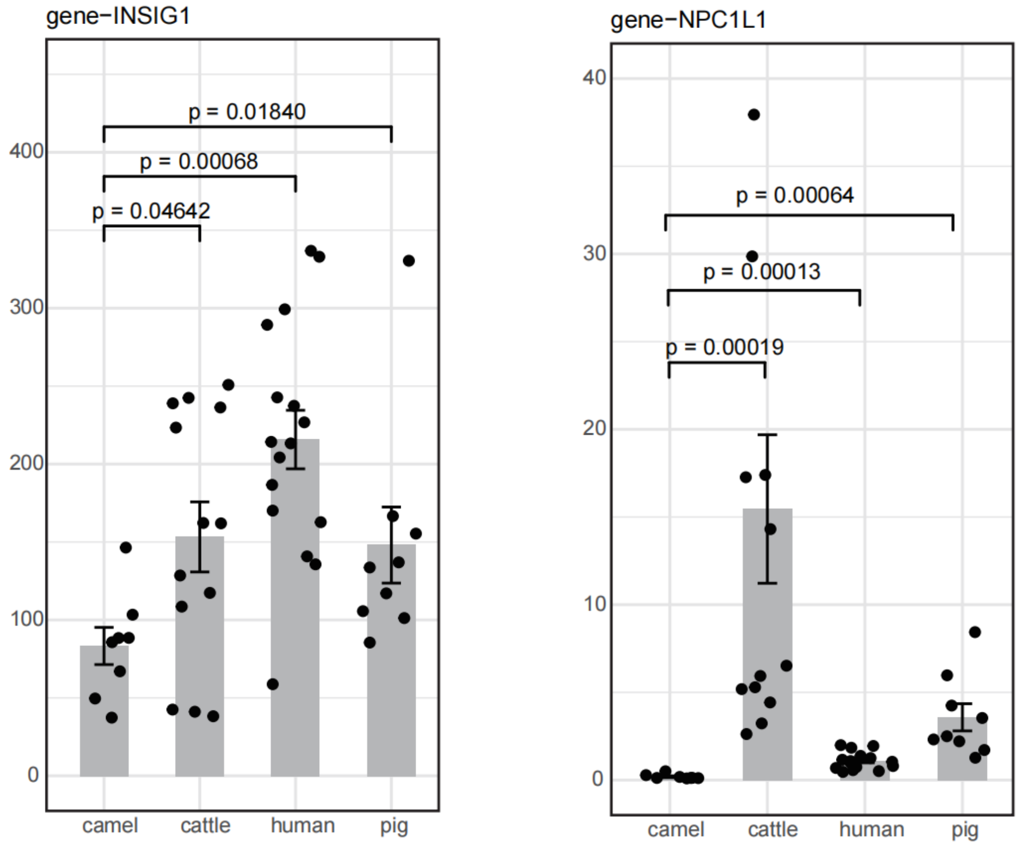
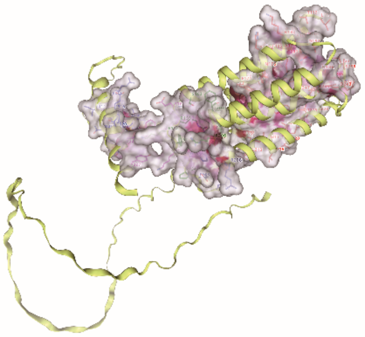


Fig. S7 Expression level of *NPC1L1* (left) and *INSIG1* (right) in camel, cattle, human, and pig.

WT

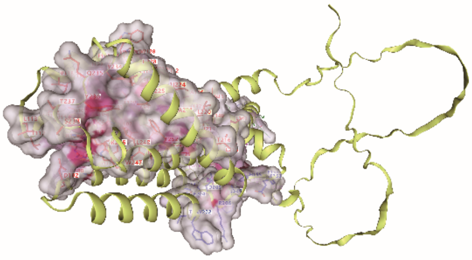
```
Query  .........10.........20.........30.........40.........50.........60.........70.........80.........90
A      MPRLHDHFWS CSCAHSARRR GPPRASAAGL AAKVGEMINV SVSGPSLLAA HGAPDADPAP RGRSAAMSGP EPGSPYPNTW HHRLLQRSLV
       .........100.........110.........120.........130.........140.........150.........160.........170....
       LFSVGVVLAL  VLNLLQIQRN  VTLFPEEVIA  TIFSSAWWVP  PCCGTAAAVV  GLLYPCIDSH  LGEPHKFKRE  WASVMRCIAV  FVGI
       .....180.........190.........200.........210.........220.........230.........240.........250.......
       NHASAK  LDFANNVQLS  LTLAALSLGL  WWTFDRSRSG  LGLGITIAFL  ATLITQFLVY  NGVYQYTSPD  FLYIRSWLPC  IFFSGGVT
       .260.........270.........280.........290.........300.........310.........320.........330
       VG  NIGRQLAMLI  PFCEELNLKT  TWLFHKTRSN  YRVFLKSPIV  IESSKPPILR  ARKILEENLT  VDYDKDYLFS
```



CurPocket ID	Cavity volume	Center (x, y, z)	Cavity size (x, y, z)
C1	4347	-6, 6, 6	30, 18, 27
C2	549	0, -29, -19	14, 11, 15
C3	319	8, -18, -10	12, 7, 10
C4	208	11, -13, 0	8, 9, 11
C5	134	17, -5, -3	5, 9, 5

H-R

```
Query  .........10.........20.........30.........40.........50.........60.........70.........80.........90
A      MPRLHDHFWS CSCAHSARRR GPPRASAAGL AAKVGEMINV SVSGPSLLAA HGAPDADPAP RGRSAAMSGP EPGSPYPNTW HHRLLQRSLV
       .........100.........110.........120.........130.........140.........150.........160.........170....
       LFSVGVVLAL  VLNLLQIQRN  VTLFPEEVIA  TIFSSAWWVP  PCCGTAAAVV  GLLYPCIDSR  LGEPHKFKRE  WASVMRCIAV  FVGI
       .....180.........190.........200.........210.........220.........230.........240.........250.......
       NHASAK  LDFANNVQLS  LTLAALSLGL  WWTFDRSRSG  LGLGITIAFL  ATLITQFLVY  NGVYQYTSPD  FLYIRSWLPC  IFFSGGVT
       .260.........270.........280.........290.........300.........310.........320.........330
       VG  NIGRQLAMLI  PFCEELNLKT  TWLFHKTRSN  YRVFLKSPIV  IESSKPPILR  ARKILEENLT  VDYDKDYLFS
```



CurPocket ID	Cavity volume	Center (x, y, z)	Cavity size (x, y, z)
C1	3832	-6, 5, 6	30, 18, 27
C2	403	8, -15, -2	13, 9, 10
C3	261	5, -17, -12	12, 7, 7
C4	174	-5, -29, -29	9, 8, 11
C5	146	-15, -36, -41	7, 7, 6

Fig. S8 The curpocket and cavity for INSIG1(H132R) and wildtype INSIG1 identified by CB-Dock2.

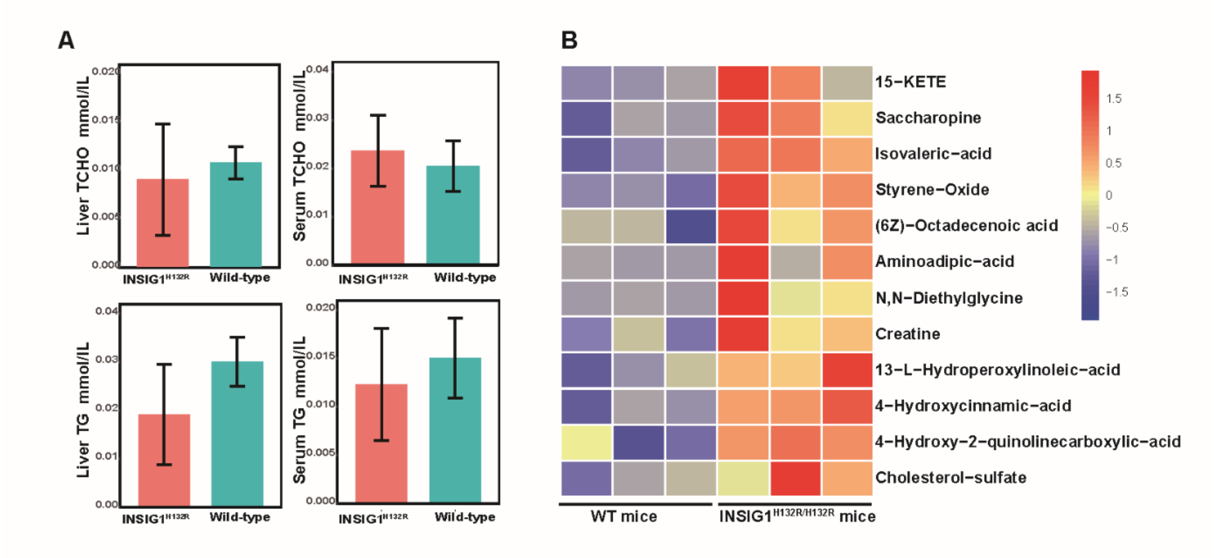


Fig. S9 The phenotypes of wild-type and INSIG1$^{132R/132R}$ gene-edited mice

(A) The cholesterol (up) and triglyceride (down) concentrations in serum and liver of mutant (n=3) and wild-type mice (n=3).

(B) Differential metabolites in the liver between mutant and wild mice.

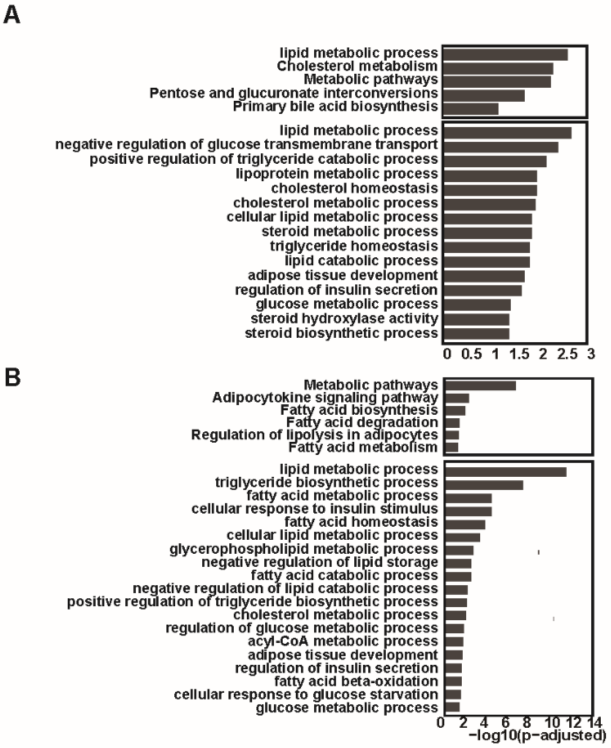
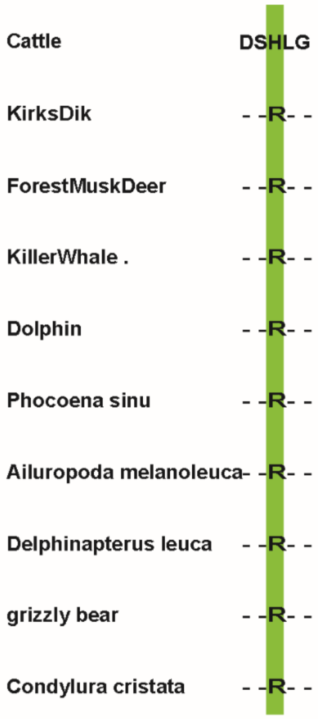


Fig.S10 The difference of transcriptome and metabolome between wild-type and gene-edited mice

(A-B) Pathways were enriched by differentially expressed genes in liver (A) and visceral fat (B) of homozygous and wild-type mice.



Cattle	DSHLG
KirksDik	- -R- -
ForestMuskDeer	- -R- -
KillerWhale .	- -R- -
Dolphin	- -R- -
Phocoena sinu	- -R- -
Ailuropoda melanoleuca	- -R- -
Delphinapterus leuca	- -R- -
grizzly bear	- -R- -
Condylura cristata	- -R- -

Fig. S11 the H-R mutation of INSIG1 in other more animals.

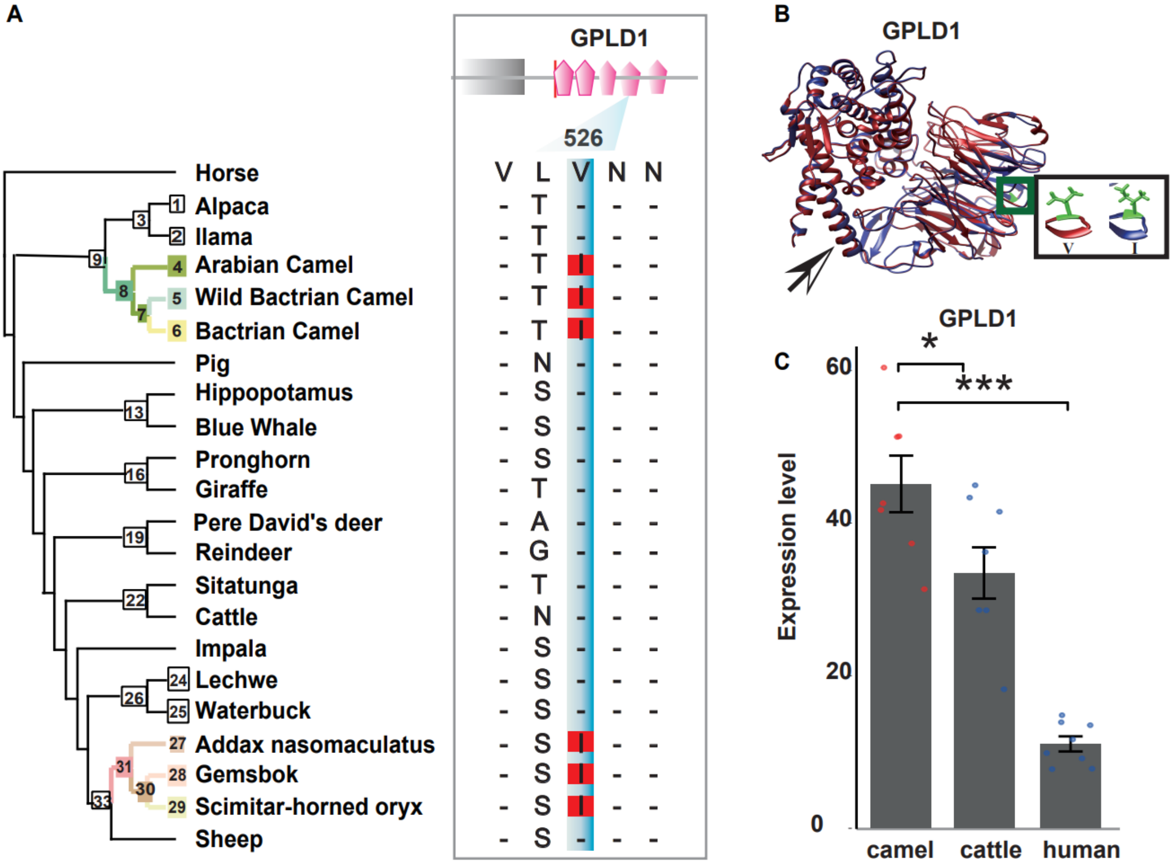


Fig. S12 (A) Parallel substitutions at conserved site of GPLD1 protein in Camelus and Hippotraginae lineages. (B) Mutation sites in the 3D protein structures of three protein are indicated in green. The red chains represent the normal protein structure, and the blue chains represent those after parallel mutations. The locations where structural changes occurred before and after the mutation are indicated by arrows. (C) Expression level of liver in camel, cattle and human. t-test, * : P < 0.05; ** : P < 0.01; *** : P < 0.001.

Table S1. De novo genome assessment.

Scaffold	Wild Bactrian camel	Arabian camel	Bactrian camel	de novo
N0	122,453,268	124,715,342	46,538,883	164,120,332
N10	121,270,977	123,797,624	19,607,279	149,680,253
N20	102,155,050	120,724,560	16,165,641	118,545,724
N30	94,313,465	97,468,828	13,872,274	105,728,113
N40	78,620,977	81,230,863	11,315,732	99,165,517
N50	76,025,729	70,369,702	8,812,066	79,915,242
N60	68,617,425	68,281,056	6,903,358	71,079,562
N70	49,638,757	52,130,679	4,942,766	60,221,085
N80	33,771,694	31,123,794	3,491,600	38,068,993
N90	27,488,461	24,767,672	1,821,536	30,245,471
N100	224	245	200	14,149

Table S2. State of 18,430 ancestral placental mammal in Camelidae.

genome	**intact genes**	**genes with inactivating mutations**	**genes with missing sequence**
De novo genome	17632	689	109
Camelus bactrianus	13498	2178	2754
Camelus dromedarius	16472	1766	192
Camelus ferus	16296	1888	246
Lama glama llama	15611	1000	1819
Lama glama chaku_ llama	16582	899	949
Lama guanicoe cacsilensis guanaco	17226	901	303
Vicugna pacos alpaca	13133	1415	3882
Vicugna pacos alpaca (vicPac2)	13181	1618	3631
Vicugna pacos huacaya alpaca	15756	964	1710
Vicugna vicugna mensalis vicugna	17246	838	346

Table S3. Genome statistics for 22 ungulates used in this manuscript.

Common name	**Species name**	**Genbank accession**
horse	*Equus caballus*	GCA_002863925.1
alpaca	*Vicugna pacos*	GCA_000164845.5
llama	*Lama glama*	GCA_013239585.1
Arabian Camel	*Camelus dromedarius*	GCA_000803125.3
wild Bactrian camel	*Camelus ferus*	GCA_009834535.1
Bactrian Camel	*Camelus bactrianus*	GCA_000767855.
pig	*Sus scrofa*	GCA_000003025.6
Hippopotamus	*Hippopotamus amphibius*	GCA_030028045.1
Blue Whale	*Balaenoptera musculus*	GCA_009873245.3
Pronghorn	*Antilocapra americana*	GCA_007570785.1
Giraffe	*Giraffa camelopardalis*	GCA_017591445.1
Pere David's deer	*Elaphurus davidianus*	GCA_021018665.1
Reindeer	*Rangifer tarandus*	GCA_949782905.1
Sitatunga	*Tragelaphus spekii*	GCA_006411015.1
Cattle	*Bos taurus*	GCA_002263795.3
Impala	*Aepyceros melampus*	GCA_006408695.1
Lechwe	*Kobus leche*	GCA_014926565.1
Waterbuck	*Kobus ellipsiprymnus*	GCA_006410655.1
Addax nasomaculatus	*Addax nasomaculatus*	GCA_019593525.1
Gemsbok	*Oryx gazella*	GCA_003945745.1
Scimitar-horned oryx	*Oryx dammah*	GCA_014754425.2
Sheep	*Ovis aries*	GCA_016772045.1

Table S17. The sequence of primers used for qPCR.

Gene	Forward
ABCA1	F: ACCCACCCTATGAACAACATGA
	R: GAGTCGGGTAACGGAAACAGG
ABCG5	F: ACTGCTTCTCCTACGTCCTG
	R: CTGTAGTTGCCAATCAGTCGG
ABCB1	F: TTGCTGCTTACATTCAGGTTTCA
	R: AGCCTATCTCCTGTCGCATTA

F, forward; R, reverse