Supplementary Information

Whole genome sequencing of giant pandas provides insights into demographic history and local adaptation

Shancen Zhao[1,2,10], Pingping Zheng[1,3,10], Shanshan Dong[2,10], Xiangjiang Zhan[1,10], Qi Wu[1,10], Xiaosen Guo[2], Yibo Hu[1], Weiming He[2], Shanning Zhang[4], Wei Fan[2], Lifeng Zhu[1], Dong Li[2], Xuemei Zhang[2], Quan Chen[2], Hemin Zhang[5], Zhihe Zhang[6], Xuelin Jin[7], Jinguo Zhang[8], Huanming Yang[2], Jian Wang[2], Jun Wang[2,9], Fuwen Wei[1]

[1]Key Lab of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.

[2]Shenzhen Key Laboratory of Transomics Biotechnologies, BGI-Shenzhen, Shenzhen, China.

[3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China.

[4]China Wildlife Conservation Society, Beijing, China.

[5]China Conservation and Research Center for the Giant Panda, Wolong, Sichuan, China.

[6]Chengdu Research Base of Giant Panda Breeding, Chengdu, Sichuan, China.

[7]Shaanxi Wild Animal Rescue and Research Center, Louguantai, Xi'an, Shaanxi, China.

[8]Beijing Zoo, Beijing, China.

[9]Department of Biology, University of Copenhagen, Copenhagen, Denmark.

[10]These authors contributed equally to this work.

Correspondence should be addressed to F. W. (weifw@ioz.ac.cn) or Ju. W. (wangj@genomics.org.cn).

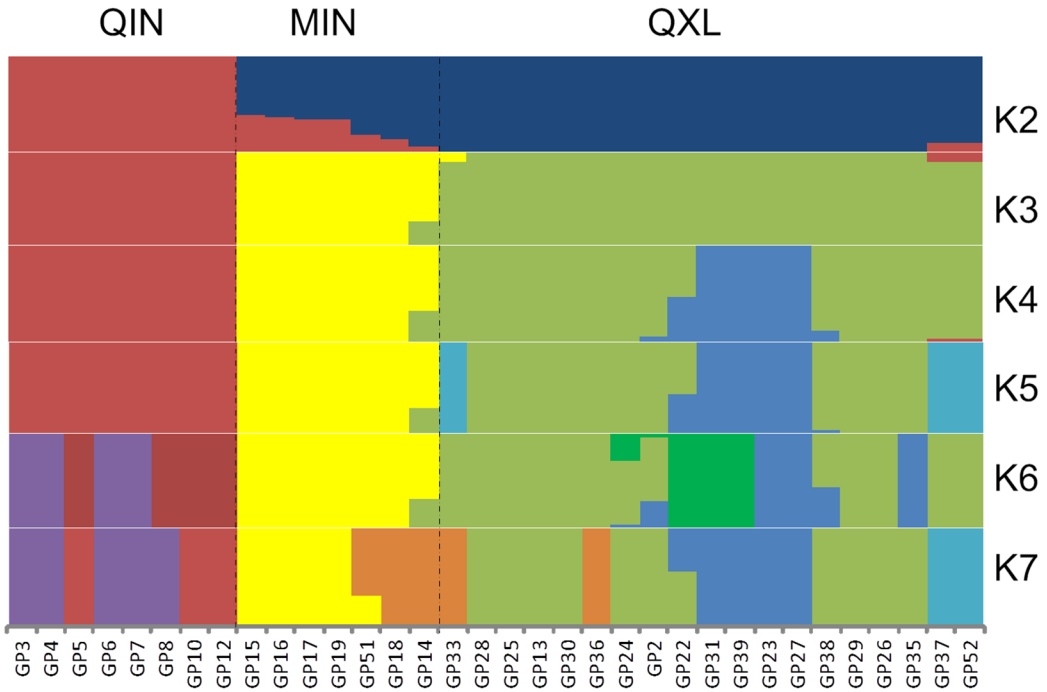
Nature Genetics doi:10/1038/ng.2494

Table of contents

Nature Genetics doi:10/1038/ng.2494

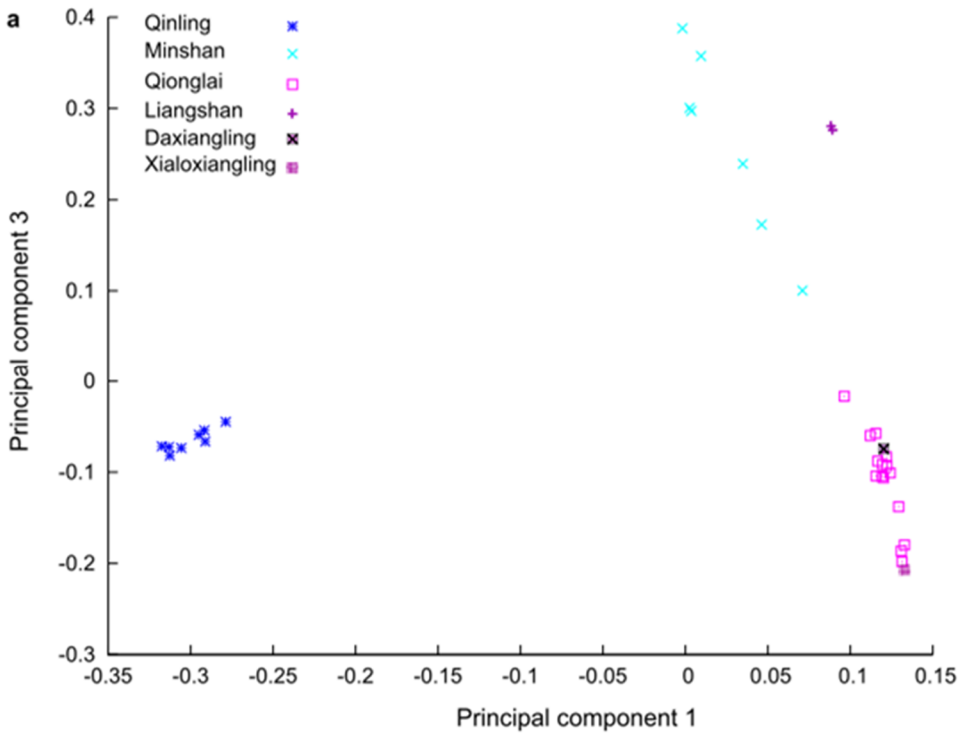
Supplementary Figures

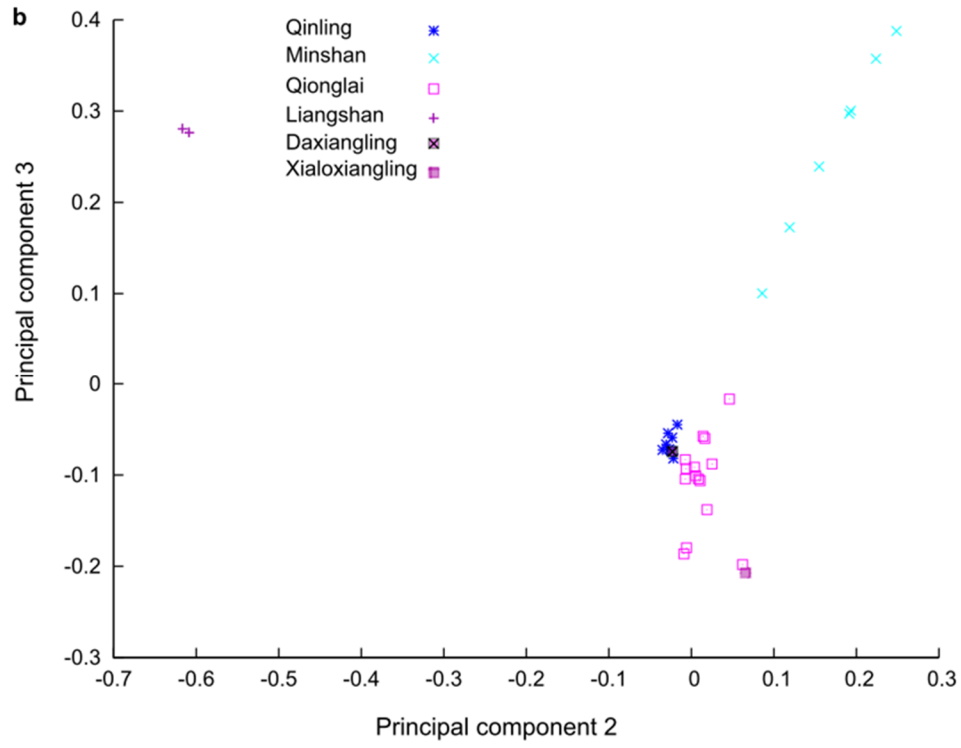
Supplementary Figure 1. The genetic structure of studied panda populations inferred by the ADMIXTURE analysis. The number of genetic clusters *K* was defined from 2 to 7. When *K* = 4, two subpopulations within QXL were inferred: one comprised Xiaoxiangling and some individuals from Qionglai, and the other included Daxiangling, Liangshan and the remaining Qionglai.



Nature Genetics doi:10/1038/ng.2494

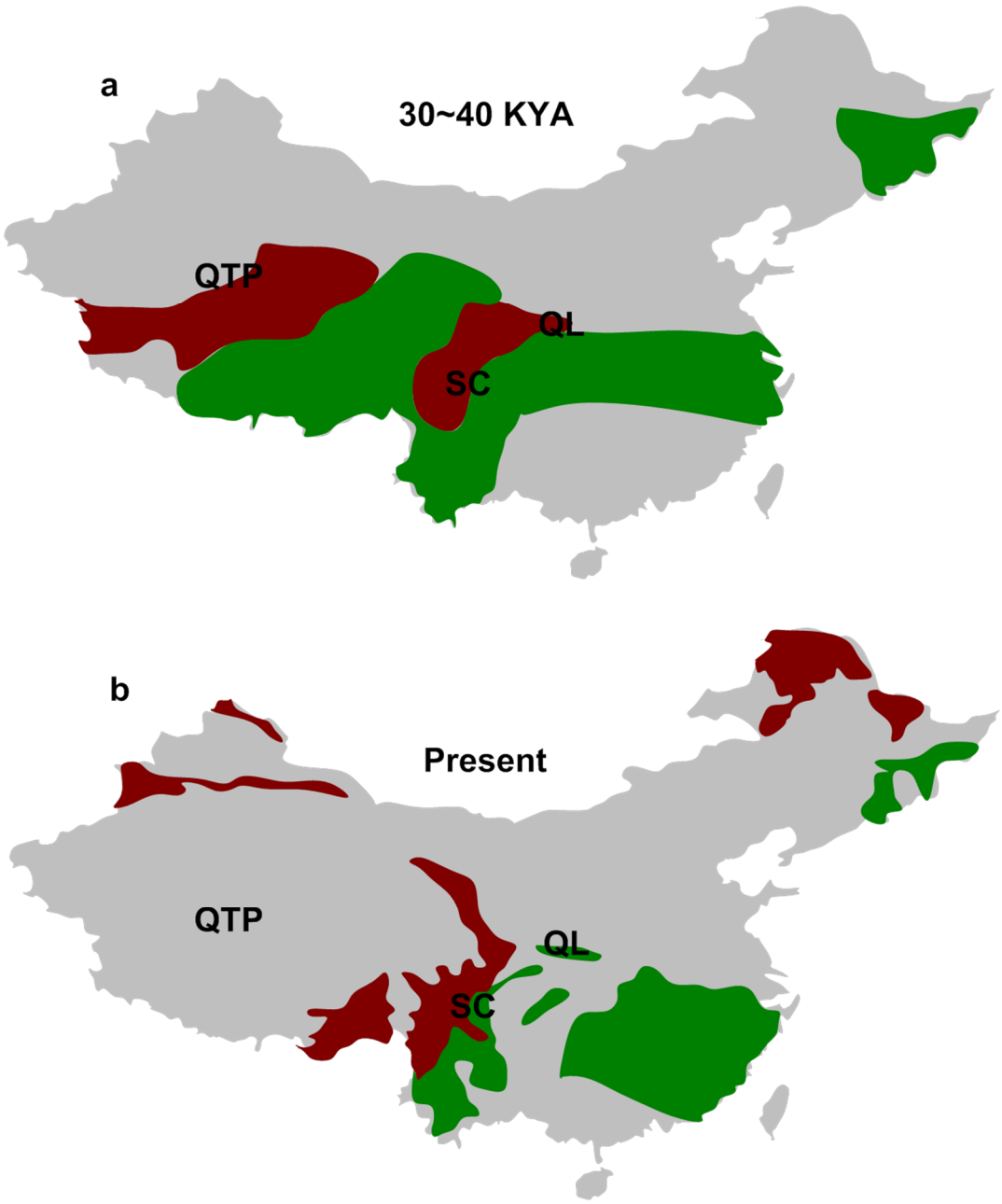
Supplementary Figure 2. Principal Component Analysis on the 34 wild pandas. (**a**) Principal components 1 and 3. (**b**) Principal components 2 and 3.





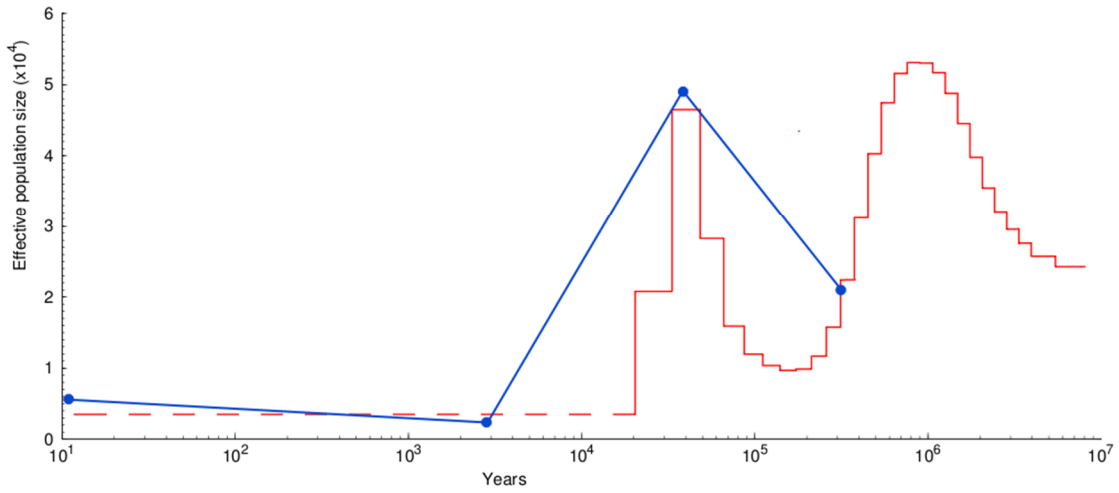
Nature Genetics doi:10/1038/ng.2494

Supplementary Figure 3. Conifer (red) and broadleaf-conifer (green) forest at (**a**) 30~40 KYA and (**b**) present, modified from the report[1]. QL: Qinling Mountains, SC: Sichuan, QTP: Qinghai-Tibetan Plateau. QIN pandas occupy the Qingling Mountains whereas non-QIN pandas occupy the five mountain ranges in Sichuan (i. e. Minshan, Qionglai, Daxiangling, Xiaoxiangling and Liangshan).



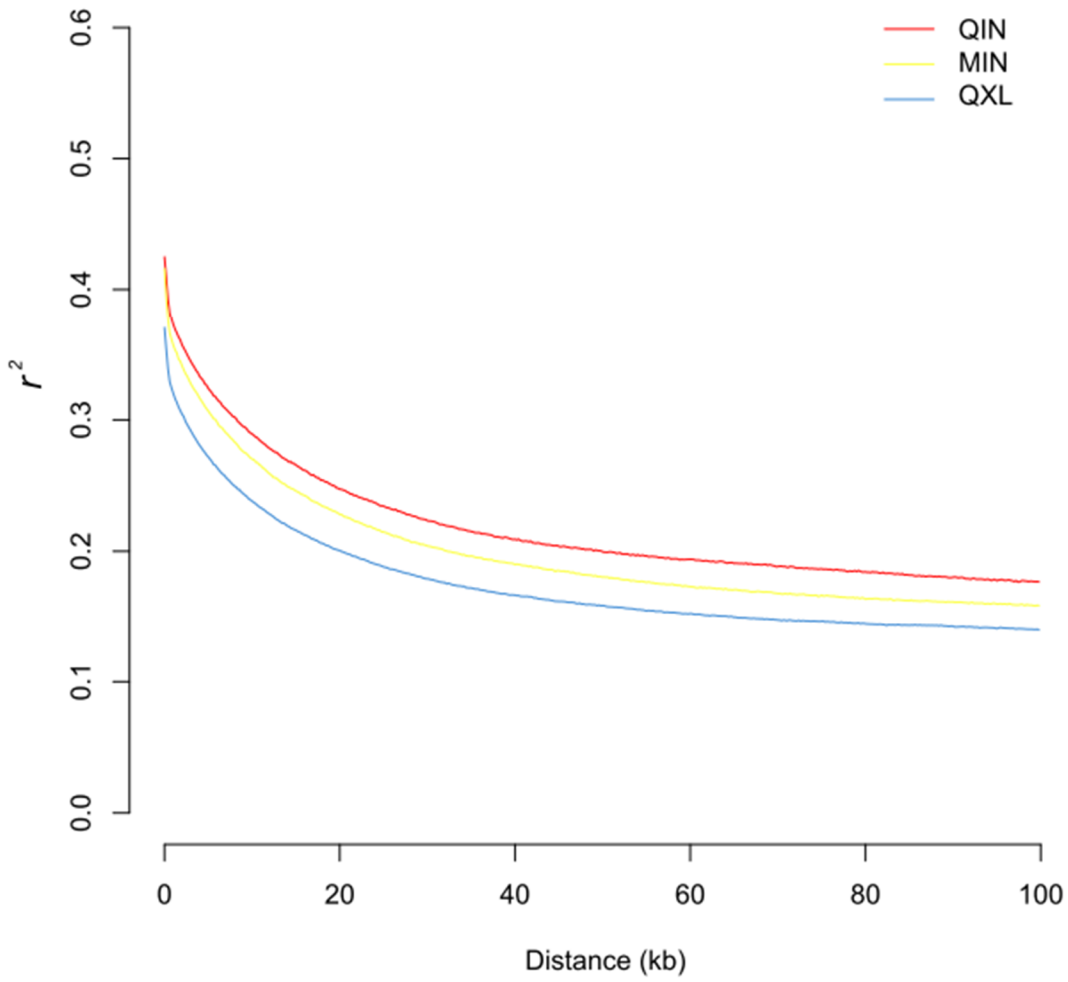
Nature Genetics doi:10/1038/ng.2494

Supplementary Figure 4. Comparision of effective population size changes between PSMC and ∂a∂i approaches. The red and blue lines show N_e of PSMC and ∂a∂i results, respectively. Note that PSMC simulation cannot detect population changes more recent than 20 KYA (red dashed line). The population peak at 30~50 KYA and the subsequent bottleneck were uncovered by both methods.



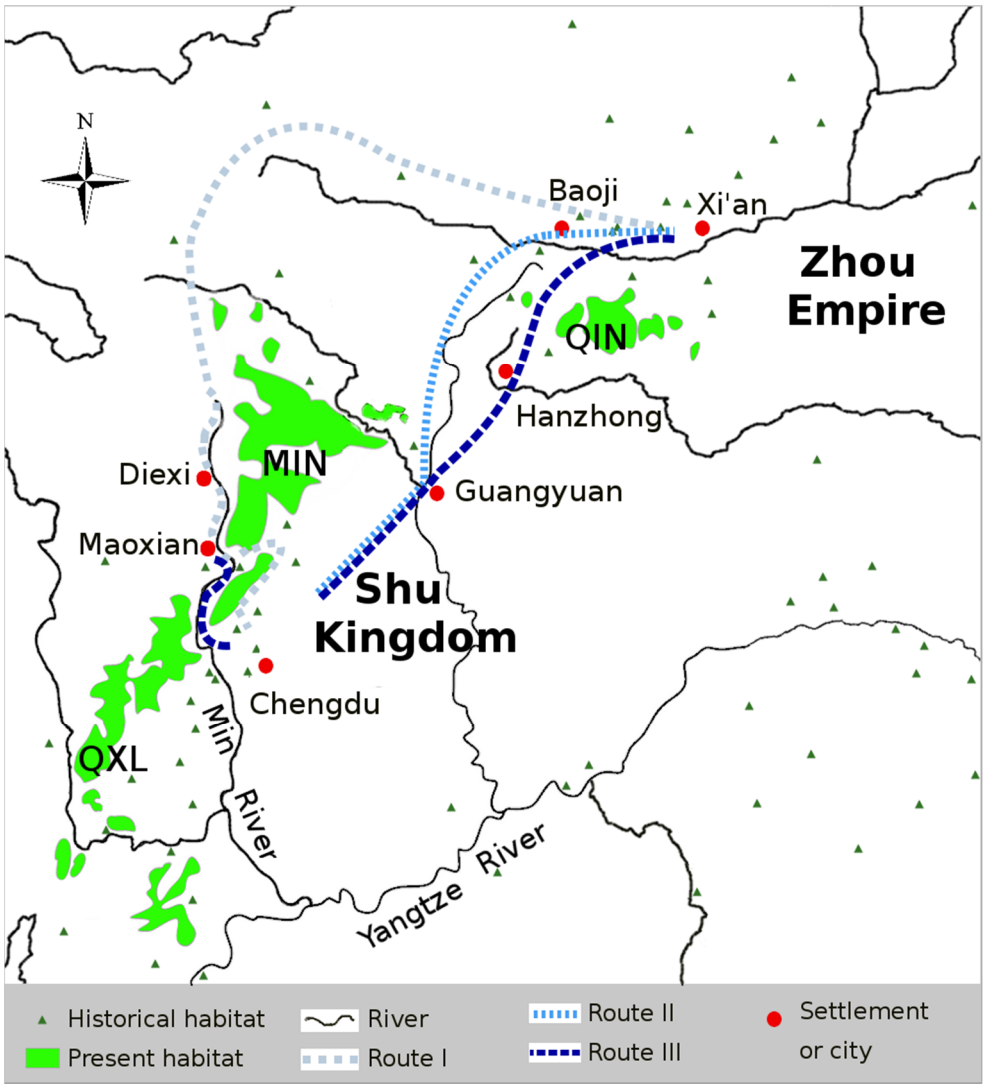
Nature Genetics doi:10/1038/ng.2494

Supplementary Figure 5. Linkage disequilibrium (LD) patterns for the three panda populations. To minimize the influence of sample size, we used similar sample size for each population for the LD analysis with all individuals in QIN (n=8), all individuals in MIN (n=7), and eight randomly selected individuals from QXL (GP2, GP13, GP22, GP25, GP26, GP27, GP39, GP52). QIN population exhibits higher LD than others. X axis: physical distances between two SNPs marked in kb; Y axis: r^2 used to measure linkage disequilibrium.



Nature Genetics doi:10/1038/ng.2494

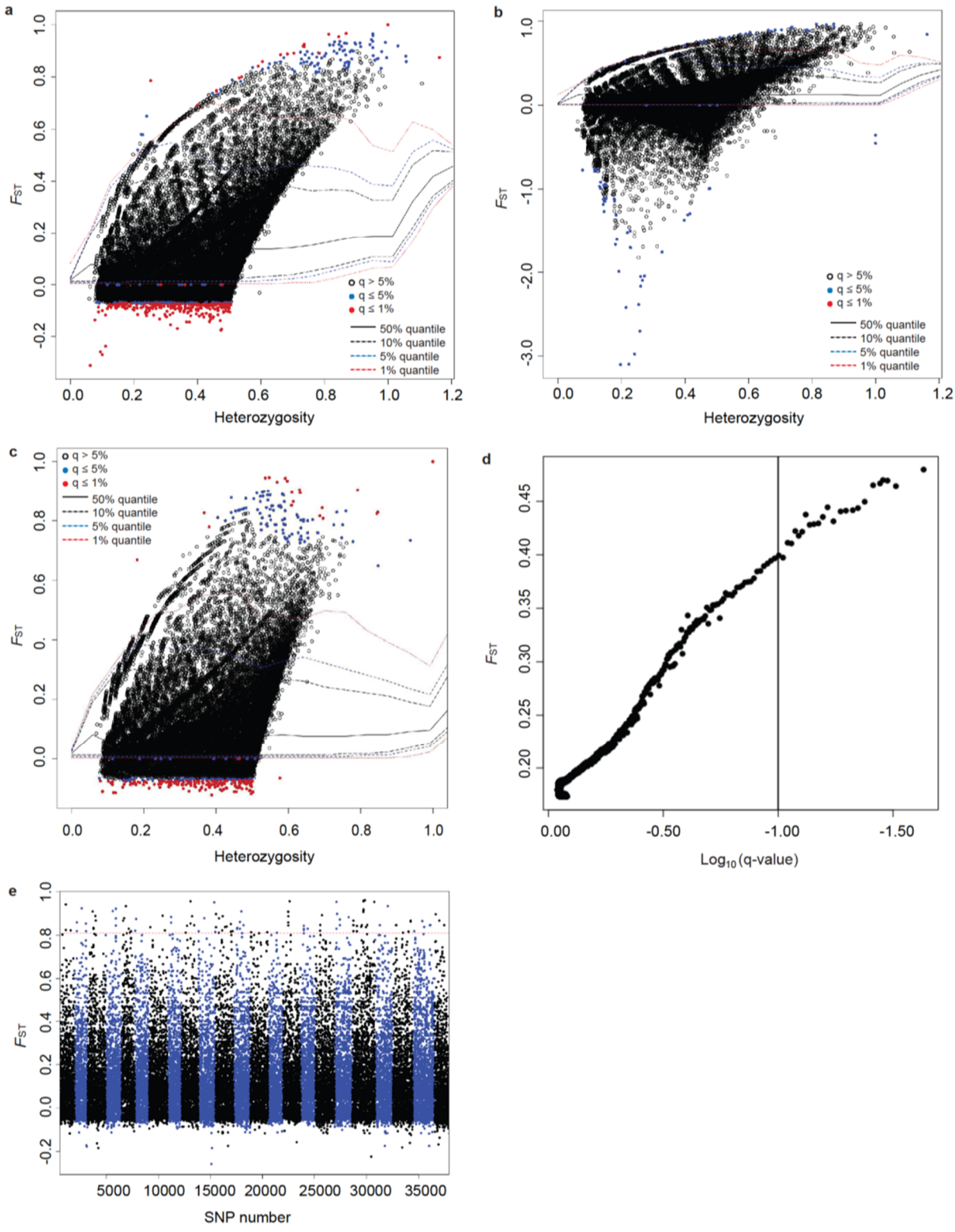
Supplementary Figure 6. Influences of human activities on the habitats of giant panda in the last three thousand years. The present panda habitats and recorded historical sites within the last 3000 years are shown in light and dark green, respectively. Three main routes are shown as the blue dashed lines (detailed description in **Supplementary Note)**. The red dots are key settlements or cities along the routes connecting ancient Shu Kingdom with the outside. Note: 1) many dynasties existed in the northern area in history, but only Zhou Empire is shown; 2) many settlements and cities were renamed for several times in history and only current names are shown on the map. For example, it was recorded that the place of Diexi, was once named Canling. Before the name of Canling, it had been a key settlement of ancient Shu for thousands of years. However, this place was devastated by a gigantic earth quake in 1933.



Nature Genetics doi:10/1038/ng.2494

Supplementary Figure 7. Detection of loci under selection of QIN and non-QIN populations. The simulated joint distribution of Heterozygosity-F_{ST} of SNPs from QIN and non-QIN using Arlequin: (**a**) Detection of loci under selection based on F_{ST} under hierarchical model; (**b**) Detection of loci under selection based on F_{CT} under hierarchical model; (**c**) Detection of loci under selection based on F_{ST} under finite island model. The upper red and blue dots are outliers subjected to directional selection (False Discovery Rate $q < 0.05$), whereas the lower red and blue dots represent candidate loci under balancing selection ($q < 0.05$). (**d**) Detection of loci under selection using the Bayescan. Log_{10}(q-value): logarithm of q-values to base 10 for the model including selection. Outliers with $q < 0.1$ were shown at the right side of the black vertical line. (**e**) Detection of loci under selection using the global F_{ST} test. The loci with top 1% of F_{ST} were ranked above the red dashed line ($F_{ST} = 0.808464$).

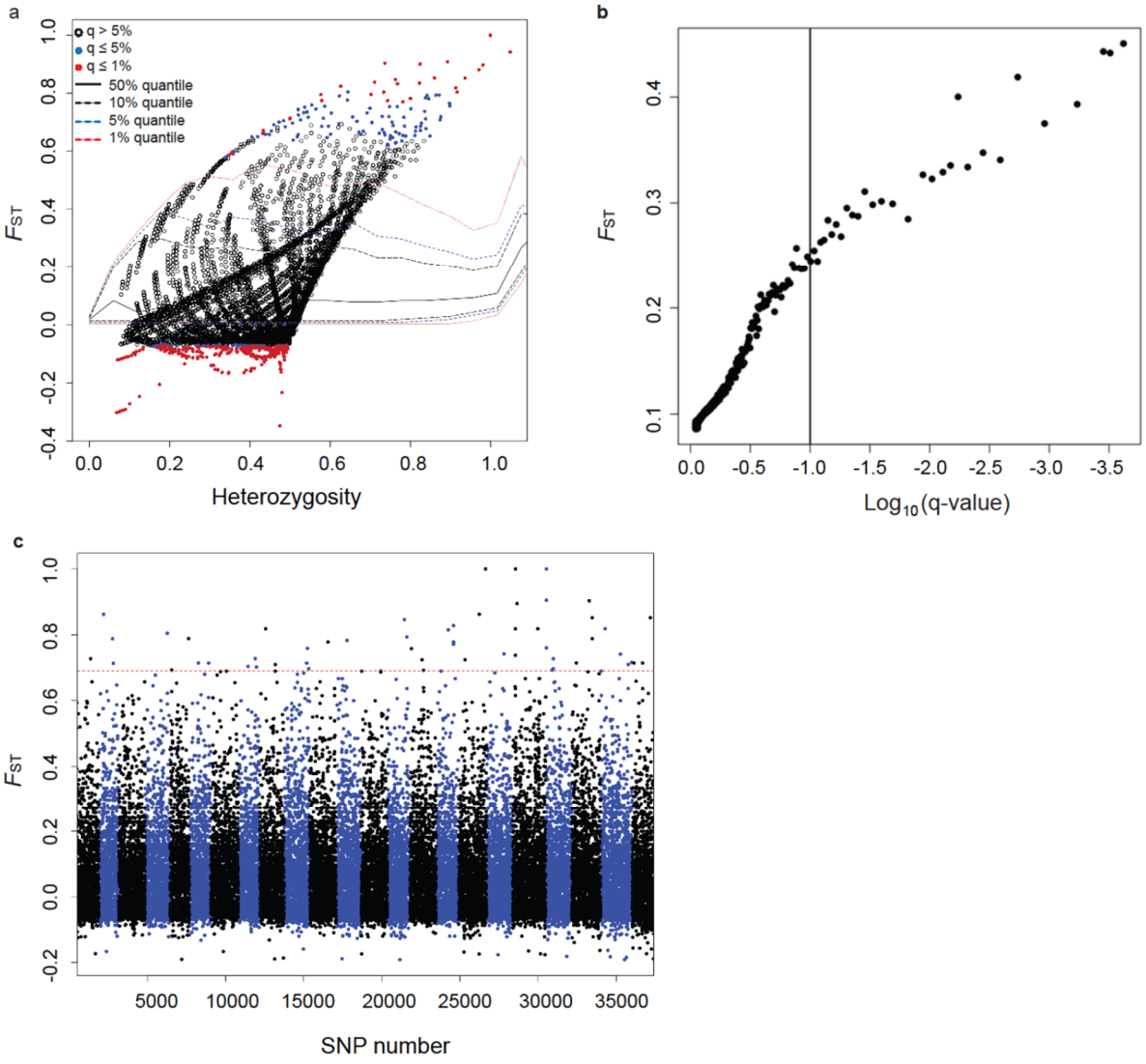
Nature Genetics doi:10/1038/ng.2494



119

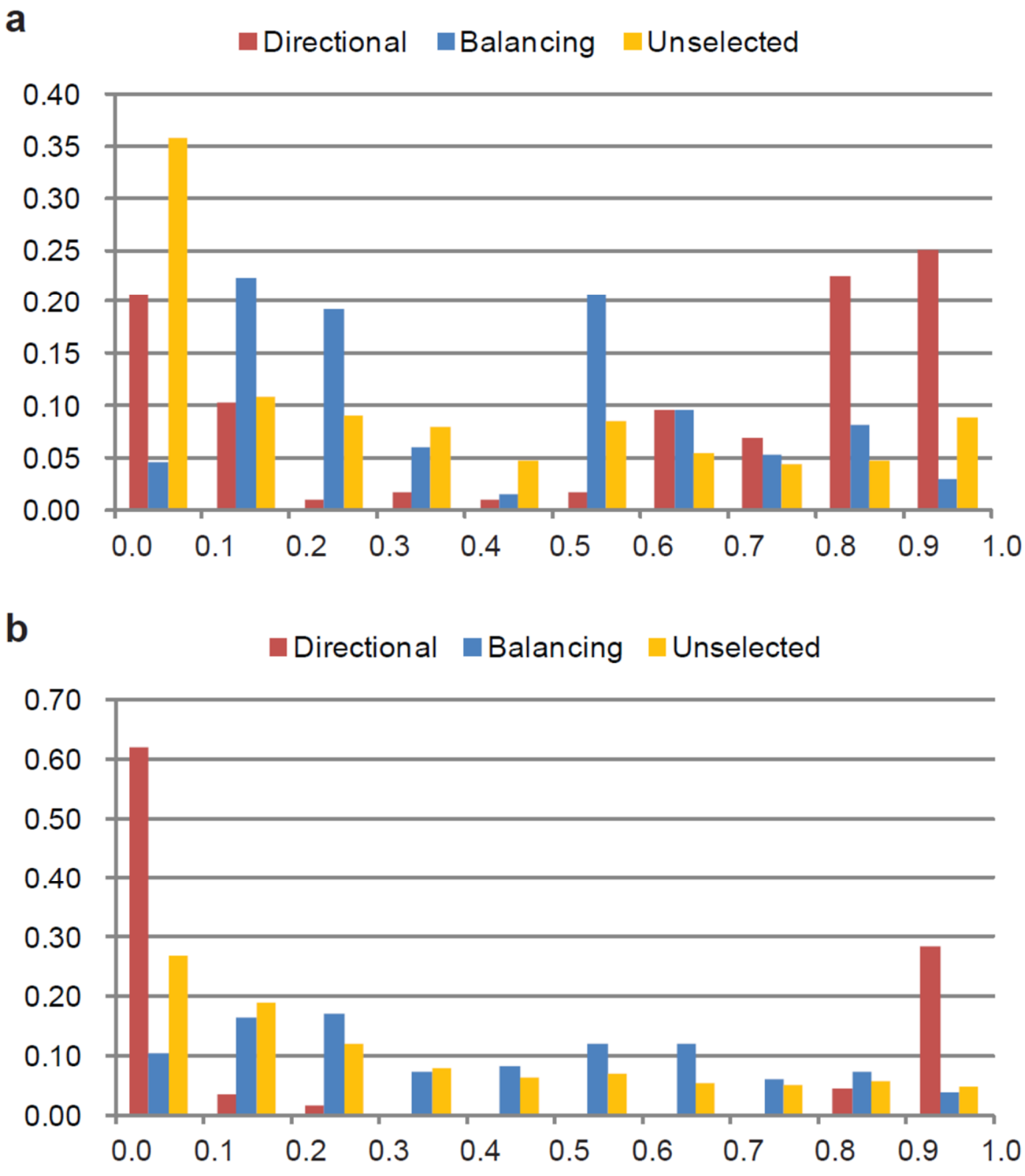
Nature Genetics doi:10/1038/ng.2494

Supplementary Figure 8. Detection of loci under selection of MIN and QXL populations. The simulated joint distribution of Heterozygosity-F_{ST} of SNPs from MIN and QXL using Arlequin: (**a**) Detection of loci under selection based on F_{ST} under finite island model. The upper red and blue dots are outliers subjected to directional selection ($q < 0.05$), whereas the lower red and blue dots under balancing selection ($q < 0.05$). (**b**) Detection of loci under selection using the Bayescan. Log_{10}(q-value): logarithm of q-values to base 10 for the model including selection. Outliers with $q < 0.1$ were shown at the right side of the black vertical line. (**c**) Detection of loci under selection using the global F_{ST} test. The loci with top 1% of F_{ST} were ranked above the red dashed line ($F_{ST} = 0.689655$).



Nature Genetics doi:10/1038/ng.2494

Supplementary Figure 9. Derived allele frequency distribution for the alleles under balancing selection, under directional selection, and the remaining unselected alleles in (**a**) QIN population and (**b**) non-QIN population. The X-axis represents the derived allele frequency bins(that is 0.0-0.1, 0.1-0.2, ..., 0.9-1.0), and the Y-axis represents the frequency of that allele-frequency bin in the studied panda population.



Nature Genetics doi:10/1038/ng.2494

Supplementary Tables

Supplementary Table 1. Sample list of the sequenced giant pandas.

Sample ID	Locality	Mountain range	Origin	Type of samples
GP2	Baoxing, Sichuan	QIO	Wild-born	Blood
GP3	Foping, Shaanxi	QIN	Wild-born	Blood
GP4	Taibai, Shaanxi	QIN	Wild-born	Blood
GP5	Foping, Shaanxi	QIN	Wild-born	Blood
GP6	Yangxian, Shaanxi	QIN	Wild-born	Blood
GP7	Taibai, Shaanxi	QIN	Wild-born	Blood
GP8	Taibai, Shaanxi	QIN	Wild-born	Blood
GP10	Foping, Shaanxi	QIN	Wild-born	Muscle
GP12	Foping, Shaanxi	QIN	Wild-born	Blood
GP13	Baoxing, Sichuan	QIO	Wild-born	Blood
GP14	Qingchuan, Sichuan	MIN	Wild-born	Blood
GP15	Wenxian, Gansu	MIN	Wild-born	Blood
GP16	Pingwu, Sichuan	MIN	Wild-born	Blood
GP17	Pingwu, Sichuan	MIN	Wild-born	Blood
GP18	Beichuan, Sichuan	MIN	Wild-born	Blood
GP19	Qingchuan, Sichuan	MIN	Wild-born	Blood
GP22	Baoxing, Sichuan	QIO	Wild-born	Blood
GP23	Baoxing, Sichuan	QIO	Wild-born	Blood
GP24	Lushan, Sichuan	QIO	Wild-born	Blood
GP25	Wenchuan, Sichuan	QIO	Wild-born	Blood
GP26	Baoxing, Sichuan	QIO	Wild-born	Blood
GP27	Baoxing, Sichuan	QIO	Wild-born	Blood
GP28	Chongzhou, Sichuan	QIO	Wild-born	Blood
GP29	Baoxing, Sichuan	QIO	Wild-born	Blood
GP30	Lushan, Sichuan	QIO	Wild-born	Blood
GP31	Baoxing, Sichuan	QIO	Wild-born	Blood
GP33	Wenchuan, Sichuan	QIO	Wild-born	Blood
GP35	Baoxing, Sichuan	QIO	Wild-born	Blood
GP36	Baoxing, Sichuan	QIO	Wild-born	Blood
GP37	Leibo, Sichuan	LS	Wild-born	Blood
GP38	Luding, Sichuan	DXL	Wild-born	Blood
GP39	Mianning, Sichuan	XXL	Wild-born	Blood
GP51	Anxian, Sichuan	MIN	Wild-born	Blood
GP52	Ebian, Sichuan	LS	Wild-born	Blood
GP53	Chengdu BC	QIO × LS	Captive-born (Admixed)	Blood
GP54	Wolong GPCRC	QIO × LS	Captive-born (Admixed)	Blood
GP57	Wolong GPCRC	QIO × LS	Captive-born (Admixed)	Blood
GP58	Wolong GPCRC	QIO × MIN	Captive-born (Admixed)	Blood
GP60	Wolong GPCRC	QIO × MIN	Captive-born (Admixed)	Blood
GP61	Beijing Zoo	QIO × MIN	Captive-born (Admixed)	Blood
GP64	Wolong GPCRC	QIO × MIN	Captive-born (Admixed)	Blood
GP65	Chengdu BC	MIN × LS	Captive-born (Admixed)	Blood
GP66	Chengdu BC	MIN × LS	Captive-born (Admixed)	Blood
GP67	Chengdu BC	MIN × LS	Captive-born (Admixed)	Blood

Nature Genetics doi:10/1038/ng.2494

GP68	Chengdu BC	MIN × LS	Captive-born (Admixed)	Blood
GP70	Wolong GPCRC	QIN × QIO	Captive-born (Admixed)	Blood
GP71	Chengdu BC	QIN × QIO	Captive-born (Admixed)	Blood
GP72	Chengdu BC	QIN × QIO	Captive-born (Admixed)	Blood

Note: Wolong GPCRC: China Conservation and Research Center for the Giant Panda in Wolong; Chengdu BC: Chengdu Research Base of Giant Panda Breeding; QIO: Qinonglai; QIN: Qinling; MIN: Minshan; DXL: Daxiangling; XXL: Xiaoxiangling; LS: Liangshan.

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 2. Pairwise F_{ST} values among three panda populations.

	Sample size	QIN	MIN	QXL
QIN	8	-	0.2281561	0.2218929
MIN	7	0.1190304	-	0.1508209
QXL	19	0.1158682	0.0924764	-

Note: Mean F_{ST} values are presented in the top right and standard deviations are shown in the bottom left of the matrix.

Nature Genetics doi:10/1038/ng.2494

144 **Supplementary Table 3.** θ_π and θ_w for three panda populations.

	Sample size	θ_π ($\times 10^{-3}$)	θ_w ($\times 10^{-3}$)
QIN	8	1.133221737	1.043086946
MIN	7	1.224627512	1.165239912
QXL	19	1.366216636	1.297351139

145

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 4. Parameters and results inferred by ∂a∂i simulations.

Events	Item symbols	Values	Confidence Interval (95%)	Units
Ancestral population size	nuA	20850	-	individuals
Population size of QIN, at T1	nuQ1*	4951	(819, 6102)	individuals
Population size of QIN, at T2	nuQ2*	981	(117, 1654)	individuals
Population size of QIN, at T3	nuQ3*	1686	(202, 2079)	individuals
Population size of QIN, at T4 (present)	nuQ4*	169	(21, 250)	individuals
Population size of non-QIN at T1	nuN1	15899	(14747, 20031)	individuals
Population size of non-QIN at T2	nuN2*	47868	(15985, 147271)	individuals
Population size of non-QIN at T3	nuN3*	695	(89, 1032)	individuals
Population size of MIN, at T3	nuMA*	480	(58, 676)	individuals
Population size of QXL, at T3	nuXA	215	(29, 281)	individuals
Population size of MIN, at T4 (present)	nuM*	847	(116, 998)	individuals
Population size of QXL, at T4 (present)	nuX*	3726	(437, 4849)	individuals
Migration from non-QIN into QIN between T1 and T2	mQN1*	0.003	(0.000, 0.004)	individuals/year
Migration from QIN into non-QIN between T1 and T2	mNQ1*	0.617	(0.054, 2.832)	individuals/year
Migration from non-QIN into QIN between T2 and T3	mQN2*	0.055	(0.019, 0.093)	individuals/year
Migration from QIN into non-QIN between T2 and T3	mNQ2*	0.091	(0.082, 0.117)	individuals/year
Migration from MIN into QIN, after T3	mQM*	0.014	(0.013, 0.017)	individuals/year
Migration from QIN into MIN, after T3	mMQ*	0.066	(0.054, 0.072)	individuals/year
Migration from QXL into QIN, after T3	mQX*	0.000	(0.000, 0.000)	individuals/year
Migration from QIN into QXL, after T3	mXQ*	0.053	(0.041, 0.070)	individuals/year
Migration from QXL into MIN, after T3	mMX*	0.108	(0.097, 0.136)	individuals/year
Migration from MIN into QXL, after T3	mXM*	0.373	(0.290, 0.479)	individuals/year
Time 1: the first split, QIN and non-QIN	T1*	304664	(125171, 708756)	years
Time 2: the total population size peak, between two splits	T2*	38879	(4882, 58877)	years
Time 3: the second split, MIN and QXL	T3*	2777	(377, 4082)	years
Time 4: present	T4	-	-	-

Note: There are 22 free parameters in the model applied in ∂a∂i, which are shown with asterisks (*) in the column of "Item symbols".

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 5. The number of selected SNPs and genes containing selected loci across panda populations.

	QIN and non-QIN		MIN and QXL	
	SNP number	Gene number	SNP number	Gene number
Directional selection	134	111	52	44
Balancing selection	212	152	0	0

Note that we defined loci that were identified as under selection by two or more methods (**Supplementary Tables 11** and **12**) to be true selected loci in order to minimize the false positives.

Nature Genetics doi:10/1038/ng.2494

154 **Supplementary Table 6.** SNPs under directional selection in the analysis of QIN and non-QIN pandas.

SNP ID	BayeScan (q<0.1)	Arlequin (q<0.05)			Global F_{ST} (top 0.01)	DAF*	Gene ID	Gene name	Gene description	Ligands to taste receptors
		Finite island model	Hierarchical island model							
		F_{ST}	F_{ST}	F_{CT}						
ChrNew11_68822509	×	√	√	√	√	-	Ame_R009913	*TAS2R3*	taste receptor type 2 member 3-like;Bitter taste receptor	T2R3: cromolyn† T2R49: chloroquine†
ChrNew13_15102794	×	×	√	√	×	QIN	Ame_R013354	*TAS2R49*	taste receptor type 2 member 49-like, Bitter taste receptor	
ChrNew18_73304281	√	√	×	×	√	QIN	Ame_R014718	*OR51B5*	Olfactory receptorClass I family 51	
ChrNew1_2522266	×	√	√	×	×	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor Class I, family 51	
ChrNew1_2522308	×	√	√	×	×	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor Class I, family 51	
ChrNew21_70881518	×	×	√	√	×	-	Ame_R019632	*OR1J4*	Olfactory receptor, Class II family 1	
ChrNew17_34223446	√	√	×	×	√	-	Ame_R010062	*Olfr237-ps1, MOR261-4, Olfr438, OR2A2*	Olfactory receptor Class II family 2	
ChrNew10_92288817	×	√	√	×	√	-	Ame_R003118	*Olr37, Olfr521,MOR101-2,OR2AT4*	Olfactory receptor Class II family 2	
ChrNew22_7947187	×	√	√	×	√	-	Ame_R015495	*Olfr1094,MOR179-7,OR5T1*	Olfactory receptor, Class II family 5	
ChrNew22_7947186	×	√	√	×	√	-	Ame_R015495	*Olfr1094,MOR179-7,OR5T1*	Olfactory receptor, Class II family 5	
ChrNew22_24480089	×	√	√	×	×	-	Ame_R015884	*Olfr1406,MOR267-5*	Olfactory receptor Class II family 10	
ChrNew6_52879057	×	√	√	√	×	-	Ame_R018269	*OR11A1*	Olfactory receptor, Class II family 11	
ChrNew6_52879430	×	√	√	×	×	-	Ame_R018269	*OR11A1*	Olfactory receptor, Class II family 11	
ChrNew15_39099245	√	√	√	×	√	non-QIN	Ame_R016404	*ACSBG2*	acyl-CoA synthetase bubblegum family member 2	
ChrNew12_42854253	×	√	×	×	√	QIN	Ame_R003963	*ADAM12*	ADAM metallopeptidase domain 1	
ChrNew18_13933934	√	√	×	×	√	QIN	Ame_R006591	*ADAM18*	disintegrin and metalloproteinase domain-containing protein	
ChrNew21_85106806	×	×	√	√	×	non-QIN	Ame_R014828	*ADAMTSL3*	a disintegrin and metalloproteinase with thrombospondin motifs 17	
ChrNew19_48607413	×	√	√	×	×	non-QIN	Ame_R008412	*AKAP12*	A kinase (PRKA) anchor protein 12	
ChrNew3_26776604	×	×	√	√	×	-	Ame_R003995	*AMPD3*	adenosine monophosphate deaminase 3	
ChrNew22_81741605	×	√	√	×	√	QIN	Ame_R018985	*APOB*	apolipoprotein B	

Nature Genetics doi:10/1038/ng.2494

ChrNew11_15057064	×	√	√	×	√	QIN	Ame_R003874	*ARHGAP9*	Rho GTPase activating protein 9
ChrNew1_72959361	×	×	√	√	×	QIN	Ame_R017919	*ATM*	ataxia telangiectasia mutated
ChrNew4_63559904	√	√	×	×	√	non-QIN	Ame_R006113	*ATP5A1*	F-type H+-transporting ATPase subunit alpha
ChrNew5_87815484	×	×	√	√	×	QIN	Ame_R006622	*C1orf114*	Uncharacterized protein C1orf114
ChrNew5_27451189	×	√	√	√	×	QIN	Ame_R009487	*C3orf25*	EF-hand domain-containing protein C3orf25
ChrNew5_27485974	√	√	√	√	√	non-QIN	Ame_R009489	*CAND2*	cullin-associated and neddylation-dissociated 2 (putative)
ChrNew12_39718279	√	√	√	×	√	QIN	Ame_R009441	*CASS4*	breast cancer anti-estrogen resistance
ChrNew1_10536358	×	×	√	√	×	QIN	Ame_R007314	*CELSR2*	cadherin EGF LAG seven-pass G-type receptor 2-like
ChrNew16_26734362	×	×	√	√	×	non-QIN	Ame_R016468	*CHD8*	chromodomain helicase DNA binding protein 8
ChrNew9_53180750	√	√	√	×	√	QIN	Ame_R011104	*CLDN16*	claudin-16-like
ChrNew9_53180723	×	√	√	×	√	non-QIN	Ame_R011104	*CLDN16*	claudin-16-like
ChrNew15_32716869	×	√	√	×	√	QIN	Ame_R016217	*CRYM*	crystallin, mu
ChrNew5_76495105	×	√	√	×	×	non-QIN	Ame_R016965	*DLGAP2*	discs, large (Drosophila) homolog-associated protein 2
ChrNew15_32818766	×	√	√	×	√	QIN	Ame_R016221	*DNAH3*	DNAH3; dynein, axonemal, heavy chain 3
ChrNew15_32835250	×	√	√	×	√	non-QIN	Ame_R016221	*DNAH3*	DNAH3; dynein, axonemal, heavy chain 3
ChrNew15_32806488	×	√	×	×	√	QIN	Ame_R016221	*DNAH3*	DNAH3; dynein, axonemal, heavy chain 3
ChrNew15_32810598	×	√	√	×	√	QIN	Ame_R016221	*DNAH3*	DNAH3; dynein, axonemal, heavy chain 3
ChrNew18_10592532	√	√	√	×	√	QIN	Ame_R011133	*DNHL1*	dynein heavy chain, axonemal
ChrNew5_37375737	×	√	√	×	×	non-QIN	Ame_R005426	*DPYS*	dihydropyrimidinase
ChrNew1_19487540	×	×	√	√	×	QIN	Ame_R006993	*DST*	dystonin
ChrNew23_2284944	×	√	√	√	×	QIN	Ame_R015987	*DUSP27*	dual specificity phosphatase 27
ChrNew23_5617468	×	√	√	×	×	QIN	Ame_R019520	*ELMO2*	engulfment and cell motility protein
ChrNew2_63588493	√	√	√	×	√	QIN	Ame_R004145	*ELOVL4*	elongation of very long chain fatty acids protein 4-like
ChrNew1_69277347	×	√	√	×	×	non-QIN	Ame_R017604	*ETS2*	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
ChrNew20_66205685	√	√	√	√	√	QIN	Ame_R014467	*FAM135A*	family with sequence similarity 135, member A
ChrNew6_8492307	×	√	×	×	√	non-QIN	Ame_R010626	*FMNL3*	formin-like 3
ChrNew17_24784368	×	√	√	√	×	non-QIN	Ame_R018842	*GCN1L1*	GCN1 general control of amino-acid synthesis 1-like 1

Nature Genetics doi:10/1038/ng.2494

ChrNew19_16019060	×	√	√	×	√	non-QIN	Ame_R013045	*GDAP2*	ganglioside induced differentiation associated protein 2
ChrNew19_16046609	×	√	×	×	√	QIN	Ame_R013045	*GDAP2*	ganglioside induced differentiation associated protein 2
ChrNew13_5049362	×	√	×	×	√	QIN	Ame_R003791	*GLT6D1*	glycosyltransferase 6 domain containing 1
ChrNew13_5049439	×	√	√	×	×	non-QIN	Ame_R003791	*GLT6D1*	glycosyltransferase 6 domain containing 1
ChrNew11_87909862	×	√	√	×	×	QIN	Ame_R002437	*GOLGA4*	golgi autoantigen, golgin subfamily a, 4
ChrNew23_27914672	×	√	√	×	√	QIN	Ame_R019377	*GSTA2*	glutathione S-transferase
ChrNew9_24941043	×	×	√	√	×	QIN	Ame_R011255	*GSTA3*	glutathione S-transferase
ChrNew17_22569208	×	√	√	×	√	QIN	Ame_R002735	*HEMGN*	hemogen
ChrNew11_40894894	×	√	√	×	√	-	Ame_R006066	*hypothetical protei*	Putative uncharacterized protein
ChrNew9_22892853	×	√	×	×	√	QIN	Ame_R006371	*hypothetical protei*	hypothetical protein;chitin elicitor receptor kinase 1
ChrNew22_83045145	×	√	√	×	√	-	Ame_R016847	*hypothetical protei*	hypothetical protein
ChrNew5_15883840	×	√	√	×	√	QIN	Ame_R011984	*hypothetical protein*	hypothetical protein
ChrNew23_44608974	×	√	√	√	√	QIN	Ame_R018589	*hypothetical protein*	hypothetical protein
ChrNew3_41624189	×	×	√	×	√	non-QIN	Ame_R016305	*hypothetical protein*	KIAA0947
ChrNew5_66479295	×	√	×	×	√	non-QIN	Ame_R006158	*hypothetical protein*	SET domain containing 5
ChrNew4_4114112	×	√	√	√	×	non-QIN	Ame_R002161	*hypothetical protein*	hypothetical protein
ChrNew10_60776413	×	×	√	√	×	non-QIN	Ame_R005983	*hypothetical protein*	hypothetical protein
ChrNew10_98762469	×	√	×	×	√	QIN	Ame_R013070	*ITGA10*	integrin alpha 10
ChrNew9_35892755	×	√	√	×	×	QIN	Ame_R008020	*KANK1*	KN motif and ankyrin repeat domains 1
ChrNew17_23116167	×	√	×	×	√	QIN	Ame_R002743	*KIAA1529*	C9orf174 - KIAA1529
ChrNew3_34377876	×	√	√	√	√	-	Ame_R000807	*KLHL1*	kelch-like protein
ChrNew7_75485516	×	√	√	×	√	-	Ame_R005661	*L1TD1*	LINE-1 type transposase domain-containing protein 1-like
ChrNew11_37848584	×	√	√	×	×	QIN	Ame_R010574	*LARP6*	La ribonucleoprotein domain family, member 6
ChrNew8_29639620	×	√	×	×	√	non-QIN	Ame_R000120	*LGALS9C*	lectin, galactoside-binding, soluble, 9C
ChrNew19_41020781	×	√	√	×	√	QIN	Ame_R016428	*LIPG*	lipase, endothelial;triacylglycerol lipase
ChrNew19_59364967	√	√	√	×	√	QIN	Ame_R011719	*LMBRD2*	LMBR1 domain containing 2
ChrNew11_16195805	×	√	×	×	√	QIN	Ame_R003897	*LRIG3*	leucine-rich repeats and immunoglobulin-like domains 3
ChrNew11_1917567	√	√	√	×	√	non-QIN	Ame_R010726	*LRRC8C*	leucine rich repeat containing 8 family, member C

Nature Genetics doi:10/1038/ng.2494

ChrNew11_88135131	×	×	√	√	×	QIN	Ame_R002439	*LRRFIP2*	leucine rich repeat (in FLII) interacting protein 2
ChrNew9_78249344	×	√	√	×	√	QIN	Ame_R017483	*LY9*	lymphocyte antigen 9
ChrNew22_22234107	×	×	√	√	×	QIN	Ame_R016952	*LYST*	lysosomal trafficking regulator
ChrNew8_21770308	×	√	√	×	×	QIN	Ame_R005022	*MAP3K19*	mitogen-activated protein kinase kinase kinase 9
ChrNew14_60042598	×	√	×	×	√	non-QIN	Ame_R005515	*MARCO*	macrophage receptor with collagenous structure
ChrNew3_56445985	×	√	√	×	√	QIN	Ame_R004324	*mRNA*	KRAB domain-containing zinc finger protein
ChrNew2_90411464	×	√	√	×	√	QIN	Ame_R011044	*MYO18B*	myosin XVIIIB
ChrNew2_22928866	×	√	√	×	√	QIN	Ame_R016715	*MYO3A*	MYO3A; myosin IIIA
ChrNew19_43825414	×	√	√	×	√	QIN	Ame_R010455	*NAG*	neuroblastoma amplified sequence
ChrNew13_89522611	×	√	√	×	√	non-QIN	Ame_R015725	*NCAM1*	neural cell adhesion molecule 1
ChrNew22_95637334	×	×	√	√	×	QIN	Ame_R017853	*NPHP4*	nephronophthisis 4
ChrNew4_50340954	√	√	√	×	√	-	Ame_R000305	*NR3C1*	nuclear receptor subfamily 3, group C, member 1
ChrNew23_4941648	×	√	√	×	√	QIN	Ame_R017629	*PACS2*	phosphofurin acidic cluster sorting protein 2
ChrNew21_52252578	×	√	√	×	√	QIN	Ame_R017825	*PDILT*	protein disulfide isomerase-like, testis expressed
ChrNew1_46803774	√	√	√	×	√	non-QIN	Ame_R003936	*PLCE1*	phospholipase C, epsilon 1
ChrNew10_23573833	×	√	×	×	√	QIN	Ame_R008818	*PTK2*	PTK2 protein tyrosine kinase 2
ChrNew4_63848342	√	√	√	×	√	QIN	Ame_R006116	*RNF165*	ring finger protein 165
ChrNew19_84110344	×	×	√	√	×	QIN	Ame_R013129	*RORB*	RAR-related orphan receptor beta
ChrNew1_95737132	×	×	√	√	×	QIN	Ame_R003759	*RTTN*	rotatin
ChrNew3_55152056	×	√	√	×	√	non-QIN	Ame_R004309	*SDK1*	sidekick homolog 1, cell adhesion molecule (chicken)
ChrNew4_40252953	×	√	√	×	×	QIN	Ame_R003148	*SH3BP5*	SH3-domain binding protein 5 (BTK-associated)
ChrNew14_21346992	×	√	√	√	×	non-QIN	Ame_R004638	*SLC12A8*	solute carrier family 12 (potassium/chloride transporters), member 8
ChrNew19_95400019	√	√	√	×	√	QIN	Ame_R013331	*SLC4A3*	solute carrier family 4 (anion exchanger), member 3
ChrNew22_56991758	×	×	√	√	×	QIN	Ame_R016491	*SLC5A9*	solute carrier family 5 (sodium/glucose cotransporter), member 9
ChrNew5_59584336	×	×	√	√	×	QIN	Ame_R001837	*SORL1*	sortilin-related receptor, L(DLR class) A repeats-containing
ChrNew5_10706846	×	√	√	×	√	non-QIN	Ame_R001735	*STEAP4*	metalloreductase STEAP

Nature Genetics doi:10/1038/ng.2494

ChrNew12_55220708	×	√	√	√	√	QIN	Ame_R006746	*T*	BRA, T; brachyury protein
ChrNew10_42495344	×	×	√	√	×	QIN	Ame_R003640	*TCF19*	Transcription factor 19 (Transcription factor SC1)
ChrNew14_78193702	×	√	√	×	×	non-QIN	Ame_R010587	*THNSL2*	hreonine synthase-like 2
ChrNew5_84831787	×	√	×	×	√	QIN	Ame_R008608	*TMC5*	transmembrane channel-like 5
ChrNew19_72179033	×	√	×	×	√	non-QIN	Ame_R010955	*TNKS1BP1*	tankyrase 1 binding protein 1, 182kDa
ChrNew20_25236651	×	√	√	√	×	non-QIN	Ame_R014671	*TOMM7*	translocase of outer mitochondrial membrane 7 homolog
ChrNew19_4966537	×	√	√	√	√	QIN	Ame_R019157	*TRG@*	t-cell receptor gamma-2 chain C region isoform 1
ChrNew19_4966570	×	√	√	√	√	QIN	Ame_R019157	*TRG@*	t-cell receptor gamma-2 chain C region isoform 1
ChrNew19_4966524	×	√	√	√	√	QIN	Ame_R019157	*TRG@*	t-cell receptor gamma-2 chain C region isoform 1
ChrNew19_4966595	×	√	√	×	√	QIN	Ame_R019157	*TRG@*	t-cell receptor gamma-2 chain C region isoform 1
ChrNew21_88952750	√	×	×	×	√	non-QIN	Ame_R017872	*TSEN15*	tRNA splicing endonuclease 15 homolog (S. cerevisiae)
ChrNew20_19822022	×	√	√	√	×	QIN	Ame_R002795	*TTC32*	tetratricopeptide repeat domain 32
ChrNew12_22361355	×	√	×	×	√	non-QIN	Ame_R012834	*TTF1*	transcription termination factor 1-like
ChrNew10_999815	×	√	×	×	√	QIN	Ame_R006883	*TTN*	titin
ChrNew10_992952	×	√	√	×	×	non-QIN	Ame_R006883	*TTN*	titin
ChrNew10_1223250	×	×	√	√	×	QIN	Ame_R006883	*TTN*	titin-like
ChrNew11_37982360	×	√	√	√	×	QIN	Ame_R010575	*UACA*	uveal autoantigen with coiled-coil domains and ankyrin repeatS
ChrNew3_56897576	×	√	√	√	×	QIN	Ame_R004335	*USP42*	ubiquitin specific peptidase 42
ChrNew20_25068040	×	√	√	√	√	non-QIN	Ame_R014667	*VWA3A*	von Willebrand factor A domain containing 3A
ChrNew20_25068079	×	√	√	√	×	QIN	Ame_R014667	*VWA3A*	von Willebrand factor A domain containing 3A
ChrNew19_52740920	×	√	√	√	√	non-QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52745625	√	√	√	×	√	QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52738268	√	√	√	×	√	QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52740908	×	√	√	×	√	non-QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52760186	√	√	×	×	√	non-QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52734234	√	√	√	√	√	non-QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52751619	√	√	√	×	√	QIN	Ame_R018965	*VWF*	von Willebrand factor

Nature Genetics doi:10/1038/ng.2494

ChrNew19_52731692	√	√	√	×	√	non-QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew19_52738103	√	√	√	×	√	non-QIN	Ame_R018965	*VWF*	von Willebrand factor
ChrNew9_65636461	×	√	√	×	√	QIN	Ame_R004649	*VWF*	von Willebrand factor
ChrNew14_56236211	×	×	√	√	×	QIN	Ame_R012304	*WAC*	WW domain containing adaptor with coiled-coil
ChrNew20_19873960	×	×	√	√	×	QIN	Ame_R002796	*WDR35*	WD repeat domain 35
ChrNew11_94083515	×	√	√	×	√	-	Ame_R006929	*ZNF839*	zinc finger protein 839

Note: The symbol " √ " means that SNPs were detected to be under directional selection by this method. *The population shown in the column of "DAF" is the population that harbors higher derived allele frequency at this locus. †The ligands to both taste receptors were synthetic bitter compounds identified in a previous report[2].

Nature Genetics doi:10/1038/ng.2494

158 **Supplementary Table 7.** SNPs under balancing selection in the analysis of QIN and non-QIN pandas.

SNP ID	BayeScan (q<0.1)	Arlequin (q<0.05)			Gene ID	Gene name	Gene description	Ligands to olfactory receptors
		Finite island model	Hierarchical island model					
		F_{ST}	F_{ST}	F_{CT}				
ChrNew23_58888944	×	√	√	×	Ame_R020747	*Olfr624*	Olfactory receptor, Class I family 51	
ChrNew14_62441286	×	√	√	×	Ame_R020209	*OR52D1*	Olfactory receptor, Class I family 52	
ChrNew22_38984697	×	√	√	×	Ame_R020273	*OR52E2*	Olfactory receptor, Class I family 52	
ChrNew22_38984706	×	√	√	×	Ame_R020273	*OR52E2*	Olfactory receptor, Class I family 52	
ChrNew22_38984528	×	√	√	×	Ame_R020273	*OR52E2*	Olfactory receptor, Class I family 52	
ChrNew23_35721322	×	√	√	×	Ame_R019672	*OR52R1, MOR30-1**	Olfactory receptor, Class I family 52	MOR30-1*: *Heptanoic acid*†*; Octanoic acid*†; Decanal; Decanoic acid; Nonanal; Nonanoic acid
ChrNew23_35721321	×	√	√	×	Ame_R019672	*OR52R1, MOR30-1**	Olfactory receptor, Class I family 52	MOR30-1*: *Heptanoic acid*†*; Octanoic acid*†; Decanal; Decanoic acid; Nonanal; Nonanoic acid
ChrNew1_44681904	×	√	√	×	Ame_R019568	*OR10H1*	Olfactory receptor, Class II family 10	
ChrNew14_97213960	×	√	√	×	Ame_R020795	*OR11G2*	Olfactory receptor, Class II family 11	
ChrNew14_97213965	×	√	√	×	Ame_R020795	*OR11G2*	Olfactory receptor, Class II family 11	
ChrNew11_74812407	×	√	√	×	Ame_R018387	*OR13C4*	Olfactory receptor, Class II family 13	
ChrNew11_74838207	×	√	√	×	Ame_R018388	*OR13F1*	Olfactory receptor, Class II family 13	
ChrNew11_74838212	×	√	√	×	Ame_R018388	*OR13F1*	Olfactory receptor, Class II family 13	
ChrNew18_33134801	×	√	√	×	Ame_R012645	*OR4A15*	Olfactory receptor, Class II family 4	
ChrNew17_10511961	×	√	√	×	Ame_R019846	*OR4A47*	Olfactory receptor, Class II family 4	
ChrNew1_88826806	×	√	√	×	Ame_R019141	*OR4C11*	Olfactory receptor, Class II family 4	
ChrNew10_21181535	×	√	√	×	Ame_R018858	*OR4C11*	Olfactory receptor, Class II family 4	
ChrNew22_75816288	×	√	√	×	Ame_R018196	*OR4D2*	Olfactory receptor, Class II family 4	
ChrNew21_82094708	×	√	√	×	Ame_R020256	*OR6C75*	Olfactory receptor, Class II family 6	
ChrNew21_82094691	×	√	√	×	Ame_R020256	*OR6C75*	Olfactory receptor, Class II family 6	
ChrNew21_82094695	×	√	√	×	Ame_R020256	*OR6C75*	Olfactory receptor, Class II family 6	
ChrNew10_82353890	×	√	√	×	Ame_R019859	*OR6M1*	Olfactory receptor, Class II family 6	
ChrNew17_40015449	×	√	√	×	Ame_R018714	*OR7A10,OLF4*	Olfactory receptor, Class II family 7	

Nature Genetics doi:10/1038/ng.2494

ChrNew17_40015433	×	√	√	×	Ame_R018714	*OR7A10,OLF4*	Olfacotry receptor, Class II family 7
ChrNew17_40015444	×	√	√	×	Ame_R018714	*OR7A10,OLF4*	Olfacotry receptor, Class II family 7
ChrNew17_40015420	×	√	√	×	Ame_R018714	*OR7A10,OLF4*	Olfacotry receptor, Class II family 7
ChrNew17_40015424	×	√	√	×	Ame_R018714	*OR7A10,OLF4*	Olfacotry receptor, Class II family 7
ChrNew17_40015423	×	√	√	×	Ame_R018714	*OR7A10,OLF4*	Olfacotry receptor, Class II family 7
ChrNew23_59602996	×	√	√	×	Ame_R020638	*OR7A17*	Olfacotry receptor, Class II family 7
ChrNew23_59602998	×	√	√	×	Ame_R020638	*OR7A17*	Olfacotry receptor, Class II family 7
ChrNew17_39965097	×	√	√	×	Ame_R018711	*OR7C1*	Olfacotry receptor, Class II family 7
ChrNew1_58428350	×	√	√	×	Ame_R019942	*OR7E5P*	Olfacotry receptor, Class II family 7
ChrNew19_64709779	×	√	√	×	Ame_R020084	*OR7E5P*	Olfacotry receptor, Class II family 7
ChrNew20_54410780	×	√	√	×	Ame_R019373	*OR7E5P*	Olfacotry receptor, Class II family 7
ChrNew22_63103053	×	√	√	×	Ame_R020375	*Olfr1394, MOR280-1,OR2M3*	Olfacotry receptor, Class II family2
ChrNew17_39796647	×	√	√	×	Ame_R018704	*OR7A10,OLF4*	Olfacotry receptor, Class II family7
ChrNew19_71747462	×	√	√	×	Ame_R010940	*OR9G9, Olfr1013, MOR213-2*	Olfacotry receptor, Class II family9
ChrNew19_71747471	×	√	√	×	Ame_R010940	*OR9G9, Olfr1013, MOR213-2*	Olfacotry receptor, Class II family9
ChrNew19_71747474	×	√	√	×	Ame_R010940	*OR9G9, Olfr1013, MOR213-2*	Olfacotry receptor, Class II family9
ChrNew9_79919106	×	√	√	×	Ame_R010298	*AKAP13*	A kinase (PRKA) anchor protein 13
ChrNew2_46652107	×	√	√	×	Ame_R014444	*ACTN4*	actinin, alpha 4
ChrNew17_63448765	×	√	√	×	Ame_R010887	*ADAM29*	ADAM metallopeptidase domain 29
ChrNew17_55576954	×	√	√	×	Ame_R007671	*ACAN*	aggrecan
ChrNew19_9288371	×	√	√	×	Ame_R003468	*AKR7A3*	aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)
ChrNew19_9288375	×	√	√	×	Ame_R003468	*AKR7A3*	aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)
ChrNew19_9288372	×	√	√	×	Ame_R003468	*AKR7A3*	aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)
ChrNew18_96699929	×	√	√	×	Ame_R009720	*ALMS1*	Alstrom syndrome 1
ChrNew19_1018711	×	√	√	×	Ame_R009990	*MKI67*	antigen identified by monoclonal antibody Ki-67
ChrNew19_39762831	×	√	√	×	Ame_R015102	*ASTL*	astacin-like metallo-endopeptidase (M12 family)
ChrNew4_39246929	×	√	√	×	Ame_R007613	*BUD13*	BUD13 homolog (S. cerevisiae)
ChrNew20_92806715	×	√	√	×	Ame_R017281	*BTN1A1*	butyrophilin, subfamily 1, member A1
ChrNew2_44612544	×	√	√	×	Ame_R004346	*CISD2*	CDGSH iron sulfur domain 2

Nature Genetics doi:10/1038/ng.2494

ChrNew1_30975358	×	√	√	×	Ame_R004259	*CEP250*	centrosomal protein 250kDa
ChrNew19_70349251	×	√	√	×	Ame_R008786	*CLN6*	ceroid-lipofuscinosis, neuronal 6, late infantile, variant
ChrNew11_82499324	×	√	√	×	Ame_R003505	*CLCN2*	chloride channel 2
ChrNew12_56335994	×	√	√	×	Ame_R006757	*hypothetical protein*	copine I
ChrNew3_28982574	×	√	√	×	Ame_R004031	*CYB5R2*	cytochrome b5 reductase 2
ChrNew8_65691412	×	√	√	×	Ame_R002325	*CYTH4*	cytohesin 4
ChrNew4_54038729	×	√	√	×	Ame_R008460	*DAPK1*	death-associated protein kinase 1
ChrNew4_12404103	×	√	√	×	Ame_R001357	*DSPP*	dentin sialophosphoprotein
ChrNew12_67497935	×	√	√	×	Ame_R017578	*DPCR1*	diffuse panbronchiolitis critical region 1
ChrNew2_82192254	×	√	√	×	Ame_R000924	*DUOX2*	dual oxidase 2
ChrNew16_48992687	×	√	√	×	Ame_R012122	*DYRK1B*	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B
ChrNew20_7709177	×	√	√	×	Ame_R020685	*ENPP7*	ectonucleotide pyrophosphatase/phosphodiesterase 7
ChrNew20_7709366	×	√	√	×	Ame_R020685	*ENPP7*	ectonucleotide pyrophosphatase/phosphodiesterase 7
ChrNew20_7709180	×	√	√	×	Ame_R020685	*ENPP7*	ectonucleotide pyrophosphatase/phosphodiesterase 7
ChrNew20_7709183	×	√	√	×	Ame_R020685	*ENPP7*	ectonucleotide pyrophosphatase/phosphodiesterase 7
ChrNew20_7709185	×	√	√	×	Ame_R020685	*ENPP7*	ectonucleotide pyrophosphatase/phosphodiesterase 7
ChrNew23_83240933	×	√	√	×	Ame_R020760	*FAM22G*	FAM22G isoform 1
ChrNew23_83240960	×	√	√	×	Ame_R020760	*FAM22G*	FAM22G isoform 1
ChrNew23_83240932	×	√	√	×	Ame_R020760	*FAM22G*	FAM22G isoform 1
ChrNew22_66413334	×	√	√	×	Ame_R018605	*FAM169B*	family with sequence similarity 169, member B
ChrNew13_85822209	×	√	√	×	Ame_R010659	*FAM169B*	family with sequence similarity 169, member B
ChrNew3_64566295	×	√	√	×	Ame_R011210	*FAM169B*	family with sequence similarity 169, member B
ChrNew17_30586942	×	√	√	×	Ame_R012023	*FAM169B*	family with sequence similarity 169, member B
ChrNew22_49783060	×	√	√	×	Ame_R017412	*FAM169B*	family with sequence similarity 169, member B
ChrNew2_93232213	×	√	√	×	Ame_R001159	*FAM169B*	family with sequence similarity 169, member B
ChrNew17_30586663	×	√	√	×	Ame_R012023	*FAM169B*	family with sequence similarity 169, member B
ChrNew7_46616425	×	√	√	×	Ame_R008004	*QRICH2*	glutamine rich 2
ChrNew20_92540467	×	√	√	×	Ame_R017249	*HIST1H1C*	histone cluster 1, H1c
ChrNew15_99657191	×	√	√	×	Ame_R001969	*HIPK2*	homeodomain interacting protein kinase 2

Nature Genetics doi:10/1038/ng.2494

ChrNew21_98531185	×	×	√	√	Ame_R020248	*hypothetical protein*	hypothetical protein
ChrNew10_67739106	×	√	√	×	Ame_R011529	*hypothetical protein*	hypothetical protein
ChrNew9_62708766	×	√	√	×	Ame_R007278	*hypothetical protein*	hypothetical protein
ChrNew13_2156381	×	√	√	×	Ame_R014805	*hypothetical protein*	hypothetical protein
ChrNew23_1613092	×	√	√	×	Ame_R016958	*hypothetical protein*	hypothetical protein
ChrNew21_98531177	×	√	√	×	Ame_R020248	*hypothetical protein*	hypothetical protein
ChrNew13_28725823	×	√	√	×	Ame_R002825	*hypothetical protein*	hypothetical protein
ChrNew20_13283492	×	√	√	×	Ame_R008248	*hypothetical protein*	hypothetical protein
ChrNew2_10160010	×	√	√	×	Ame_R019012	*hypothetical protein*	hypothetical protein
ChrNew21_91627453	×	√	√	×	Ame_R020093	*hypothetical protein*	hypothetical protein
ChrNew10_43799304	×	√	√	×	Ame_R003664	*hypothetical protein*	hypothetical protein
ChrNew23_51279921	×	√	√	×	Ame_R019813	*hypothetical protein*	hypothetical protein
ChrNew10_67739078	×	√	√	×	Ame_R011529	*hypothetical protein*	hypothetical protein
ChrNew14_19525056	×	√	√	×	Ame_R004607	*hypothetical protein*	hypothetical protein
ChrNew14_57276286	×	√	√	×	Ame_R020112	*hypothetical protein*	hypothetical protein
ChrNew22_57289561	×	√	√	×	Ame_R016494	*hypothetical protein*	hypothetical protein
ChrNew23_51280104	×	√	√	×	Ame_R019813	*hypothetical protein*	hypothetical protein
ChrNew11_40894676	×	√	√	×	Ame_R006066	*hypothetical protein*	hypothetical protein
ChrNew11_40894695	×	√	√	×	Ame_R006066	*hypothetical protein*	hypothetical protein
ChrNew2_93342079	×	√	√	×	Ame_R001160	*hypothetical protein*	hypothetical protein
ChrNew3_69145234	×	√	√	×	Ame_R010801	*hypothetical protein*	hypothetical protein
ChrNew11_99115838	×	√	√	×	Ame_R007895	*hypothetical protein*	hypothetical protein
ChrNew2_93342160	×	√	√	×	Ame_R001160	*hypothetical protein*	hypothetical protein
ChrNew13_28725800	×	√	√	×	Ame_R002825	*hypothetical protein*	hypothetical protein
ChrNew3_69145158	×	√	√	×	Ame_R010801	*hypothetical protein*	hypothetical protein
ChrNew8_85590710	×	√	√	×	Ame_R011062	*hypothetical protein*	hypothetical protein
ChrNew23_1613208	×	√	√	×	Ame_R016958	*hypothetical protein*	hypothetical protein
ChrNew11_40894682	×	√	√	×	Ame_R006066	*hypothetical protein*	hypothetical protein
ChrNew11_40894665	×	√	√	×	Ame_R006066	*hypothetical protein*	hypothetical protein
ChrNew12_99793839	×	√	√	×	Ame_R008887	*hypothetical protein*	hypothetical protein

Nature Genetics doi:10/1038/ng.2494

ChrNew3_69145529	×	√	√	×	Ame_R010801	*hypothetical protein*	hypothetical protein
ChrNew3_6327465	×	√	√	×	Ame_R006187	*hypothetical protein*	hypothetical protein
ChrNew9_69817809	×	√	√	×	Ame_R018429	*hypothetical protein*	hypothetical protein
ChrNew22_12949916	×	√	√	×	Ame_R020569	*IGLV1-44*	immunoglobulin lambda variable 1-44
ChrNew22_69140192	×	√	√	×	Ame_R020506	*IGLV1-50*	immunoglobulin lambda variable 1-50 (non-functional)
ChrNew23_50313630	×	√	√	×	Ame_R020781	*IGLV2-14*	immunoglobulin lambda variable 2-14
ChrNew20_71817292	×	√	√	×	Ame_R020609	*IGLV2-14*	immunoglobulin lambda variable 2-14
ChrNew23_57669140	×	√	√	×	Ame_R020546	*IGLV3-21*	immunoglobulin lambda variable 3-21
ChrNew23_35750698	×	√	√	×	Ame_R020592	*IGLV5-45*	immunoglobulin lambda variable 5-45
ChrNew23_28125615	×	√	√	×	Ame_R020394	*IGLV5-45*	immunoglobulin lambda variable 5-45
ChrNew21_29334874	×	√	√	×	Ame_R020278	*IGLV5-45*	immunoglobulin lambda variable 5-45
ChrNew21_29334872	×	√	√	×	Ame_R020278	*IGLV5-45*	immunoglobulin lambda variable 5-45
ChrNew23_58445107	×	√	√	×	Ame_R020741	*IGLV5-48*	immunoglobulin lambda variable 5-48 (non-functional)
ChrNew23_59201226	×	√	√	×	Ame_R020727	*IGLV5-49*	immunoglobulin lambda variable 9-49
ChrNew9_48370084	×	√	√	×	Ame_R001044	*IL28RA*	interleukin 28 receptor, alpha (interferon, lambda receptor)
ChrNew21_3992310	×	√	√	×	Ame_R019357	*IVL*	involucrin
ChrNew21_3992487	×	√	√	×	Ame_R019357	*IVL*	involucrin
ChrNew5_65976111	×	√	√	×	Ame_R006157	*KIF26B*	kinesin family member 26B
ChrNew5_65975254	×	√	√	×	Ame_R006157	*KIF26B*	kinesin family member 26B
ChrNew5_65975654	×	√	√	×	Ame_R006157	*KIF26B*	kinesin family member 26B
ChrNew23_58818373	×	√	√	×	Ame_R020541	*KRTAP9-2*	kinesin family member 26B
ChrNew5_88613037	×	√	√	×	Ame_R001279	*LRTM1*	leucine-rich repeats and transmembrane domains 1
ChrNew23_50245993	×	√	√	×	Ame_R020263	*LILRA3*	leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 3
ChrNew12_24050761	×	√	√	×	Ame_R016641	*LILRB5*	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 5
ChrNew13_73438568	×	√	√	×	Ame_R008695	*MST1R*	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)
ChrNew23_54815176	×	√	√	×	Ame_R020367	*MAN2B2*	mannosidase, alpha, class 2B, member 2
ChrNew21_73925225	×	√	√	×	Ame_R018378	*MRPL15*	mitochondrial ribosomal protein L15
ChrNew7_44370375	×	√	√	×	Ame_R010136	*MUC16*	mucin 16, cell surface associated
ChrNew20_13330264	×	√	√	×	Ame_R008250	*MUC5B*	mucin 5B, oligomeric mucus/gel-forming

Nature Genetics doi:10/1038/ng.2494

ChrNew20_13329960	×	√	√	×	Ame_R008250	*MUC5B*	mucin 5B, oligomeric mucus/gel-forming
ChrNew14_20375207	×	√	√	×	Ame_R004626	*MUC4*	mucin-4-like
ChrNew22_69890448	×	√	√	×	Ame_R019098	*MYOM1*	myomesin 1, 185kDa
ChrNew3_62388896	×	√	√	×	Ame_R011730	*MYO19*	myosin XIX
ChrNew19_17975987	×	√	√	×	Ame_R008218	*NAV2*	neuron navigator 2
ChrNew11_24295877	×	√	√	×	Ame_R012485	*NOLA2*	NHP2 ribonucleoprotein homolog (yeast
ChrNew6_15298839	×	√	√	×	Ame_R015259	*NMT2*	N-myristoyltransferase 2
ChrNew10_71840661	×	√	√	×	Ame_R002454	*PDZRN3*	PDZ domain containing ring finger 3
ChrNew12_66366023	×	√	√	×	Ame_R012719	*PVRL2*	poliovirus receptor-related 2 (herpesvirus entry mediator B)
ChrNew12_25347974	×	√	√	×	Ame_R009657	*PKD1L1*	polycystic kidney disease 1 like 1
ChrNew17_5433179	×	√	√	×	Ame_R007147	*PRMT2*	protein arginine methyltransferase 2
ChrNew11_10852223	×	√	√	×	Ame_R012186	*PDIA4*	protein disulfide isomerase family A, member 4
ChrNew19_12534776	×	√	√	×	Ame_R012983	*PRUNE2*	prune homolog 2 (Drosophila)
ChrNew23_77532352	×	√	√	×	Ame_R020818	*AE000661.1*	Putative uncharacterized protein
ChrNew21_98547473	×	√	√	×	Ame_R020250	*hypothetical protein*	Putative uncharacterized protein
ChrNew12_22881229	×	√	√	×	Ame_R012845	*RALGDS*	ral guanine nucleotide dissociation stimulator
ChrNew13_48596016	×	√	√	×	Ame_R007491	*RGS16*	regulator of G-protein signaling 16
ChrNew17_19453603	×	√	√	×	Ame_R018602	*RPTN*	repetin
ChrNew17_19454038	×	√	√	×	Ame_R018602	*RPTN*	repetin
ChrNew22_80577727	×	√	√	×	Ame_R020621	*ARHGAP17*	Rho GTPase activating protein 17
ChrNew22_80577843	×	√	√	×	Ame_R020621	*ARHGAP17*	Rho GTPase activating protein 17
ChrNew22_80577613	×	√	√	×	Ame_R020621	*ARHGAP17*	Rho GTPase activating protein 17
ChrNew22_80577807	×	√	√	×	Ame_R020621	*ARHGAP17*	Rho GTPase activating protein 17
ChrNew22_80577823	×	√	√	×	Ame_R020621	*ARHGAP17*	Rho GTPase activating protein 17
ChrNew21_87544710	×	√	√	×	Ame_R018610	*ARHGEF11*	Rho guanine nucleotide exchange factor (GEF) 11
ChrNew21_87541081	×	√	√	×	Ame_R018610	*ARHGEF11*	Rho guanine nucleotide exchange factor (GEF) 11
ChrNew21_87552818	×	√	√	×	Ame_R018610	*ARHGEF11*	Rho guanine nucleotide exchange factor (GEF) 11
ChrNew21_87541102	×	√	√	×	Ame_R018610	*ARHGEF11*	Rho guanine nucleotide exchange factor (GEF) 11
ChrNew21_87544200	×	√	√	×	Ame_R018610	*ARHGEF11*	Rho guanine nucleotide exchange factor (GEF) 11

Nature Genetics doi:10/1038/ng.2494

ChrNew23_36878512	×	√	√	×	Ame_R020436	*RNASE6*	ribonuclease
ChrNew13_50112420	×	√	√	×	Ame_R007062	*RPL29*	ribosomal protein L29
ChrNew10_25855578	×	√	√	×	Ame_R013604	*RRBP1*	ribosome binding protein 1 homolog 180kDa (dog)
ChrNew6_99038396	×	√	√	×	Ame_R003007	*SET*	SET nuclear oncogene
ChrNew15_60573147	×	√	√	×	Ame_R006868	*SIRPA*	signal-regulatory protein alpha
ChrNew20_13150037	×	√	√	×	Ame_R008243	*hypothetical protein*	similar to mucin 6, gastric
ChrNew20_13149863	×	√	√	×	Ame_R008243	*hypothetical protein*	similar to mucin 6, gastric
ChrNew23_53763903	×	√	√	×	Ame_R020064	*SPRR1A*	small proline-rich protein 1A
ChrNew20_65526042	×	√	√	×	Ame_R019829	*SLC7A1*	solute carrier family 7 (cationic amino acid transporter, y+ system), member 1
ChrNew20_86672563	×	√	√	×	Ame_R007694	*SON*	SON DNA binding protein
ChrNew20_86672578	×	√	√	×	Ame_R007694	*SON*	SON DNA binding protein
ChrNew5_23212698	×	√	√	×	Ame_R001918	*SPTB*	spectrin, beta, erythrocytic
ChrNew9_48223885	×	√	√	×	Ame_R001041	*FUSIP1*	splicing factor, arginine/serine-rich 13A
ChrNew9_40230672	×	√	√	×	Ame_R010108	*SYNC1*	syncoilin, intermediate filament protein
ChrNew14_2813494	×	√	√	×	Ame_R016738	*TNN*	tenascin-N-like
ChrNew14_2813479	×	√	√	×	Ame_R016738	*TNN*	tenascin-N-like
ChrNew18_17825222	×	√	√	×	Ame_R013713	*TOMM22*	translocase of outer mitochondrial membrane 22 homolog (yeast)
ChrNew7_87437871	×	√	√	×	Ame_R012475	*TMEM178*	transmembrane protein 178
ChrNew13_3285567	×	√	√	×	Ame_R017757	*TRIM33*	tripartite motif-containing 33
ChrNew13_3310538	×	√	√	×	Ame_R017757	*TRIM33*	tripartite motif-containing 33
ChrNew21_93142749	×	√	√	×	Ame_R019085	*TRIM26*	Tripartite motif-containing protein 26 (Zinc finger protein 173)(Acid finger protein)(AFP)(RING finger protein 95)
ChrNew21_93142840	×	√	√	×	Ame_R019085	*TRIM26*	Tripartite motif-containing protein 26 (Zinc finger protein 173)(Acid finger protein)(AFP)(RING finger protein 95)
ChrNew10_84816959	×	√	√	×	Ame_R011740	*TROVE2*	TROVE domain family, member 2
ChrNew10_84822694	×	√	√	×	Ame_R011740	*TROVE2*	TROVE domain family, member 2
ChrNew19_95081940	×	√	√	×	Ame_R013316	*TUBA1*	tubulin, alpha 4a
ChrNew19_95082612	×	√	√	×	Ame_R013316	*TUBA1*	tubulin, alpha 4a
ChrNew3_70715967	×	√	√	×	Ame_R016229	*GALNT2*	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)
ChrNew19_41616975	×	√	√	×	Ame_R010233	*C14orf138*	Uncharacterized protein C14orf138

Nature Genetics doi:10/1038/ng.2494

ChrNew12_64939391	×	√	√	×	Ame_R009586	*C19orf52*	Uncharacterized protein C19orf52
ChrNew12_64939266	×	√	√	×	Ame_R009586	*C19orf52*	Uncharacterized protein C19orf52
ChrNew4_7757913	×	√	√	×	Ame_R001530	*C2orf16*	Uncharacterized protein C2orf16
ChrNew18_38551585	×	√	√	×	Ame_R019040	*C6orf10*	Uncharacterized protein C6orf10
ChrNew11_45782302	×	√	√	×	Ame_R014369	*ZNF236*	zinc finger protein 236
ChrNew9_98673623	×	√	√	×	Ame_R016695	*ZNF267*	zinc finger protein 267
ChrNew23_63153166	×	√	√	×	Ame_R020634	*ZNF324*	zinc finger protein 324
ChrNew23_63153230	×	√	√	×	Ame_R020634	*ZNF324*	zinc finger protein 324
ChrNew5_99189625	×	√	√	×	Ame_R005813	*ZNF473*	zinc finger protein 473
ChrNew10_69026463	×	√	√	×	Ame_R019976	*ZNF695*	zinc finger protein 695
ChrNew10_69025951	×	√	√	×	Ame_R019976	*ZNF695*	zinc finger protein 695
ChrNew23_45901521	×	√	√	×	Ame_R020472	*ZNF720*	zinc finger protein 720
ChrNew23_45855414	×	√	√	×	Ame_R020657	*ZNF729*	zinc finger protein 729
ChrNew23_53340482	×	√	√	×	Ame_R020548	*ZNF840*	zinc finger protein 840 (pseudogene)
ChrNew23_53340506	×	√	√	×	Ame_R020548	*ZNF840*	zinc finger protein 840 (pseudogene)

Note: The symbol " √ " means that SNPs were detected to be under directional selection by this method. *Receptor genes whose ligands have been identified by previous literature[3]; †Odors have been detected in scent markers secreted by giant pandas[4].

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 8. SNPs under directional selection in the analysis of MIN and QXL pandas.

SNP ID	Bayescan (q<0.1)	Finite island model of Arlequin (q<0.05)	Global F_{ST} (top0.01)	DAF*	Gene ID	Gene name	Gene description
ChrNew1_2522278	√	√	√	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor, Class I, family 51
ChrNew1_2522287	√	√	√	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor, Class I, family 51
ChrNew1_2522308	√	√	√	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor, Class I, family 51
ChrNew1_2522266	√	√	√	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor, Class I, family 51
ChrNew1_2522260	√	√	√	-	Ame_R019625	*OR51L1, Olr61, MOR11-1*	Olfactory receptor, Class I, family 51
ChrNew23_36714564	×	√	√	MIN	Ame_R020324	*OR51I2,MOR19-1,Olr109*	Olfactory receptor, Class I, family 51
ChrNew21_63109974	×	√	√	-	Ame_R020177	*OR7A17*	Olfactory receptor, Class II, family 7
ChrNew21_71942033	×	√	√	-	Ame_R020338	*OR7A5*	Olfactory receptor, Class II, family 7
ChrNew21_71942069	×	√	√	-	Ame_R020338	*OR7A5*	Olfactory receptor, Class II, family 7
ChrNew20_3655972	√	√	√	-	Ame_R020292	*OR7A17*	Olfactory receptor, Class II, family 7
ChrNew20_3655968	×	√	√	-	Ame_R020292	*OR7A17*	Olfactory receptor, Class II, family 7
ChrNew20_3665430	×	√	√	-	Ame_R020293	*OR7A17*	Olfactory receptor, Class II, family 7
ChrNew11_74741902	×	√	√	-	Ame_R018383	*OR13F1,Olfr275,MOR262-2 cOR10H10,OR10H4, Olfr55,*	Olfactory receptor, Class II, family 13
ChrNew19_78502004	√	√	√	-	Ame_R019690	*MOR267-17*	Olfactory receptor, Class II, family 10
ChrNew16_44308776	×	√	√	QXL	Ame_R014491	*AGPAT3*	1-acylglycerol-3-phosphate O-acyltransferase 3;lysophosphatidic acid acyltransferase / lysophosphatidylinositol acyltransferase
ChrNew20_27845380	√	×	√	MIN	Ame_R004672	*ARL5C*	ADP-ribosylation factor-like 5C
ChrNew19_9274910	√	√	√	QXL	Ame_R003467	*AKR7A3*	aldo-keto reductase family 7 (aflatoxin aldehyde reductase)
ChrNew17_93852606	√	√	√	QXL	Ame_R015154	*ABCA3*	ATP-binding cassette sub-family A member 3-like
ChrNew5_99480305	√	√	√	MIN	Ame_R005829	*BCL2L12*	BCL2-like 12 (proline rich)
ChrNew15_52268762	√	√	√	QXL	Ame_R007142	*FAM44A*	biorientation of chromosomes in cell division 1-like
ChrNew22_94909015	×	√	√	QXL	Ame_R019860	*FAM48A*	family with sequence similarity 48, member A
ChrNew19_6402612	×	√	√	MIN	Ame_R016862	*BRD4*	bromodomain containing 4
ChrNew10_84572823	√	×	√	QXL	Ame_R011736	*CDC73*	cell division cycle 73, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)
ChrNew1_69277347	√	√	√	QXL	Ame_R017604	*ETS2*	c-ets proto-oncogene protein
ChrNew15_8243255	√	√	√	MIN	Ame_R015202	*PDE3*	cGMP-inhibited 3',5'-cyclic phosphodiesterase
ChrNew4_90030129	√	√	√	QXL	Ame_R000044	*CNGB1*	cyclic nucleotide gated channel beta 1
ChrNew5_9845804	×	√	√	QXL	Ame_R001723	*DMTF1*	cyclin D binding myb-like transcription factor 1
ChrNew16_18944219	×	√	√	QXL	Ame_R014377	*DNAH11*	Dynein heavy chain family protein

Nature Genetics doi:10/1038/ng.2494

ChrNew20_24974976	×	√	√	QXL	Ame_R014665	*EEF2K*	eukaryotic elongation factor-2 kinase
ChrNew23_3683377	×	√	√	QXL	Ame_R019814	*ECSIT*	evolutionarily conserved signaling intermediate in Toll pathways
ChrNew22_68822213	×	√	√	MIN	Ame_R018140	*GSDMA*	gasdermin A
ChrNew10_90923274	×	√	√	QXL	Ame_R003100	*GUCY2D*	guanylate cyclase 2D;olfactory guanylyl cyclase GC-D-like
ChrNew18_65371945	×	√	√	MIN	Ame_R004743	*HSP90AA5P*	heat shock 90kDa protein; molecular chaperone HtpG
ChrNew9_22843718	√	√	√	MIN	Ame_R006370	*HGSNAT*	heparan-alpha-glucosaminide N-acetyltransferase
ChrNew14_91973667	√	√	√	-	Ame_R002946	*hypothetical gene*	hypothetical protein
ChrNew13_2157729	√	×	√	MIN	Ame_R014805	*hypothetical gene*	hypothetical protein
ChrNew19_6513806	√	√	√	QXL	Ame_R016866	*ILVBL*	ilvB (bacterial acetolactate synthase)-like; acetolactate synthase-like protein
ChrNew2_29986744	×	√	√	MIN	Ame_R000203	*LHCGR*	lutropin-choriogonadotropic hormone receptor-like
ChrNew23_54816570	√	√	√	MIN	Ame_R020367	*MAN2B2*	mannosidase, alpha, class 2B, member 2;epididymis-specific alpha-mannosidase
ChrNew17_70454787	√	√	√	MIN	Ame_R011997	*NLE1*	notchless homolog 1
ChrNew8_82175410	×	√	√	QXL	Ame_R011009	*LPIN1*	phosphatidate phosphatase LPIN
ChrNew15_45440354	√	√	√	QXL	Ame_R003234	*PKHD1*	polycystic kidney and hepatic disease 1 (autosomal recessive)
ChrNew16_69297617	√	√	√	MIN	Ame_R012549	*RPL23*	ribosomal protein L23-like
ChrNew16_69297636	√	√	√	MIN	Ame_R012549	*RPL23*	ribosomal protein L23-like
ChrNew16_69297602	√	√	√	MIN	Ame_R012549	*RPL23*	ribosomal protein L23-like
ChrNew19_6627632	√	√	√	QXL	Ame_R016871	*SLC1A6*	solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6
ChrNew2_84021829	×	√	√	MIN	Ame_R000965	*STARD9*	StAR-related lipid transfer (START) domain containing 9
ChrNew14_78193702	√	√	√	QXL	Ame_R010587	*THNSL2*	threonine synthase-like 2
ChrNew12_35488634	×	√	√	-	Ame_R015755	*TNNI1*	troponin I, slow skeletal muscle
ChrNew22_11522281	√	√	√	QXL	Ame_R019134	*UHRF1BP1L*	UHRF1 binding protein 1-like
ChrNew2_83122683	√	√	√	MIN	Ame_R000939	*WDR76*	WD repeat domain 76
ChrNew8_87044960	×	√	√	MIN	Ame_R010469	*ZNF804A*	zinc finger protein 804A

Note: The symbol " √ " means that SNPs were detected to be under directional selection by this method. *The population shown in the column of "DAF" is the population that harbors higher derived allele frequency at this locus.

Nature Genetics doi:10/1038/ng.2494

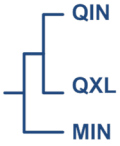
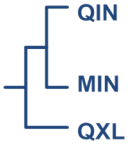
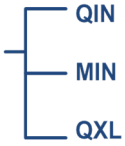
Supplementary Table 9. Statistics for the first four eigenvalues in PCA of wild pandas.

Eigenvectors	Eigenvalue	*Twstat*	*P*-value [a]
1	7.284983	16.627	*1.64E-21*
2	4.000275	6.789	*2.42E-07*
3	3.744156	7.833	*1.29E-08*
4	3.059057	0.137	0.144585

Note: [a]The italic numbers in the column "*P*-value" are significant based on Tracy-Widom test[5] (*P*<0.05).

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 10. Comparison of four demographic models using ∂a∂i.

Alternative models	QIN diverged first; then MIN, QXL	MIN diverged first; then QIN, QXL	QXL diverged first; then QIN, MIN	QIN, MIN, QXL emerged simultaneously
Patterns				
Log-likelihood	-17229.45	-18969.54	-22519.28	-18929.8

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 11. The number of outlier SNPs detected between QIN and non-QIN populations using different analysis methods.

	Arlequin (q < 0.05)			Bayescan (q < 0.1)	Global F_{ST} (top 1%)
	Hierarchical island model		Finite island model		
	F_{ST}	F_{CT}	F_{ST}		
Directional	164	50	125	25	123
Balancing	386	57	289	0	-

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 12. The number of outlier SNPs detected between MIN and QXL populations using different analysis methods.

	Finite Island Model in Arlequin (q < 0.05)	Bayescan (q < 0.1)	Global F_{ST} (top 1%)
Directional	176	30	65
Balancing	759	0	-

Nature Genetics doi:10/1038/ng.2494

Supplementary Table 13. Annotation of the loci under selections using the KEGG pathway database.

Category of KEGG pathway		Gene number		
		QIN and non-QIN		MIN and QXL
		Directional	Balancing	Directional
Cellular Processes	Cell Communication	7	5	0
	Cell Growth and Death	2	0	1
	Cell Motility	3	1	0
	Transport and Catabolism	2	2	1
Environmental Information Processing	Membrane Transport	0	0	1
	Signal Transduction	7	2	3
	Signaling Molecules and Interaction	8	4	1
Genetic Information Processing	Folding, Sorting and Degradation	2	3	2
	Replication and Repair	0	1	1
	Transcription	0	1	0
	Translation	0	3	1
Metabolism	Amino Acid Metabolism	1	0	0
	Carbohydrate Metabolism	1	1	0
	Energy Metabolism	1	0	0
	Glycan Biosynthesis and Metabolism	1	3	2
	Lipid Metabolism	3	1	1
	Metabolism of Cofactors and Vitamins	1	0	0
	Metabolism of Other Amino Acids	3	0	0
	Nucleotide Metabolism	2	1	2
	Xenobiotics Biodegradation and Metabolism	3	0	0
Organismal Systems	Circulatory System	0	1	0
	Development	2	5	0
	Digestive System	1	3	0
	Endocrine System	1	0	2
	Environmental Adaptation	2	1	1
	Excretory System	1	1	0
	Immune System	7	1	1
	Nervous System	0	0	1
	Sensory System	11	23	10

Nature Genetics doi:10/1038/ng.2494

178 **Supplementary Table 14.** Annotation of the loci under selections using the KEGG Brite

179 function database.

		Gene number		
		QIN and non-QIN		MIN and QXL
Category of KEGG Brite protein family		Directional	Balancing	Directional
Cellular Processes	Cell adhesion molecules (CAMs)	3	3	0
	CAM ligands	7	3	0
	Cellular antigens	3	18	0
	Cytokine receptors	0	2	0
	Enzyme-linked receptors	0	1	0
	G protein-coupled receptors	12	23	9
	Glycan binding proteins	1	1	0
	GTP-binding proteins	0	0	1
	Ion channels	0	2	1
	Nuclear receptors	3	0	0
Environmental Information Processing	Cytoskeleton proteins	5	13	1
	Solute carrier family	3	1	1
	Transporters	0	0	1
Genetic Information Processing	Chaperones and folding catalysts	4	1	1
	Chromosome	4	4	1
	DNA repair and recombination proteins	1	1	1
	DNA replication proteins	1	2	0
	Proteasome	1	0	1
	Ribosome	0	2	1
	Ribosome biogenesis	0	1	0
	Spliceosome	0	4	0
	Transcription factors	7	9	2
	Transcription machinery	0	1	1
	Transfer RNA biogenesis	0	1	0
	Translation factors	1	0	0
	Ubiquitin system	2	2	1
Metabolism	Enzymes	21	20	9
	Glycosyltransferases	1	2	0
	Lipid biosynthesis proteins	2	0	1
	Peptidases	4	2	0
	Protein kinases	6	5	1
	Proteoglycans	1	1	0

180

Nature Genetics doi:10/1038/ng.2494

Supplementary Note

Additional sampling information

We also sequenced 14 captive admixed pandas (**Supplementary Table 1**) to improve the quality of SNP calling for the studied wild pandas (see section of **Dataset used)**. The fourteen admixed individuals were derived by crossing two different geographic populations: three from Qinling × Qionglai, three from Qionglai × Liangshan, four from Qionglai × Minshan, and four from Minshan × Liangshan.

Population SNP calling

Sequence Alignment/Map (SAM) format files were first imported to Samtools[6] for sorting and merging. 'rmdup' command was used to remove potential PCR duplicates: if multiple read pairs have identical external coordinates, only the pair with highest mapping quality was retained. We then used a method previously described[7] to call the SNPs in the following way: calculated genotype likelihoods from reads for each individual at each genomic location, and estimated the allele frequencies in the sample using a Bayesian approach, which was applied jointly to all individuals in a population. The method improves the population based SNP calling quality by calibrating the biases and possible false positives caused by low-coverage sequencing. We first estimated allele frequencies for each site (Site Frequency Spectrum) following the report[7], and then selected all sites with $p_0 \leq 0.01$ to make sure the sites with a probability $\geq 99\%$ of being SNPs.

Datasets used

1) We obtained an average of 10.5 Gb mapped reads for each wild panda and 9.8 Gb for each admixed panda (See the table of **Sequencing data and mapping summary**). According to the approach described above, we performed population SNP calling for 48 individuals (34 wild and 14 admixed pandas). After the strict filtering of coverage depth (≥ 50 and ≤ 300),

Nature Genetics doi:10/1038/ng.2494

copy number (< 1.1), and rank sum test (RST, $P > 0.001$), a total of 16,085,433 high quality SNPs were identified. We then performed SNP calling using the same approach for 34 wild pandas and filtered the SNP set with coverage depth (≥ 50), copy number (< 1.1), and RST ($P > 0.001$). Because large sample size is expected to improve the accuracy of population SNP inference, we extracted SNPs that were in the intersection of two datasets of 34 wild individuals and 48 individuals (wild plus admixed pandas). Finally, we identified 13,020,055 high quality SNPs for the wild pandas (See the figure of **SNP distribution for the 34 wild pandas**). For the genetic populations defined in this study, we also performed SNP calling for each wild population and retained those loci that also appeared in the data set of all wild individuals (See the table named **Number of SNPs in the whole genome and gene region**). The SNPs in each population were used for the analyses of genetic diversity, population differentiation, recent population history, and local adaptation.

2) The genome sequence of the giant panda[8] from GigaScience (Supplementary URLs) was used as reference for SNP calling and the dataset for demographic history inference with the PSMC approach.

3) Considering that X chromosomes differ from autosomes in inferring population genetic information for being hemizygous in males, with reduced genetic variation, at ¾ of effective population size of autosomes, with lower mutation rate, and the genes on X chromosome being under rapid evolution[9,10], we excluded SNPs located in X chromosomal regions and kept autosomal SNPs for further analyses of the giant panda. We aligned the scaffolds of the panda genome to the sex chromosomes of human (genome hg18) using BLASTZ[11]. If a scaffold shared more than 60% of sequence similarity with the aligned human sex chromosomes, we excluded the scaffolds in subsequent analyses.

Nature Genetics doi:10/1038/ng.2494

Figure: SNP distribution for the 34 wild pandas. The standard normal distribution and the real distribution are showed in blue line and red vertical bars, respectively.

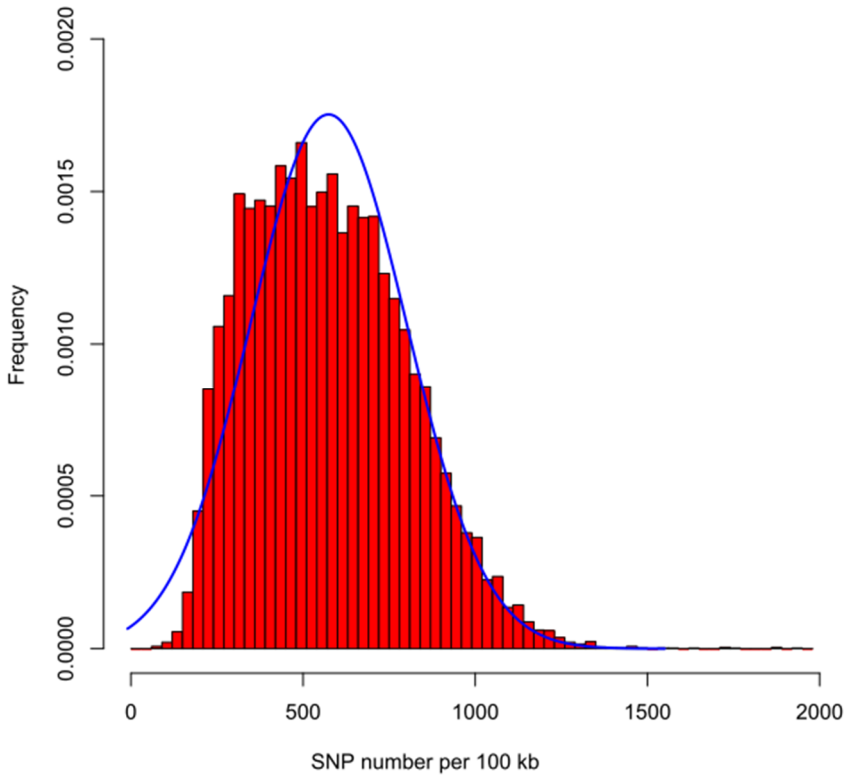


Table: Sequencing data and mapping summary

	Sample size	Raw data (Gb)	Mapped bases (Gb)	Mapped rate (%)	Mismatch rate (%)	%Genome coverage	Effective depth (×)
Wild (mean ± s.d.)	34	13.98 ± 1.34	10.46 ± 2.15	74.32 ± 10.7	0.44 ± 0.06	91.48 ± 4.24	4.65 ± 0.96
Admixed (mean ± s.d.)	14	13.12 ± 1.87	9.76 ± 1.64	74.38 ± 6.63	0.46 ± 0.05	91.43 ± 3.91	4.34 ± 0.73
Total (mean ± s.d.)	48	13.73 ± 1.54	10.25 ± 2.02	74.34 ± 9.6	0.45 ± 0.06	91.46 ± 4.11	4.56 ± 0.90

Note: s.d.= standard deviation.

Table: Number of SNPs in the whole genome and gene region

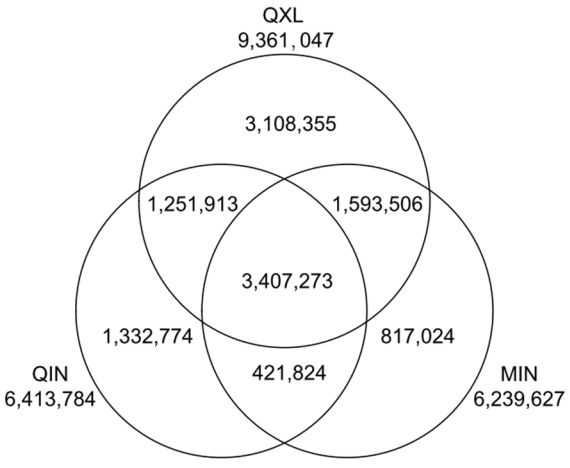
Population	Sample size	Whole genome	Gene region	CDS	Intron
QIN	8	6,413,784	1,161,726	28,957	1,132,769
MIN	7	6,239,627	1,105,020	26,695	1,078,325
QXL	19	9,361,047	1,747,663	48,677	1,698,986
Total	34	13,020,055	2,504,468	75,799	2,428,669

Nature Genetics doi:10/1038/ng.2494

234 **Genetic diversity**

235 Among three identified panda populations, the total number of SNPs was 6,413,784 for QIN,
236 6,239,627 for MIN, and 9,361,047 for QXL, with 3,407,273 shared among populations (The
237 table of **Number of SNPs in the whole genome and gene region** and the figure named **A**
238 **Venn diagram showing the number of unique and shared SNPs among three panda**
239 **populations**). The small proportion of shared SNPs (26%) among populations indicated that
240 the populations might have diverged long ago. To assess the effects of polymorphism in the
241 genes we predicted large-effect SNPs of 34 individuals: 56 were expected to alter initiation
242 methionine residues, six inferred to disrupt stop codons, 990 predicted to induce premature
243 stop codons, and 1,108 expected to disrupt splicing donor or acceptor sites (Table of **Number**
244 **of large-effect SNPs for three panda populations**). Among different gene families, the
245 distribution of large-effect SNPs was not random, with olfactory receptor gene family
246 possessing the highest level (0.74 SNP/10 kb, Figure of **Annotation of large-effect SNPs for**
247 **the wild pandas**).

248 **Figure: A Venn diagram showing the number of unique and shared SNPs among three**
249 **panda populations** (QXL, QIN and MIN). The numbers outside the circles are the total
250 number of SNPs for each population. The numbers in circles show unique SNPs for each
251 population, shared SNPs between any two populations, or among three populations.



252

Nature Genetics doi:10/1038/ng.2494

Figure: Annotation of large-effect SNPs for the wild pandas. Large-effect SNPs were annotated according to Interpro gene families (Supplementary URLs). The large-effect SNP number of every gene family was calculated. Only those families with more than 30 SNPs that provide sufficient statistical power are retained for the chi-square test. The observed class is the large-effect to non-large-effect ratio in each family, while the expected one is the large-effect to non-large-effect ratio of total families. The gene families with significantly abundant large-effect SNPs (χ^2 test, $P < 0.01$) were shown. The number of large-effect SNPs per 10 kb was in proportion to the bar length.

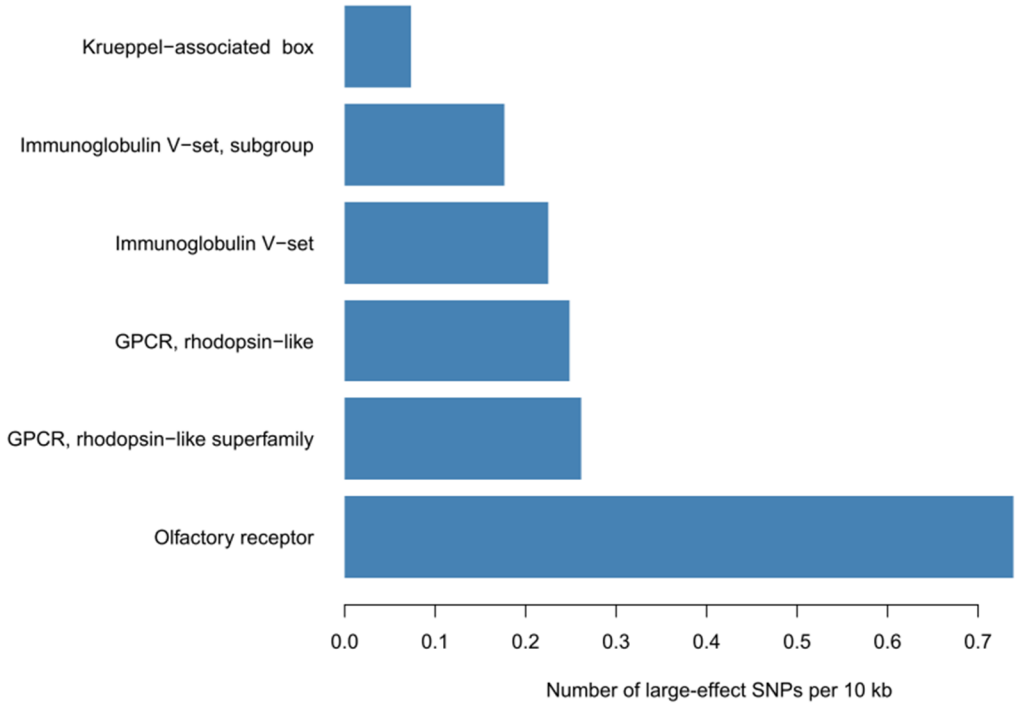


Table: Number of large-effect SNPs for three panda populations

	Sample size	Disrupt_start codon	Disrupt_stop codon	Introduce_stop codon	Alter_splice site	Total number
QIN	8	18	2	266	310	596
MIN	7	13	2	239	266	520
QXL	19	37	4	515	618	1174
Total	34	56	6	990	1108	2160

Nature Genetics doi:10/1038/ng.2494

Estimation of divergence time and sequence between the giant panda and polar bear

We used TreeFam[12,13] to identify concatenated coding DNA sequences (CDS) of single-copy orthologous genes from nine species (i.e. giant panda *Ailuropoda melanoleuca*, polar bear *Ursus maritimus*, human *Homo sapiens*, dog *Canis lupus familiaris*, cow *Bos taurus*, cat *Felis catus*, pig *Sus scrofa*, rhesus monkey *Macaca mulatta*, chimpanzee *Pan troglodytes*). We performed multiple alignments of peptide sequences for each gene family using MUSCLE software[14], and converted the peptide alignments into CDS alignments using a *PERL* script. Five fossil calibration times (Table of **Five fossil calibration times used for estimation of the divergence time between giant panda and polar bear**) were incorporated to estimate divergence time based on the concatenated CDSs with more than 80% of non-gap sites using PAML mcmctree program[15,16] under the "Correlated molecular clock" and "JC69" model. Category number of discrete gamma distribution among sites was set to 4, and "usedata" was set to 0 while "alpha" was calculated by PhyML[17] from the same sequence set. After a burn-in of 10,000 iterations, the Markov chain Monte Carlo process was run with sample numbers and sample frequency being set to 100,000 and 2, respectively. Acceptance proportion was set as from 0.15 to 0.7. Other parameters were set to their default values. Two independent runs were performed and the results were the same. The divergence time was estimated to be 16.4 MYA (95% confidence interval: 10.5 - 22.8 MYA) between the polar bear and giant panda.

In addition, we estimated the sequence divergence between the panda and polar bear to be 3.53% based on the autosomal syntenic regions identified by LASTZ (Supplementary URLs).

Table: Five fossil calibration times used for estimation of the divergence time between giant panda and polar bear.

Species 1 - Species 2	Divergence time (MYA)
Human - Chimpanzee	(~ - 10)[a]
Human - Rhesus monkey	(23 - 33.9)[a]

Nature Genetics doi:10/1038/ng.2494

Dog - Cat	(50 - 63)[b]
Pig - Cow	(48.3 - 53.5)[a]
Human - Dog	(95.3 - 113)[a]

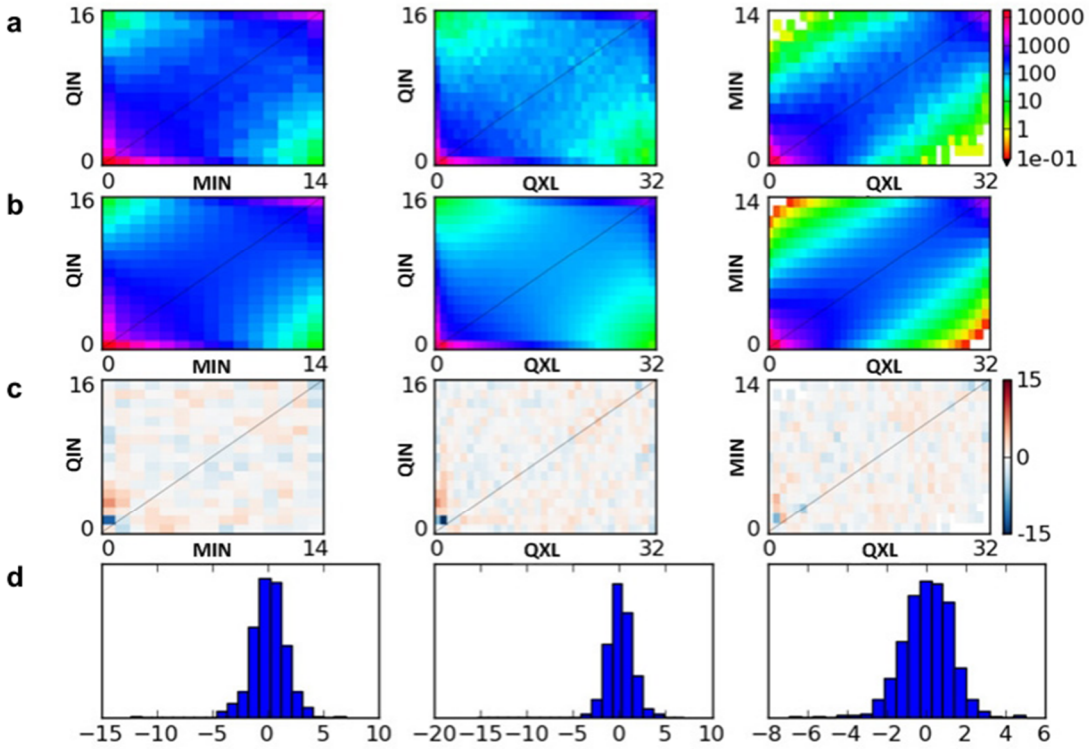
Note: [a]the times estimated in a previous report on tree of life[18]; [b]the time was derived from the divergence date between feliform and caniform[19].

Comparison of four alternative divergence models simulated by ∂a∂i

For three identified panda populations (QIN, MIN, and QXL), four divergence models were considered: 1) (QIN (MIN, QXL)), that is, the ancestral population split into QIN and non-QIN population and then the non-QIN population separated into MIN and QXL; 2) (MIN (QIN, QXL)); 3) (QXL(QIN, MIN)); 4) the ancestral population split into three populations simultaneously. We simulated the four models with the same dataset under the three-population model in ∂a∂i[20] independently and calculated the likelihoods to compare the fit of each model. As shown in **Supplementary Table 10**, the first model gained the maximum log-likelihood value and was thus chosen as the optimal one for the giant panda. The simulation results of the optimal model were shown in the figure named **Comparisons of allele frequency spectra (AFS) between the model and real data of three panda populations using ∂a∂i**.

Figure: Comparisons of allele frequency spectra (AFS) between the model and real data of three panda populations using ∂a∂i. (**a**) Marginal AFS of the real data for each pair of populations. (**b**) AFS of the maximum-likelihood model simulated based on the real data. The residuals between the model and real data are shown in heat maps (**c**) and bar graphs (**d**).

Nature Genetics doi:10/1038/ng.2494



304

305 **Human activities during the past six thousand years in the habitats of giant panda**

306 The giant pandas occupy the northern and western mountains circling the Sichuan Basin in
307 China. As shown in **Supplementary Fig. 6**, three routes went across the mountains and
308 connected ancient Shu Kingdom (i. e. Sichuan) with the outside, especially the Central Plains
309 in China[21]. The three routes were replaced one by another temporally. Route I in the west,
310 where human activities could be traced back to as early as ~6 KYA, was the earliest and
311 longest. Route II, which was shorter but more difficult to build than Route I, was in use
312 during Shang and Zhou Dynasties (~1600 - 256 BC)[21]. Route III, built across the high
313 mountains (e. g. Qinling) by Qin Emperor in his preparation for unifying the whole China
314 (around 400 BC)[21], was the most recent, which is also highly utilized today.

315 **Chinese historical records and literatures cited**

316 1) During the early stage of the Spring-and-Autumn Period (770~486 BC), Chinese people

Nature Genetics doi:10/1038/ng.2494

317 began to use cast iron tools[22], domesticated animals to pull plows[22], established large-scale
318 projects to harness rivers[23], and developed water conservation projects[23].
319 2) Deforestation caused by construction of palaces in the Qin Dynasty was described in
320 "*Rhapsody on E-Pang Palace*", a well-known work written by Mu Du (803 - 852 AD) in the
321 Tang dynasty.
322 3) As mentioned in "*Rhapsody on the Imperial Park*" (*Shanglin Fu*), a masterpiece of
323 Xiangru Sima (179 - 118 BC) who was a prominent essayist and poet in the West Han
324 Dynasty, the giant panda was raised in the royal garden (Shanglin Garden, ~138 BC - 20 AD)
325 near the imperial capital (Chang'an, i.e. nowadays Xi'an). The Shanglin Garden was also
326 recorded in the *Han Jiu Yi* written by Hong Wei (about 25 - 57 AD).
327 4) As recorded in the chapters of "Biographies of Junji Hou and Wanche Xue" of the *Old*
328 *Book of Tang*, the panda skins were used to reward the distinguished ministers.
329 5) A panda skull was excavated in a royal mausoleum of the Han Dynasty (155 BC)[24].

330 **Derived allele frequency (DAF) test**

331 We used F_{ST}-based methods to measure the level of population differentiation of one locus
332 between two populations and detected the SNP outliers under selections. The frequency of
333 loci under positive selections varied among different panda populations. DAF was used to
334 estimate the extent of deviation in populations from ancestral state and positive selections
335 usually result in the increase of DAF in a population[25,26]. As for those SNP loci detected to be
336 under the positive selection, we compared their DAF values between two populations to
337 localize selections to certain populations. A higher frequency of derived alleles (for given loci)
338 in one population might indicated that it was adapted to that local environment.

Nature Genetics doi:10/1038/ng.2494

Supplementary URLs

Genomic data from the giant panda, http://dx.doi.org/10.5524/100004; Interpro gene familie, http://www.ebi.ac.uk/Tools/pfa/iprscan/; LASTZ, http://www.bx.psu.edu/miller_lab/.

Supplementary References

1. Zhan, X. *et al.* Molecular evidence for Pleistocene refugia at the eastern edge of the Tibetan Plateau. *Mol. Ecol.* **20**, 3014-3026 (2011).
2. Meyerhof, W. *et al.* The molecular receptive ranges of human TAS2R bitter taste receptors. *Chem. Senses* **35**, 157-70 (2010).
3. Saito, H. *et al.* Odor coding by a mammalian receptor repertoire. *Sci. Signal* **2**, ra9 (2009).
4. Hagey, L. & MacDonald, E. Chemical cues identify gender and individuality in giant pandas (Ailuropoda melanoleuca). *J. Chem. Ecol.* **29**, 1479-1488 (2003).
5. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
6. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
7. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75-78 (2010).
8. Li, R. *et al.* The complete genome sequence of the giant panda. *Nature* **463**, 311-317 (2010).
9. Hartl, D.L. & Clark, A.G. *Principles of Population Genetics*, 495 (Sinauer Associates, Inc. Publishers, Sunderland, 2007).
10. Nachman, M.W. & Crowell, S.L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297-304 (2000).
11. Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103-107 (2003).

Nature Genetics doi:10/1038/ng.2494

12. Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic. Acids Res.* **34**, D572-D580 (2006).
13. Ruan, J. *et al.* TreeFam: 2008 Update. *Nucleic. Acids Res.* **36**, D735-D740 (2008).
14. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic. Acids Res.* **32**, 1792-1797 (2004).
15. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-1591 (2007).
16. Yang, Z. & Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212-226 (2006).
17. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321 (2010).
18. Benton, M.J. & Donoghue, P.C. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26-53 (2007).
19. Springer, M.S., Murphy, W.J., Eizirik, E. & O'Brien, S.J. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* **100**, 1056-1061 (2003).
20. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
21. Ren, N. *Illustrations and Annotations of Huayang Guo Zhi*. (Shanghai Ancient Books Publishing House, Shanghai, 1987).
22. Meyer, M.W. *China: A Concise History* (Rowman & Littlefield Publishers, Lanham, 1994).
23. Needham, J., Wang, L. & Lu, G. Civil Engineering and Nautics. in *Science and*

Nature Genetics doi:10/1038/ng.2494

Civilization in China (ed. Needham, J.) **4**, Part 3 (Cambridge University Press, Cambridge, 1971)

24. Wang, X. The initial Archaeological excavations of burial pit of Nanling mausoleum of Han Dynasty and related questions about the giant panda skull and rhino skeleton. *Wenwu* **11**, 24-29 (1981).

25. Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918 (2007).

26. Sabeti, P.C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614-1620 (2006).

Nature Genetics doi:10/1038/ng.2494