Supporting Online Material and Supplementary Information: Altitude adaptation in Tibet caused by introgression of Denisovan-like DNA

Emilia Huerta-Sánchez[1,2,3*], Xin Jin[1, 4*], Asan[1,5,6*], Zhuoma Bianba[7*], Benjamin Peter[2], Nicolas Vinckenbosch[2], Yu Liang[1,5,6], Xin Yi[1,5,6], Mingze He[1,8], Mehmet Somel[9], Peixiang Ni[1], Bo Wang[1], Xiaohua Ou[1], Huasang[1], Jiangbai Luosang[1], Zha Xi Ping Cuo[10], Guoyi Gao[11], Ye Yin[1], Wei Wang[1], Xiuqing Zhang[1,12,13], Xun Xu[1], Huanming Yang[1,14,15], Yingrui Li[1], Jian Wang[1,15#], Jun Wang[1,14,16,17,18#], Rasmus Nielsen[1,2,19,20#]

[1]BGI-Shenzhen, Shenzhen, 518083, China

[2] Department of Integrative Biology, University of California, Berkeley, CA

[3]School of Natural Sciences, University of California, Merced, CA

[4]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, 510006, China

[5]Binhai Institute of Gene Technology, BGI-Tianjin, Tianjin, 300308, China

[6]Tianjin Medical Genomics Engineering Center, BGI-Tianjin, Tianjin, 300308, China

[7] The People's Hospital of Lhasa, Lhasa, 850000, China

[8] Bioinformatics and Computational Biology Program, Iowa State University

[9] Department of Biological Sciences, Middle East Technical University, Ankara, Turkey

[10]The No.2 people's hospital of Tibet Autonomous Region,850000, China

[11]The hospital of XiShuangBanNa Dai Nationalities, Autonomous Jinghong 666100, Yunnan, China.

[12] The Guangdong Enterprise Key Laboratory of Human Disease Genomics, BGI-Shenzhen,Shenzhen,China

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

[14]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia

[15] James D. Watson Institute of Genome Science, Hangzhou, China

[16] Department of Biology, University of Copenhagen, Ole MaaløesVej 5, 2200 Copenhagen, Denmark

[17]Macau University of Science and Technology, AvenidaWai long, Taipa, Macau 999078, China

[18]Department of Medicine, University of Hong Kong, Hong Kong

[19]Department of Statistics, University of California, Berkeley, CA

[20]Department of Biology, University of Copenhagen, Copenhagen, Denmark

* These authors contributed equally to this work.

co-corresponding: E-mail: wangj@genomics.cn (Ju.W.); wangjian@genomics.cn (Ji.W.); rasmus_nielsen@berkeley.edu

Estimate of effective population size

We obtained an estimate of the Tibetan population's effective size by fitting a demographic model of constant population size to the synonymous site frequency spectrum of the exome data from Yi *et al.* 2010 (S1). We used the software *dadi* (S2) to fit the model to the data, and we estimated an effective population size of *N=7000*.

Estimate of the mutation rate

We estimated the mutation rate in the *EPAS1* region by using an Approximate Bayesian Computation (ABC) framework following Peter *et al.* 2012 (S3). Using the same set of summary statistics, we obtained posterior estimates for the mutation rate under three different models: a neutral model, a model of selection on a de novo mutation, and a model of selection on standing variation. The summary statistics used are both based on haplotype patterns and based on the site frequency spectrum. In particular we used iHS (S4), EHH (S5), and EHH in a 10 kb, 20 kb and 50 kb window, centered on the putative selected site. For the frequency based statistics, we used Tajima's D (S6), Fay & Wu's H (S7), the average number of pairwise differences π, and the number of segregating sites S as statistics. All these statistics are calculated for three regions: a central region of 20 kb around the selected site, an intermediate region consisting of all sites 20–50 kb away from the selected site, and a faraway region consisting of all sites further than 50 kb away from the selected site. For the neutral model, the mean and median of the posterior distribution of the mutation rate is 1.97e-08 and

1.98e-08. For the selection on a *de novo* mutation model, the mean and median of the posterior distribution of the mutation rate is 2.059e-08 and 2.055e-08. For a model of the selection on standing variation, the mean and median of the posterior distribution of the mutation rate is 2.09e-08 and 2.07e-08. These estimates are very similar to phylogenetic estimates of 2.0e-08 (S8).

Distribution of human-chimp differences in 32.7kb regions

We computed the number of human-chimp differences in contiguous 32.7kb non-overlapping regions across the entire human genome. As the human-chimp alignment will have gaps, the 32.7kb sequence may have start and stop coordinates that exceed a length of 32.7kb. Therefore, to minimize selecting regions with poor alignment, we imposed a condition to keep only those windows whose total length was less than 33.2kb (i.e less than 500 base pairs of gaps). Note, however, that even then the number of fixed differences is assessed over exactly 32700 bases. The mean and median of this distribution are 417.4 and 410 respectively. The observed number of human-chimp differences within the 32.7kb region of *EPAS1* is 417, which lies right in the center of the genome-wide distribution. Assuming a human-chimp divergence time of 6.5 million years, and a generation time of 20 years, the mutation rate estimate for the *EPAS1* 32.7kb region is 1.96x10-8 per base pair per generation, again very close to the phylogenetic estimates (S8).

Extended Data Figure 5b

The observed haplotype structure observed in Figure 3 is compatible with the introgression hypothesis. As expected under the introgression hypothesis, the Tibetan haplotype is more distant to the non-Tibetan haplotypes than the Denisovan haplotype because, after the admixture event, the introgressed haplotype accumulated extra mutations. In contrast, the Denisovan haplotype, being the product of DNA extracted from an ancient specimen, did not have time to accumulate a similar number of mutations. For example, the divergence between the introgressed haplotype (the Tibetan (Tib) haplotype) and the Yoruban haplotype would be larger than between the observed Denisovan haplotype and the Yoruban (YRI) haplotype (see Extended Data Fig. 5b):

$$T^{Tib-YRI} = a + k + b + f + e$$
$$T^{Denisovan-YRI} = a + i + b + f$$

Here the letters (*a, b, e, f, i, k*) represent branch lengths (see Extended Data Fig. 5b). $T^{Tib\text{-}YRI}$ is larger than $T^{Denisovan\text{-}YRI}$ because $i < (k + e)$. We note that different regions in the genome have different genealogies but for an introgressed region we would expect a genealogy such as the one in this figure.

Haplotype networks constructed using other criteria, Extended Data Figure 6a and 6b

Figure 3 plots the network of the 40 most common haplotypes. Alternatively, we also used the 20 sites such that the frequency difference between Tibetans and each of the 14 populations from the 1000 Genomes project (S9) is at least 0.65 to construct a haplotype network (Extended Data Fig. 6a). The resulting region

spanned by these SNPs is the same 32.7kb region as previously identified by considering sites that are the most differentiated between Tibetans and Han (Table S3). Extended Data Fig. 4 plots the allele frequencies in all of the populations in the 20 sites within this region. We limited ourselves to these SNPs to reduce the number of haplotypes to display in a single figure, while keeping the SNPs that are most informative regarding the relationship between the selected haplotype and other haplotypes. Including all SNPs would produce an un-manageable number of haplotypes that could not be meaningfully displayed in a single figure. We inspected all the haplotypes defined by these 20 SNPs in all populations. Of the possible 2264 haplotypes (2×(number of individuals in the 1000 Genomes Project) + 80 Tibetan chromosomes), we identified 43 unique haplotypes. We used the R (S10) software package "*pegas*" (S11) to build a tree that connects haplotypes based on pairwise differences (see Extended Data Fig. 6a). The Denisovan individual is homozygous at all the 20 sites. Extended Data Fig. 6b plots the number of pairwise differences between each of the 42 haplotypes and the Denisovan haplotype. Note, Figures 6a-b do not include the Iberians (IBS) for clarity (to reduce the number of colors needed for the plot). We decided to exclude this population because only a small number (18) of Iberians was sequenced, and the haplotypes of these individuals are identical to other European haplotypes.

Number of pairwise differences

We computed the number of pairwise differences between a modern human

haplotype and the two Denisovan haplotypes as follows:

$H_kD = \sum_i^{allsites} \frac{\Delta_{H_kD}(i)}{2}$, where $\Delta_{H_kD}(i) = 0$ if Denisovan (*D*) is homozygous for the modern human allele at site *i*, $\Delta_{H_kD}(i) = 1$ if *D* is heterozygote, and $\Delta_{H_kD}(i) = 2$ if *D* is homozygotes for another allele. We divided by 2 because the Denisovan is not phased, and we weight accordingly.

D statistics under models of no gene flow

As another approach to assess the probability of an ancestral lineage having given rise to the 32.7kb haplotype we observe in Tibetans in the absence of gene flow, we compared D-statistics between human populations under simulations of several demographic models. We used all the 4 models of Sankararaman *et al.* 2012 (S12) that do not include gene flow but do contain combinations of ancestral structure and an out-of-Africa bottleneck : a model with ancestral structure and a bottleneck (asm_b), a model with ancestral structure and no bottleneck (asm_nb), a model with no ancestral structure and with a bottleneck (ngf_b), and a model with no ancestral structure and no bottleneck (ngf_nb). Our model included the Denisovans (DEN) along with four human populations: Tibetans (TIB), Han-Chinese (CHB), Europeans (CEU) and Yorubans (YRI). For Denisovans, we sampled a single diploid sample; for all other populations we sampled 80 chromosomes. Because there is no single accepted demographic history between Tibetans and Han, we assumed split times of 3kya, 5kya and 10kya. For the split times of the other populations, we used the Sankararaman *et al.* 2012 (S12) estimates, which are 21kya for the split between Han/Tibet and

Europeans and 350kya for the split between Humans and Denisovans. In the no-gene flow model, we assumed that Africans split from non-Africans 62.5kya ago, and for the ancestral structure model, we assumed the structure lasted until 225kya. We assumed an effective population size of 10,000 individuals and a generation time of 25 years, and we used the hapmap recombination map to estimate the local recombination rate in the *EPAS1* gene. All simulations were done in *ms* (Hudson 2002 (S13)), and the particular commands used were (for a 3kya split time for Han/Tibetans, but we also used a Han/Tibetan divergence time of 5kya and 10kya):

```
#no gene flow with bottleneck
./ms 322 100000 -t 32.7 -r 30.8 32700 -I 5 80 80 80 80 2 0 -ej
0.003000 1 2 -ej 0.021000 2 3 -en 0.025000 3 .01 -en 0.025500 3 1
-ej 0.062500 3 4 -ej 0.350000 4 5

#no gene flow without bottleneck
./ms 322 100000 -t 32.7 -r 30.8 32700 -I 5 80 80 80 80 2 0 -ej
0.003000 1 2 -ej 0.021000 2 3 -ej 0.062500 3 4 -ej 0.350000 4 5

#ancestral structure with bottleneck
./ms 322 100000 -t 32.7 -r 30.8 32700 -I 5 80 80 80 80 2 0 -m 1 2
10.000000 -m 2 3 10.000000 -m 3 4 10.000000 -m 2 4 10.000000 -m 2
1 10.000000 -m 3 2 10.000000 -m 4 3 10.000000 -m 4 2 10.000000 -
ej 0.003000 1 2 -ej 0.021000 2 3 -en 0.025000 3 .01 -en 0.025500
3 1 -ej 0.225000 3 5 -em 0.225000 4 5 5 -em 0.225000 5 4 2 -ej
0.350000 4 5

#ancestral structure without bottleneck
./ms 322 100000 -t 32.7 -r 30.8 32700 -I 5 80 80 80 80 2 0 -m 1 2
10.000000 -m 2 3 10.000000 -m 3 4 10.000000 -m 2 4 10.000000 -m 2
1 10.000000 -m 3 2 10.000000 -m 4 3 10.000000 -m 4 2 10.000000 -
ej 0.003000 1 2 -ej 0.021000 2 3 -ej 0.225000 3 5 -em 0.225000 4
5 5 -em 0.225000 5 4 2 -ej 0.350000 4 5
```

We compared these simulations with human data from the 1,000 Genomes project (S9), where we used 80 sample chromosomes per population to avoid

any bias due to sample sizes. D-statistics were calculated according to eq. 2 in Durand *et al.* 2011 (S14). The two modern human populations used in computing D-statistics are Tibetans and either CHB, CEU or YRI. The results from the findings are displayed in Extended Data Fig. 9 which assumes a Tibetan-Han divergence time of 3000 years, but the results are also significant (p-value < 0.001) if we assume a larger divergence time of 5000years or 10000years. Within the 32.7kb region, we find 48 SNPs that are derived in Denisovans. From this data, we estimate that D(TIB, CHB, DEN, Chimp) = -0.9584960, D(TIB, CEU, DEN, Chimp) = -0.8990739, and D(TIB, YRI, DEN, Chimp) = -0.8818433. Thus, D is very low regardless of whether we use CHB, CEU or YRI. Comparing the D values to the simulations under the null model, we find that the observed D statistics are in the extreme tail, with $p < 0.001$ for all comparisons (see Tables S8 to S10). In fact, for the ancestral structure model, none of the simulations were able to produce D values as low as what we observed. We also find that the distribution of D becomes wider with increasing split time or when we include a bottleneck, and as a consequence, we find the highest *p*-values when we calculate D(TIB,YRI,DEN,Chimp). Thus, these simulations suggest that it is very unlikely that D values that low are due to incomplete lineage sorting in the ancestral population, consistent with our claim of introgression.

Supplementary References

S1. Yi, X. *et al*., Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* **329**,75-78 (2010).

S2. Gutenkunst, R.N. *et al.* Inferring the joint demographic history of multiple populations from multidimensional SNP data. *PLoS Genetics* 5:e1000695 (2009)

S3. Peter B. *et al.* Distinguishing between Selective Sweeps from Standing Variation and from a *De Novo* Mutation. PLoS Genetics 8(10): e1003011 (2012).
S4. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4: e72 doi:10.1371/journal.pbio.0040072.
S5. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837. doi: 10.1038/nature01140v
S6. Tajima F (1989) Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123: 585–595.
S7. Fay JC, Wu C-I (2000) Hitchhiking Under Positive Darwinian Selection. Genetics 155: 1405–1413.
S8. Scally A. and Durbin R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet.* 13, 745-53 (2012).
S9. McVean, G. *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
S10. R (http://r-cran.org)
S11. Paradis. E. Pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*. 26, 419-420 (2010).
S12. Sankararaman, S. et al., 2012. The Date of Interbreeding between Neandertals and Modern Humans. *PLoS Genet*, 8(10), p.e1002947.
S13. Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics (Oxford, England)*, **18**, 337–338 (2002).
S14. Durand, E.Y. *et al.* Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* **28**, 2239–2252 (2011).
S15. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 388, 222-226 (2012).
S16. Prufer K. et al., The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43-49 (2014).

Supplementary Tables

Table S1 The first 238 SNPs in *EPAS1* (out of 477)

46358021	46366087	46374208	46387234	46397320	46404972	46411824
46358466	46366545	46374220	46387724	46397399	46405027	46411936
46358835	46366806	46375111	46387842	46398115	46405339	46412034
46358881	46367079	46377419	46388002	46398158	46405551	46412045
46359412	46367187	46377519	46388691	46398742	46405681	46412777
46359580	46367600	46377754	46388789	46399043	46405848	46412975
46359642	46367949	46378084	46389040	46399106	46405962	46412981
46360148	46368021	46378895	46389060	46399122	46406105	46413280
46360207	46368077	46381044	46389112	46399227	46406430	46413588
46361168	46368411	46381320	46389383	46399820	46406635	46413618
46361928	46368427	46381791	46389445	46400062	46407353	46414442
46362045	46368782	46382173	46389679	46400494	46408037	46414693
46362345	46369143	46382195	46389699	46400575	46408150	46414706
46362500	46369325	46382305	46390356	46400630	46408551	46414813
46362729	46369670	46382465	46390540	46400998	46408725	46415114
46362869	46369864	46382792	46390560	46401144	46408767	46415289
46363161	46370096	46382793	46391093	46401384	46408848	46415320
46363224	46370323	46382942	46391108	46401587	46409011	46416260
46363225	46371633	46383343	46391312	46401613	46409174	46416579
46363398	46371939	46383410	46391314	46401854	46409849	46416783
46363650	46372006	46383582	46391844	46402399	46409904	46416800
46363785	46372009	46384165	46392298	46402410	46409966	46416896
46363837	46372042	46384435	46392352	46402544	46410137	46416963
46364243	46372398	46384562	46393696	46402559	46410265	46417642
46364282	46372566	46384907	46394382	46402893	46410431	46418258
46364296	46372897	46385230	46394649	46403037	46411073	46418501
46364908	46372929	46385306	46394807	46403636	46411104	46418595
46364955	46373095	46385310	46395343	46403771	46411148	46419026
46365183	46373244	46385677	46395811	46403832	46411252	46419364
46365275	46373557	46386112	46395912	46403996	46411461	46420168
46365353	46373879	46386859	46395953	46404222	46411466	46420523

Table S2 The rest of the SNPs in *EPAS1* (239 out of 477)

46421653	46430803	46437257	46446028	46454321	46460714	46470622
46422184	46431148	46437557	46446311	46454398	46461837	46470686
46422310	46431261	46437563	46446577	46454616	46462046	46470811
46422521	46431312	46437875	46446803	46454696	46462168	46471311
46423274	46431903	46438077	46446970	46454766	46462549	46471869
46423562	46432417	46438230	46447040	46454918	46463076	46472086
46423846	46432428	46438356	46447429	46455000	46463399	46472285
46424212	46432777	46438363	46447626	46455359	46463461	46472426
46424521	46432852	46439387	46447779	46455371	46463470	46472464
46424629	46432913	46439392	46448104	46455755	46463475	46472505
46424814	46433162	46440538	46448606	46455923	46463620	46472959
46424939	46433179	46440601	46448760	46456124	46463679	46473935
46425145	46433193	46441523	46448832	46456182	46463716	46474490
46425953	46433234	46441835	46448867	46456226	46463983	46474591
46425987	46433385	46441992	46449937	46456669	46464109	46475380
46426708	46433495	46442452	46450292	46456706	46464184	46475749
46426714	46433570	46442536	46450455	46456764	46464286	46475834
46426803	46433948	46442608	46450800	46456942	46464491	46475977
46426899	46433978	46442683	46451085	46457175	46464975	46476366
46427468	46434578	46442715	46451260	46457547	46467027	46476377
46427484	46434957	46442799	46451331	46458016	46467706	46477018
46428425	46435048	46442827	46451374	46458052	46467739	46477115
46428827	46435147	46442991	46451529	46458087	46467802	46477781
46428892	46435236	46443838	46451737	46458097	46467939	46478082
46429676	46435241	46443888	46451807	46458185	46467967	46478097
46429746	46435886	46443960	46452072	46458210	46469451	46478103
46429769	46436395	46444229	46452534	46458310	46469459	46478227
46429992	46436567	46444328	46452778	46458361	46469569	46478448
46430077	46436666	46444364	46452877	46459163	46469718	46479112
46430398	46436691	46444376	46452964	46459439	46469820	46483002
46430666	46436725	46444948	46453534	46459458	46469980	46484000
46430716	46436739	46445066	46453862	46459986	46470186	46484327
46430755	46437085	46445268	46454163	46460001	46470444	46484933
46430769	46437097	46445510	46454165	46460292	46470494	46486982
						46487171

Table S3 SNPs in *EPAS1* with the largest frequency differences between Tibetans and Han individuals.

SNP LABEL	Position (hg18)	Position (hg19)	Tibetan derived freq (out of 80)	Han derived freq (out of 80)
1	46421420	46567916	67	0
2	46422184	46568680	66	0
3	46422521	46569017	66	0
4	46423274	46569770	66	0
5	46423846	46570342	67	0
6	46424521	46571017	66	0
7	46424939	46571435	67	0
8	46428892	46575388	70	10
9	46430716	46577212	10	72
10	46430755	46577251	69	1
11	46430769	46577265	10	73
12	46430803	46577299	70	8
13	46432777	46579273	71	11
14	46432913	46579409	10	75
15	46433193	46579689	16	79
16	46433978	46580474	12	75
17	46435147	46581643	69	7
18	46435236	46581732	69	7
19	46437085	46583581	67	1
20	46438363	46584859	67	1
21	46441523	46588019	67	1
22	46441835	46588331	67	1
23	46442536	46589032	66	1
24	46442827	46589323	14	79
25	46446311	46592807	64	1
26	46447040	46593536	16	76
27	46447626	46594122	64	1
28	46451260	46597756	65	1
29	46451529	46598025	65	1
30	46453534	46600030	65	1
31	46453862	46600358	65	1
32	46454165	46600661	65	1

Table S4 Haplotypes at the five SNPs of interest in *EPAS1* in Tibetans (Extended Data Figure 3), in the genomes from Meyer *et al.* 2012 (S15) and from the published high coverage Altai Neanderthal genome (S16). Note we present the most common haplotype in Tibetans only. The data from Meyer *et al.* (2012) is not phased but since most individuals are homozygous, we can infer the haplotypes corresponding to these SNPs.

Position (hg19)	**46567916**	**46568680**	**46569017**	**46569770**	**46570342**
Tibetan	A	G	G	A	A
Denisovan	A	G	G	A	A
Altai Neanderthal	G	G	G	A	G
Dinka	G	A	A	G	G
French	G	A	A	G	G
Papuan	G	A	A	G	G
Sardinian	G	A	A	G	G
Han	G	A	A	G	G
Yoruba	G	A	A	G	G
Karitiana	G	A	A	G	G
San	G	A	A	G	G
Mandeka	G	A	A	G	G
Dai	G	A	A	G	G
Mbuti	A	A	A	G	G
CEU	G	A	A	G	G
YRI	G	A	A	G	G
YRI	A	A	A	G	G
Vindija	G	A	M	G	G
Vindija	A	G	M	A	G

Table S5. Frequency of 43 haplotypes defined by the 20 most differentiated sites between Tibetan and the 1000 Genomes project populations. Table S6 contains the frequencies of these haplotypes. The first haplotype (AGGAAGCCGGGCGTGCGAAC) is the Denisovan haplotype.

Haplotype	Haplotype
1. AGGAAGCCGGGCGTGCGAAC	23. GAAGGCGTAGACGTAACAAA
2. GAAGGCGCAGAGGTGCGGAA	24. AGGGAGCCGAGGCTGCGGGC
3. AGGAAGCCGAGGCTAACAAA	25. GAAGGCGTAGACGTAACAAC
4. GAAGGCGCGAGGCTGCGGGC	26. GGGGGCGTAGACGCAACAAA
5. AAAGGCGTAGACGCAACAAA	27. GAAGGCGTAGACGCACCAAA
6. AAAGGCGTAGACGCAACAAC	28. AGGAGCGTAGACGCAACAAA
7. AGGAAGCCAGACGCAACAAA	29. GGGGGCGTAGACGCAACAAC
8. GAAGGCGTAGACGCAACGAA	30. AGGAAGCCAGGGCTGCGGGC
9. AGGAAGCCGAGGCTGCGGGC	31. AAAGGCGTAGACGCAACGAA
10. GAAGGCGTAGACGCAACAAC	32. AGGGGCGTAGACGTAACAAA
11. AAAAACCCAAGGCTGCGGGC	33. AGGAAGCTAGACGCAACAAA
12. GAAGGCGCGAGGCCAACAAA	34. AGGAGCGTAGACGTAACAAA
13. GAAGGCGTAGACGCGACAAA	35. AGGAGCGTAGACGTAACAAC
14. GAAAGCGTAGACGTAACGAC	36. AGGAACGCAGACGTGCGGAC
15. GAAAGGGTAGACGCAACAAA	37. GAAGGCGTAGACGTGCGGAC
16. GAAGGCGTAGACGCAACAAA	38. GGGAACGTAGACGCAACAAA
17. AAAGGCGTAGACGTAACAAA	39. GAAGGCGTAAAGCCAACAAA
18. GAAGGGGTAGACGCAACAAC	40. GGGAACGTAGACGCAACAAC
19. GAAGGGGTAGACGCAACAAA	41. AGGAAGCCAAGGCTGCGGGC
20. AGGAAGGTAGACGCAACAAA	42. AGGAAGCCGAACGCACGGGC
21. GAAAGCGTAGACGTAACAAC	43. AGGGGCGTAGACGCAACAAA
22. GAAAGCGTAGACGTAACAAA	

Table S6 The frequency of each haplotype (from Table S5) by population. TIB and DEN are abbreviations for Tibetan and Denisovan. The other abbreviations correspond to the populations from the 1000 Genomes project.

	GBR	FIN	CHS	PUR	CLM	CEU	YRI	CHB	JPT	LWK	ASW	MXL	TSI	TIB	DEN
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
5	1	0	0	3	0	0	6	0	0	3	5	0	0	0	0
6	0	0	0	1	0	0	11	0	0	10	3	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	4	0	0	0	2	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	57	0
10	0	0	0	1	0	0	3	0	0	9	1	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0
16	167	180	196	100	108	158	102	190	178	133	84	126	180	10	0
17	3	0	0	1	1	2	0	0	0	0	2	0	3	0	0
18	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
23	2	1	2	2	0	3	22	0	0	9	11	0	3	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
25	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
26	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
27	5	5	0	2	10	7	0	0	0	0	3	6	9	0	0
28	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
31	0	0	0	0	0	0	12	0	0	7	4	0	0	0	0
32	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
34	0	0	0	0	0	0	4	0	0	2	2	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
36	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

	GBR	FIN	CHS	PUR	CLM	CEU	YRI	CHB	JPT	LWK	ASW	MXL	TSI	TIB	DEN
37	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
38	0	0	0	0	1	0	7	0	0	0	0	0	1	0	0
39	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
43	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Table S7. Column 2 corresponds to the positions of the 32 SNPs that show the largest frequency difference between the Han and the Tibetans (see Table S3). The third and fourth columns are the high coverage Altai Neanderthal (S16) alleles at those sites. The fifth and the sixth columns are the Denisovan alleles. The seventh column is the list of alleles corresponding to the most common Tibetan haplotype at these SNPs. The last column is the frequency of the derived alleles at these sites in Tibetans (out of 80 chromosomes). The yellow highlighting represents sites that are not SNPs in the 1000 Genomes project. The blue highlighting represents the 20 sites that are only highly differentiated between Tibetans and all populations from the 1000 Genomes project. The green highlighting represents sites where the Tibetans are different from the Denisovan.

SNP LABEL	hg19	Nea	Nea	Denisovan	Denisovan	Tibetan most common haplotype	Tibetan derived frequency
1	46567916	G	G	A	A	A	67
2	46568680	G	G	G	G	G	66
3	46569017	G	G	G	G	G	66
4	46569770	A	A	A	A	A	66
5	46570342	G	G	A	A	A	67
6	46571017	C	C	G	G	G	66
7	46571435	G	G	C	C	C	67
8	46575388	A	A	G	G	G	70
9	46577212	C	C	T	T	T	10
10	46577251	T	T	C	C	C	69
11	46577265	C	C	A	A	A	10
12	46577299	A	A	G	G	G	70
13	46579273	A/G	A/G	G	G	G	71
14	46579409	G	G	A	A	A	10
15	46579689	A	A	G	G	G	16
16	46580474	T	T	T	T	T	12
17	46581643	C	C	C	C	C	69
18	46581732	C	C	C	C	C	69
19	46583581	G	G	G	G	A	67
20	46584859	A	A	G	G	G	67
21	46588019	C	C	C	C	G	67
22	46588331	G	G	G	G	C	67
23	46589032	C	C	T	T	T	66
24	46589323	C	C	A	A	A	14
25	46592807	C	C	T	T	T	64
26	46593536	A	A	A	A	A	16
27	46594122	A	A	G	G	G	64
28	46597756	C	C	C	C	C	65
29	46598025	C	C	G	G	G	65
30	46600030	A	A	A	A	G	65
31	46600358	A	A	A	A	G	65
32	46600661	A	A	C	C	C	65

Table S8: P-values for D under the four models simulated assuming a Tibetan-Han (TIB-CHB) split of 3,000 years ago. Each entry gives the proportion of simulations with a D value lower than the one observed out of 100,000 simulations.

3,000y divergence	TIB-CHB	TIB-CEU	TIB-YRI
Ngf_b	0.00016	0.00024	0.00088
Ngf_nb	0.00009	0.00014	0.00045
Asm_b	<1e-5	<1e-5	<1e-5
Asm_nb	<1e-5	<1e-5	<1e-5

Table S9: P-values for D(TIB, CEU, Den, Chimp) under the four models simulated. Each entry gives the proportion of simulations with a D value lower than the one observed out of 100,000 simulations.

5,000y divergence	TIB-CHB	TIB-CEU	TIB-YRI
Ngf_b	0.00008	0.00019	0.00094
Ngf_nb	0.00008	0.00008	0.00041
Asm_b	<1e-5	<1e-5	<1e-5
Asm_nb	<1e-5	<1e-5	<1e-5

Table S10: P-values for D(TIB, YRI, Den, Chimp) under the four models simulated. Each entry gives the proportion of simulations with a D value lower than the one observed out of 100,000 simulations.

10,000y divergence	TIB-CHB	TIB-CEU	TIB-YRI
Ngf_b	0.00021	0.00023	0.00078
Ngf_nb	0.00007	0.00012	0.00042
Asm_b	<1e-5	<1e-5	<1e-5
Asm_nb	<1e-5	<1e-5	<1e-5

Table S11. The characterization of the 14 sites where the Denisovan differs from the most common Tibetan haplotype. In the main text we quote 12 differences because if the Denisovan is heterozygous we consider that to contribute only 0.5 (see Supplemetary Information titled "Number of pairwise differences"). Five of the mutations occurred on the Denisovan lineage (light blue). Nine of the mutations occur on the modern human lineage (light orange, except position 46600226 is also derived in Denisovans). In pink we show the mutations that are unique to Tibetans and that presumably have landed on the Denisovan-introgressed haplotype after introgression.

Positions (hg19)	Denisovan Alleles	Chimp allele	Tibetan allele	Humans (1000Genome samples and Tibetan samples)
46578392	A/C	A	A	A
56579659	A	T	T	T
46580956	C/G	C	C	C
46594186	G/T	G	G	G
46595734	G/C	G	G	G
46581074	C	C	G	C/G; G on Tibetan haplotype
46583581	G	G	A	G/A; only Tibetans have the derived allele A at high frequency and the Han have it at small frequency
46588019	C	C	G	C/G only Tibetans have the derived allele G at high frequency and the Han have it at small frequency
46588331	G	G	C	G, only Tibetans have the derived allele C at high frequency and Han have it at small frequency
46596433	A	A	G	A/G; G on Tibetan haplotype
46597870	T	T	C	T/C; C on Tibetan haplotype
46600030	A	A	G	A/G; G on Tibetan haplotype
46600226	G	T	T	T/G; T on Tibetan hapltoye
46600358	A	A	G	A/G, only Tibetans have the derived allele G at high frequency. All the 1000Genome samples have A.