Supplementary Information 1

The Ust'-Ishim locality: Geology, palaeontology, and radiocarbon dating.

Yaroslav V. Kuzmin*, Aleksei A. Bondarev, Pavel A. Kosintsev, , T. Higham, K. Douka and Bence Viola
* To whom correspondence should be addressed (kuzmin@fulbrightmail.org)

The Ust'-Ishim femoral shaft was found in 2008 by N.V. Peristov, an ivory carver and local historian who collected fossil materials on the banks of the Irtysh River, in Ust'-Ishim and Tara counties of Omsk Oblast, Russian Federation (Russia). As Peristov collects the material for carving and exhibitions, not for scientific purposes, he did not record the exact locality of the specimen. In 2010, A. Bondarev identified the bone as a hominin, presumably *Homo sapiens*, femur.

Our information on the locality of the discovery comes from the recollections of N.V. Peristov, and thus has to be taken with caution. The region of the middle course of the Irtysh River in Western Siberia is covered by thick Quaternary deposits. Along the Irtysh River, late Middle Pleistocene and Late Pleistocene sediments are widely distributed[1], so it is also possible that the femur was originally deposited at another nearby site than where it was collected by N.V. Peristov.

Based on the recollections of Peristov, the femur had eroded from alluvial deposits of the Irtysh River, in the large bluff on the river's left bank, north of the village of Ust'-Ishim, between the Ust'-Ishim airstrip and the Nikolsky ferry crossing (Figure S1.1). The geographic coordinates of the find are: 57°43′ northern latitude, and 71°10′ eastern longitude. The bluff where the Late Pleistocene sediments are exposed is about 8–10 m above the water level (Figure S1.2).

The late Middle Pleistocene deposits in the vicinity of Ust'-Ishim are represented by the interglacial Tobol Suite. It consists of alluvial sediments up to 30–35 m thick. Middle Pleistocene mammal fossils belonging to the Vyatkino faunal complex (including *Archidiscodon* ex. gr. *trogontherii*, *Alces latifrons*, and *Equus* ex. gr. *mosbachensis*) were found in the Tobol Suite. Carpological finds show a boreal forest vegetation (dominated by conifer trees) at that time. The Late Pleistocene deposits consist of several stratigraphic horizons. The interglacial deposits of the Kazantsevo Horizon, Oxygen Isotope Stage 5e (OIS 5e) cover the alluvium of the older fourth river terrace. The Kazantsevo vegetation was represented by boreal forests. Both the Kazantsevo Horizon and the Ermakovo glacial horizon (OIS 4) constitute the body of the third terraces of the Ob and Irtysh rivers, with total thickness of ca. 25–30 m (including up to 10 m of alluvium). The numerous mammal bones from deposits of the third terrace belong to the Upper Palaeolithic faunal complex. Palynological data for the Ermakovo Horizon allow reconstructing the existence of cold steppe vegetation. The interstadial Karginsky [Karginian] Horizon (OIS 3), dated to ca.

50–25 ka ago, constitutes the alluvium of the second Irtysh River terrace, and the overlying deposits (loess-like loams) on the watersheds, with total thickness up to 10–20 m. Bones of the representatives of the Upper Palaeolithic complex are numerous in this horizon. The lower parts of the second terrace are especially rich in animal fossils (see Figure S1.2) where mammal bones are deposited in the sand and gravel layers. Some bones were likely redeposited from the older horizons. These are so-called Buginka and Zagvozdino layers dated to ca. 40–24.5 ka ago[2]. The vegetation during the Karginsky interstadial (OIS 3) in the central West Siberian Plain was of forest tundra type[3]. The Ust'-Ishim human femur probably eroded from the Buginka–Zagvozdino layers of the second Irtysh River terrace.

Quaternary mammal fossils are plentiful on the banks of the Irtysh River around the Ust'-Ishim outcrop. The majority (up to 30%) belong to mammoth (*Mammuthus* sp.), with woolly mammoth (*M. primigenius*) as the most common species, and with rare *M. chosaricus* and *M. trogontherii*. About 25% of the fossils belong to Pleistocene bison (*Bison priscus*). Horse (*Equus* sp.) fossils constitute about 20% of the assemblage, and consist predominantly of *E.* cf. *gallicus*, with rare finds of *E. mosbachensis*. The woolly rhinoceros (*Coelodonta antiquitatis*) comprises about 15% of the total fossil assemblage. Other mammals are quite rare (less than 10% in total), and include the following species and taxa (in order of decreasing abundance): *Alces* ex gr. *alces-americanus*, *Cervus elaphus*, *Ursus arctos*, *Rangifer tarandus*, *Saiga tatarica*, *Alces (Cervalces) latifrons, Megaloceros giganteus*, *Ovibos moschatus*, *Ursus savini*, *Panthera spelaea*, *Canis lupus*, *Praeovibos* sp., *Gulo gulo* and *Stephanorhinus kirchbergensis*[4].

Radiocarbon dating

We used direct AMS radiocarbon dating to clarify the age of the human femur. Two samples of Ust'-Ishim 1 were submitted to the Oxford Radiocarbon Accelerator Unit (ORAU), Research Laboratory of Archaeology and the History of the Art, Oxford University (Oxford, UK). The pretreatment of the bone followed the usual protocol at the ORAU, using ultrafiltration[5]. Both samples yielded abundant and well preserved collagen, satisfying the requirements of good collagen preservation[6]. For details of collagen yield and other analytical parameters see Table SI1.1.

The two direct dates on the femur are identical. The first sample (taken in 2011) is 41,400 ± 1300 BP (OxA-25516), while the second sample (taken and dated in 2014) is 41,400 ± 1400 BP (OxA-30190). As the two samples come from the same individual, we combined them using R-combine function of the OxCal platform[7], yielding a combined age of 41410 ± 960 BP (Fig. SI1.3.). This value corresponds approximately to a calendar age of 45,770–44,010 cal BP (68.2% probability) or 46,880–43,210 cal BP (95.4% probability), using the IntCal 13 curve[8] in OxCal v4.2.3[7] (for individually calibrated ages and $F^{14}C$ activity ratios see Table SI1.2., and Fig. SI1.4.).

References

1. Kaplyanskaya, F. A. & Tarnogradsky, V. D. in *Stratigrafiya SSSR. Chetvertichnaya Systema. Polutom 2 [Stratigraphy of the USSR. The Quaternary System. Half-Volume 2]* (eds Krasnov, I. I.) 227-270 Moscow, 1984).
2. Krivogonov, S. K. *Stratigrafiya i paleogeografiya Nizhnego Priirtyshya v epokhu poslednego oledeneniya (po karpologicheskim dannym) [The stratigraphy and palaeogeography of the lower Irtysh in the last glaciation (based on carpological data)]* (Nauka, Novosibirsk, 1988).
3. Zynovyev, E. in *Back to the roots and back to the future. Towards a new synthesis amongst taxonomic, ecological and biogeographical approaches in carabidology* (eds Peney, L., Erwin, T. & Assman, T.) 241-254 (Pensoft, Sofia, Moscow, 2008).
4. Kosintsev, P. A. & Bobkovskaya, N. E. in *Chervertichnaya Paleozoologiya na Urale [The Quaternary Paleozoology in the Urals]* (eds Smirnov, N. G.) 226-232 Yekaterinburg, 2003).
5. Brock, F., Higham, T., Ditchfield, P. & Ramsey, C. B. Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon* **52,** 103-112 (2010).
6. Brock, F., Wood, R., Higham, T. F., Ditchfield, P., *et al.* Reliability of nitrogen content (% N) and carbon: Nitrogen atomic Ratios (C: N) as indicators of collagen preservation suitable for radiocarbon dating. *Radiocarbon* **54,** 879-886 (2012).
7. Bronk Ramsey, C., Dee, M., Lee, S., Nakagawa, T. & Staff, R. Developments in the calibration and modelling of radiocarbon dates. *Radiocarbon* **52,** 953 (2010).
8. Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. *Radiocarbon* **55,** 1869-1887 (2013).

Table SI1.1. Analytical parameters established during the AMS ^{14}Cdating of the Ust'-Ishim femur. Collagen yield, stable isotopes, %C, C/N ratios, all fall within expected values. We should also note that the stable isotope values of the two samples dated at the ORAU are identical to the values measured at Leipzig in the palaeodietary reconstruction of the Ust'-Ishim individual.

OxA	Material	Protocol	Bone used (mg)	Collagen yield (mg)	% Coll. yield	% C	δ^{13}C	δ^{15}N	C:N
25516	bone	Ultrafiltration	890	68.8	7.7	45.8	-19.2	14.2	3.2
30190	bone	Ultrafiltration	450	45.2	10	43.4	-19.2	14.2	3.3

Table SI1.2. Radiocarbon determinations from the Ust'-Ishim femur, shown here as radiocarbon years BP and F^{14}C activity ratios, and also calibrated using the IntCal13 calibration curve and the OxCal software. The dates were performed on 2 separate samples from the same human bone, extracted and analysed in 2011 (OxA-25516) and in 2014 (OxA-30190) at the Oxford Radiocarbon Accelerator Unit (ORAU).

Site	OxA	Date BP	±	F^{14}C	±	Calibrated years BP	
						68.2 %	95.4%
Ust-Ishim	25516	41400	1300	0.0058	0.00097	46120-43580	48120-42810
Ust-Ishim	30190	41400	1400	0.00574	0.00098	46240-43620	48290-42840

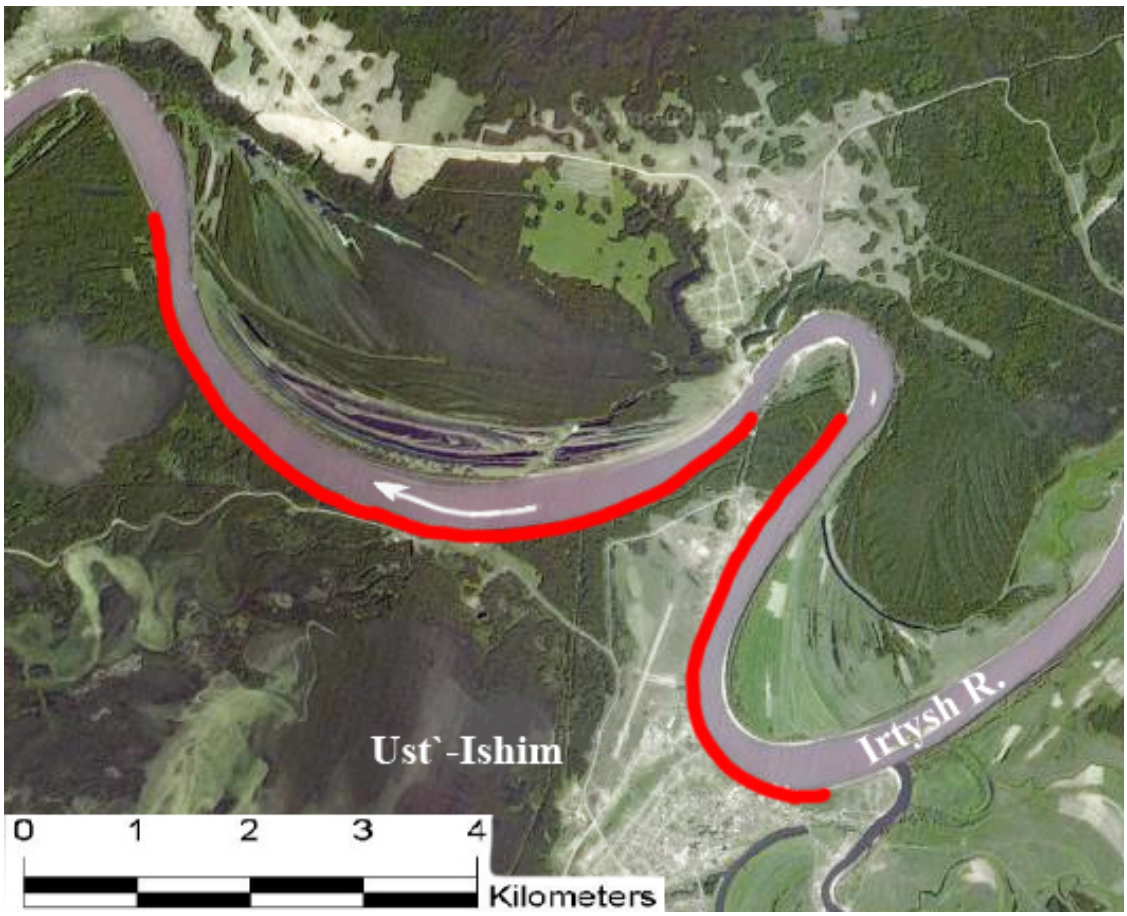


Figure S1.1: The location of the Ust'-Ishim outcrop (bluff) indicated by the red lines. Image by A.A. Bondarev, 2014; imagery from http://kosmosnimki.ru

Figure S1.2: The Ust'-Ishim outcrop (indicated by the arrow); fossils are found mostly on the eroded bank of the second terrace of the Irtysh River. Photo by A.A. Bondarev, 2013.

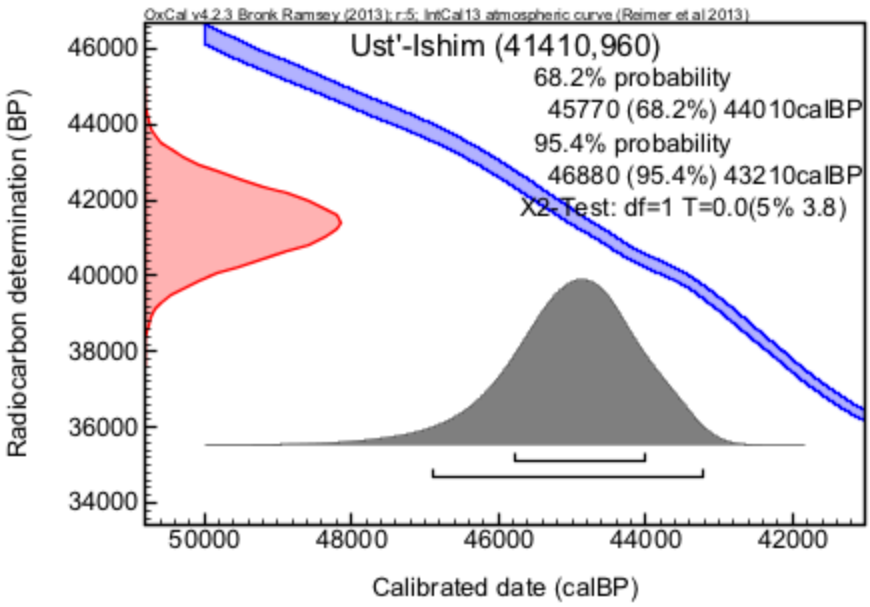


Figure S1.3.: The two radiocarbon determinations are combined prior to calibration using the R-combine function of the OxCal platform. This is a justified practise given that the dates derive from the same individual. The new combined value is 41410 ±960 BP, or 45770-44010 cal BP (68.2%) / 46880-43210 cal BP (95.4%) after calibration.

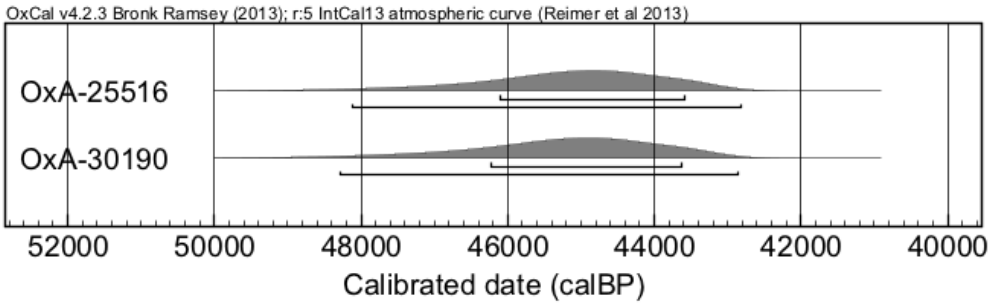


Figure S1.4.: Calibration of the two results from Ust'-Ishim.

Supplementary Information 2

Early modern humans in Eurasia

Susan G. Keates*, Dmitry I. Razhev*, Sergei M. Slepchenko, Yaroslav V. Kuzmin and Bence Viola
* To whom correspondence should be addressed (archres2010@t-online.de & rajevd0@gmail.com)

1. The earliest modern humans in Eurasia. The dating of the earliest modern human fossils in Eurasia is of crucial importance for the understanding of the modern human migration out of Africa. In order to estimate the age of these fossils, two main approaches have been used. The first one was to determine the age of a human fossil by direct radiometric dating, applying the radiocarbon (^{14}C) and electron spin resonance (ESR) methods. The second approach was to apply radiometric dating methods to associated material (charcoal, animal bones, etc.), or use statistical models of several associated dates to infer the age of the fossil[1,2]. Obviously, the first approach is more reliable than the second one. In the early 2000s, numerous supposedly Upper Palaeolithic fossils were redated using direct dating, and a large number proved to derive from the Holocene[3,4]. Therefore, the ages of human fossils based only on dates of associated material should be treated with caution.

If one only considers the more reliable direct dates, then, the earliest modern humans in Eurasia are from Tianyuan Cave and Peştera cu Oase, ca. 39,500–39,100 cal BP (Table SI2.1, Fig. SI2.1). Two other finds, Kostenki 14 and Kent's Cavern 4 are slightly younger, at ca. 37,800–35,400 cal BP.

In terms of indirect dating of early modern humans in Eurasia, the Asian fossils from Laibin and Baoyiyan Cave are of similar antiquity as the directly ^{14}C-dated Tianyuan Cave. In Europe, two deciduous teeth from Uluzzian layers of the Grotta di Cavallo were proposed to have modern human affinities, and date to about 43-44 ka cal BP based on ^{14}C dates on associated shell carbonates[1]. Similarly, it has been proposed that the Kent's Cavern 4 maxilla could date to 42-43 ka cal BP using Bayesian modeling of dates on associated animal remains[2], even though this date was called into question on the base of stratigraphic and documentation issues with the excavation of the site in the 1920s[5].

In Asia, most early modern human remains are dated indirectly. The majority of dates based on materials associated with human fossils are insecure because of apparent ambiguity of association between human fossils and the dated material (see Shiyu, Ziyang, Liujiang, Yuanyang, Lianhua Cave, and Nanshan Cave sites). In other cases, unknown or uncertain provenance (see Xiaohui Cave, Lianhua Cave, Huanglong Cave, Xianren Cave, Bailian Cave, and Zhangkou Cave sites) or large age discrepancy (see Tubo site) indicate that most or all of the earliest dates cannot be accepted at face value (for details see [6]).

A mandibular symphysis, and other fragments from Zhiren Cave in South China are another possible early modern human from China[7]. The fossils have been indirectly dated by U-series to more than 106,000 years ago based on material located above the human finds. The specimens show a mixture of modern and archaic features, and thus their fully modern status is questionable.

Apparently modern human fossils were also described from Tam Pa Ling Cave in Laos, with an age estimate of more than 46,000 years[8]. A direct U/Th date of 61 ka was obtained on the human remains, but the indirect ^{14}C and OSL dates of between 45 and 50 ka on the deposits are considerably younger. The dates have been questioned[9], but the Tam Pa Ling specimen seems to be one of the earliest modern humans in mainland Asia.

In Table S2.1 and Figure S2.1 we list direct, and persuasive indirect ^{14}C dates for modern humans predating 30 ka cal BP from Eurasia. The Ust'-Ishim fossil is the oldest directly ^{14}C-dated modern human in Northern Eurasia, and among the oldest modern humans dated directly or indirectly.

References

1. Benazzi, S., Douka, K., Fornai, C., Bauer, C. C., *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479,** 525-528 (2011).
2. Higham, T., Compton, T., Stringer, C., Jacobi, R., *et al.* The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* **479,** 521-524 (2011).
3. Keates, S. G., Hodgins, G. W. L., Kuzmin, Y. V. & Orlova, L. A. First direct dating of a presumed Pleistocene hominid from China: AMS radiocarbon age of a femur from the Ordos Plateau. *J Hum Evol* **53,** 1-5 (2007).
4. Street, M., Terberger, T. & Orschiedt, J. A critical review of the German Paleolithic hominin record. *J Hum Evol* **51,** 551-579 (2006).
5. White, M. & Pettitt, P. Ancient digs and modern myths: The age and context of the Kent's Cavern 4 maxilla and the earliest *Homo sapiens* specimens in Europe. *European Journal of Archaeology* **15,** 392-420 (2012).
6. Keates, S. G. The chronology of Pleistocene modern humans in China, Korea, and Japan. *Radiocarbon* **52,** 428-465 (2010).
7. Liu, W., Jin, C.-Z., Zhang, Y.-Q., Cai, Y- J., *et al.* Human remains from Zhirendong, South China, and modern human emergence in East Asia. *Proc Natl Acad Sci U S A* **107,** 19201-19206 (2010).
8. Demeter, F., Shackelford, L. L., Bacon, A. M., Duringer, P., *et al.* Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc Natl Acad Sci U S A* **109,** 14375-14380 (2012).
9. Pierret, A., Zeitoun, V. & Forestier, H. Irreconcilable differences between stratigraphy and direct dating cast doubts upon the status of Tam Pa Ling fossil. *Proc Natl Acad Sci* **109,** E3523 (2012).

Table S2.1. Ages of directly ^{14}C dated modern human fossils from Eurasia outside the Levante, calibrated using the IntCal13 curve[16] in OxCal 4.2.3 [17]

Specimen	LabNo	Age (14C BP)	Cal BP (68%)	Cal BP (95.4%)	Reference
Cavallo C (associated)	OxA-19242	39990 ± 340	43950 - 43259	44329 - 43015	[1]
Kents Cavern 4 model		Bayesian model	43110 - 41890	44180 - 41530	[2]
Kent's Cavern 4	OxA-1621	30900 ± 900	35800 - 34050	37431 - 33407	[3]
La Crouzade VI	Erl-9415	30640 ± 640	35192 - 34029	36020 – 33655	[4]
Paviland 1	OxA-16412	28870 ± 180			[5]
Paviland 1	OxA-16502	28400 ± 320	33460 – 33065 (group calibration)	33608 – 32872 (group calibration)	[5]
Paviland 1	OxA-16503	28820 ± 340			[5]
Paviland 1	OxA-16413	29490 ± 210			[5]
Vilhonneur 1	Beta-216141	27110 ± 210	31241 - 30983	31375 - 30840	[6]
Peştera cu Oase 1	GrA-22810	34290 + 970 - 870	39986 - 37678	40968 - 36532	[7]
Mladeč 9a	VERA-3076A	31500 +420 -400	35810 - 34960	36250 - 34656	[8]
Mladeč 2	VERA-3074	31320 + 410 -390	35622 - 34816	36110 - 34532	[8]
Mladeč 1	VERA-3073	31190 +400 -390	35503 - 34717	35995 - 34396	[8]
Mladeč 8	VERA-3075	30680 + 380 -360	34948 - 34261	35380 - 33979	[8]
Mladeč 25c	VERA-2736	26330 ± 170	30839 - 30477	30980 - 30233	[8]
Obłazowa Cave	OxA-4586	31000 ± 550	35508 - 34433	36104 - 34016	[9]
Peştera Muierii 1	LuA-5229	30150 ± 800	35025 - 33524	36080 - 32654	[10]
Peştera Muierii 2	OxA-16252	29110 ± 190	33590 - 33107	33755 - 32841	[10]
Peştera Cioclovina	LuA-5229	29000 ± 700	33855 - 32250	34405 - 31469	[10]
Kostenki 14	Oxa-X-2395-15	33250 ± 500	38211 - 36816	38684 - 36262	[11]
Kostenki 1	OxA-15055	32070 ± 190	36195 - 35762	36374 - 35529	[12]
Buran-Kaya III	GrA-37938	31900 ± 230	36085 - 35559	36296 - 35262	[13]
Sungir 2	OxX-2395-6 Hyp	30100 ± 550	34655 - 33740	35283 - 33185	[11]
Sungir 3	OxX-2395-7 Hyp	30000 ± 550	34596 - 33665	35154 - 33031	[11]
Ust'-Ishim 1	OxA-25516 & 30190	41410 ± 960	45470 - 44010	46880 - 43210	This paper
Pokhrovka / Maly Log 2	OxA-19850	27740 ± 150	31595 - 31305	31822 - 31182	[14]
Tianyuan Cave	BA-03222	34430 ± 510	39556 - 38451	40254 - 37761	[15]

1. Benazzi, S., Douka, K., Fornai, C., Bauer, C. C., *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479,** 525-528 (2011).
2. Higham, T., Compton, T., Stringer, C., Jacobi, R., *et al.* The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* **479,** 521-524 (2011).
3. Hedges, R. E. M., Housley, R. A., Law, I. A. & Bronk, C. R. Radiocarbon dates from the Oxford AMS system: Archaeometry datelist 9. *Archaeometry* **31,** 207-234 (1989).
4. Henry-Gambier, D. & Sacchi, D. La Crouzade V-VI (Aude, France): Un des plus anciens fossiles d'anatomie moderne en Europe occidentale. *Bulletins et Mémoires de la Société d'Anthropologie de Paris* (2008).
5. Jacobi, R. M. & Higham, T. F. The "Red Lady" ages gracefully: new ultrafiltration AMS determinations from Paviland. *J Hum Evol* **55,** 898-907 (2008).
6. Henry-Gambier, D., Beauval, C., Airvaux, J., Aujoulat, N., *et al.* New hominid remains associated with Gravettian parietal art (Les Garennes, Vilhonneur, France). *J Hum Evol* 53, 747-750 (2007).
7. Trinkaus, E., Moldovan, O., Milota, S., Bîlgăr, A., *et al.* An early modern human from the Peştera cu Oase, Romania. *Proc Natl Acad Sci U S A* **100,** 11231-11236 (2003).
8. Wild, E. M., Teschler-Nicola, M., Kutschera, W., Steier, P., *et al.* Direct dating of early Upper Palaeolithic human remains from Mladec. *Nature* **435,** 332-335 (2005).
9. Valde-Nowak, P., Nadachowski, A., Madeyska, T., *et al. Obłazowa cave* (Institute of Archaeology and Ethnology, Polish Academy of Science, Krakow, 2003).
10. Soficaru, A., Doboş, A. & Trinkaus, E. Early modern humans from the Peştera Muierii, Baia de Fier, Romania. *Proc Natl Acad Sci U S A* **103,** 17196-17201 (2006).

11. Marom, A., McCullagh, J. S. O., Higham, T. F. G., Sinitsyn, A. A. & Hedges, R. E. M. Single amino acid radiocarbon dating of Upper Paleolithic modern humans. *Proc Natl Acad Sci U S A* **109,** 6878-6881 (2012).
12. Higham, T. F. G., Jacobi, R. M. & Bronk Ramsey, C. AMS radiocarbon dating of ancient bone using ultrafiltration. *Radiocarbon* **48,** 179-195 (2006).
13. Prat, S., Péan, S. C., Crépin, L., Drucker, D. G., *et al.* The oldest anatomically modern humans from far Southeast Europe: Direct dating, culture and behavior. *PLoS One* **6,** e20834 (2011).
14. Akimova, E. V., Higham, T., Stasyuk, I., Buzhilova, A., *et al.* A new direct radiocarbon AMS date for an Upper Palaeolithic human bone from Siberia. *Archaeometry* **52,** 1122-1130 (2010).
15. Shang, H., Tong, H., Zhang, S., Chen, F. & Trinkaus, E. An early modern human from Tianyuan Cave, Zhoukoudian, China. *Proc Natl Acad Sci U S A* **104,** 6573-6578 (2007).
16. Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. *Radiocarbon* **55,** 1869-1887 (2013).
17. Bronk Ramsey, C., Dee, M., Lee, S., Nakagawa, T. & Staff, R. Developments in the calibration and modelling of radiocarbon dates. *Radiocarbon* **52,** 953 (2010).

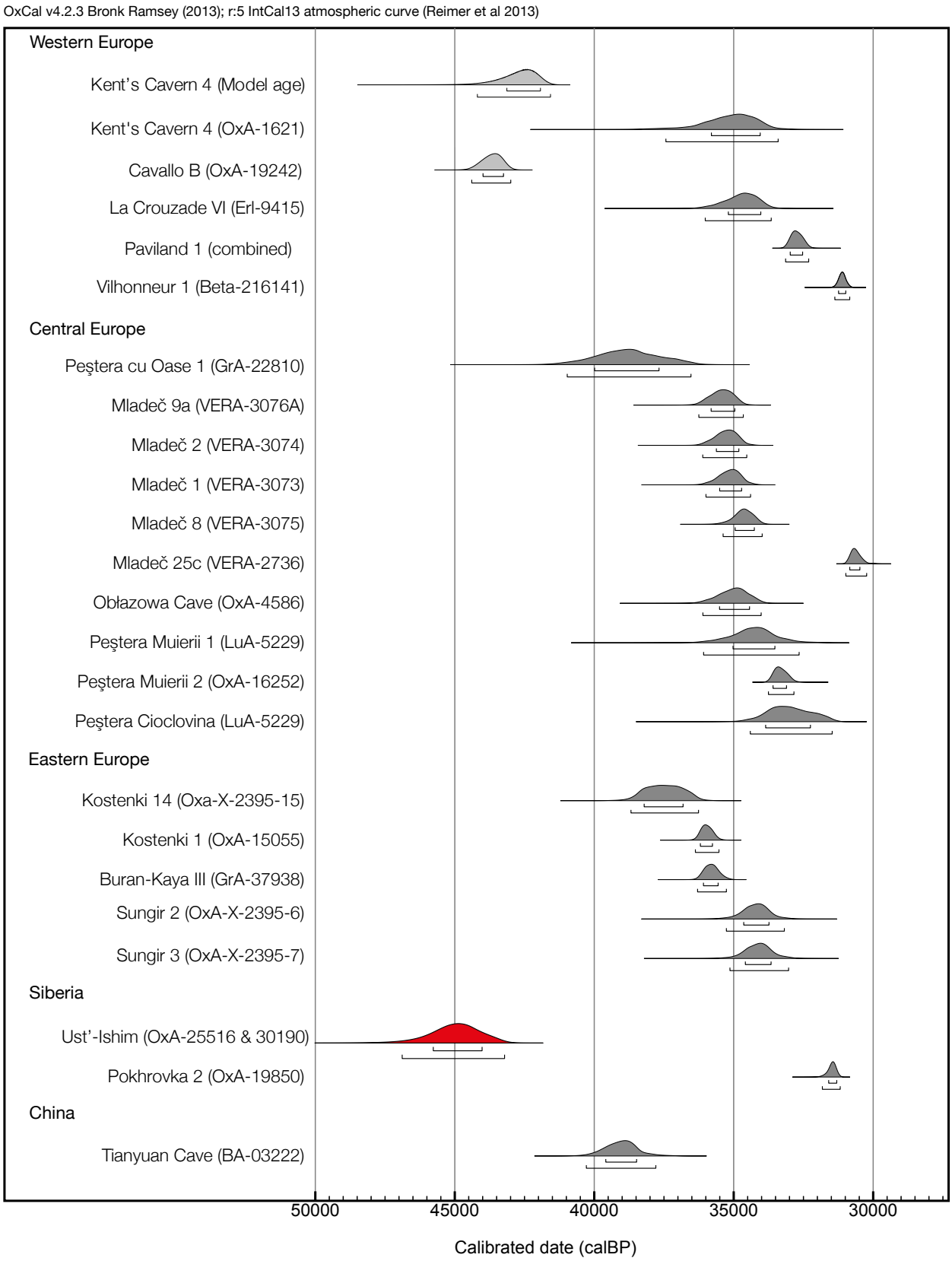


Figure S2.1: Comparison of calibrated radiocarbon ages of the earliest modern human fossils in Northern Eurasia. Specimens in light grey are only indirectly dated. For references see Table S2.1.

Supplementary Information 3

The morphology of the Ust'-Ishim femur

Bence Viola*, Sergey Slepchenko, Dimitry Razhev and Jean-Jacques Hublin
* To whom correspondence should be addressed (bence.viola@eva.mpg.de)

Ust'-Ishim 1 is a relatively complete left femoral diaphysis, with a maximum preserved length of about 328 mm (Fig. S3.1.). Proximally, the femur is broken at the level of the lesser trochanter. On the posterior side only the base of the lesser trochanter is preserved, while anteriorly the fracture is slightly higher, just below the level of the femoral tubercle. The distal fracture line is about 75 mm distal of the separation of the medial and lateral lips of the *linea aspera* on the posterior surface, anteriorly the shaft is preserved about 20 mm further distally.

The preservation of the bone surface is relatively good on the medial half of the diaphysis, but the lateral half of the shaft shows large (up to 5 mm) pits, especially near the distal end and in the area of the gluteal buttress. These pits are likely due to fluvial transport of the specimen.

Morphology. The gluteal tuberosity is clearly visible on the posterolateral surface of Ust'-Ishim 1. This area is strongly damaged by pitting, making the observation of the rugosity and fine morphology hard, but at about the level of the broken lesser trochanter, a distinct oval raised area denotes a small third trochanter (14 mm si x 6 mm ml). From here a raised ridge, the gluteal ridge runs distally, joining the lateral lip of the *linea aspera* after about 69 mm. There is a marked gluteal buttress in Ust'-Ishim 1 laterally of the gluteal tuberosity, separated by a slight sulcus that can be palpated in the middle course of the gluteal ridge. The spiral line is well expressed from the level of the lesser trochanter distally, while the pectineal line is not discernible.

The strong gluteal buttress makes the shaft appear slightly curved mediolaterally. The shaft is relatively strongly bent anteroposteriorly, even though this cannot be quantified due to the position of the distal break.

The midshaft shows a large pilaster on the posterior surface. Both lips of the *linea aspera* are marked, and separated by about 5 mm proximally of the midshaft and about 7.5 mm near the nutrient foramen. Laterally, the pilaster is delimited by a wide sulcus near the midshaft, which becomes less distinct distally but is present until the distal end of the fragment. On the medial side, a slight narrow sulcus is present in the midshaft area. A nutrient foramen pierces the pilaster about 40 mm distally of the midshaft.

At the distal fracture end, the anteromedial surface of the shaft is flattened, already leading up to the patellar surface. Medially the shaft is slightly flaring and the medial supracondylar line curves strongly medially, indicating that the fragment reaches the top of the popliteal surface.

Comparative morphology. The morphology of the proximal shaft in Ust'-Ishim 1 is similar to that seen in Upper Palaeolithic modern humans. While Neandertals and Early Anatomically Modern Humans (Early AMH, consisting of the Skhul and Qafzeh samples) tend to have lesser development of the gluteal buttress, and thus more rounded shafts[1], Upper Palaeolithic femora are mediolaterally expanded[2]. The meric index, based on the mediolateral and anteroposterior diameters of the shaft at the maximum extent of the gluteal buttress shows Ust'-Ishim 1 near the mean of Upper Palaeolithic modern humans, but about one standard deviation below Neandertals and Early AMH (Table S3.1. , Fig. S3.2.).

It has been shown[3] that though femoral curvature does not distinguish between Neandertals and Early anatomically modern humans or early Upper Palaeolithic humans, all of these populations have stronger femoral curvature than present day populations, probably due to increased activity levels in these groups. Thus, the relatively high curvature of Ust'-Ishim 1 does not give any indications of its affinities. The morphology of the midshaft is rather distinct in Neandertals and modern humans. Neandertal femora have a relatively rounded cross section without pilaster, with the two lips of the linea aspera frequently indistinct[4]. In contrast to this, most Upper Palaeolithic humans and Early AMH have in general a teardrop shaped cross section, with a marked pilaster[2]. Ust'-Ishim 1 conforms closely to the Upper Palaeolithic morphology. Its pilastric index is at the upper limit of the Neandertal range of variation, slightly higher than the earlier Upper Palaeolithic sample, near the average of the post-LGM Upper Palaeolithic sample, and somewhat lower than the Early AMHs (Table S3.1., Fig. S3.3.).

The cross sectional geometry of the midshaft, reflecting habitual loading of the diaphysis[5,6] also shows strong differences between Neandertals and modern humans. Taking the second moment of area as an approximation of bending strength, Neandertals in general having shafts that resist mediolateral bending better, while modern human femoral shafts are more resistant to anteroposterior bending[1,2]. Ust'-Ishim 1 is similar to Upper Palaeolithic and Early AMHs in showing an increased anteroposterior bending strength (Table S3.2, Fig. S3.4.).

Length estimation. The total length of the specimen cannot be easily estimated using regressions, as no clear landmarks are preserved distally. Its general size and morphology, as well as the proximodistal contours of the shaft in comparison to other femora allow us to get a reasonable estimate of its approximate length.

We used a geographically varied sample of 20 femora from the collections of the Department of Anthropology, University of Vienna and of the Natural History Museum Vienna, that were CT-scanned and/or surface scanned for another project. Using Avizo 7.1 (VSG) we fitted the femora of the comparative sample to Ust'-Ishim 1, and scaled them using affine scaling to fit the fossil as closely as possible. Differences in the anteroposterior curvature of the femora influence the estimated length, but a maximum length of between 430 and 460 mm, and a biomechanical length of about 410-430 mm seems most likely.

Methods. Osteometric measurements were taken following Martin's system[7]. For the measurement of the subtrochanteric diameters we followed Sládek and colleagues' recommendation[8], taking the maximum diameter at the most developed point of the gluteal buttress as the mediolateral diameter, and the anteroposterior diameter perpendicularly to it.

The specimen was µCT-scanned using a BIR Actis 225/300 µCT at the Max-Planck-Institute of Evolutionary Anthropology, and the scans were reconstructed to a resolution of 0.078 mm isovoxel. Avizo 7.1 was used to create sections through the midshaft, analogous to the methodology of Ruff and Hayes[5]. Cross sectional properties were measured on exported .tif slices in ImageJ 1.47 using MomentMacro 1.4.

Comparative data for the external and cross sectional measurements came primarily from the literature.

Comparative samples

Neandertals: Amud[9,10]; Abri Bourgeois-Delaunay[11]; Font de Forêt[10,12], Hortus[13], Krapina[13], La Chapelle aux Saints[13], La Ferrassie[13], La Quina 5[14], Les Pradelles[15], Feldhofer Grotte[13], Rocher de Villeneuve[4], Shanidar[1,9], Spy[10,13], St. Césaire[4], Tabun[9,13,16].

Early AMH: Qafzeh[16], Skhul[13,16].

Pre-LGM Upper Palaeolithic: Arene Candide[17], Baosso da Torre[17], Barma Grande[13], Brno[13], Byci Skála[13], Caviglione[17], Cro-Magnon[13,17], Dolní Vestonice[8], Grotte des Enfants[13,17], La Rochette[13], Le Roc[13], Mladeč[2], Nahal ein Gev[17], Nazlet Khater[18], Paglicci[17], Parabita[17], Pataud[13], Paviland[19], Pavlov[8], Predmost[13], Sordes[13], Sungir[20], Willendorf [21].

Post-LGM Upper Palaeolithic: Arene Candide[17], Bichon[17], Bruniquel[17], Cap Blanc[13], Chancelade[13], Grotte des Enfants[17], La Madeleine[13,17], Laugerie Basse[17], Le Peyrat[17], Le Placard[17], Neussing[17], Obercassel[13], Riparo Continenza[17], Riparo Tagliente[17], Rochereil[17], Romanelli[17], Romito[17], San Teodoro[17], St. Germain[17], Veyrier[13].

Figure S3.1. The Ust'-Ishim 1 femur in ventral (a), lateral (b), dorsal (c) and medial (d) view. Cross section at 50% biomechanical length (e).

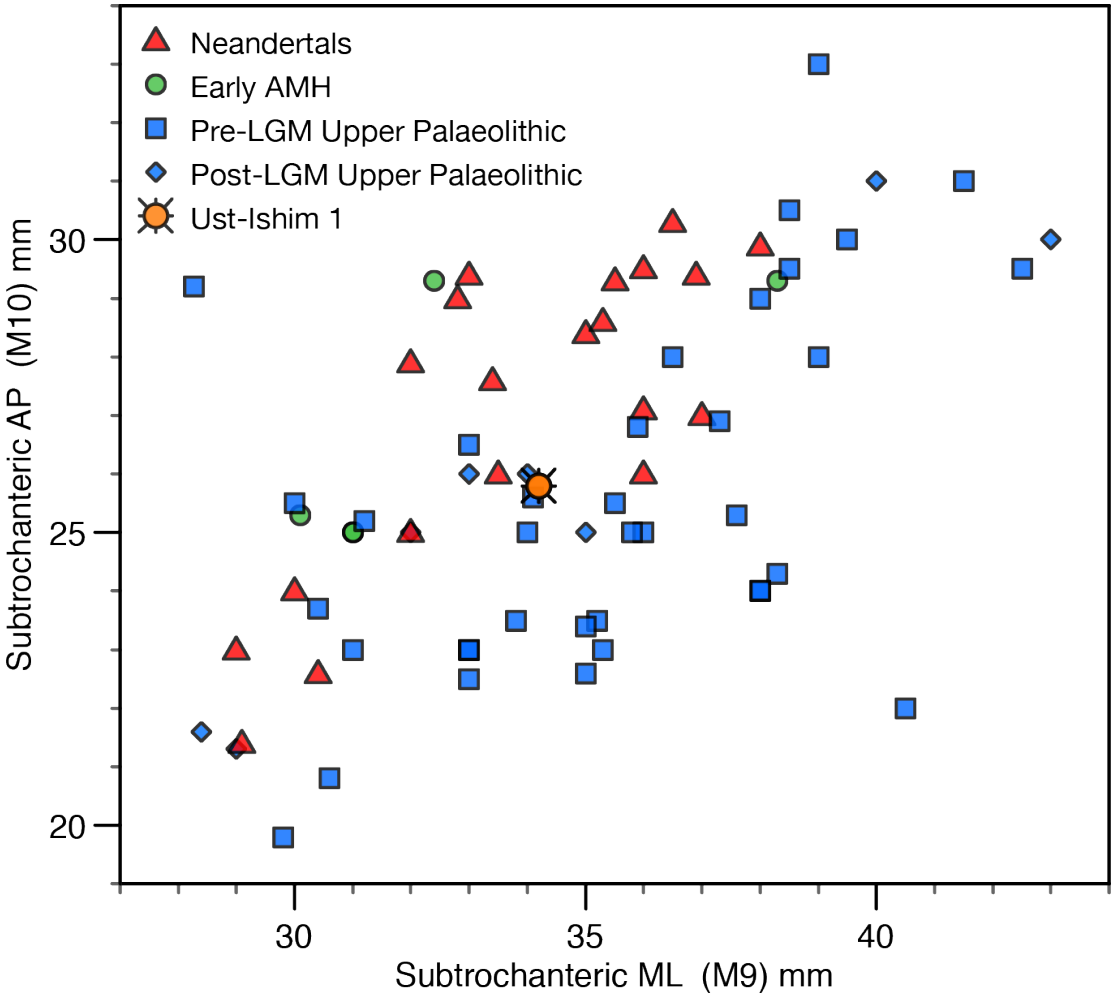


Figure S3.2. Biplot of the anteroposterior and mediolateral diameters at the subtrochanteric level in Neandertals, Early AMH, Post and Pre-LGM Upper Palaeolithic humans compared to Ust'-Ishim 1.

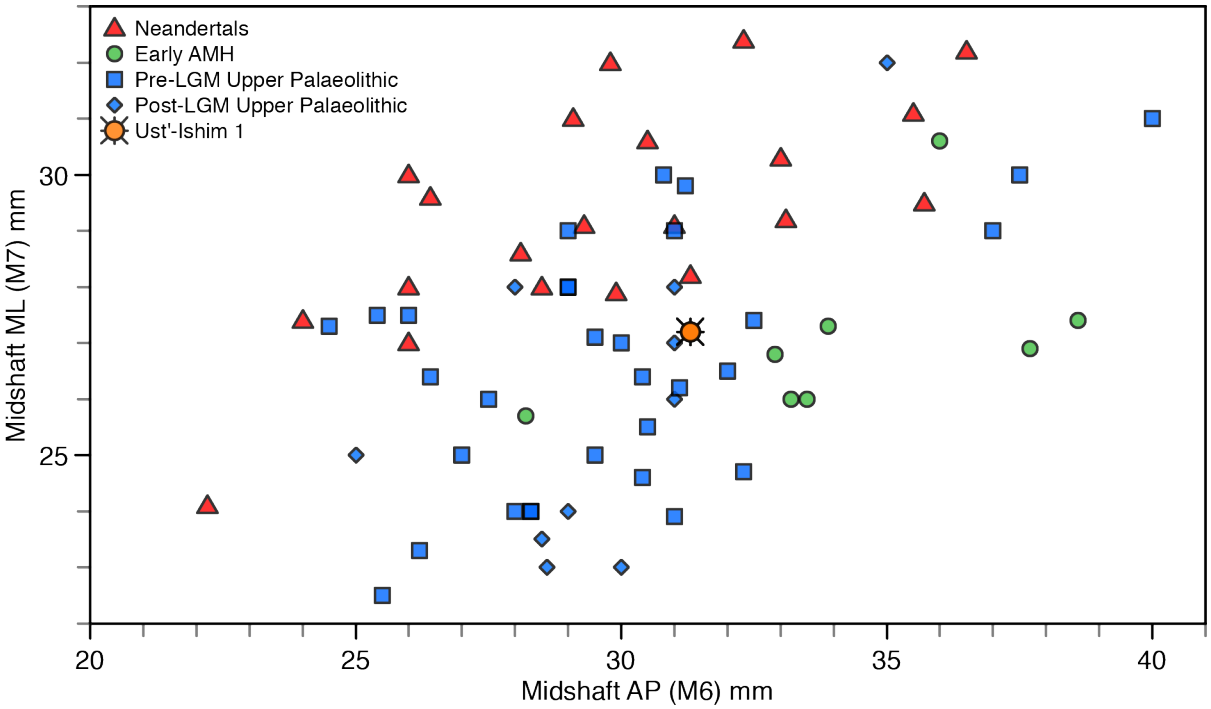


Figure S3.3. Biplot of the anteroposterior and mediolateral diameters of the midshaft in Neandertals, Early AMH, Post and Pre-LGM Upper Palaeolithic humans compared to Ust'-Ishim 1.

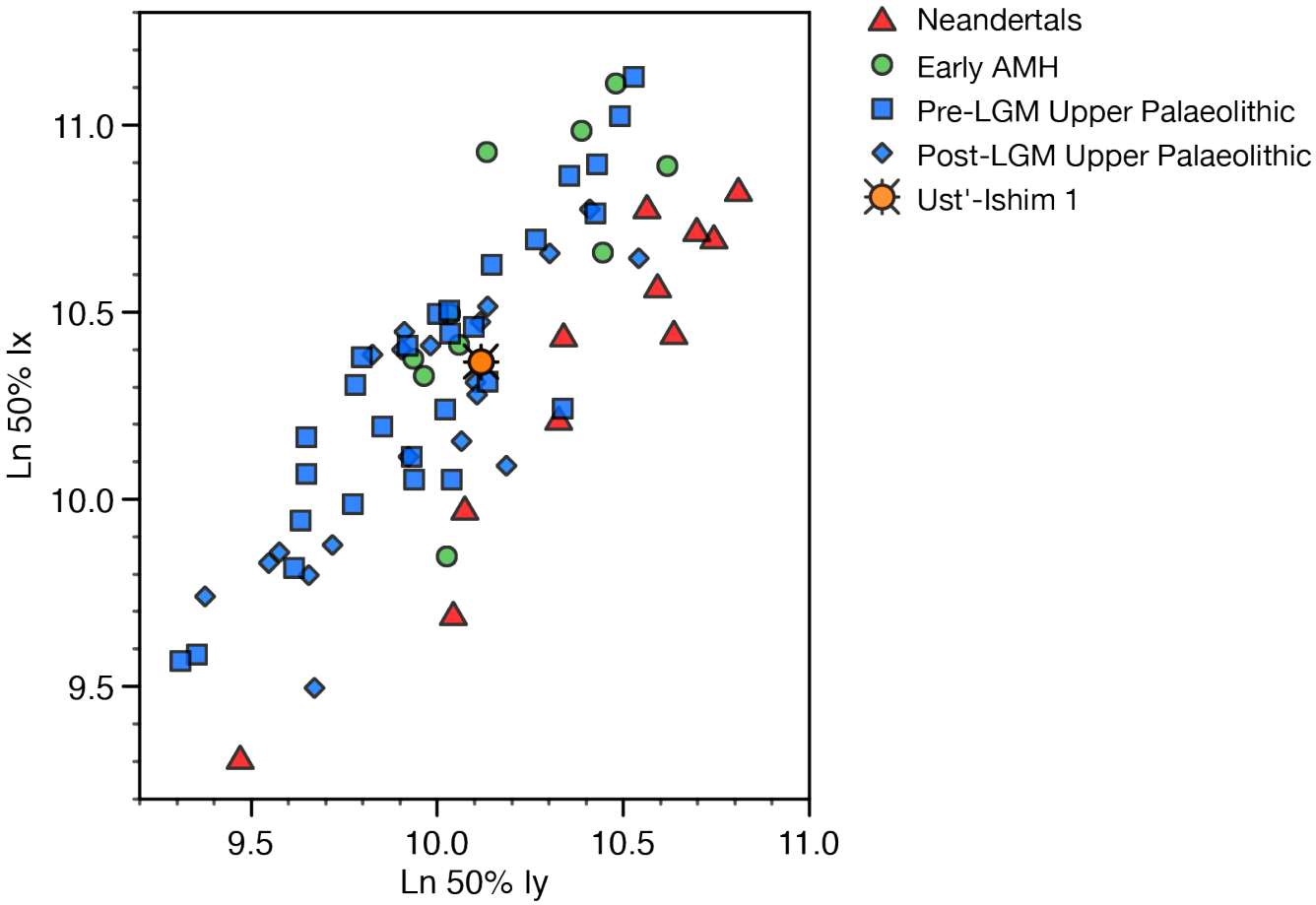


Figure S3.4. Biplot of the natural logarithms of the anteroposterior and mediolateral second moments of area of the midshaft in Neandertals, Early AMH, Post and Pre-LGM Upper Palaeolithic humans compared to Ust'-Ishim 1.

Table S3.1. Metric comparisons of the Ust'-Ishim 1 femur to Neandertals, Early anatomically modern humans, Early Upper Palaeolithic modern humans (pre-Last Glacial Maximum) and Late Upper Palaeolithic modern humans (post-LGM).

	Midshaft AP (M6)	Midshaft ML (M7)	Pilastric index (M6/M7)	Subtrochanteric ML (M9)	Subtrochanteric AP (M10)	Meric index (M10/M9)
Neandertals	29.7±3.8 (21)*	29.6±2.3 (22)	101.3±9.8 (21)	33.9±2.8 (20)	27.1±2.7 (22)	79.9±4.9 (20)
Early AMH	34.3±3.3 (8)	27.1±1.6 (8)	126.5±10.5 (8)	32.6±3.3 (5)	26.8±2.3 (5)	82.5±5.2 (5)
Early UP	30±3.5 (34)	26.6±2.1 (35)	112.4±10.8 (33)	35.0±3.9 (41)	25.6±3 (39)	73.9±10.3 (39)
Late UP	29.7±2.6 (10)	26±2.9 (10)	115.1±10.1 (10)	34.3±5.1 (8)	25.7±3.5 (8)	75.2±3.3 (8)
Ust'-Ishim 1	31.3	27.2	115.1	34.2	25.8	75.4

*$(\bar{x} \pm sd(n))$

Table S 3.2. Cross-sectional geometric properties of Ust'-Ishim 1, Neandertals, Early Anatomically modern humans, Early Upper Palaeolithic modern humans (pre-Last Glacial Maximum) and Late Upper Palaeolithic modern humans (post-LGM). a: Cross sectional properties at the 50% level; b: Cross sectional properties at the 80% level

a.

	50% TA	50% CA	50% CA/TA	50% Ix	50% Iy	50% J
Neandertals	652±129 (11)*	532±104 (11)	81,6±4,1 (11)	33656±13244 (11)	34658±11377 (11)	66763±26533 (11)
Early AMH	657±95(10)	510±74 (10)	77,8±4,4 (10)	42881±15247 (10)	27900±7210 (10)	61793±20421 (10)
Early UP	587±99 (28)	440±89 (26)	75,8±7,2 (26)	33162±13533 (28)	22765±7229 (28)	55928±20272 (28)
Late UP	558±83(20)	448±84 (20)	80,0±6,8 (20)	28833±9577 (20)	21926±6670 (20)	50760±15589 (20)
Ust'-Ishim 1	596	498	83,4	31807	24785	56592

*$(\bar{x} \pm sd(n))$

b.

	80% TA	80% CA	80% CA/TA	80% Imax	80% Imin	80% J
Neandertals	796±150 (5)	587±104 (5)	74,7±10,7 (5)	57637±18258 (5)	39629±12981(5)	97266±30349 (5)
Early AMH	696±120 (3)	574±116 (3)	82,2±4,4 (3)	49921±24685 (3)	29218±6664 (3)	79139±30745 (3)
Early UP	650±107 (26)	481±95 (26)	74±6,4 (26)	44550±15576 (26)	23541±8326 (26)	68087±22811 (26)
Late UP	618±96 (18)	480±70 (18)	78,3±9,7 (18)	40319±12473 (18)	21344±6737 (18)	61680±18107 (18)
Ust-Ishim 1	762	539	70,7	51379	34594	85974

*$(\bar{x} \pm sd(n))$

References

1. Trinkaus, E. & Ruff, C. B. Diaphyseal cross-sectional geometry of Near Eastern Middle Palaeolithic humans: The Tibia. *Journal of Archaeological Science* **26,** 1289-1300 (1999).
2. Trinkaus, E., Smith, F. H., Stockton, T. C. & Shackelford, L. L. in *Early Modern Humans at the Moravian Gates* (ed. Teschler-Nicola, M.) 385-446 (Springer, Vienna, 2007).
3. Shackelford, L. L. & Trinkaus, E. Late Pleistocene human femoral diaphyseal curvature. *Am J Phys Anthropol* **118,** 359-370 (2002).
4. Beauval, C., Maureille, B., Lacrampe-Cuyaubère, F., Serre, D., *et al.* A late Neandertal femur from Les Rochers-de-Villeneuve, France. *Proc Natl Acad Sci U S A* **102,** 7085-7090 (2005).
5. Ruff, C. B. & Hayes, W. C. Cross-sectional geometry of Pecos Pueblo femora and tibiae - a biomechanical investigation: I. Method and general patterns of variation. *Am J Phys Anthropol* **60,** 359-381 (1983).
6. Ruff, C. B. Biomechanical analyses in archaeological human skeletons. In *Biological Anthropology of the Human Skeleton* (eds Katzenberg, M. A. & Saunders, S. R.) 183-206 (Wiley-Liss, Hoboken, 2008).
7. Bräuer, G. in *Anthropologie* (ed Knussmann, R.) 160-192 (Gustav Fischer Verlag, Stuttgart, 1988).
8. Sládek, V., Trinkaus, E., Hillson, S., Holliday, T. W. & Svoboda, J. *The People of the Pavlovian: Skeletal Catalogue and Osteometrics of the Gravettian Fossil Hominids from Dolní Vestonice and Pavlov* (Acad. of Sciences of the Czech Republic, Inst. of Archaeology, Brno, 2000).
9. Trinkaus, E. *The Shanidar Neanderthals* (Academic Press, New York, 1983).
10. Trinkaus, E. & Ruff, C. Diaphyseal cross-sectional morphology and biomechanics of the Fond-de-Forêt 1 femur and the Spy 2 femur and tibia. *Bulletin de la Societé Royale Belge d'Anthropologie et de Prehistoire* **100,** 33-42 (1989).
11. Condemi, S. *Les Néanderthaliens de La Chaise (Abri Bourgeois-Delaunay)* (Éditions du Comité des Travaux Historiques et Scientifiques, Paris, 2001).
12. Twiesselmann, F. *Le fémur néanderthalien de Fond-de-Forêt (Province de Liège)* (Institut Royal des Sciences Naturelles de Belgique, 1961).
13. Trinkaus, E. The evolution of the hominid femoral diaphysis during the Upper Pleistocene in Europe and the Near East. *Zeitschrift für Morphologie und Anthropologie* **67,** 291-319 (1976).
14. Martin, H. *Recherches sur l'évolution du Moustérien dans le gisement de La Quina (Charente) III: L'homme fossile* (Libr. Octave Doin, Paris, 1923).
15. Mussini, C., Crevecoeur, I., Garralda, M. D., Mann, A. & Maureille, B. A new Neandertal femoral diaphysis from Les Pradelles (Marillac-le-Franc, Charente, France). *Periodicum biologorum* **114,** 117-123 (2012).
16. Trinkaus, E. & Ruff, C. B. Diaphyseal cross-sectional geometry of Near Eastern Middle Palaeolithic humans: The femur. *Journal of Archaeological Science* **26,** 409-424 (1999).
17. Shackelford, L. L. *Regional variation in the postcranial robusticity of Late Upper Palaeolithic humans* (PhD Thesis, Washington University St. Louis, St. Louis, MO, 2005).
18. Crevecoeur, I. *Étude anthropologique du squelette du Paléolithique supérieur de Nazlet Khater 2 (Égypte)* (Leuven University Press, Leuven, 2008).
19. Trinkaus, E. & Holliday, R. in *Paviland Cave and the 'Red Lady': a definitive report* (ed. Aldhouse-Green, S. H. R.) (Western Academic and Specialist Press, Bristol, 2000).
20. Mednikova, M. & Trinkaus, E. Femoral midshaft diaphyseal cross-sectional geometry of the Sunghir 1 and 4 Gravettian human remains. *Anthropologie* **39,** 103-109 (2001).
21. Teschler-Nicola, M. & Trinkaus, E. Human remains from the Austrian Gravettian: the Willendorf femoral diaphysis and mandibular symphysis. *J Hum Evol* **40,** 451-465 (2001).

Supplementary Information 4

Stable isotope analyses of the Ust'-Ishim femur

Domingo C. Salazar-García*, Michael P. Richards
* To whom correspondence should be addressed (domingo_carlos@eva.mpg.de)

1. Palaeodietary isotopic reconstructions

Carbon and nitrogen stable isotope analysis is a commonly used method for the reconstruction of past human and animal diets. This technique is based on the principle that the isotopic composition of the food eaten by both animals and humans is recorded in their body tissues after a predictable isotope fractionation[1-3].

The consumption of terrestrial (^{13}C depleted) and marine (^{13}C enriched) foods can be distinguished by $\delta^{13}C$, the relative abundance of stable carbon isotopes ^{13}C and ^{12}C[4]. This isotopic system also helps to define the input on the diet of C_3 (^{13}C depleted) and C_4 resources (^{13}C enriched)[5]. Since there are no edible C_4 plants reported in Siberia during the Palaeolithic[6], the main use of carbon stable isotopes is in this case for estimating the consumption of marine protein: a $\delta^{13}C$ value of around -20 ‰ would indicate a total terrestrial diet, while one about -12 ‰ would indicate that almost all ingested protein was marine. As there is an increase of between 3-5 ‰ in nitrogen isotopes of body tissues between consumer tissues and the protein consumed, $\delta^{15}N$ stable isotope ratios are useful for detecting the presence of high trophic level aquatic resources in the diet[7], as well as for distinguishing animal-rich diets from plant-rich diets[8-10].

Bone collagen is the preferred substrate for carbon and nitrogen stable isotope analysis. Accepted quality indicators are used to assess the preservation quality of the extracted collagen[11,12]. Because of the relatively slow rate of collagen turnover, collagen stable isotope values may reflect the average isotope values of food consumed over many years of life[13]. In addition, they reflect the isotopic signals of the main dietary protein sources rather than that of a diet as a whole[14].

2. Material and Methods

We removed a 300 mg sample from the distal end of the Ust'-Ishim 1 femur for carbon and nitrogen stable isotope analysis of bone collagen (Table SI4.1). Prior to analysis, visible contaminants were removed with aluminium oxide powder abrasion. Collagen extraction proceeded following[15], with the addition of an ultrafiltration step[16]. To summarize, the sample was demineralized in 0.5M HCl solution at 5 ºC over the course of two weeks, and was then rinsed three times with deionized water until the pH became neutral. This was followed by gelatinization over 48 hours at 70 ºC, and later by filtering with a 5 μm EZEE© filter and ultrafiltering with >30 kDa Amicon© ultrafilters. The purified solution was finally frozen and lyophilized before being weighed into tin capsules and loaded into the mass spectrometers.

The carbon and nitrogen isotope ratios in collagen were measured using a Delta XP continuous-flow isotope ratio mass spectrometer after being combusted in an elemental analyzer Flash EA 2112 that was interfaced with it (Thermo-Finnigan©, Bremen, Germany) at the Department of Human Evolution of the Max-Planck Institute for Evolutionary Anthropology (Leipzig, Germany). The sample was measured in duplicate. Stable carbon isotope ratios were expressed relative to the VPDB scale (Vienna PeeDee Belemnite) and stable nitrogen isotope ratios were measured relative to the AIR scale (atmospheric N_2), using the delta notation (δ) in parts per thousand (‰). Repeated analysis of internal and international standards determined an analytical error better than 0.2 ‰ (1σ) for $\delta^{13}C$ and $\delta^{15}N$.

3. Results

The stable isotope results for Ust'-Ishim 1 are presented in Table SI 4.1. The sample yielded enough collagen at the >30 kDa fraction, and met published quality controls[12,14]. Unfortunately, as the context of the human is unknown it was not possible to obtain contemporary fauna from the site to construct a comparative dataset to better understand the human isotope values.

The $\delta^{13}C$ and $\delta^{15}N$ values of Ust'-Ishim 1 are -19.0 ‰ and 14.2 ‰ respectively. The carbon value indicates the diet was based on terrestrial C_3 plants and also animals that consumed them. The nitrogen isotope value is relatively high compared to many early modern humans from Europe[17]. The high nitrogen isotope value could indicate that an important proportion of dietary protein was from aquatic foods, namely freshwater fish, as was suggested for Mesolithic humans from the Iron Gates gorge region of SE Europe[18] and for an early modern human from Peştera cu Oase, Romania[17]. As expected due to the find location of this individual, there is no indication from the carbon isotope value of any marine protein. When compared to other Palaeolithic modern human specimens (Table SI 4.2 and Fig. 4.), this individual has similar isotopic values to a number of specimens from Eastern Europe and Siberia, especially Kostenki 1 and 18[17,19]. This isotope result is intriguing, and could indicate a particular dietary adaptation to freshwater fish in the Eurasian Steppe region during this time period. However, in order to better assess the possible influence of climatic fluctuations on the isolated human nitrogen value as suggested for other regions of Europe [20], additional contemporary faunal isotopic baseline data is needed. This would allow to make more definitive conclusions about this individual's diet.

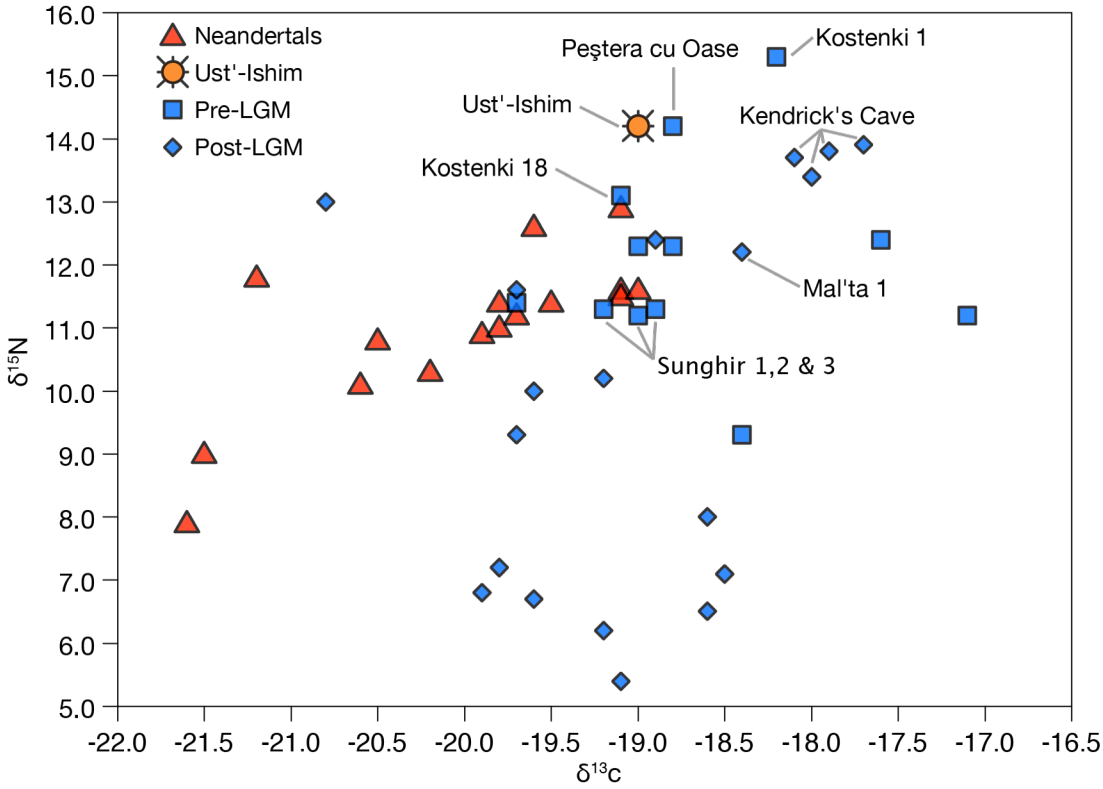


Figure S4.1. Carbon and nitrogen isotope values for Neandertals and Upper Palaeolithic modern humans in comparison with Ust'-Ishim 1. Samples from the Eurasian plain are labeled, as well as Late Palaeolithic specimens from Kendrick's cave that relied on marine protein.

Table SI 4.1. $\delta^{13}C$ and $\delta^{15}N$ values, collagen control indicators (%C, %N, C:N), S-EVA number

S-EVA	Species	Bone	$\delta^{13}C$ (‰)	$\delta^{15}N$ (‰)	%C	%N	C:N
28704	Human	femur	-19.0	14.2	43.3	15.7	3.2

Table SI 4.2. Radiocarbon dates, $\delta^{13}C$ and $\delta^{15}N$ values of other Palaeolithic modern humans from the Eurasian Steppe.

Site / individual	Radiocarbon date	$\delta^{13}C$ (‰)	$\delta^{15}N$ (‰)
Kostenki 1	32600 ± 1,100 BP	-18.2	15.3
Kostenki 18	21020 ± 180 BP	-19.1	13.1
Mal'ta 1	19880 ± 160 BP	-18.4	12.2
Sunghir 1	22930 ± 200 BP	-19.2	11.3
Sunghir 2	23830 ± 220 BP	-19.0	11.2
Sunghir 3	24100 ± 240 BP	-18.9	11.3

References

1. Katzenberg, M. A. in *Biological Anthropology of the Human Skeleton* (eds Katzenberg, M. A. & Saunders, S. R.) 305-328 (Wiley-Liss, New York, 2000).
2. Lee-Thorp, J. A. On isotopes and old bones. *Archaeometry* **50,** 925-950 (2008).
3. Sealy, J. in *Handbook of Archaeological Sciences* (eds Brothwell, D. R. & Pollard, A. M.) 269-279 (Wiley, Chichester, 2001).
4. Chisholm, B. S., Nelson, D. E. & Schwarcz, H. P. Stable-carbon isotope ratios as a measure of marine versus terrestrial protein in ancient diets. *Science* **216,** 1131-1132 (1982).
5. van der Merwe, N. J. & Vogel, J. C. ^{13}C content of human collagen as a measure of prehistoric diet in woodland North America. *Nature* **276,** 815-816 (1978).
6. Sage, R. F., Wedin, D. A. & Li, M. in *C4 plant biology* (eds Sage, R. F. & Monson, R. K.) 313-373 (Academic Press, 1999).
7. Schoeninger, M. J., DeNiro, M. J. & Tauber, H. Stable nitrogen isotope ratios of bone collagen reflect marine and terrestrial components of prehistoric human diet. *Science* **220,** 1381-1383 (1983).
8. DeNiro, M. J. & Epstein, S. Influence of diet on the distribution of nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta* **45,** 341-351 (1981).
9. Minagawa, M. & Wada, E. Stepwise enrichment of ^{15}N along food chains: further evidence and the relation between ^{15}N and animal age. *Geochimica et Cosmochimica Acta* **48,** 1135-1140 (1984).
10. Schoeninger, M. J. & DeNiro, M. J. Nitrogen and carbon isotopic composition of bone collagen from marine and terrestrial animals. *Geochimica et Cosmochimica Acta* **48,** 625-639 (1984).
11. DeNiro, M. J. Postmortem preservation and alteration of in vivo bone collagen isotope ratios in relation to palaeodietary reconstruction. *Nature* **317,** 806-809 (1985).
12. Van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *Journal of Archaeological Science* **26,** 687-695 (1999).
13. Hedges, R. E., Clement, J. G., Thomas, C. D. & O'Connell, T. C. Collagen turnover in the adult femoral mid-shaft: Modeled from anthropogenic radiocarbon tracer measurements. *Am J Phys Anthropol* **133,** 808-816 (2007).
14. Ambrose, S. & Norr, L. in *Prehistoric Human Bone: Archaeology at the Molecular Level* (eds Lambert, J. B. & Grupe, G.) 1-37 (Springer, Berlin, 1993).
15. Richards, M. P. & Hedges, R. E. M. Stable isotope evidence for similarities in the types of marine foods used by late Mesolithic humans at sites along the Atlantic coast of Europe. *Journal of Archaeological Science* **26,** 717-722 (1999).
16. Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen extraction by modified Longin method. *Radiocarbon* **30,** 171-177 (1988).
17. Richards, M. P. & Trinkaus, E. Isotopic evidence for the diets of European Neanderthals and early modern humans. *Proc Natl Acad Sci U S A* **106,** 16034-16039 (2009).
18. Nehlich, O., Borić, D., Stefanović, S. & Richards, M. P. Sulphur isotope evidence for freshwater fish consumption: a case study from the Danube Gorges, SE Europe. *Journal of Archaeological Science* **37,** 1131-1139 (2010).
19. Richards, M. P., Pettitt, P. B., Stiner, M. C. & Trinkaus, E. Stable isotope evidence for increasing dietary breadth in the European mid-Upper Paleolithic. *Proc Natl Acad Sci U S A* **98,** 6528-6532 (2001).

Supplementary Information 5

The Ust'-Ishim modern human: archaeological implications

Nicolas Zwyns*, Yaroslav Kuzmin, Bence Viola

* To whom correspondence should be addressed (nzwyns@ucdavis.edu)

The Ust'-Ishim remains were collected in the Western Siberian plain, a region that stretches from West to East between the Ural and the Altai ranges, and North to South between the Arctic Ocean and the Kazakh steppe. With about half of the plain below 100 m asl, it is one of the world's largest unbroken lowlands. Its flat landscape features the Vasyugan Swamp, the largest peat land in the Northern Hemisphere. The peculiar geographical setting of Western Siberia combined with a harsh continental climate and a lack of available raw material allegedly account for a visible scarcity of Palaeolithic settlements relative to the neighboring regions.

Until now, only 30 localities are known for a region that covers about a third of Siberia[1]. Most of these localities (Shikaevka II, Gari, Novyi Tartas, Vengerovo-5, Chernoozerie II, Mogochino I, Tomskaya, Achinskaya and Shestakovo) are attributed to the late Sartan period (OIS 2). The site of Voronino-Yaya is the exception and would indicate a human presence circa 28 ka[1-3].

Earlier settlements can be found between the Ob and the Yenisei Rivers where the Western plain meets the Altai lowlands and the Krasnoyarsk Basin. The site of Ust'-Izhul represents a Middle Palaeolithic workshop dated by Infrared-stimulated luminescence (IRSL) to the Last Interglacial (OIS 5e) [4]. Aryshevskoe-1 has yielded an assemblage potentially attributed to the early phases of the Upper Palaeolithic but for which the date is unclear. Two radiocarbon dates obtained on humic soil samples from the lower archaeological layer indicate an age older than 40 ka and a piece of charcoal derived from the upper archaeological layer has yielded an age of 33,630 +-995 (SOAN-4180) [3].

The most relevant archaeological record for the GI12 time interval is found in the Northwestern and Central Altai where the OH5-OH6 assemblages from Kara-Bom represent the earliest reliable evidence for Upper Palaeolithic assemblages in Siberia[5,6]. It is defined based on a series of peculiar technological features clearly expressed in the production of blades. Large and robust blade blanks are produced using stone-hammers and a semi-circular asymmetrical core reduction system from two opposed platforms. Thick side blades are regularly detached for the management of the core convexities. The latter are selected, snapped and further reduced to produce smaller blades using the burin-core method[7]. Along with specific retouched tools[8], this exceptional combination of features assigns these assemblages to a united techno-complex and not only to a generic Upper Palaeolithic classification[9].

In the Altai, it has been recognized at Ust'-Karakol 1 (sector 1, OH5.5 and OH5.4), Kara-Tenesh, and Maloyalomanskaya Cave (for a review see the Kara-Bom variant in Derevianko, 2012). The IUP can be traced eastward in the Cis-Baikal (e.g., Makarovo-4)[8,10], Trans-Baikal (e.g. Khotyk, Kamenka A and C and Podzvonkaya)[11-14] Mongolia (e.g. Tolbor 4)[15,16] and China (Shuidonggou)[17-19]. The IUP complex also marks the first appearance of the use of pigments, body ornaments, formal bone tools and even musical instruments in Siberia[20,21].

The quantity and quality of chronological data associated with the IUP vary depending on the region (for a review of the labs and methods see[22]). Large standard deviation and occasional infinite ages are frequently associated with estimation close to the limit of the method. Younger results are not reported here as they may reflect differences between measurements (Conventional/AMS) or pretreatment methods, and/or contaminations by younger material. They could also indicate that in some regions the IUP may have lasted a few thousands of years.

The ages presented below are the earliest and finite AMS dates associated with the IUP as defined by Zwyns[9]. They show a clear temporal overlap between the onset of IUP during the GI12 and the date obtained from the Ust'-Ishim 1 individual.

	layer/industry	Material	^{14}C	sigma	calib. BP	sigma	Lab number
Kara Bom	OH5-6	Charcoal	43200	1500	46931	1995	GX-17597
Kara Bom	OH5-6	Charcoal	43300	1600	47025	2052	GX-17596
Kamenka	A	Bone	40500	3800	44848	3587	AA-26743
Khotyk	3	Bone	38200	2800	42039	2598	AA-60267
Podzvonkaya	3	Bone	38900	3300	42755	3132	AA-26741
Tolbor 4	OH6	Bone	37400	2600	41354	2475	AA-79314

From the archaeological material, the IUP can be clearly distinguished from the Middle Palaeolithic assemblages from Siberia. From a chronological point of view, these two complexes overlap with the UI human remains. The Mousterian of the Sibiryachikha variant is defined based on two cave sequences, Okladnikov Cave and Chagyrskaya Cave[23,24]. Radiocarbon dates available for Okladnikov place the Mousterian occupation between 44,000 ± 4000 BP (stratum 7) to 33,500 ± 700 BP (stratum 1)[24] suggesting that it overlaps with the appearance of the IUP complex [25] whereas new results from Chagyrskaya would indicate an age beyond the limit of the radiocarbon range, perhaps during (OIS4-early OIS3)[24]. Both sequences are now clearly associated with easternmost Neandertal human remains[26-28].

In Siberia, the IUP would last circa 5,000 years before being replaced by a full-fledged Early Upper Palaeolithic (EUP)[9]. The latter corresponds to a series of behavioral changes in terms of knapping techniques and methods. A notable switch toward the use of soft-hammer concomitantly occurs with a genuine bladelet production (sometimes also referred to as microblades[29]), and is indicative of a change in weaponry (composite spears), also observed in Europe around the same time. Although a possible continuity between IUP and EUP cannot be ruled out, it is generally assumed that these two complexes correspond to distinct human groups. As it is the case in Europe, the EUP would correspond to a permanent human settlement in Siberia.

1. Zenin, V. N. Major stages in the human occupation of the West Siberian Plain during the Paleolithic. *Archaeology, Ethnology & Anthropology of Eurasia* **3,** 22-44 (2002).
2. Derevianko, A. P. & Markin, S. V. in *The Paleolithic of Siberia - New Discoveries and Interpretation* (eds Derevianko, A. P., Shimkin, D. B. & Powers, W. R.) 84-106 (University of Illinois Press, Urbana and Chicago, 1998).
3. Vasil'ev, S. A., Kuzmin, Y. V., Orlova, L. A. & Dementiev, V. N. Radiocarbon-based chronology of the Paleolithic in Siberia and its relevance to the peopling of the New World. *Radiocarbon* **44,** 503-530 (2002).
4. Chlachula, J., Drozdov, N. I. & Ovodov, N. D. Last Interglacial peopling of Siberia: the Middle Palaeolithic site Ust' Izhul', the upper Yenisei area. *Boreas* **32,** 506-520 (2003).
5. Derevianko, A. P., Nikolayev, S. V. & Petrin, V. T. The dating of the paleolithic Kara-Bom site with physical methods (C14 and EPR). *Altaica* **3,** 2-8 (1993).
6. Goebel, Derevianko, P. & Petrin, T. Dating the Middle-to-Upper-Paleolithic transition at Kara-Bom. *Curr Anthropol* **34,** 452-458 (1993).
7. Zwyns, N., Rybin, E. P., Hublin, J. J. & Derevianko, A. P. Burin-Core Technology & Laminar Reduction sequence in the Initial Upper Paleolithic from Kara-Bom (Gorny-Altai, Siberia). *Quaternary International* **259,** 33-47 (2012).
8. Rybin , E.P., Tools, beads, and migrations: Specific cultural traits in the Initial Upper Paleolithic of Southern Siberia and Central Asia, Quaternary International (2014), http://dx.doi.org/10.1016/j.quaint.2014.04.031.
9. Zwyns, N. *Laminar technology and the onset of the Upper Palaeolithic in the Altai, Siberia* (Leiden University Press, Leiden, 2012).
10. Goebel, T. & Aksenov, M. Accelerator radiocarbon dating of the initial Upper Palaeolithic in southeast Siberia. *Antiquity* **69,** 349-357 (1995).
11. Lbova, L. Chronology and paleoecology of the Early Upper Paleolithic in the Transbaikal region (Siberia). *Eurasian Prehistory* **5,** 109-114 (2008).

12. Orlova, L., Kuzmin, Y. & Lbova, L. in *Paleoliticheskie Kultury Zabaikalya I Mongolii (Novye Fakty, Metody I Gipotezy) [Palaeolithic cultures of the Transbaikal and Mongolia (New facts, methods and hypotheses)]* 88-92 (Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences Press, Novosibirsk, 2005).
13. Tashak, V. in *Arkheologiya I Kulturnaya Antropologiya Dalnego Vostoka I Tsentralnoi Azii (Archaeology and cultural anthropology of the Far East and of Central Asia)* 25-33 2002).
14. Tashak, V. Plastinchatiie industrii Zabaikalya (Laminar Industries from Transbaikal). *Stratum plus* **1,** 79-93 (2009).
15. Derevianko, A. P., Zenin, A. N., Rybin, E. P., Gladyshev, S. A., *et al.* The technology of early Upper Paleolithic lithic reduction in Northern Mongolia: The Tolbor-4 site. *Archaeology, Ethnology and Anthropology of Eurasia* **29,** 16-38 (2007).
16. Gladyshev, S. A., Olsen, J. W., Tabarev, A. V. & Kuzmin, Y. V. Chronology and Periodization of Upper Paleolithic Sites in Mongolia. *Archaeology, Ethnology and Anthropology of Eurasia* **38,** 33-40 (2010).
17. Brantingham, P., Gao, X., Madsen, D., Bettinger, R. & Elston, R. in *The Early Upper Palaeolithic beyond Western Europe* (eds Brantingham, P. J., Kuhn, S. L. & Kerry, K. W.) 223-241 (University of California Press, Berkeley, 2004).
18. Li, F., Kuhn, S. L., Gao, X. & Chen, F. Y. Re-examination of the dates of large blade technology in China: a comparison of Shuidonggou Locality 1 and Locality 2. *J Hum Evol* **64,** 161-168 (2013).
19. Morgan, C., Barton, L., Xing, G., Mingjie, Y., *et al.* Re-Dating Shuidonggou Locality 1 and Implications for the Initial Upper Paleolithic in East Asia. *Radiocarbon* **56,** 165-179 (2014).
20. Derevianko, A. P. & Rybin, E. P. The earliest representations of symbolic behaviour by Paleolithic humans in the Altai Mountains. *Archaeology, Ethnology and Anthropology of Eurasia* **4,** 27-50 (2003).
21. Lbova, L. Evidence of Modern Human Behavior in the Baikal Zone during the Early Upper Paleolithic Period. *Bulletin of the Indo-Pacific Prehistory Association* **30,** 9-13 (2011).
22. Kuzmin & Orlova Radiocarbon Chronology of the Siberian Paleolithic. *Journal of World Prehistory* **12,** 1-53 (1998).
23. Derevianko, A. P. & Markin, S. V. in *Characteristic features of the Middle to Upper Palaeolithic Transition in Eurasia* (eds Derevianko, A. P. & Shunkov, M. V.) 40-50 (Publishing Department of the Inst. f. Archaeology and Ethnography of the SB RAS, Novosibirsk, 2011).
24. Derevianko, A. P., Markin, S. V. & Shunkov, M. V. The Sibiryachikha facies of the Middle Paleolithic of the Altai. *Archaeology, Ethnology and Anthropology of Eurasia* **41,** 89-103 (2013).
25. Kuzmin, Y. V. in *The Early Upper Palaeolithic beyond Western Europe* (eds Brantingham, P. J., Kuhn, S. L. & Kerry, K. W.) 196-206 (University of California Press, Berkeley, 2004).
26. Viola, B. T., Markin, S. V., Buzhilova, A. P., Mednikova, M. B., *et al.* New Neanderthal remains from Chagyrskaya Cave (Altai Mountains, Russian Federation). *American Journal of Physical Anthropology* **147,** 293-294 (2012).
27. Viola, T. B. *New Hominid Remains from Central Asia and Siberia: the Easternmost Neanderthals?* (PhD Thesis, University of Vienna, Vienna, 2009).
28. Viola, T. B., Markin, S. V., Zenin, A., Shunkov, M. V. & Derevianko, A. P. in *Characteristic features of the Middle to Upper Palaeolithic Transition in Eurasia* (eds Derevianko, A. P. & Shunkov, M. V.) 207-213 (Publishing Department of the Inst. f. Archaeology and Ethnography of the SB RAS, Novosibirsk, 2011).
29. Kuzmin, Y. V. in *Origin and Spread of Microblade Technology in Northern Asia and North America* (eds Kuzmin, Y. V., Keates, S. & Shen, C.) 115-124 (Archaeology Press, Simon Fraser University, Burnaby, BC, 2007).

Supplementary Information 6

Sampling, Library Preparation and Sequencing

Qiaomei Fu*, Ayinuer Aximu Petri, Matthias Meyer
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

A human femur from Ust'-Ishim

The Ust'-Ishim sample is a relatively complete femoral shaft found in 2008 by N.V. Peristov, on the banks of the Irtysh River, in Ust'-Ishim and Tara counties of Omsk Province, Russian Federation (Russia). In 2010 A. Bondarev identified the bone as a hominin femur.

Sampling and DNA Extraction

The Ust'-Ishim femur was sampled under clean-room conditions. A dentistry drill was used to remove the surface and to recover between 41 and 130 mg of bone powder (Table S6.1) from nine locations within a 2 cm x 2 cm section of the distal end of the femur shaft (Figure S6.1). For each sample this bone powder was used to prepare 100 μL of DNA extract using the method described in Dabney et al. 2013 [1].

Library preparation

To evaluate DNA preservation, one library (B5271) was produced from 20 μl of the 100 μl extract E1122 using a double stranded library preparation method [2] without enzymatic removal of deaminated cytosines. A 4-bp 'clean-room key' sequence (5'-GTCT-3') was added to the P5 and P7 library adapters to allow the identification of contaminating sequences derived from modern DNA libraries [3]. In addition, sample-specific indexes were introduced into both the P5 and P7 adaptors during library amplification [2], which was carried out in two stages. First, the library was first amplified for 10 cycles using AmpliTaq Gold DNA Polymerase (Applied Biosystems), which is able to incorporate nucleotides across uracils and thus preserves deamination-induced damage patterns. The PCR product was then purified

using the MinElute PCR purification kit (Qiagen, Hilden, Germany) and used as template for a second round of amplification with Herculase II Fusion DNA polymerase (Agilent), a proof-reading enzyme, using primers IS5 and IS6 [4] as described in Dabney and Meyer 2012[5]. MtDNA was captured using a hybridization capture approach [6]. The resulting library was again amplified with primers IS5 and IS6, and the concentration of the purified library was determined using a Bioanalyzer DNA 1000 chip.

After DNA preservation was confirmed for the specimen, nine additional libraries were produced from 5µl of extract E1122 as well as 8 additional extracts using single stranded library preparation [7] (Table S6.2). Miscoding DNA damage was removed during library preparation by treatment with uracil-DNA-glycosylase (UDG) isolated from *Archaeoglobus fulgidus* (*Afu*) and endonuclease (Endo VIII) as described elsewhere [8]. These libraries were PCR amplified once using AccuPrime Pfx DNA polymerase (Life Technologies) [5] using the double-index scheme described above [2]. Library concentrations were determined using a DNA-1000 chip on the Bioanalyzer 2100.

The extract that showed the highest proportion of endogenous human DNA (E1159) was used to produce 8 additional libraries (Table S6.2) for deep sequencing, each using 8 µl of extract. Since a substantial proportion of the library molecules generated with the single-stranded method are short, a size fractionation step using acrylamide gels ("gel cut") was carried out to remove library molecules with inserts shorter than approximately 35 bp as described elsewhere [7]. The size fractionated libraries were quantified by qPCR and amplified using AccuPrime Pfx DNA polymerase (Life Technologies) with primers IS5 and IS6. Concentrations of the libraries were determined using a DNA-1000 chip on the Bioanalyzer 2100.

Shotgun sequencing

One library (B5271) (Table S6.2) was sequenced on a twentieth of one lane of the Illumina MiSeq (MS-102-1001 MiSeq Reagent Kit (300-cycles - PE) (version 1)). Libraries B5339, B5340, B5341, B5342, B5343, B5344, B5345, B5346 and B5347 were pooled and sequenced on one MiSeq lane (see Table S6.2). Sequencing was performed for 76 cycles from both ends of the libraries, and two additional seven base

pair index reads were included [2]. An indexed control ϕX 174 library was added to the sample libraries to yield 2-3% control reads (index 5'-TTGCCGC-3').

Library B5347 was additionally sequenced on one Illumina HiSeq 2000 flowcell. Libraries B3899, B3901, B3902, B3903, B3904, B3905, B3906 and B3907 were sequenced as a pool using ten Illumina HiSeq 2000 flowcells (80 lanes). All HiSeq sequences were produced using a paired-end setup with 95 cycles and two seven base pair index reads, following the instructions from the manufacturer for multiplex sequencing on the HiSeq platform with a TruSeq PE Cluster Kit v3 - cBot – HS cluster generation kit and a TruSeq SBS Kit v3 sequencing chemistry. An indexed control library of ϕX 174 was spiked into the pool library prior to sequencing, contributing to ~0.5% of the resulting sequences [2].

Sequence processing, mapping and genotyping

For all sequences longer than 35bp, the initial sequence processing including base-calling, alignment (to the human reference GRCh37/1000 Genomes), duplicate removal, indel realignment, and genotyping) were as described in [9] (Table S6.3). To reduce the impact of cytosine deamination on genotyping we reduced the base quality score of any 'T' in the first base or last two bases of sequences from all single-stranded, UDG-treated libraries prior to genotyping.

Table S6.1. DNA extracts made from the Ust'-Ishim femur.

Extract	Powder (mg)	%Sequences mapped
E1122	41	3.32
E1152	101	1.82
E1153	101	4.31
E1154	106	3.46
E1155	109	8.16
E1156	107	5.46
E1157	110	6.74
E1158	110	5.06
E1159	130	9.95

Table S6.2 Overview of extracts, libraries and sequences generated from the Ust'-Ishim femur.

Library	Extract	Extract used for library (µl)	Library type	Sequenced lanes
B5271	E1122	20	Double strand+no UDG	1/20 Miseq
B5339	E1122	5	Single strand +UDG	1/9 Miseq
B5340	E1152	5	Single strand +UDG	1/9 Miseq
B5341	E1153	5	Single strand +UDG	1/9 Miseq
B5342	E1154	5	Single strand +UDG	1/9 Miseq
B5343	E1155	5	Single strand +UDG	1/9 Miseq
B5344	E1156	5	Single strand +UDG	1/9 Miseq
B5345	E1157	5	Single strand +UDG	1/9 Miseq
B5346	E1158	5	Single strand +UDG	1/9 Miseq
B5347	E1159	5	Single strand +UDG	1/9 Miseq +8 Hiseq
B3899	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3901	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3902	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3903	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3904	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3905	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3906	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq
B3907	E1159	8	Single strand +UDG+gel cut (35bp)	10 Hiseq

Table S6.3 Sequencing and mapping statistics for the DNA libraries prepared from extract E1159.

Library	Sequence aligned to GRCh37/1000 Genomes
B5347	90,644,828
B3899	268,364,156
B3901	266,963,357
B3902	215,617,911
B3903	233,642,262
B3904	265,305,973
B3905	225,937,647
B3906	222,561,107
B3907	250,450,192

Figure S6.1. The Ust'-Ishim 1 femur. Samples for DNA extraction were removed from the part indicated by a red rectangle.

References

1 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15758-15763, doi:10.1073/pnas.1314445110 (2013).

2 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, e3, doi:10.1093/nar/gkr771 (2012).

3 Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897, doi:10.1038/nature08976 (2010).

4 Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380, doi:10.1038/nature03959 (2005).

5 Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87-94, doi:10.2144/000113809 (2012).

6 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

7 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

8 Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* **8**, 737-748, doi:10.1038/nprot.2013.038 (2013).

9 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

Supplementary Information 7

Data Quality

Qiaomei Fu, Philip L.F. Johnson, Matthias Meyer, Michael Lachmann, Janet Kelso*

* To whom correspondence should be addressed (kelso@eva.mpg.de)

Summary

We calculate basic statistics and estimate contamination for sequences from each of the libraries prepared from the Ust'-Ishim femur. The sequences from the libraries made from the extract E1159 yield an average coverage of 42.4-fold of the autosomal genome, 21.9-fold of the X chromosome, 21.7-fold of the Y chromosome, and 4,439-fold of the mtDNA. We estimate the over-all human mitochondrial DNA contamination to 0.50% (95% confidence interval (CI): 0.26%-0.94%) and the autosomal DNA contamination to be less than 0.13% (95% CI: 0-0.13%).

DNA library characteristics and genomic coverage

Proportion of endogenous DNA and fragment length distribution

The proportion of fragments mapped to the human reference genome (GRCh37/1000 Genomes release) ranges from ~1% to ~10% among the nine extracts prepared from the Ust'-Ishim bone (Table S6.1; Figure S7.1A). The extract that yielded the highest fraction of endogenous fragments (E1159) was used to produce 8 additional libraries using the single-strand library preparation method and UDG treatment (Table S6.2). A random subset of the merged sequences from these libraries was used to calculate the fragment length distribution before and after mapping (map quality $\geq$ 30) (Figure S7.1B).

Base composition and nucleotide misincorporation patterns

Base compositions for all aligned sequences from the single-stranded libraries prepared from E1159 as a function of fragment lengths are shown in Figure S7.2. The patterns of nucleotide misincorporations in mitochondrial and autosomal sequences were evaluated using data from a single library only (double-stranded library B5271 from extract E1122, see Table S6.2) that was

prepared without enzymatic removal of uracils (Figure S7.3A). Treatment of all single stranded libraries with endonuclease (Endo VIII) and uracil-DNA-glycosylase (UDG) as described in Supplementary Note 6 reduced the frequency of deaminated cytosines at the 5'- and 3'-ends of the molecules from over 30% in the untreated libraries to less than 3% (Fig S7.3 B).

Genomic coverage and GC dependence

We computed coverage for the uniquely mappable regions of the genome (defined as genomic regions where all possible 35 base pair reads overlapping a given position align only to this position allowing up to two mismatches) from the VCF files generated from alignments of the Ust'-Ishim sequences to the human reference sequence. Mean coverage is 42.4-fold and median coverage is 42-fold. The genomic coverage of Ust'-Ishim is similar to that generated for 14 present-day human genomes from around the world in the so-called "panel B" [1] (Figure S7.4) which is used for most analyses here.

We used the number of GC nucleotides in the 50 bases flanking each position (25 on either side) to divide the genome into bins of different GC content. As previously observed for the Neandertal and Denisovan genomes [1,2], the coverage is negatively correlated with GC content (Figure S7.5).

Data filtering

Filters for autosomal data

We applied a minimal set of filters to the Ust'-Ishim genome sequence, as well as to all other genomes used in this study. The primary purpose is to reduce the influence of mapping errors. Briefly, we restricted our analysis to unique regions in the genome, using Tandem Repeat Finder (for hg19) and mappability tracks (map35_100%) [1]. We required that sequences align with a minimum root-mean-square mapping quality (MQ field in VCFs) of 30. For modern genomes we used only sites that are between the 2.5% and 97.5% quartiles of the coverage distribution. For ancient genomes we applied a GC-corrected coverage filter due to the dependence of coverage on GC-content (Figure S7.5). We determined the coverage distribution in each bin of different GC content, and excluded sites that fall within the 2.5% and 97.5% quartiles in each GC bin.

The human reference genome assembly (GRCh37/1000 Genomes release) contains a total 2.68 gigabases (Gb) (excluding N's). Of these, 1.65 Gb (62%) remain for Ust'-Ishim after applying all filters and 0.66 Gb pass the filters in all individuals (Ust'-Ishim, Neandertal, Denisovan, and 25 present-day humans from the A and B panels).

Filters for the X and Y chromosomes

The filters for the X and Y chromosomes are similar to those for the autosomes. We filtered bases that are in the upper and lower 2.5% coverage quartiles (Table S7.2) and do not apply the GC-corrected coverage on the Ust'-Ishim Y chromosome data since the variation of GC content on the Y chromosome is less than on the autosomes and X chromosome (Figure S7.5B).

After filtering, 81.8 Mb of the X chromosome and 4.61 Mb of the Y chromosome remain for the Ust'-Ishim individual. When restricting to the positions covered in all individuals (the Ust'-Ishim individual and the modern humans in Table S7.2), 32.1 Mb of the X chromosome and 1.8 Mb of the Y chromosome remain.

Contamination estimates

Mitochondrial contamination estimate

We quantified the frequency of C > T and G > A substitutions, which are typical of ancient DNA, in the alignments of the non-UDG-treated sequence fragments against the mtDNA reference. We found that between 30 and 35% cytosine residues close to the ends of the DNA fragments appear as thymine residues in the sequences (Figure S7.3C). This is consistent with the majority of the DNA stemming from an ancient source [3,4].

For each of the UDG-treated libraries created from E1159 we calculated the agreement between the aligned fragments and the mitochondrial consensus for each position in the assembled mitochondrial genome (Fig. S7.6). All libraries have more than 500-fold average coverage of the mtDNA (Table S7.1). The lowest coverage at any position is 77-fold and all fragments support the consensus base at this position. The average support for the consensus base is $\geq$99.7% for all libraries (Table S7.1). The lowest consensus support is 95.3% at position 5734 with 172 fold coverage (rCRS co-ordinates) in B3904 (5C, 3G, 164T). The differences at this position are

unlikely to be the result of contamination since all known modern human mtDNA genomes have the same nucleotide as the consensus sequence at this position. Sequence differences at this position are therefore likely to represent sequencing and alignment errors and damage.

The observation of only a single DNA sequence from deep sequencing of the mitochondrial genome provides evidence that the mtDNAs in the libraries come from a single individual.

The Ust'-Ishim mtDNA differs at only one site from all members of a panel of 311 present-day humans, which means that there is little power to derive a confident contamination estimate using the approach using fixed differences that we have previously applied to the Denisovan and Neandertal genomes [5]. Instead we used a likelihood-based method [6] that uses 311 present-day human mtDNA sequences as the potentially contaminating population and generates a combined estimate of contamination and error. The average contamination for the pool of libraries sequenced for this genome is estimated to be 0.50% (95% CI: 0.26%-0.94%) (Table S7.1), which may represent an upper bound as some errors are likely to be misinterpreted as contamination.

Autosomal contamination estimate

We applied a maximum-likelihood based method from reference [5] to co-estimate sequence error and contamination, along with two nuisance population parameters. We found that this model fits the data poorly (compare discrepancy between observed data in Figure S7.7A and the distribution expected under the model in S7.7B). Upon closer inspection, this poor fit arises due to a significant number of sites where Ust'-Ishim appears homozygous for the derived allele but has more non-derived alleles than expected from sequencing error under the model (e.g. sites with >40 derived but >7 non-derived alleles). Since the model from reference [5] assumes that all sites have the same sequence error rate, we hypothesized that the poor fit may arise from a small proportion of sites with a significantly higher error rate. We therefore modified the model to consider a distribution of error rates. We detail the new likelihood function below.

Let $\Omega = \{c, p_{ad}, p_{dd}, \mathsf{E}, f\}$ denote the set of all parameters, where:

- $c \rightarrow$ contamination rate. A given read will be from a (contaminating) human with probability c and endogenous with probability $1-c$.
- $p_{ad} \rightarrow$ probability of the ancient individual being heterozygous, given that all 1000 Genome humans have the same allele, which differs from the chimpanzee genome at this site.
- $p_{dd} \rightarrow$ probability of the ancient individual being homozygous for the human allele, given that all 1000 Genome humans have the same allele, which differs from the chimpanzee genome differ at this site.
- $\mathsf{E} \rightarrow$ distribution of error probabilities among sites. We assume error rates vary between sites but that a given site has a single, fixed error rate --- perhaps as a function of its local nucleotide context. Further, we assume error rates at different sites are independent. Note that since we are only looking at variable sites (i.e. these sites are spread out), this assumption of independence seems quite reasonable.

We write the probability of the observed numbers of derived alleles (n_d) as the product of the probabilities of the L individual sites, conditional on the number of reads (n) sampled at each site:

$$lik(\Omega) = \Pr(n_{1,d}, \ldots, n_{L,d} \mid n_1, \ldots, n_L, \Omega) = \prod_i \Pr(n_{i,d} \mid n_i, \Omega) \tag{1}$$

Dropping the subscript i for ease of notation, we condition on the true derived allele frequency, t, and assume contamination and sequencing error occur independently:

$$\Pr(n_d \mid n, \Omega) = \sum_{t=0}^{2} \Pr(t \mid p_{ad}, p_{dd}) \int_{\varepsilon} \Pr(\varepsilon \mid \mathsf{E}) \Pr(n_d \mid t, n, c, \varepsilon, f) \tag{2}$$

In equation (2):

- The first term inside the sum (the probability of the truth) is a simple function of the parameters:

$$\Pr(t \mid p_{ad}, p_{dd}) = \begin{cases} 1 - p_{ad} - p_{dd} & t = 0 \\ p_{ad} & t = 1 \\ p_{dd} & t = 2 \end{cases} \tag{3}$$

- The second term (just inside the integral) gives the probability of a site having a particular error rate, ε, which is drawn from the error rate distribution, E.
- The final term, which gives the probability of the observed number of derived alleles, follows a binomial distribution:

$$\Pr(n_d \mid t, n, c, \varepsilon, f) = \binom{n}{n_d} q_t^{n_d} (1 - q_t)^{n - n_d} \tag{4}$$

where

$$\begin{aligned} q_2 &= cf(1-\varepsilon) + c(1-f)\varepsilon/3 + (1-c)(1-\varepsilon) \\ q_1 &= cf(1-\varepsilon) + c(1-f)\varepsilon/3 + (1-c)(1-\varepsilon)/2 + (1-c)\varepsilon/6 \\ q_0 &= cf(1-\varepsilon) + c(1-f)\varepsilon/3 + (1-c)\varepsilon/3 \end{aligned} \tag{5}$$

The overall likelihood of the data given the our parameters can be calculated from (1) and substituting (2), (3), (4), and (5) in turn.

Finally we estimate our primary parameter of interest (c) by maximizing the likelihood of the data over all parameters, including the nuisances $\{c, p_{ad}, p_{dd}, \mathsf{E}, f\}$. We reduce our optimization dimensions by restricting the error distribution to be a simple two category discrete distribution:

$$\Pr(\varepsilon \mid \mathsf{E}) = \begin{cases} e_{1p} & if \ \varepsilon = e_1 \\ 1 - e_{1p} & if \ \varepsilon = e_2 \end{cases}$$

Confidence intervals for c can be generated using a likelihood ratio test of the global maximum likelihood to the profile likelihood ($\ell(c) = \max_{p_{ad}, p_{dd}} [lik(\Omega)]$) and comparing to a χ^2 distribution with 1 degree of freedom.

Our modified model results in a vastly improved fit to the Ust'-Ishim data (Figure S7.7C) and a 95% confidence interval for the estimate of present-day human contamination of 0-0.13%. The modified model also better fits the data in reference [5]; however, due to other features of those data arising as a result of the greater evolutionary separation from modern humans, the contamination estimate in those data remains unchanged when the new model is applied.

Table S7.1 MtDNA contamination, coverage and average support of the mitochondrial consensus sequence generated from libraries made from extract E1159.

	Contamination				
		CI			Average
Library	Mean	2.50%	97.50%	Coverage	support
B5347	0.1%	0.0%	0.3%	395	99.8%
B3899	0.7%	0.4%	1.3%	558	99.8%
B3901	0.5%	0.3%	0.9%	552	99.7%
B3902	0.0%	0.0%	0.4%	461	99.7%
B3903	1.0%	0.5%	1.6%	476	99.7%
B3904	0.4%	0.2%	0.7%	544	99.7%
B3905	0.7%	0.3%	1.2%	472	99.7%
B3906	0.8%	0.5%	1.4%	464	99.7%
B3907	0.3%	0.1%	0.7%	517	99.7%

Table S7.2. 95% coverage quantiles for each sample in uniquely mappable regions of the Y-chromosome. The coverage of the positions in the range of 2.5% and 97.5% coverage quartiles are remained for Ust'-Ishim, and the modern humans from Panel A and B.

Samples	IDS	2.50%	97.50%
Ust'-Ishim	Ust'-Ishim	9	39
Dinka_A	DNK02	6	30
Dinka_B	DNK07	9	29
Mandenka_A	HGDP01284	5	22
Mandenka_B	HGDP01286	8	31
Mbuti_A	HGDP00456	5	22
Mbuti_B	HGDP00982	8	31
San_A	HGDP01029	7	29
San_B	HGDP01036	9	38
Yoruba_A	HGDP00927	7	27
Yoruba_B	HGDP00936	9	32
Karitiana_A	HGDP00998	5	23
Karitiana_B	HGDP01015	8	29
Dai_A	HGDP01307	5	24
Dai_B	HGDP01308	9	31
Han_A	HGDP00778	6	24
Han_B	HGDP00775	8	29
French_A	HGDP00521	6	23
French_B	HGDP00533	10	35
Sardinian_A	HGDP00665	5	22
Sardinian_B	HGDP01076	9	31
Papuan_A	HGDP00542	5	23
Papuan_B	HGDP00546	10	35
Australian_B2	WON,M	10	35

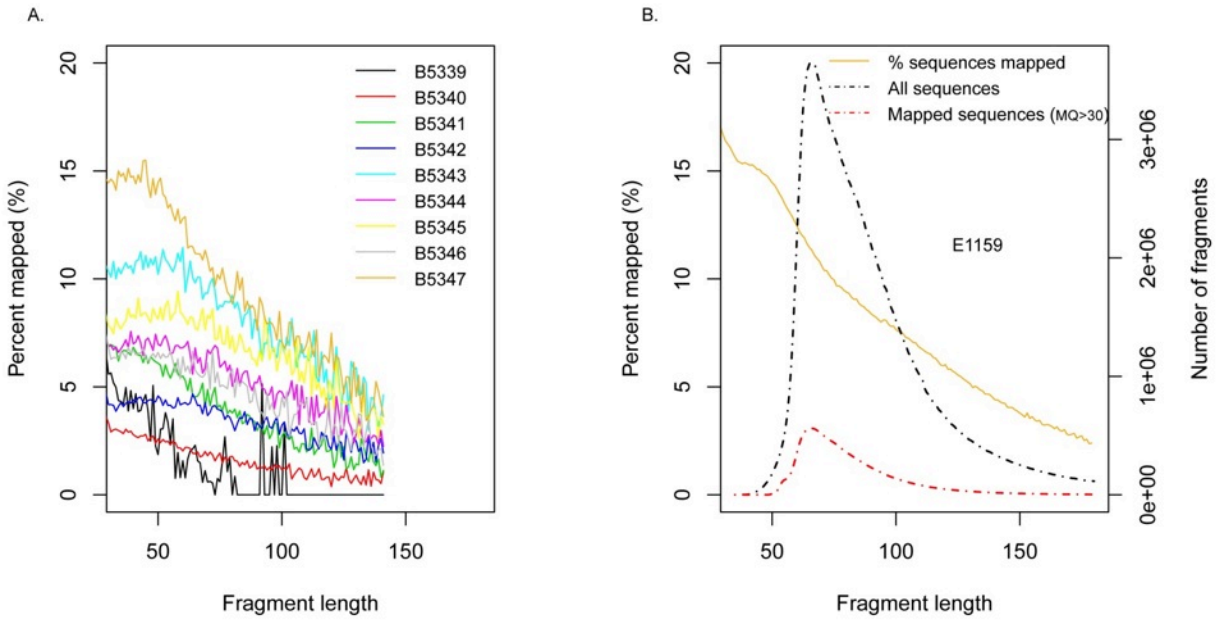


Figure S7.1: Fragment length distribution and proportion of mapped sequences. Analyses were performed with a map quality 30 filter and restricted to merged sequences only. A) Proportion of mapped sequences from libraries generated from extracts prepared from different parts of the Ust'-Ishim bone. The library B5347 from extract E1159 has the highest number of mapped sequences. B) Fragment length distribution and proportion of all mapped sequences from the eight libraries (except B5347) (see Table S7.1) prepared from the extract E1159. The results of each of the libraries (not shown) are similar to each other.

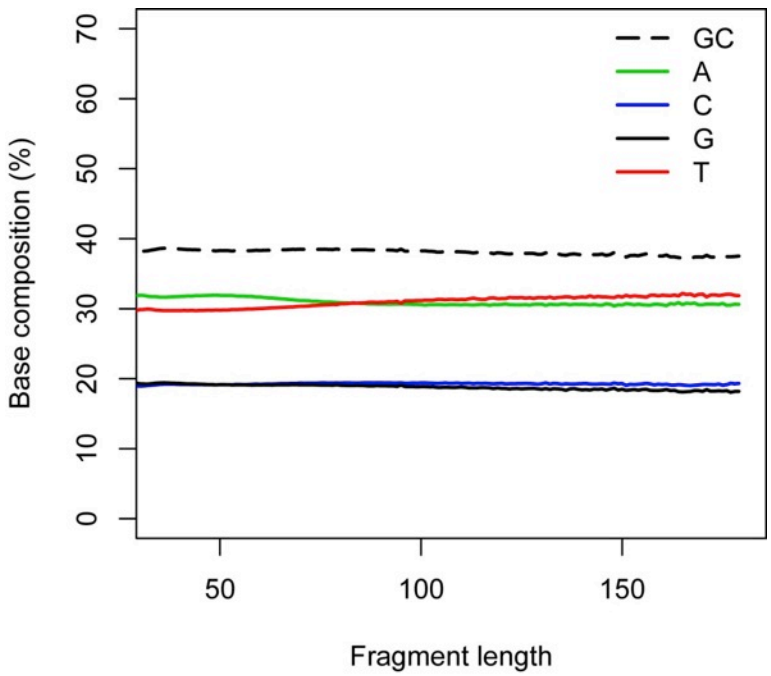


Figure S7.2: Base composition as a function of fragment-length. The data from eight libraries prepared from the Ust'-Ishim bone are shown (map quality >=30).

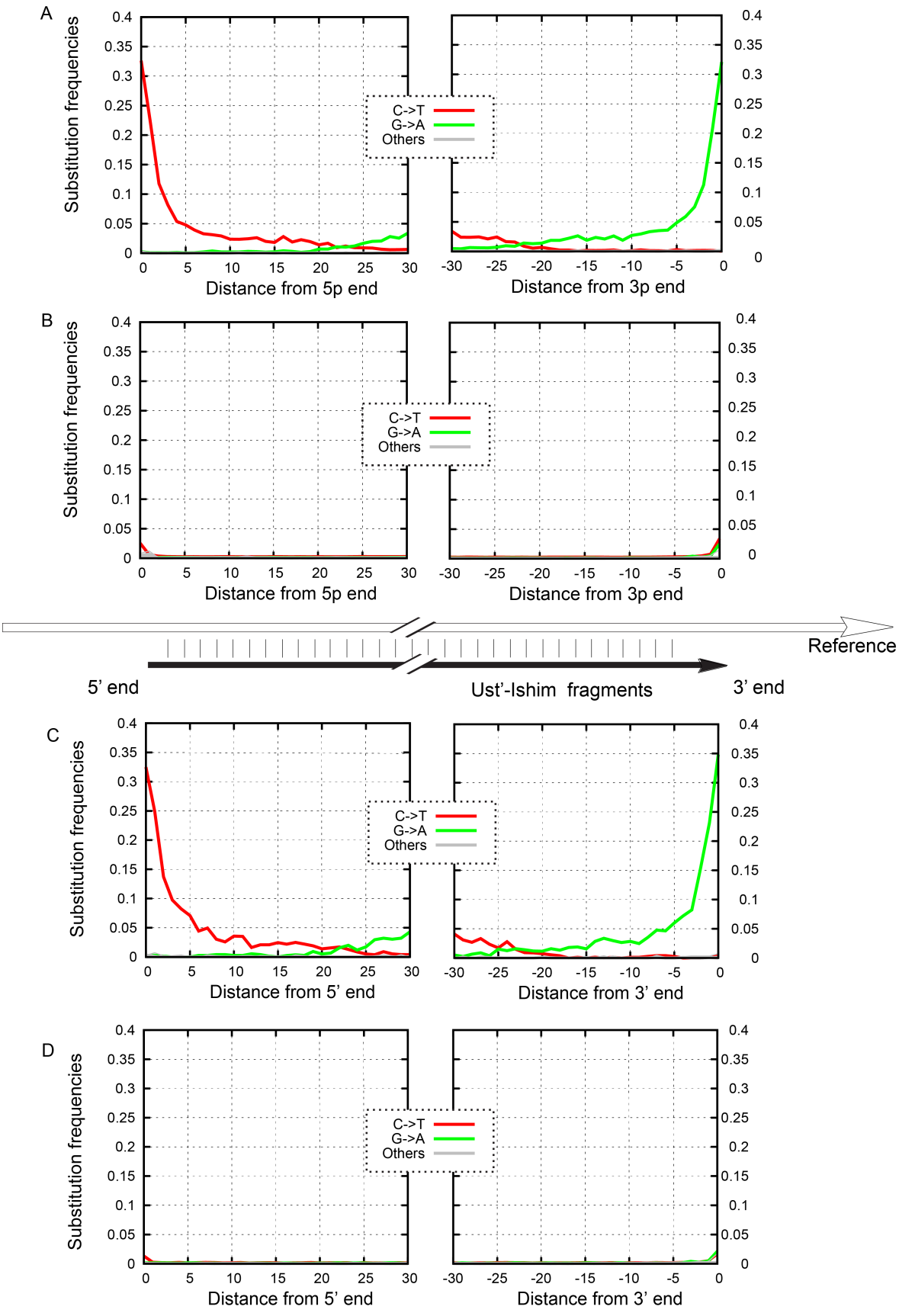

A
Substitution frequencies
C->T
G->A
Others
Distance from 5p end
Distance from 3p end
B
Substitution frequencies
C->T
G->A
Others
Distance from 5p end
Distance from 3p end
Reference
5' end
Ust'-Ishim fragments
3' end
C
Substitution frequencies
C->T
G->A
Others
Distance from 5' end
Distance from 3' end
D
Substitution frequencies
C->T
G->A
Others
Distance from 5' end
Distance from 3' end

Figure S7.3: Substitution patterns in DNA libraries from the Ust'-Ishim individual. The frequency of C > T and G > A differences to the human genome (above) or mtDNA reference sequence (below) is shown as a function of the position at 5' end and 3' end of the sequences. All substitution changes other than C > T and G > A are shown as 'Others'. Substitution patterns of mapped autosomal and mitochondrial DNA fragments (A) and mtDNA fragments (C) in double-stranded Ust'-Ishim libraries without enzymatic removal of deaminated cytosines (uracils). Increased frequency of C > T and G > A substitutions close to the fragment ends indicate the presence of cytosine deamination. Mapped DNA fragments from autosomal and mitochondrial DNA fragments (B) and mtDNA DNA fragments (D) from single-stranded, UDG-treated libraries from the Ust'-Ishim individual show low levels of C > T and G > A substitutions at both at 5'- and 3'-ends. The low level of C > T and G > A substitutions observed is presumably partly due to CpG methylation.

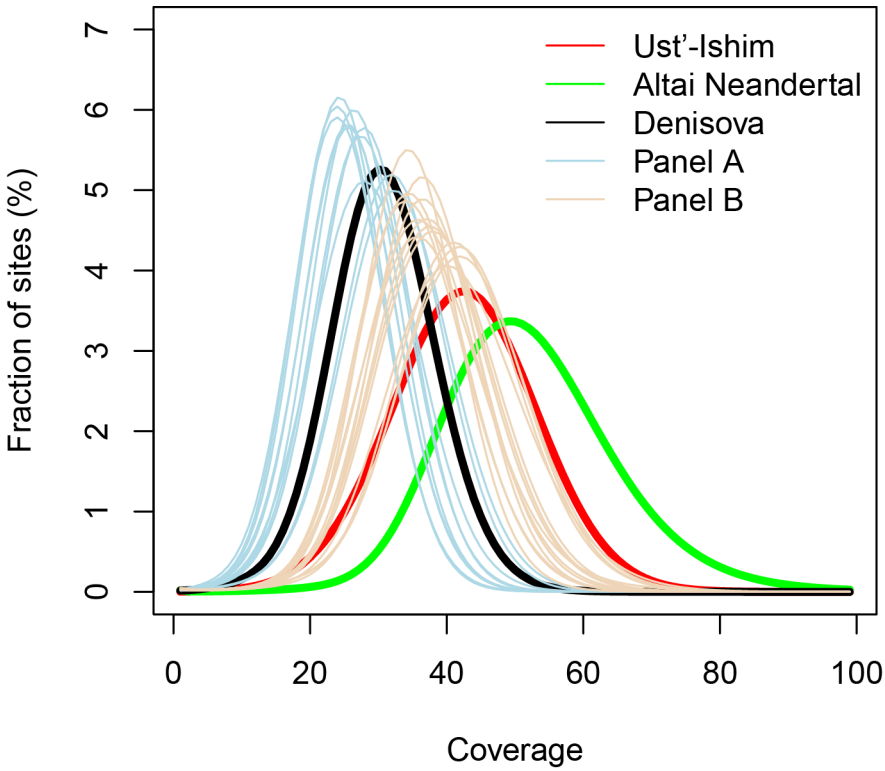


Figure S7.4: Coverage distributions in the uniquely mappable parts of the autosomal genome for the Ust'-Ishim, Neandertal, Denisovan individuals and 25 present-day human genomes from panels A and B 1.

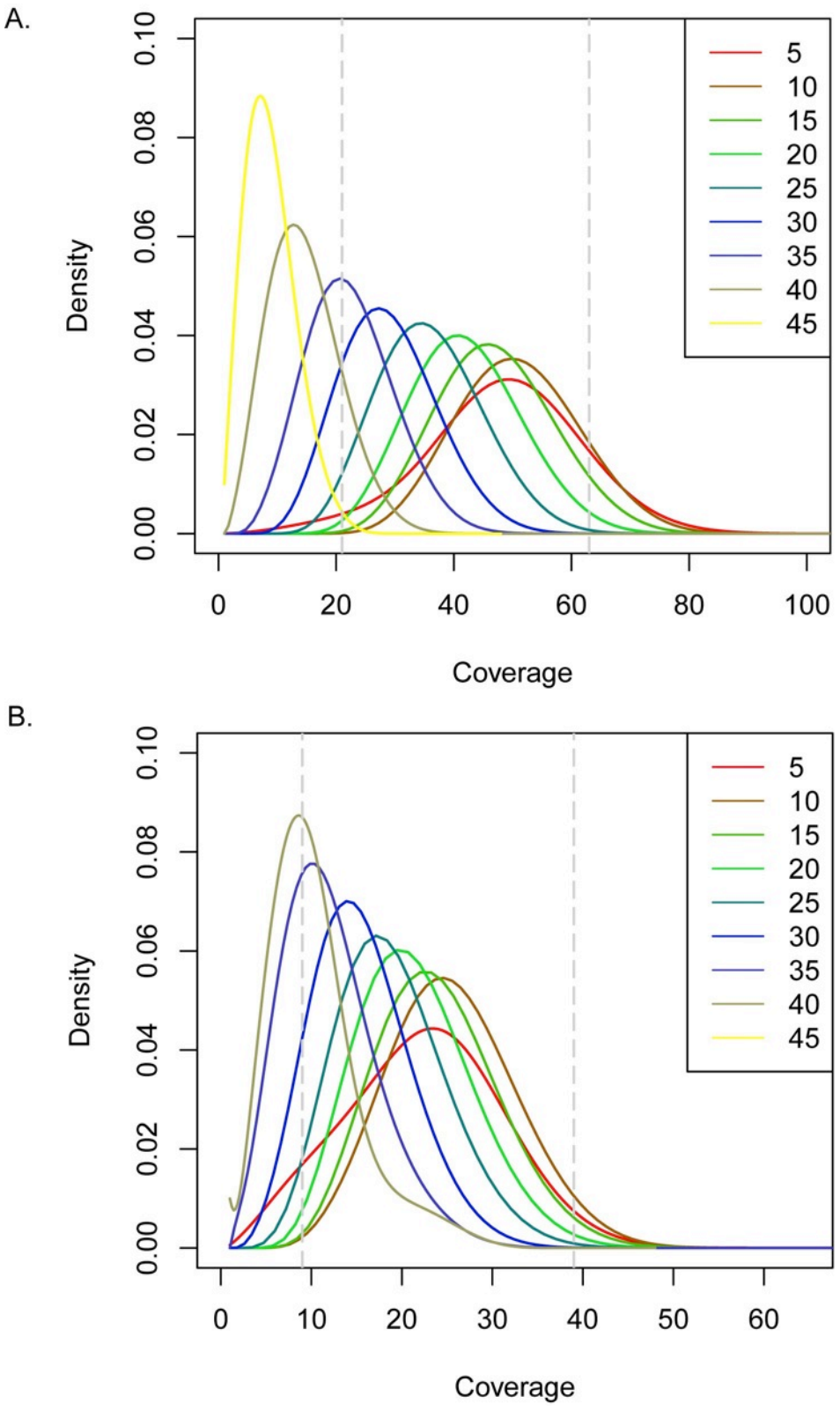


Figure S7.5: Coverage distribution in GC bins (given as the number of bases in a 51 basepair window centered on the analyzed nucleotide) (A) for the whole genome, and (B) for chromosome Y. Thus, the bin

(red curve) contains regions with 5 to 9 GC bases in a 51 basepair window etc. Overall, 95% of reads have GC content within the range delimited by the dashed vertical gray lines.

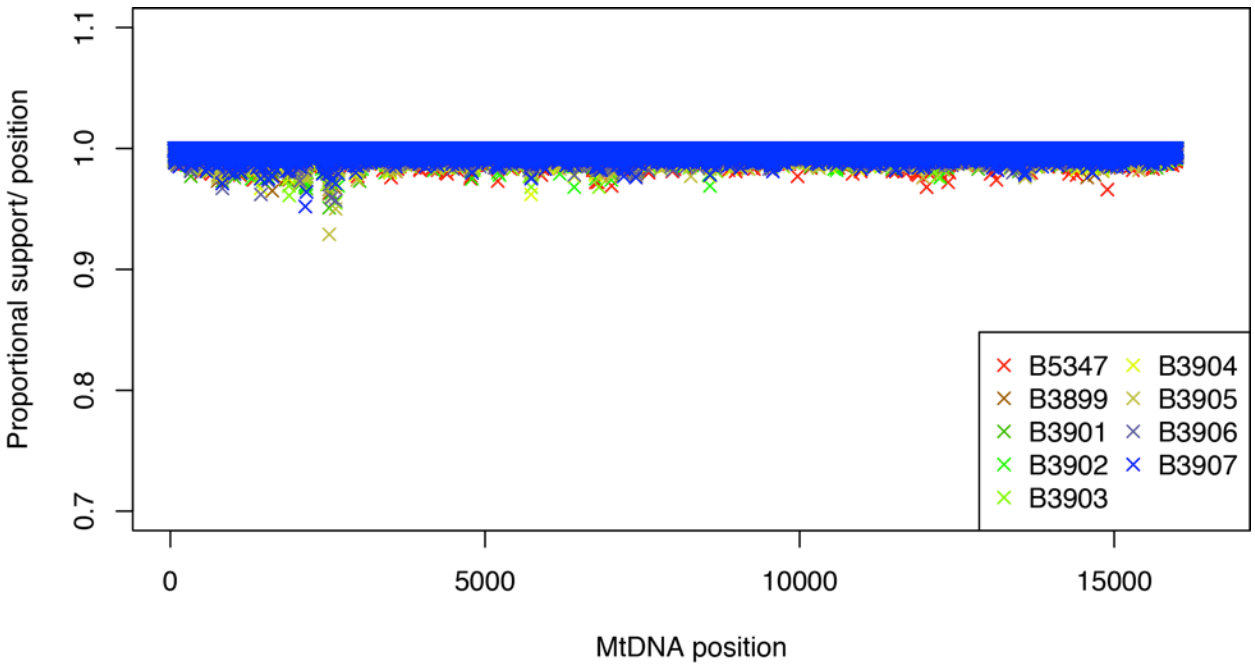


Figure S7.6: Fraction of bases at each site that agree with the consensus sequence of the mitochondrial DNA sequence from Ust'-Ishim.

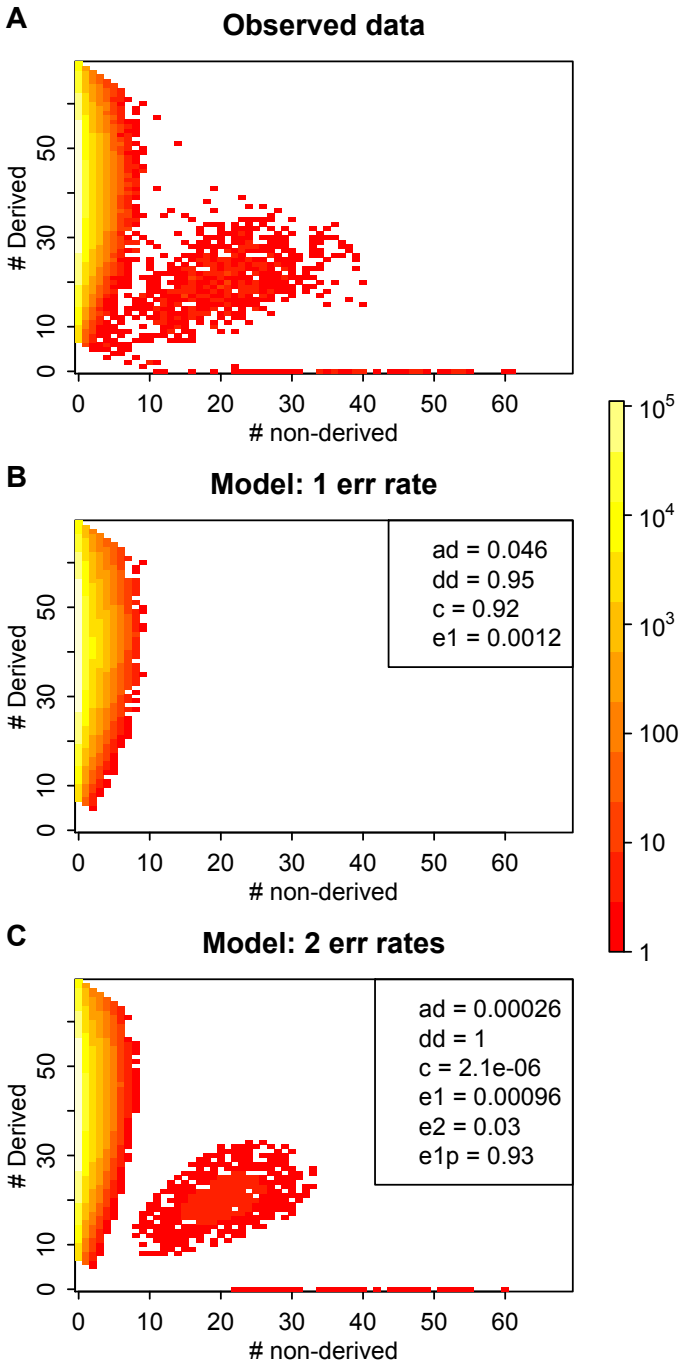


Figure S7.7: Ust'-Ishim autosomal contamination estimate using two different error rate models. We visually compare the Ust'-Ishim data (A) with best-fit model parameters for two models. Plots use heat-map colors to denote the number of sites bearing a given number of non-derived alleles (X-axis) and derived alleles (Y-axis). We analyze only sites where individuals sequenced in the 1000 genomes project all carry the derived allele (compared to great ape outgroups). A) the Ust'-Ishim observed data; B) expected data under best-fit parameter with single error rate model from reference 5; C) expected data under best-fit parameter with the model including two error rates. Legends show best-fit (maximum likelihood) estimates for parameters as defined in the text.

References

1 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

3 Burbano, H. A. *et al.* Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* **328**, 723-725, doi:328/5979/723 [pii]10.1126/science.1188046 (2010).

4 Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).

5 Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, doi:10.1126/science.1224344 (2012).

6 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* **23**, 553-559, doi:10.1016/j.cub.2013.02.044 (2013).

Supplementary Information 8

Mitochondrial Genome Analysis

Qiaomei Fu* and Janet Kelso
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Mitochondrial assembly

To identify sequences from the mitochondrial genome (mtDNA), we aligned all sequences using BWA to the mtDNA from the reference *hg19* assembly from the 1000 Genomes Project.

We also retrieved any sequences that could be aligned to known NUMTs in the nuclear genome. We identified the NUMTs by circularizing the revised Cambridge Reference Sequence (rCRS) (NC_012920.1 [1]) and cutting it into 25mers. We then mapped these 25mers to the human genome with BWA without a seed and retrieved all hits with up to two mismatches. We identified a total of 255 hits. We combined the identified regions on the same chromosome into contiguous sequences when the gap between them was no longer than 200 bp and the unmatched region between them was no longer than 2 kb. Any sequence that could be aligned to these NUMTs was then retrieved and included in the set of potential mtDNA fragments used for mitochondrial assembly. All potential mtDNA fragments were then realigned to the revised Cambridge Reference Sequence (rCRS) (NC_012920.1 [1]) using an iterative mapping assembler (MIA) [2] to generate a mtDNA consensus sequence.

Assessment of the mtDNA relationship with present-day humans

The Ust'-Ishim mtDNA consensus sequence was added to a multiple sequence alignment of 311 human mtDNAs [3]. A phylogenetic tree (Figure S8.1) was constructed in a Bayesian framework using MrBayes [4], running 30,000,000 iterations of the Markov Chain Monte Carlo with the first 3,000,000 iterations discarded as burn-in. The General Time Reversible sequence evolution model was applied with a fraction of invariable sites (GTR+I) determined by the best-fit model approach of Modeltest and PAUP*[5].

We identified the haplogroup for each mtDNA using HaploGrep [6] based on the Phylotree database (based on Phylotree.org, build 15). The Ust'-Ishim mitochondrial sequence shares a common ancestor with the sequences from 154 individuals belonging to the R haplogroup (0.88 posterior support). (Figure S8.1). The Ust'-Ishim individual carries the following substitutions that define the R haplogroup (compared to rCRS sequence) : 73G, 263G, 750G, 1438G, 2706G, 3107d, 4769G, 7028T, 8860G, 11719A, 14766T, 15326G as well as one specific substitution not seen in any present-day population (rCRS pos: 16150). Further, it lacks any of the additional substitutions that define sub-clades of haplogroup R, suggesting that it is related to the mtDNA ancestral to present-day haplogroup R.

Haplogroup R is a large and diverse haplogroup and represents the most common macro-haplogroup in West Eurasia. The 6 most common European haplogroups (H, V, J, T, U, K) [7], and half of the most common eastern Eurasian mtDNA haplogroups [8,9] belong to the R haplogroup. Though the phylogeny is not strongly supported, the Ust'-Ishim mtDNA appears to be most closely related to the direct R sub-clades R* (P, B, F, T, J) which are also found among Papuans, indigenous Australians, and mainland Eurasians.

Dating the Ust'-Ishim femur and re-estimating the mtDNA mutation rate

The carbon date for the Ust'-Ishim femur is approximately 45,000 years (46,880–43,210 cal BP (95.4% probability)). Using ancient mtDNA sequences from securely dated archaeological samples as calibration points provides an alternative method for estimating fossil dates and the mitochondrial mutation rate [10,11]. We investigated whether this date is consistent with the age that we estimate by counting missing mutations in its mtDNA sequence, combined with a previously reported estimate of the mtDNA mutation rate.

We estimated the tip date of the Ust'-Ishim branch using the Bayesian phylogenetic method implemented in BEAST [12], taking advantage of nine ancient modern humans that are directly carbon dated [10,11] as multiple calibration points. In addition, we investigated two different models of rate variation: a strict clock and an uncorrelated log-normal-distributed relaxed clock. We used a constant size and a Bayesian skyline coalescent as tree priors. For each model, we carried out two Markov Chain Monte

Carlo (MCMC) runs with 30,000,000 iterations each, sampling every 1000 steps. The first 6,000,000 iterations were discarded as burn-in. For each model both independent runs were combined resulting in 48,000,000 iterations. The final model was chosen based on the Bayes Factors [13]. The relaxed clock model and the constant size coalescent fit the data better than the constant population size model (log10 BF =16.2) and the Bayesian skyline coalescent (log10 BF =183.1). Using the constant size model and the relaxed clock model we estimate that the Ust'-Ishim femur dates to 48,741 BP (95% highest posterior density (HPD): 31,177-65,907 BP) which is consistent with the calibrated carbon date.

We also calculated an independent estimate of the mtDNA mutation rate using the carbon date obtained for the Ust'-Ishim femur together with the mtDNAs of 311 present-day humans. When repeating the above model comparisons and MCMC with BEAST, we obtain a mutation rate of 2.53×10^{-8} substitutions per site per year (95% HPD: $1.76\text{-}3.23 \times 10^{-8}$) for the complete mtDNA. This substitution rate is in agreement with a previous study [10] which estimated a rate of 2.47×10^{-8} substitutions per site per year (95% HPD: $2.16\text{-}3.16 \times 10^{-8}$) for the complete mtDNA.

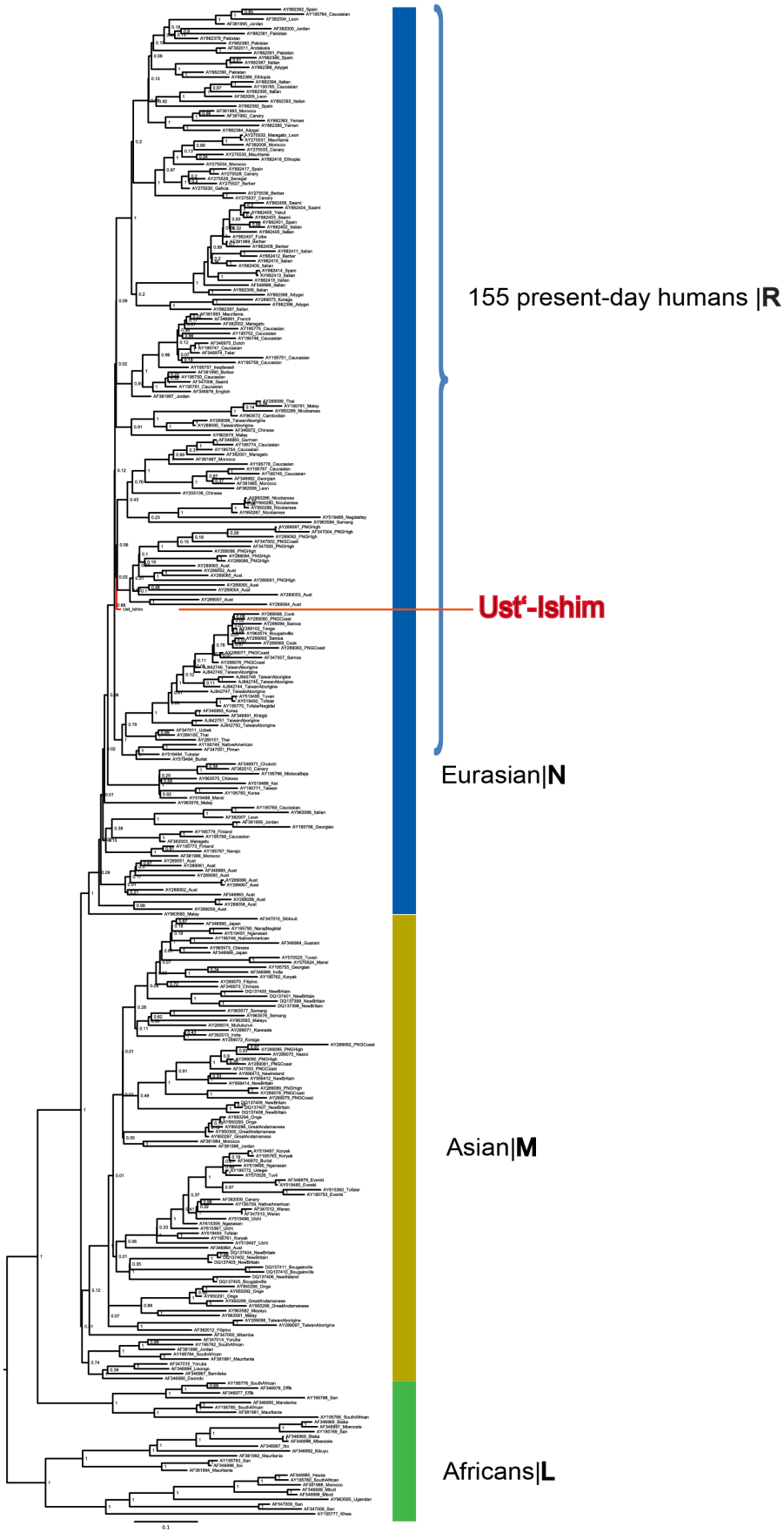


Figure S8.1 Phylogenetic tree of mtDNA from Ust'-Ishim & 311 present-day humans

References

1 Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147, doi:10.1038/13779 (1999).

2 Green, R. E. & Stenzel, U. *Mapping Iterative Assembler*, <https://github.com/udo-stenzel/mapping-iterative-assembler> (

3 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426, doi:10.1016/j.cell.2008.06.021 (2008).

4 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

5 Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818, doi:btb117 [pii] (1998).

6 Kloss-Brandstatter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* **32**, 25-32, doi:10.1002/humu.21382 (2011).

7 Herrnstadt, C. *et al.* Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups. *The American Journal of Human Genetics* **70**, 1152-1171, doi:http://dx.doi.org/10.1086/339933 (2002).

8 Schurr, T. G. & Wallace, D. C. Mitochondrial DNA diversity in Southeast Asian populations. *Hum Biol* **74**, 431-452 (2002).

9 Kong, Q. P. *et al.* Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* **73**, 671-676, doi:10.1086/377718 (2003).

10 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* **23**, 553-559, doi:10.1016/j.cub.2013.02.044 (2013).

11 Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol* **28**, 879-887, doi:10.1093/molbev/msq262 (2011).

12 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214, doi:10.1186/1471-2148-7-214 (2007).

13 Kass, R. E. & Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* **90**, 773--795 (1995).

Supplementary Information 9

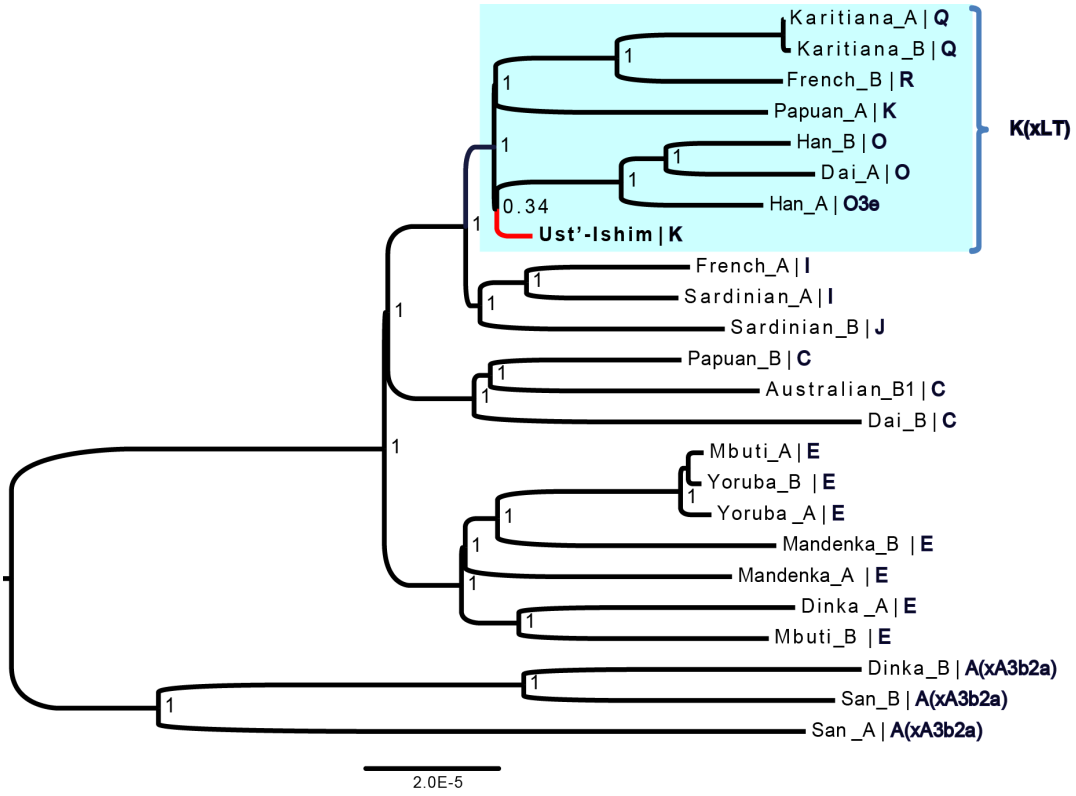
Phylogenetic reconstruction of the Ust'-Ishim Y-chromosome

Qiaomei Fu* David Reich and Janet Kelso
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Y-chromosome phylogenetic tree

After applying the Y-chromosome filters described in SI7, 1.86 Mb of the Ust'-Ishim Y chromosome and the Y-chromosomes of the 23 male present-day humans from Panels A and B were used to reconstruct a phylogenetic tree.

Figure S9.1. Phylogenetic tree of chromosome Y from Ust'-Ishim and 23 present-day humans.



To build the phylogenetic tree (Figure S9.1), we used the Bayesian framework of MrBayes [1], running 20,000,000 iterations of the Markov Chain Monte Carlo with the first 2,000,000 iterations discarded as burn-in. We applied the General Time Reversible sequence evolution model with a substitution site model (GTR+Gamma) determined by the best-fit model approach

of Modeltest and PAUP*[2]. We identified the haplogroup for each Y chromosome using the mutations described in the Y-DNA Haplogroup Tree (Y-DNA Haplogroup tree 2013 in http://www.isogg.org/tree). The Ust'-Ishim Y chromosome sequence clusters with the K(xLT) haplogroup. The Ust'-Ishim sequence shares all the mutations common to the K macro-haplogroup and has one additional specific mutation rs2033003/M526 which defines the group K(xLT) (Figure S9.1, blue part). The Ust'-Ishim Y-chromosome carries no additional mutations belonging to any of the sub-haplogroups of K(xLT); however, there are 6 additional mutations that are not observed in the 23 present-day humans to which we compare. K(xLT) [3] currently has a 54% frequency in Eurasia (Lippold et al. 2014, bioRxiv doi: 10.1101/001792).

Y-chromosome mutation rate

We used the Ust'-Ishim radiocarbon date of 45,000 years as a calibration point to estimate the Y-chromosome mutation rate. We used two different models of rate variation among branches: a strict clock and an uncorrelated log-normal-distributed relaxed clock. We used a constant size and a Bayesian skyline coalescent as two alternative tree priors. For each model, we carried out two Markov Chain Monte Carlo (MCMC) runs with 30,000,000 iterations each, sampling every 1000 steps. We discarded the first 6,000,000 iterations as burn-in. For each model, we combined both independent runs resulting in 48,000,000 iterations. We chose the final model based on the Bayes Factors [4].

The best-fit comparison of the results of the Bayesian MCMC analysis, calibrated with the Ust'-Ishim fossil age, favors the strict clock model over the relaxed clock model ($\log_{10}$ BF = 231). We cannot reject the constant size model ($\log_{10}$ BF = 0.49). With the strict clock model and the constant size model, we estimate a mutation rate of 0.76×10^{-9} substitutions per site per year (95% HPD: $0.67\text{-}0.86 \times 10^{-9}$). This estimate is consistent with the mutation rate of 0.82×10^{-9} (95% CI: $0.72 - 0.92 \times 10^{-9}$) estimated in reference [5] which used the first settlement of the Americas as a calibration point, and is in the range of the other published estimate [6] of 1.0×10^{-9} mutations/nucleotide/year (95% CI: $3.0\times10^{-10} – 2.5\times10^{-9}$) based on sequencing of a deep-rooting pedigree. The mutation rate is higher than that used by Mendez et al. (2013) [7].

In Table S9.1, we also report the most recent common ancestor (TMRCA) of each node and haplogroup sampled in the A and B panels.

Table S9.1 Most recent common ancestor (tmrca) of the node and haplogroups in Figure S9.1, expressed in thousands of years (kya) before present.

Node	Point estimate in kya	95% HPD Interval (kya)
All	153	132-175
A	136	115-157
All except A	72	63-81
All except A and E	71	62-80
KIJ	54	49-59
KxLT	50	47-55
C	53	45-61
E	53	45-62
IJ	49	43-55

References

1 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574 (2003).

2 Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818, doi:btb117 [pii] (1998).

3 Consortium, Y. C. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* **12**, 339-348, doi:10.1101/gr.217602 (2002).

4 Kass, R. E. & Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* **90**, 773--795 (1995).

5 Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562-565, doi:10.1126/science.1237619 (2013).

6 Xue, Y. *et al.* Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol* **19**, 1453-1457, doi:10.1016/j.cub.2009.07.032 (2009).

7 Mendez, F. L. *et al.* An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am J Hum Genet* **92**, 454-459, doi:10.1016/j.ajhg.2013.02.002 (2013).

Supplementary Information 10

Relationship to present-day humans inferred from PCA, ADMIXTURE, and TreeMix

Qiaomei Fu*, David Reich and Janet Kelso
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Summary

We compared the genome sequence of the Ust'-Ishim individual both to genotype data for 922 individuals from the 53 worldwide populations in the CEPH-Human Genome Diversity panel (HGDP- CEPH)[1], and to whole genome sequences from 25 worldwide individuals. The analysis suggests that Ust'-Ishim is equally closely related to most present-day Eurasians.

Principal components analysis

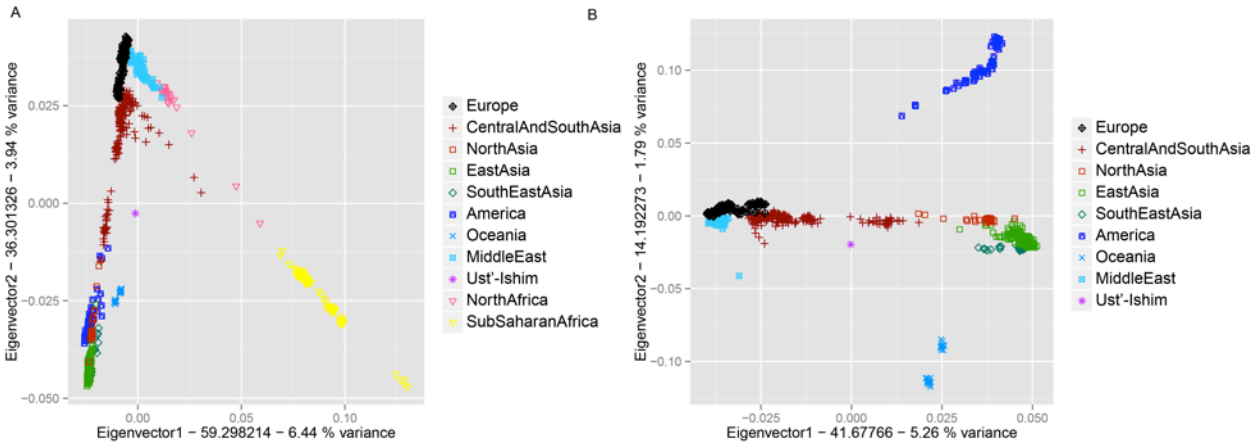
To investigate the relationship of the Ust'-Ishim individual to present-day humans worldwide, we computed pairwise allele-sharing distances [2] among all pairs of 922 HGDP-CEPH individuals representing 53 global populations [1]. We used a total of 492,308 SNPs for the analysis.

We carried out a principal components analysis (PCA) on the resulting pairwise distance matrix using SMARTPCA [2] and projected genotyping calls from Ust'-Ishim onto the PCA. We group populations by region: Europe, Central and South Asia, North Asia, East Asia, Southeast Asia, America, Oceania, Middle East, North Africa and Sub-Saharan Africa (Figure S10.1).

The first PC distinguishes Africans from non-Africans, and Ust'-Ishim clusters among the non-Africans. PC2 separates West Eurasians from eastern non-Africans (Figure S10.1A).

When we excluded the African populations from the PCA (Figure S10.1B). PC1 separates eastern non-African from European populations, while PC2 separates Oceanians and Native Americans from other populations. The PCA analysis shows that Ust'-Ishim does not group more closely with any particular present-day population. Since the Ust'-Ishim individual is at least 45,000 years old, this raises the possibility that it is related to the ancestral population of both Europeans and Asians.

Figure S10.1. PCA analysis using *SMARTPCA* [2] on HGDP SNP data with Ust'-Ishim projected.



Model-based clustering

We analyzed the same dataset using ADMIXTURE [3], which implements a model-based maximum likelihood (ML) clustering algorithm to give a ML estimate for allele frequencies in a predefined number (K) of hypothetical ancestral populations as well as a probabilistic assignment of each individual to these ancestral populations.

ADMIXTURE provides a cross validation option to define the most likely number of ancestral populations K (ADMIXTURE 1.04 Software Manual). The CV error is minimized at K=10 (Figure S10.2), and hence we use this for display of the results.

Figure S10.2. Cross-validation plot to determine the most likely number of clusters for the CEPH-HGDP data

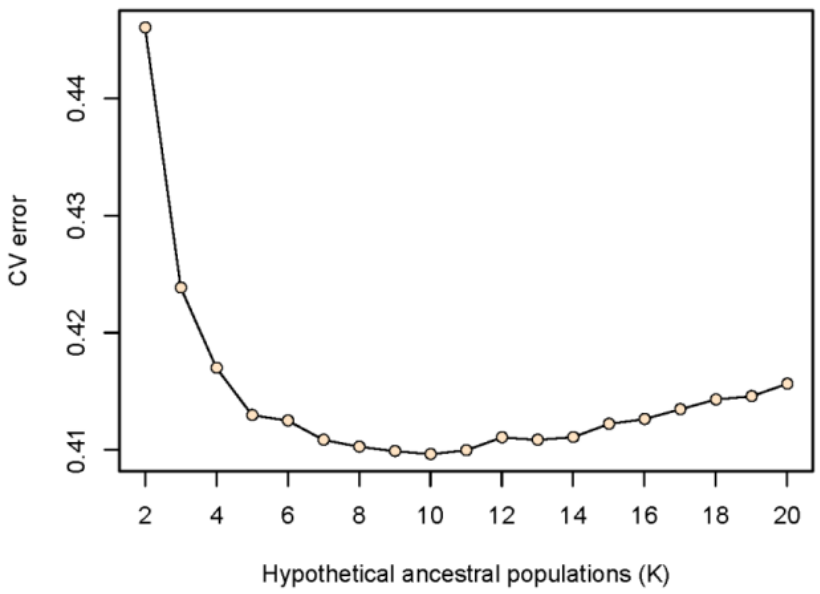
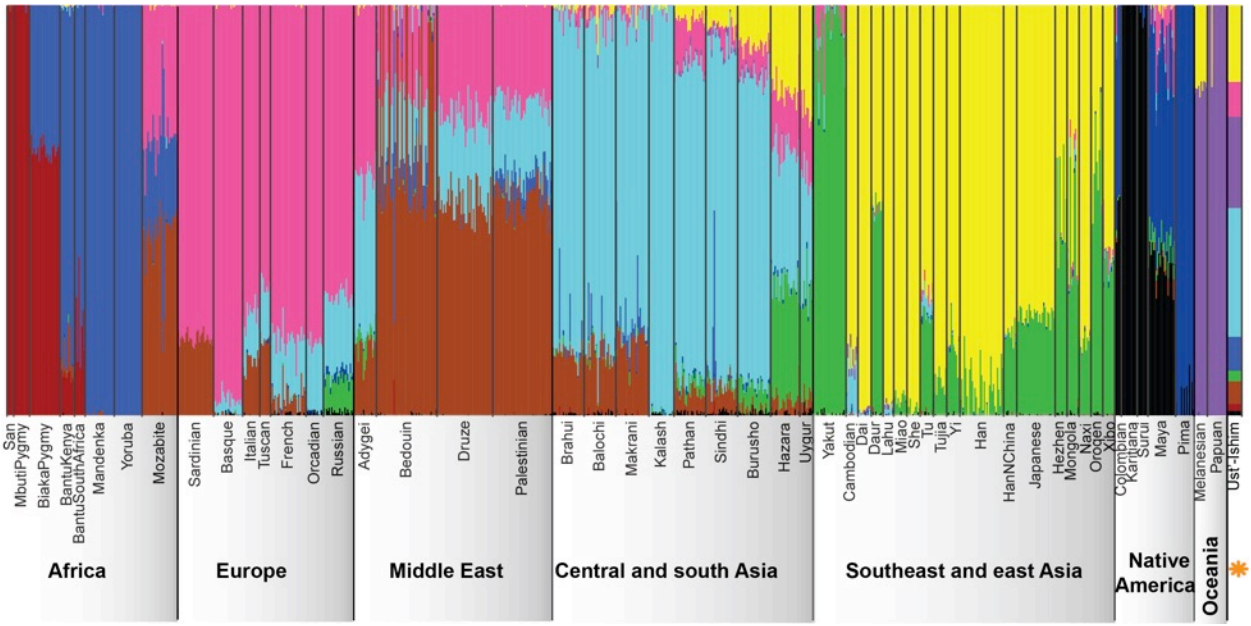


Figure S10.3 shows the ADMIXTURE results for K=10, with the column width for the Ust'-Ishim individual increased to assist in visualization. The Ust'-Ishim individual is not assigned to any of the constructed population groupings and does not appear to be particularly closely affiliated with any particular present-day population, consistent with the PCA results and the age of the individual.

Figure S10.3. Results of the ADMIXTURE analysis (K = 10) with present-day populations and Ust'-Ishim.



Phylogenetic tree incorporating admixture

To provide an overview of how Ust'-Ishim is related to present-day humans, we used TreeMix, which uses genome-wide allele frequency data and a Gaussian approximation of genetic drift to infer the population relationships using maximum likelihood trees and allowing for both population splits and potential gene flow [4]. For this analysis, we used whole genome sequence data, comparing Ust'-Ishim to the genomes of an Altai Neandertal, a Denisovan [5,6], and 25 present-day humans representing populations from Africa (Mbuti, Yoruba, Mandenka, and Dinka), Europe (French and Sardinian), Oceania (Papuan, Australian), East Asia (Dai and Han), and the Americas (Karitiana, Mixe). For each population there are two individuals sequenced except for the Mixe where there is only a single individual [5]. The allele frequencies were obtained from genotype call files described in SI6. With the filters described in SI7, there are 4,209,798 variable positions in these 28 individuals and we used these for the analysis.

When no mixture events are included TreeMix infers a tree with the same branching order as that seen for the neighbor joining trees (not shown) (Figure S10.4A). There is 100% bootstrap support for Ust'-Ishim being in the same clade as Papuan, Australian, Karitiana, Mixe, Han and

Dai. Interestingly, Ust'-Ishim branches before both the Oceanians and Asians. Figure S10.4B presents the matrix of residuals documenting how well this tree model fits the data. The positive residuals suggest three admixture events: (i) among Altai Neandertal, Denisovan, and Papuan ancestors; among (ii) French, Karitiana and Mixe; and (iii) between Altai Neandertal and Ust'-Ishim. Admixture between Denisovan and Papuan, as well as between European and Native Americans has been reported in previous studies [7,8].

Figure S10.4. (A) Maximum-likelihood tree from TreeMix. The scale bar shows ten times the average standard error of the entries in the sample covariance matrix. (B) Residual fit of the observed versus predicted squared allele frequency difference, expressed as the number of standard errors of the deviation. Colors are in the palette on the right. Residuals above zero represent populations that are more closely related to each other in the data than in the best-fit tree, and are candidates for admixture.

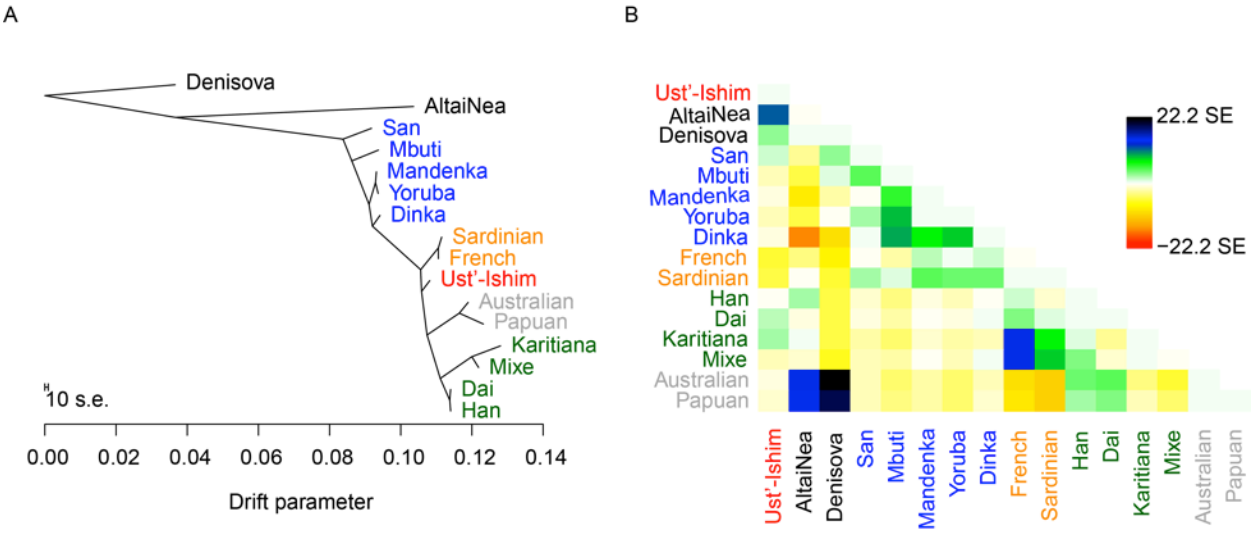
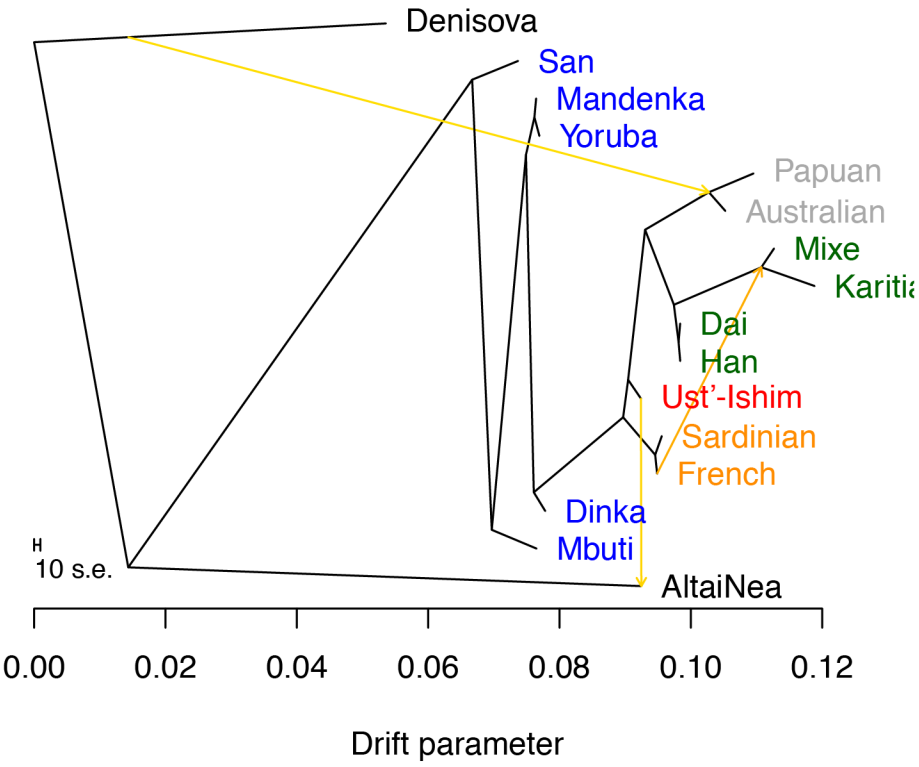


Figure S10.5. Maximum-likelihood tree from TreeMix when allowing three migrations.

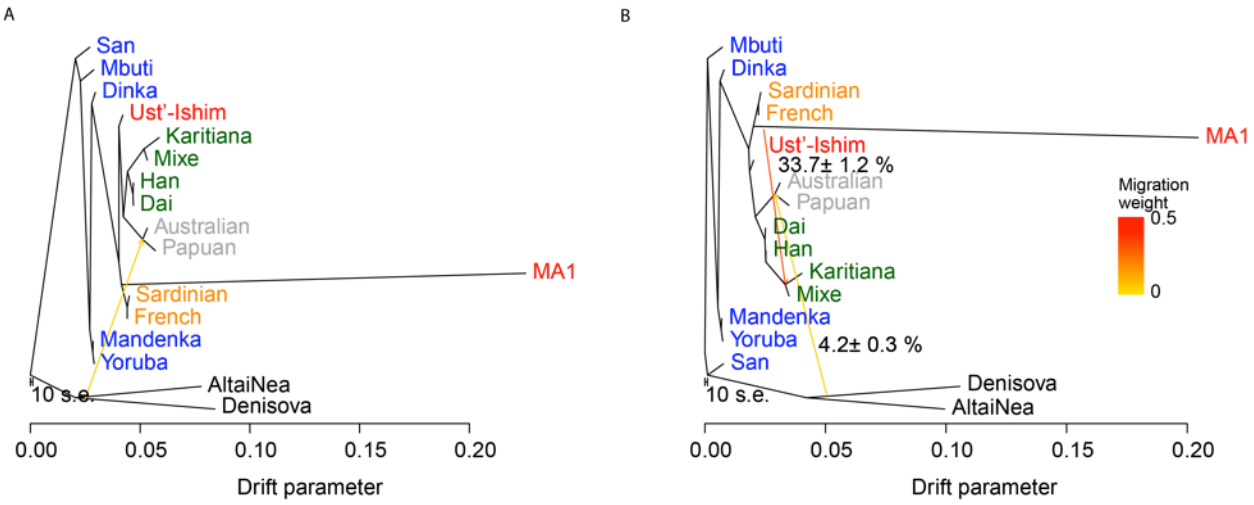


When we allow three migrations, TreeMix continues to place the Ust'-Ishim individual in the same clade as Papuan, Australian, Karitiana, Mixe, Han and Dai with 100% bootstrap support (Figure S10.5). The inferred admixture between Neandertal and Ust'-Ishim is consistent with this individual being a Eurasian that received Neandertal gene flow (Supplementary Note 16)

We next added into the TreeMix analysis an ancient modern human from Mal'ta (~1 fold genomic coverage) dating to around 24,000 BP. For the low coverage Mal'ta sample, we excluded all reads with mapping quality lower than 37. At sites covered by more than one read, we randomly sampled a read. We excluded all transition substitutions, and restricted to sites that were polymorphic in the high coverage genome sequences. There are 1,105,222 variable sites in the resulting dataset. A tree without migration events places the Ust'-Ishim individual on the branch that is basal to all Asian populations, but with only 65% bootstrap support (100 bootstrap replicates). When we allow two migration rates, the method infers: (i) 4.2% Denisovan gene flow into the ancestors of Papuans and Australians, and (ii) Native American populations

deriving ~33.7% of their ancestry from a population related to Mal'ta. Both these inferences are in agreement with previous reports [7-9]. When these are taken into account, Ust'-Ishim is resolved as being basal to all Asians with 95% bootstrap support. (Figure S10.6).

Figure S10.6 TreeMix ML trees including migration edges for the genomes of the Ust'-Ishim individual, the Mal'ta individual (MA1), 25 present-day humans, the Denisovan and the Neandertal.

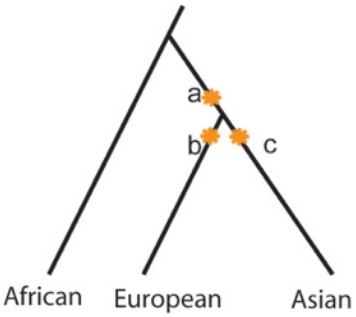


We caution that the TreeMix model is sensitive to which present-day populations are used for the analysis and the tree changes when we use different present-day human populations (Table S10.1). In particular, the analysis is sensitive to which African individuals are included. When we including only one African individual and all available Eurasian individuals, the Ust'-Ishim individual separates before the Eurasian split (position **a**) with bootstrap support between 68% and 100% depending on which population the individual is from (Table S10.1). These results suggest that recent gene flow between African and from European populations may influence the placement of Ust'-Ishim in the maximum likelihood tree. Nevertheless, the TreeMix models typically find that the Ust'-Ishim individual separates either before Eurasian split (position **a** in Figure S10.7) or just after the split and already on the eastern non-African lineage (position **c** in Figure S10.7).

Table S10.1. The bootstrap replicate support for placement of Ust'-Ishim in the phylogenetic tree varies depending on the populations used in the analysis

Combination	Population	Position a	Position b	Position c
All population from Panel A &B	French Sardinian Han Dai Karitiana Mixe Australian Papuan Dinka San Yoruba Mandenka Mbuti	0		100
No Dinka	French Sardinian Han Dai Karitiana Mixe Australian Papuan San Yoruba Mandenka Mbuti	3		97
One pop from Europe and East Asian, all Africans	French Han Dinka San Yoruba Mandenka Mbuti	68		32
One pop from Europe, East Asian and Africans	French Han Yoruba	100		
Two pop from Europe, East Asian and Africans	French Sardinian Han Dai Yoruba Mbuti	71		29
Only San from Africa	French Sardinian Han Dai Karitiana Mixe Australian Papuan Dinka	100		0
Only Mbuti from Africa	French Sardinian Han Dai Karitiana Mixe Australian Papuan Dinka	100		0
Only Mandenka from Africa	French Sardinian Han Dai Karitiana Mixe Australian Papuan Dinka	81		19
Only Yoruba from Africa	French Sardinian Han Dai Karitiana Mixe Australian Papuan Dinka	68		32
Only Dinka from Africa	French Sardinian Han Dai Karitiana Mixe Australian Papuan Dinka	76	2	22

Figure S10.7. Alternative potential phylogenetic positions of Ust'-Ishim to help in interpreting Table S10.1



References

1 Lu, T. G., S. Mallick, A. Ollmann, N. Patterson, Y. Zhan, T. Webster, D. Reich in *American Society of Human Genetics/ICHG 2011.* (ASHG).

2 Cavalli-Sforza, L. L. Genetic and cultural diversity in Europe. *Journal of Anthropological Research* **53**, 383-404 (1997).

3 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

4 Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *precedings.nature.com* http://precedings.nature.com/documents/6956/version/1 (2012).

5 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

6 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

7 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

8 Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* **89**, 516-528, doi:10.1016/j.ajhg.2011.09.005 (2011).

9 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, doi:10.1038/nature12736 (2013).

Supplementary Information 11

Relationship of Ust'-Ishim to other humans

Qiaomei Fu*, Janet Kelso, David Reich
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Computation of D-statistics

To investigate how the Ust'-Ishim genome sequence relates to other humans, both present-day and ancient, we used *D*-statistics [1-3].

If *A, B, C, D* are 4 populations, then:

$$D(A,B;C,D) = \frac{E[n_{BABA}]-E[n_{ABBA}]}{E[n_{BABA}]+E[n_{ABBA}]} \quad (6.1)$$

where E[n_{BABA}] is the expected count of alleles agreeing in populations A, C and also in B, D (but differing in A, B) and E[n_{ABBA}] is the expected count of alleles agreeing in populations A, D and also in B, C (but different in A, B). A positive *D* is an indication of more allele-sharing between either of the population pairs *(A, C)* or *(B, D)*. In contrast, a negative *D* indicates more shared alleles between either of the population pairs *(A, D)* or *(B, C)*.

We can have multiple samples representing each population (for example, for a diploid individual we have 2 samples). Defining the frequency of allele *i* in a sample from each population as a_i, b_i, c_i, and d_i, the expectations for the counts of the different sites patterns are:

$$E[n_{BABA}] = \sum_{i=1}^{n}[a_i(1-b_i)c_i(1-d_i) + (1-a_i)b_i(1-c_i)d_i] \quad (6.2)$$

$$E[n_{ABBA}] = \sum_{i=1}^{n}[(1-a_i)b_ic_i(1-d_i) + a_i(1-b_i)(1-c_i)d_i] \quad (6.3)$$

We computed standard errors on the *D*-statistics using a Weighted Block Jackknife [4] with a block size of 5 million base pairs (5 Mb).

We used *D*-statistics to test whether there is evidence that Ust'-Ishim shares more alleles with diverse pairs of other present-day populations, using the statistic *D(X, Y; Ust'-Ishim, Outgroup)*.

We computed statistics on whole-genome sequence data by identifying sites that passed the standard filters described in SI7 for both Ust'-Ishim and each of the following samples:

- The 25 present-day humans in Panels A and B.
- A 7,000 year old Mesolithic western Eurasian hunter-gatherer from Spain (La Braña) [5].
- A 24,000 year old Upper Paleolithic northern Eurasian hunter-gatherer from the Lake Baikal region of Siberia – Mal'ta (MA1) [6].

For the low coverage MA1 (~1×) and La Braña genomes (~3×), we exclude reads with a mapping quality lower than 37, and sample a random read for sites covered by more than one read. For D-statistics involving MA1 and La Braña we also restricted to transversion polymorphisms, as we were concerned that the high rate of C→T and G→A substitutions in these non-UDG-treated samples could influence results.

We also computed *D*-statistics on a set of 594,924 SNPs genotyped on the Human Origins Array (Lazaridis et al. (2014) in a set of individuals from the same populations represented in the A and B panels, as well as individuals from the Andamanese Onge population. We include the Onge as they have no evidence of the Denisovan admixture that has affected indigenous Papuans, Australians, and Philippine populations, and may therefore be a useful population for understanding the early dispersal of modern humans out of Africa. We added to this dataset randomly sampled alleles for La Braña and MA1 and genotypes from Ust'-Ishim.

Table S11.1. *D*-statistics of the form *D*(*X, Y; Ust'-Ishim, Yoruba).*

		Whole genome sequence			Human Origins SNPs	
Y	***X***	**D**	**Z-score**	**Sites used**	**D**	**Z-score**
French	Onge	n/a	n/a	n/a	0.0258	6.985
French	Karitiana	0.0279	4.216	7,811,057	0.026	6.791
French	Han	0.0166	2.413	7,781,350	0.0243	7.375
French	La Braña	0.0261	4.074	6,831,661	0.0211	4.697
French	MA1	0.0317	4.707	5,742,755	0.0283	5.88
Sardinian	Onge	n/a	n/a	n/a	0.0294	7.683
Sardinian	Karitiana	0.0418	6.131	7,823,169	0.0267	6.202
Sardinian	Han	0.0299	4.11	7,790,169	0.0294	7.389
Sardinian	La Braña	0.0399	6.139	6,843,461	0.0279	8.07
Sardinian	MA1	0.0447	6.192	5,749,878	0.0251	5.532
Karitiana	Han	-0.0122	-1.824	7,773,319	-0.0013	-0.366
Karitiana	La Braña	-0.0035	-0.491	6,828,239	-0.0055	-0.971
Karitiana	MA1	0.0028	0.389	5,736,395	0.0018	0.319
Han	La Braña	0.0066	0.915	6,800,674	-0.0043	-0.833
Han	MA1	0.0134	1.841	5,715,592	0.0029	0.532
La Braña	MA1	0.0097	1.348	5,160,808	0.0109	1.657

Table S11.2. *D*-statistics of the form *D*(*X, Y; Ust'-Ishim, Human-chimpanzee ancestor or Chimpanzee)*

		Whole genome sequence			Human Origins SNPs	
Y	***X***	**D**	**Z-score**	**Available SNPs**	**D**	**Z-score**
Mbuti	Onge	n/a	n/a	n/a	0.3523	81.735
Mbuti	Karitiana	0.384	73.093	5,780,989	0.3491	77.611
Mbuti	Han	0.3834	74.842	5,753,895	0.3511	88.604
Mbuti	La Braña	0.3812	70.895	5,093,332	0.3437	62.723
Mbuti	MA1	0.3789	67.454	4,293,995	0.3425	59.671
Yoruba	Onge	n/a	n/a	n/a	0.2768	69.276
Yoruba	Karitiana	0.3108	56.346	5,799,652	0.2733	64.712
Yoruba	Han	0.3114	57.293	5,773,067	0.2757	76.186
Yoruba	La Braña	0.3071	53.315	5,103,775	0.2686	50.582
Yoruba	MA1	0.3034	51.914	4,303,961	0.2672	47.801
Dinka	Onge	n/a	n/a	n/a	0.2501	58.334
Dinka	Karitiana	0.2787	49.342	5,762,904	0.2464	53.923
Dinka	Han	0.2785	50.525	5,739,775	0.249	63.639
Dinka	La Braña	0.276	49.09	5,075,752	0.2419	43.769
Dinka	MA1	0.2728	44.475	4,287,816	0.2412	41.001
Onge	Karitiana	n/a	n/a	n/a	-0.003	-0.625
Onge	Han	n/a	n/a	n/a	-0.0011	-0.29
Onge	La Braña	n/a	n/a	n/a	-0.0083	-1.276
Onge	MA1	n/a	n/a	n/a	-0.0091	-1.303
Karitiana	Han	-0.0017	-0.266	5,788,415	0.002	0.504
Karitiana	La Braña	-0.0062	-0.9	5,113,970	-0.0056	-0.816
Karitiana	MA1	-0.0085	-1.188	4,313,633	-0.0056	-0.8
Han	La Braña	-0.0057	-0.823	5,090,000	-0.0076	-1.202
Han	MA1	-0.0057	-0.801	4,295,352	-0.0064	-0.97
LaBrana	MA1	-0.0026	-0.371	3,921,017	-0.0007	-0.081
French	Onge	n/a	n/a	n/a	0.0205	4.929
French	Karitiana	0.0213	3.229	5,816,531	0.0181	4.155
French	Han	0.0186	2.819	5,792,641	0.0196	5.369
French	La Braña	0.0165	2.636	5,111,343	0.0132	2.348
French	MA1	0.0138	2.108	4,314,385	0.0139	2.287
Sardinian	Onge	n/a	n/a	n/a	0.024	5.605
Sardinian	Karitiana	0.0282	4.295	5,827,582	0.0215	4.751
Sardinian	Han	0.0261	3.78	5,800,680	0.0231	6.091
Sardinian	La Braña	0.0249	3.734	5,123,792	0.017	3.003
Sardinian	MA1	0.0214	3.15	4,322,387	0.0169	2.724

Results

Table S11.1 reports the D-statistics using the panel B Yoruba individual, a sub-Saharan African, as the outgroup. Table S11.2 reports the results when the outgroup is the human-chimpanzee ancestral sequence for the comparisons using whole genome data or chimpanzee for the comparisons using the Human Origins SNPs.

These analyses document the following patterns:

1. The Ust'-Ishim individual is clearly more closely related to present-day non-Africans than to present-day Africans, as statistics of the form D(non-African African; Ust'-Ishim, Outgroup) are always highly significantly positive (Z>44 standard errors from 0).
2. The D-statistics involving Ust'-Ishim and present-day non-Africans are consistent with Ust'-Ishim being an outgroup to all eastern non-Africans (Onge, Han, Karitiana) and to ancient west Eurasian (La Braña) and north Eurasian (MA1) hunter-gatherers. Specifically, we find that none of these samples shares significantly more alleles with Ust'-Ishim than any other (all |Z|-scores in Table S11.1 and Table S11.2 have values of <2). This finding is consistent with Ust'-Ishim having separated from other Eurasians around the time of the ancient divergence of West and North Eurasian hunter-gatherers (represented by La Braña and MA1), and east Eurasians (represented by Onge, Karitiana and Han). The absence of evidence for shared genetic drift with any of these groups suggests that Ust Ishim is likely to be close to the ancestor of most Eurasians, and is consistent with the old radiocarbon date.
3. The Ust'-Ishim individual shares significantly fewer alleles with present-day Europeans than with any other non-Africans we analyzed (|Z|-scores of 2.1 to 7.7, Table S11.1 and Table S11.2). We observe a similar result when we compare other present-day Europeans and Near Easterners to other Eurasians (not shown); thus, this is a general feature of present-day European populations. These results are consistent with present-day West Eurasians having inherited some of their ancestry from an ancient branch of Eurasian genetic variation that diverged prior to the separation of Ust'-Ishim from other non-Africans. This is unlikely to be an artifact of the low levels of gene flow between Yoruba and West Eurasians that have previously been documented [7], as when we recomputed the statistic with human-chimpanzee common ancestor (CA) or only chimpanzee as the outgroup instead of the Yoruba, we obtain similar results (Table S11.2). The finding of "Basal Eurasian" ancestry in West Eurasians is consistent with the report of Lazaridis et al.[8] where the finding of Basal Eurasian

ancestry in present-day West Eurasians was hypothesized to reflect ancient Near Eastern populations having a component of their ancestry from a group that was not part of the early dispersal of modern humans to Europe and northern and eastern Asia. The results with Ust'-Ishim are consistent with this idea and suggest that Ust'-Ishim may not have carried any Basal Eurasian ancestry.

References

1 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

2 Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, doi:10.1126/science.1224344 (2012).

3 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, doi:10.1534/genetics.112.145037 (2012).

4 Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* **17**, 1217-1241, doi:DOI 10.1214/aos/1176347265 (1989).

5 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228, doi:10.1038/nature12960 (2014).

6 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, doi:10.1038/nature12736 (2013).

7 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

8 Lazarides, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **In Press** (2014).

Supplementary Information 12

Heterozygosity

Cesare de Filippo*, Qiaomei Fu, Kay Prüfer
* To whom correspondence should be addressed (cesare_filippo@eva.mpg.de)

We estimated the autosomal heterozygosity of the Ust'-Ishim individual using two independent approaches: the first uses genotype calls from GATK (SI6) and the second uses the software package *mlRho* [1].

The GATK heterozygosity estimate was obtained by dividing the number of heterozygous genotypes by the total number of genotypes that passed the autosome filtering thresholds described in SI7.

The heterozygosity calculated by *mlRho* corresponds to the maximum likelihood estimate of the population mutation parameter θ under the infinite site model, which can be translated into heterozygosity if θ is small. We run *mlRho* on the sequencing reads that passed the minimal filters, but in addition we disregard sites with base quality lower than 30.

The average whole autosomal heterozygosity is very similar for both methods (Table S12.1). For simplicity we therefore describe results for only the GATK method.

The results (Table S12.1 and Figure S12.1) indicate that Ust'-Ishim has on average 8.17 heterozygous sites every 10,000 basepairs, a value that is ~4-fold higher than the heterozygosity in the archaic genomes of the Altai Neandertal and Denisovan individuals [2]. When compared to present-day humans, Ust'-Ishim heterozygosity is slightly higher than that of other non-Africans (range 5.48-7.77), but lower than the heterozygosity in Africans (range 9.64-10.65). It is worth to mention that the only significant differences between the distributions of heterozygosity across the autosomes are between Ust'-Ishim and the two American Karitiana individuals (Mann-Whitney U $P < 0.05$).

Table S12.*1 Heterozygosity estimates per 10,000 sites for Ust'-Ishim, archaic, and present-day humans*

SAMPLE	POP	Genotype calls	mlrho
Ust'-Ishim	Ust'-Ishim	8.17	7.67
Altai	Neandertal	1.75	1.67
Denisova	Denisova	2.13	1.88
HGDP01015	Karitiana_B	5.48	5.52
HGDP00998	Karitiana_A	5.72	5.65
HGDP00546	Papuan_B	5.94	5.98
MIXE0007	Mixe_B	6.08	6.1
HGDP00542	Papuan_A	6.33	6.28
WON,M	Australian1_B	6.56	6.6
BUR,E	Australian2_B	6.62	6.66
HGDP01308	Dai_B	7.15	7.15
HGDP00775	Han_B	7.19	7.18
HGDP01307	Dai_A	7.4	7.31
HGDP00778	Han_A	7.4	7.31
HGDP01076	Sardinian_B	7.3	7.34
HGDP00533	French_B	7.54	7.57
HGDP00521	French_A	7.77	7.68
HGDP00665	Sardinian_A	6.33	7.68
DNK07	Dinka_B	9.64	9.64
DNK02	Dinka_A	9.95	9.86
HGDP01286	Mandenka_B	9.97	10
HGDP00927	Yoruba_A	10.14	10
HGDP00456	Mbuti_A	10.19	10.1
HGDP00982	Mbuti_B	10.05	10.1
HGDP00936	Yoruba_B	10.02	10.1
HGDP01284	Mandenka_A	10.26	10.2
HGDP01036	San_B	10.17	10.2
HGDP0102	San_A	10.65	10.5

9

SAMPLE	POPULATION	GENOTYPE CALLS	MLRHO
Ust'-Ishim	Ust'-Ishim	8.17	7.67
Altai	Neandertal	1.75	1.67
Denisova	Denisova	2.13	1.88
HGDP01015	Karitiana_B	5.48	5.52
HGDP00998	Karitiana_A	5.72	5.65
HGDP00546	Papuan_B	5.94	5.98
MIXE0007	Mixe_B	6.08	6.1
HGDP00542	Papuan_A	6.33	6.28
WON,M	Australian1_B	6.56	6.6
BUR,E	Australian2_B	6.62	6.66
HGDP01308	Dai_B	7.15	7.15
HGDP00775	Han_B	7.19	7.18
HGDP01307	Dai_A	7.4	7.31
HGDP00778	Han_A	7.4	7.31
HGDP01076	Sardinian_B	7.3	7.34
HGDP00533	French_B	7.54	7.57
HGDP00521	French_A	7.77	7.68
HGDP00665	Sardinian_A	6.33	7.68
DNK07	Dinka_B	9.64	9.64
DNK02	Dinka_A	9.95	9.86
HGDP01286	Mandenka_B	9.97	10
HGDP00927	Yoruba_A	10.14	10
HGDP00456	Mbuti_A	10.19	10.1
HGDP00982	Mbuti_B	10.05	10.1
HGDP00936	Yoruba_B	10.02	10.1
HGDP01284	Mandenka_A	10.26	10.2
HGDP01036	San_B	10.17	10.2
HGDP01029	San_A	10.65	10.5

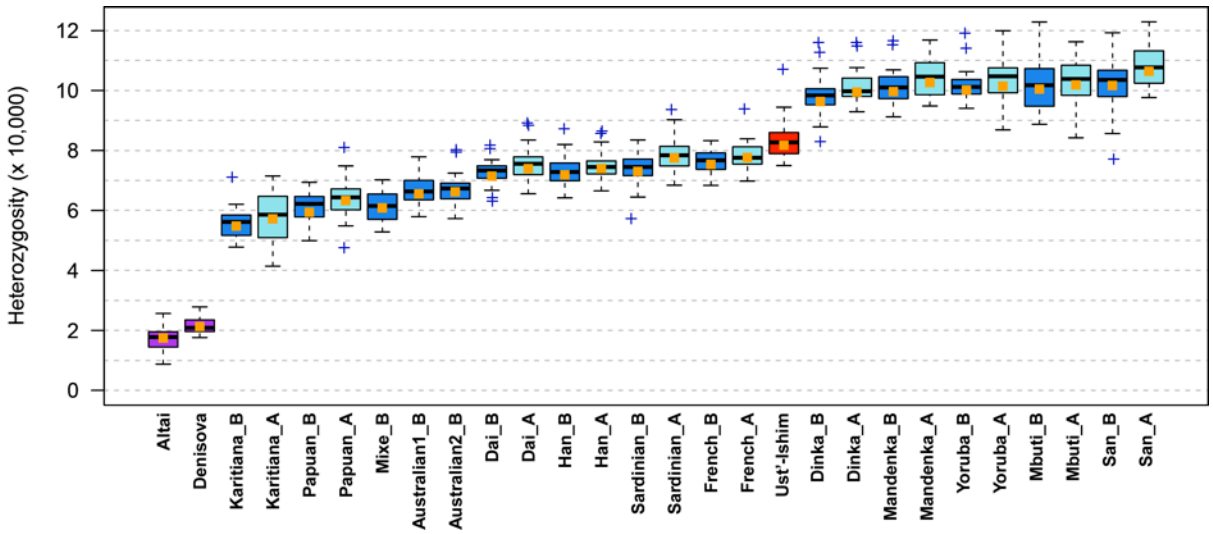


Figure S12.1: *Heterozygosity estimates on the autosomes. The yellow squares are the heterozygosity estimates using genotype calls (i.e. those in Table S12.1) for 10,000 sites and the box-and-whiskers represent the distributions across 22 autosomes given by the R-function 'boxplot' with default parameters [3]. The Ust'-Ishim individual is in red. Archaic samples are in purple. Present-day human from panel A and panel B are in light and dark blue, respectively. The samples 'Australian1_B' and 'Australian2_B' are 'WON, M' and 'BUR,E', respectively (see Table S12.1).*

References

1 Haubold, B., Pfaffelhuber, P. & Lynch, M. mlRho - a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Mol Ecol* **19 Suppl 1**, 277-284, doi:10.1111/j.1365-294X.2009.04482.x (2010).

2 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

3 Becker, R. A., Chambers, J. M. & Wilks, A. R. . *The new S language: a programming environment for data analysis and graphics*. (Chapman & Hall 1988).

Supplementary Information 13

Analysis of inbreeding in Ust'-Ishim

Qiaomei Fu*, Janet Kelso, Montgomery Slatkin and Flora Jay
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Summary

To assess inbreeding among the ancestors of the Ust Ishim individual we compared the lengths of runs of homozygosity (ROH) in the Ust'-Ishim individual to ROH in the genomes of present-day humans, the Denisovan, and the Altai Neandertal. We find no evidence of recent inbreeding among the ancestors of the Ust'-Ishim individual.

Methods

VCF files containing genotypes were called using GATK (SI6) for all autosomes. Sites were considered heterozygous if they were called heterozygous and pass the autosomal filters described in SI7. Only sites that pass the filters for Ust'-Ishim, all present-day, and all archaic individuals simultaneously were used.

For each individual we defined a "run of homozygosity" (ROH) as the distance between adjacent heterozygous sites on each chromosome. The minimum ROH length considered was 200 kb. For each individual, the number of runs of homozygosity in each of four size categories was counted and summed (Figure S13.1).

Results

Figure S13.1 gives the number of runs of homozygosity and their total length in several length categories for a set of present-day and archaic individuals. For ROH lengths larger than 200 kb, the length distribution of ROH and the total number of long runs of homozygosity in Ust'-Ishim are comparable to that seen in present day non-African individuals, indicating that the parents of the Ust'-Ishim individual are unlikely to have been closely related. This is in contrast to the Altai Neandertal where the total lengths of ROH are larger than in any other population, consistent with the parents of this Neandertal individual being closely related to each other [1].

We do not analyze runs of homozygosity shorter than 200 kb as these are likely to be identical-by-state (IBS) tracts that are not identical-by-descent (IBD), and thus not informative for inbreeding. We note that sequencing errors have an impact on the length distribution of detected ROH, in that they tend to fragment longer IBD tracts. Encouragingly, we find that IBD estimates for Panel A and Panel B individuals from the same populations are similar (Figure S13.1), despite different sequence error rates between the individuals in Panel A and Panel B [1]. This suggests that the global picture given by the total number and total length of long ROH is not substantially impacted by sequence error. To further investigate whether our results are robust to sequence error, we reanalyze the data with PLINK [2], which takes a sliding window of 2 Mb and identifies windows of homozygosity. Within a window a single heterozygous call and still be considered homozygous. A position is then defined as being in a ROH if more than 5% of windows overlapping the position were identified as homozygous. Consistent with the analysis of strict ROH (allowing no heterozygous sites), Ust'-Ishim falls in the range of other present-day humans (Figure S13.2). This remains true when the number heterozygous calls allowed per window is increased to 2 or 4. Figure S13.3 confirms that the Altai Neandertal has a higher number of long ROH than other individuals, regardless whether 1, 2 or 4 heterozygous calls are allowed.

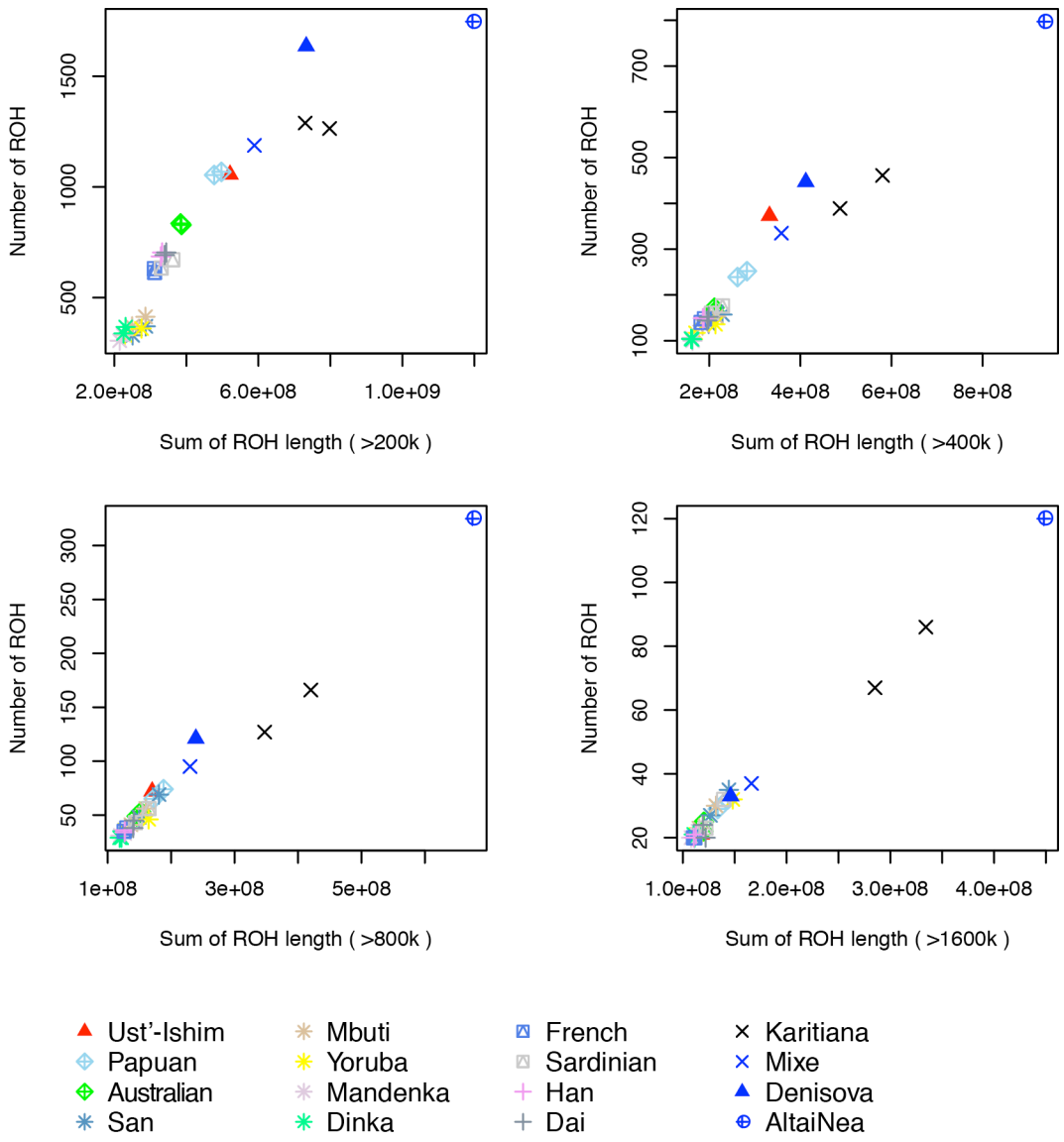


Figure S13.1. *The number of ROH compared to the total ROH length in several length categories starting from 200 kb in size for individuals from several present-day and archaic populations.*

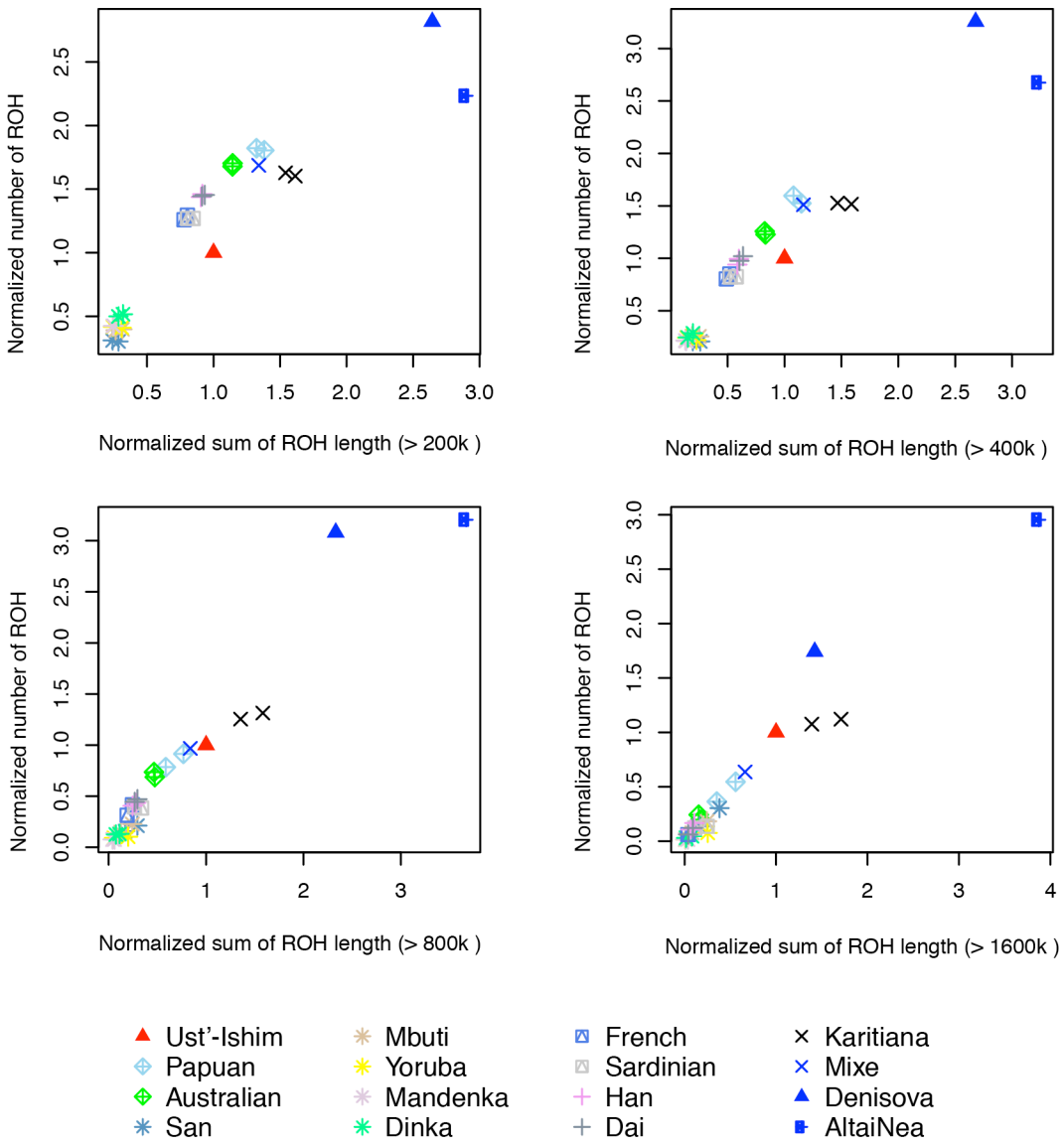


Figure S13.2. *The number of ROH compared to the total ROH length (allowing one heterozygous call in the track) normalized by the value of the Ust'-Ishim in several length categories starting from 200 kb in size for individuals from several present-day and archaic populations.*

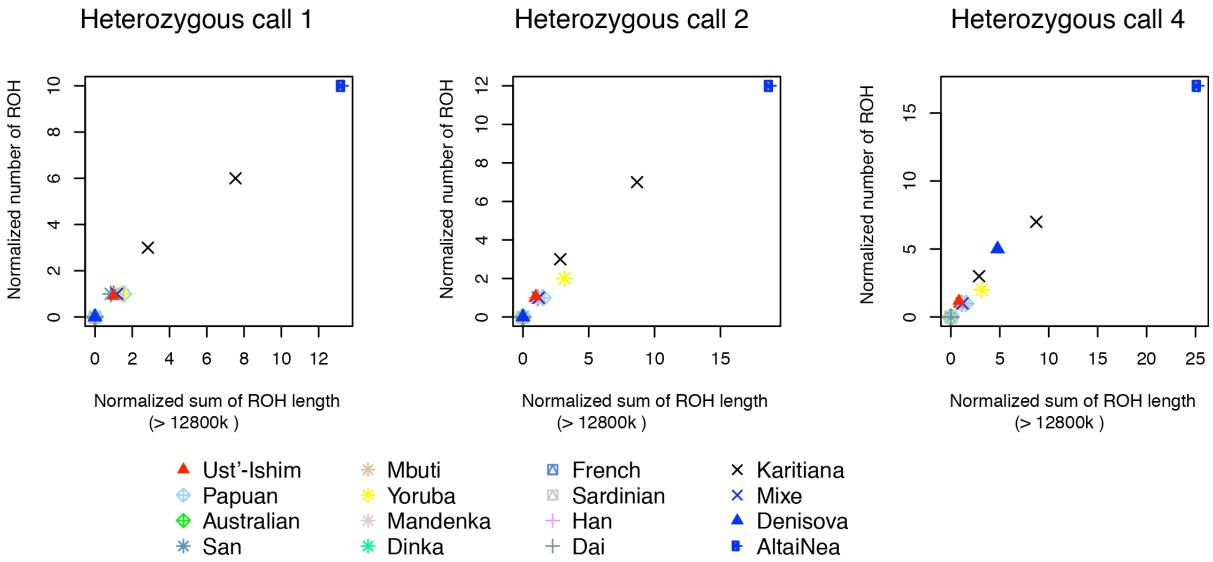


Figure S13.3. *The number of ROH compared to the total ROH length (allowing one, two or four heterozygous calls in the track) normalized by the value of the Ust'-Ishim in 12800 kb length with different for individuals from several present-day and archaic populations.*

References

1 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

Supplementary Information 14

A novel estimate of the human mutation rate

Heng Li*, Qiaomei Fu and David Reich
* To whom correspondence should be addressed (hengli@broadinstitute.org)

Summary

We used the pairwise sequentially Markovian coalescent (PSMC) method to infer the population size changes over time in the Ust'-Ishim individual, Altai Neandertal, Denisovan, and the present-day humans in Panel A and Panel B [1].

The genomic coverage for each individual is given in Table S14.1.

Table S14.1 Mean genomic sequence coverage of samples

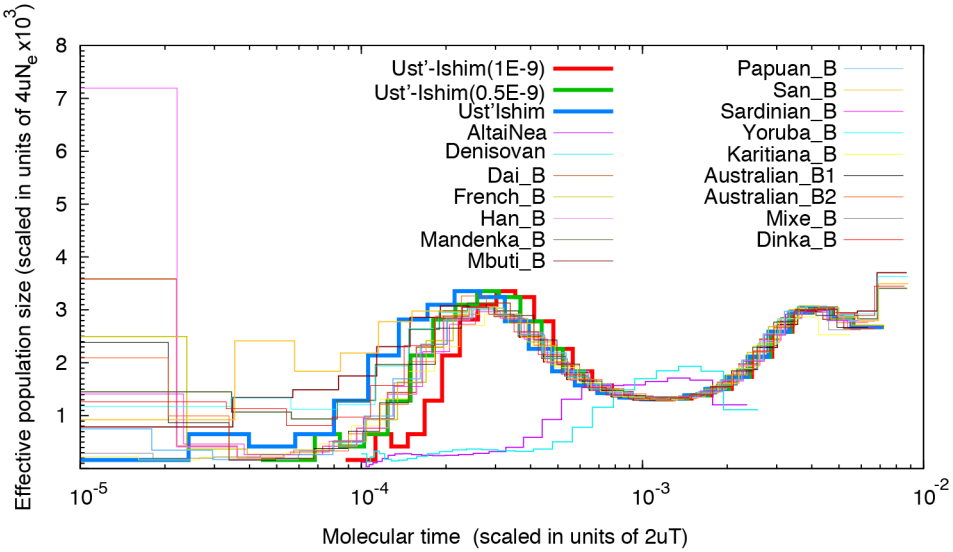
Samples	IDS	Mean Coverage
Ust'-Ishim	Ust'-Ishim	42.42
AltaiNea	AltaiNea	52.15
Denisovan	DenisovaPinky	30.74
Dinka_A	DNK02	27.98
French_A	HGDP00521	26.72
Papuan_A	HGDP00542	25.92
Sardinian_A	HGDP00665	24.67
Han_A	HGDP00778	27.74
Yoruba_A	HGDP00927	32.11
Karitiana_A	HGDP00998	26.01
San_A	HGDP01029	32.74
Mandenka_A	HGDP01284	24.5
Dai_A	HGDP01307	28.3
Mbuti_A	HGDP00456	24.34
Dai_B	HGDP01308	37.03
French_B	HGDP00533	42.21
Han_B	HGDP00775	35.18
Mandenka_B	HGDP01286	36.7
Mbuti_B	HGDP00982	37.07
Papuan_B	HGDP00546	42.44
San_B	HGDP01036	38.27
Sardinian_B	HGDP01076	38.03
Yoruba_B	HGDP00936	38.76
Karitiana_B	HGDP01015	35.05
Mixe_B	MIXE0007	41.71
Australian_B1	BUR,E	41.88
Australian_B2	WON,M	36.81

Dinka_B	DNK07	34.77

To generate a diploid consensus sequence for each individual, we used SAMtools mpileup version 0.1.18, including the extended BAQ calculation. We required (i) that read depth was between 1/3 and 2-times the average coverage (Table S14.1); (ii) that root-mean-squared mapping quality for all analyzed sites was above 10; (3) that all analyzed sites were at least 5 bp away from a predicted insertion/deletion; (4) that consensus quality was >30; and (v) that at least 18 of 35 possible overlapping 35-mers in the human reference genome sequence mapped to the correct location with zero or one mismatch. We called SNPs using SAMtools. We then used the PSMC to infer the distribution of coalescent times between two chromosomes across the autosomes.

Even without correcting for branch shortening, the Ust'-Ishim inferred population history is qualitatively similar to that of present-day modern humans (Figure S14.1, wide blue line). In particular, the Ust'-Ishim individual is inferred to have a founder event that is similar to that of East Asian and European populations. We hypothesize that it is the same historical event, likely the "out of Africa bottleneck", and the rest of this note is premised on that assumption.

Figure S14.1 Inferred population size changes over time from PSMC. *The molecular time is measured by the pairwise per-site sequence divergence d. The molecular dates of the Denisovan and the Altai Neandertal are assumed to be 10 $^{-4}$. For Ust'-Ishim, two different mutation rates (0.5 × 10^{-9} -green line, and 1.0 × 10^{-9} - red line) were tested.*



Estimating the molecular date for the Ust'-Ishim bone

We initially added 45,000 years (the age estimate of Ust'-Ishim) to the PSMC curve and assumed two different mutation rates (0.5 × 10^{-9} [2] and 1.0 × 10^{-9} [3]). It is evident that the curve obtained by assuming a mutation rate of 0.5 × 10^{-9} / bp / year matches the empirical data more closely than the curve obtained by assuming a higher rate (Figure S14.1).

Motivated by these observations, we sought to estimate the mutation rate by identifying the rate that caused the PSMC curve for Ust'-Ishim to line up most closely with the mutation rate for non-African populations, repeating this analysis for each non-African population in turn.

In what follows, we measure "molecular time" in terms of the pairwise per-site sequence divergence d. Concretely, a "molecular time" of 10^{-4} refers to the amount of time that needs to elapse to accumulate 1 mutation per 10^4 sites between a pair of sequences. Given the per-generation mutation rate μ and generation time g, we can translate molecular time d to real time t in units of years using the formula $t = d/2v$, where $v = \mu/g$ is the per-year mutation rate. Conversely, when we know that molecular time d and real time t correspond to the same event, we can estimate the per-year mutation rate $v = d/2t$.

Qualitatively, the population size history of the Ust'-Ishim sample fits with modern non-Africans after adjusting for the age of the bone. Quantitatively, however, it is not obvious how to directly quantify the similarity between curves, given that PSMC parameterizes population sizes with discrete intervals. We therefore developed an alternative approach to quantify the molecular bone age.

In order to line up the curves in a principled way, we developed a procedure that finds the molecular date for each sample that best fits the data. For a present-day non-African sample, let $\Theta_m(d)$ be the effective population size at time d, inferred by PSMC. If we make the approximation that the ancient sample has the same size history before molecular date Δ, we expect its history to be described by $\Theta_a(d|\Delta)=\Theta_m(d+\Delta)$. PSMC is a model for diploid sequences, and thus can be used to compute the probability, $P[s|\Theta(\cdot)]$, of a diploid sequence s under the size history $\Theta(d)$. We can parameterize the full PSMC model with $\Theta_a(d|\Delta)$ and use this to compute the probability of the ancient sequence under the model. Explicitly, given an ancient sequence s and a modern history $\Theta_m(\cdot)$, we can numerically compute the likelihood function of Δ with $L(\Delta)=P[s|\Theta_a(\cdot|\Delta)]=P[s|\Theta_m(\cdot+\Delta)]$. This will give us a maximum-likelihood (ML) estimate of the molecular date Δ.

Figure S14.2 shows $L(\Delta)$ inferred from diverse present-day humans. All non-Africans have a single peak around 4×10^{-5} sequence divergence. The fluctuation is likely to be caused by the PSMC discretization, so we smooth the data using a kernel smoother.

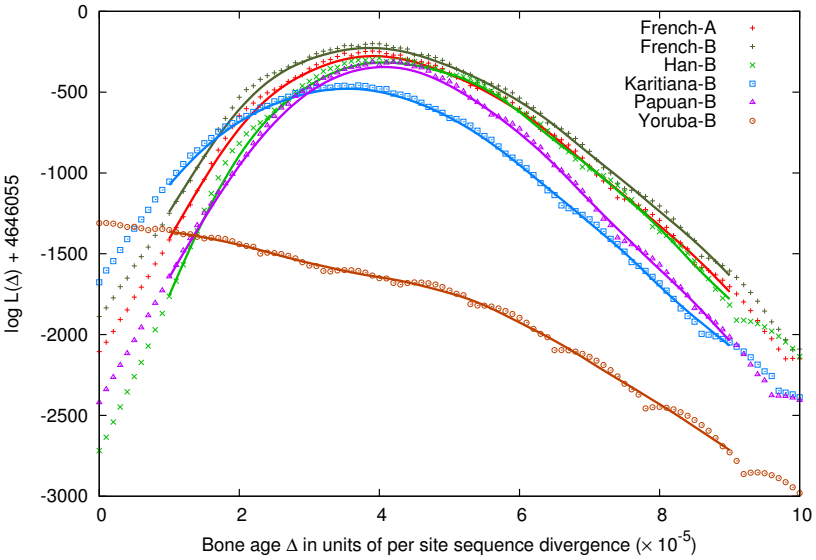


Figure S14.2 Log likelihood function of the bone age. The *probability of the ancestral sequence under the model trained from the sequence itself is -4646055. Solid lines are kernel smoothed with Epanechnikov's kernel and a radius* 10^{-5}.

Table S14.2 shows the molecular date estimate Δ for the Ust'-Ishim individual, calculated based on comparison to each non-African sample in turn. We observe that there are no systematic differences between comparisons to A- and B-panel present-day humans. Thus, the method appears to be robust to the different errors affecting the A-panel (mean of 27.4-fold coverage) and B-panel (mean of 38.3-fold coverage). This is encouraging for the method, and contrasts with direct estimates of branch shortening (which count missing mutations per base pair) in ancient genomes. Prüfer et al. [1] and the analysis in SI15 show that branch shortening computations are systematically different in their results depending on whether the comparison is to an A-panel or a B-panel human.

Table S14.2: *Ust'-Ishim molecular date from comparison to A- and B-panel individuals*

Sample	Molecular date Δ ($\times 10^{-5}$)	Mutation rate ($\times 10^{-9}$)
Dai_A	3.9	0.43
Dai_B	4.0	0.44
Han_A	3.9	0.43
Han_B	4.1	0.45
French_A	3.9	0.43
French_B	3.9	0.43
Sardinian_A	3.8	0.42
Sardinian_B	3.8	0.42
Karitiana_A	3.7	0.41
Karitiana_B	3.5	0.39
Mixe_B	3.5	0.39
Papuan_A	4.3	0.47
Papuan_B	4.1	0.45
Australian_B1	4.3	0.47
Australian_B2	4.1	0.45

Robustness of the PSMC estimate of molecular date to errors in the sequence data

When the PSMC infers population history, there are typically two types of sequencing errors that contribute to errors in the inference of how population size has changed over history.

1. The first type of error is caused by sequencing or mapping errors. These tend to be uniformly distributed, and to have their greatest effect in breaking long homozygous stretches, leading to inflated estimates of recent population size. Concretely, in PSMC inferences, such errors thus have their largest effect for time <500 generations ago, causing the inferences of population size changes in this period to be unreliable [4].
2. The second type of error is primarily caused by segmental duplications. These result in differences between copies of the genome that are paralogous and are not real point mutations at homologous positions. False-positive variants that are identified due to these duplications tend to be clustered tightly together, and can thus be falsely interpreted as short segments of the genome with very ancient coalescence. Such segments have their greatest effect for times >50,000 generations ago, causing the inferences of population size changes in this period to be unreliable [4].

To directly evaluate the effect of these errors, we carried out computer simulations of error. First, we randomly added uniform or clustered errors at an average rate 1 per 200kb to the consensus genomes sequences and evaluated the effect on our molecular age estimate. When we simulated clustered errors, we added 10 false heterozygotes evenly spaced at 100bp (i.e. spanning 1000bp) at a rate 1 per 2 Mb. The overall error rate was held constant at 1 per 200 kb in both simulation series.

Table S14.3 shows the results of these simulations using French_A as the present-day comparison sample. Our approach is reasonably robust to realistic error rates, whether the mutations are uniformly distributed or clustered. In particular, we observe no more than a ~5% fluctuation in the molecular date estimates from our simulations. In contrast, in a direct branch-shortening calculation, a 1/200 kb higher error rate in one sample than another would lead to an approximately 25% = (1/200 kb)(4×10^{-5}/2) under- or over-estimate, a much bigger effect. This greater robustness of our methodology compared with direct branch shortening estimates is in line with the observed consistency of our estimates across the A- and B-panel individuals which are known to have quite different levels of error.

Table S14.3: Robustness of the molecular date to different errors.

Errors	Δ ($\times10^{-5}$)
No added errors	3.9
Uniform errors to French_A	4.1
Uniform errors to Ust'-Ishim	3.7
Clustered errors to French_A	3.8
Clustered errors to Ust'-Ishim	3.8

Note: These analyses are based on comparison of Ust'-Ishim to the French_A sample.

Intuitively, another way to see the robustness of our method for obtaining a molecular date is that the PSMC time-stratifies the error process. Whereas the PSMC inference of population size history for periods <500 generations or >50,000 generations ago is unreliable for the reasons discussed above, in the period between 500-50,000 generations ago, the PSMC is relatively robust to errors in sequence data [4]. This is a period that includes important demographic events like the out-of-Africa bottleneck, and hence if we line up PSMC curves based on the events during this period as we do here, we are using a part of the history that is minimally affected by such errors.

Translating into a mutation rate in years while accounting for two sources of uncertainty

Up to this point, we have obtained a molecular date for the Ust'-Ishim sample in units of divergent sites per base pair (molecular time Δ). We can now translate this to a mutation rate estimate in years given an estimate of the bone age T ($v = \Delta/2T$ - quoting the equation above).

We perform this computation taking into account the uncertainty in the calendar date of the actual Ust'-Ishim bone, based on the calibrated carbon date of the sample taking into account the uncertainties of the carbon dating and recalibration of the carbon date (Figure S14.3).

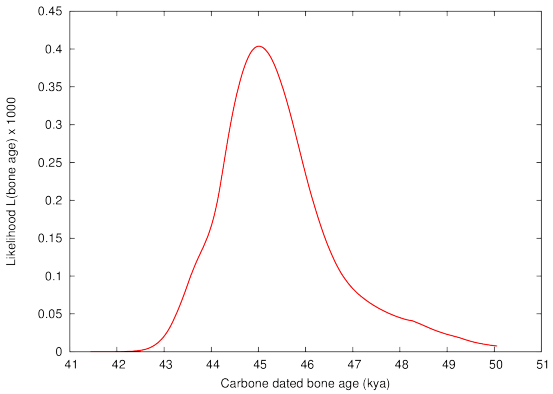


Figure S14.3 Likelihood distribution for bone age in thousands of years (calibrated C_{14} date).

To integrate both these sources of uncertainty into our estimate of a mutation rate, we used a sampling procedure to derive a posterior distribution for $\hat{v}$, the per year mutation rate. Recall that our model involves $L(T)$, the likelihood function of the carbon date, and $L_S(\Delta|r)$, the kernel-smoothed likelihood of the molecular date under radius r, taking the diploid sequence s as the modern sequence. If we assume a flat prior, we can sample s from a set of individual sequences or block bootstrapped sequences, sample r from a fixed distribution (a uniform distribution from 0 to 2×10^{-5}), and sample T and Δ from their posteriors, respectively, and derive $\hat{v}=\Delta/2T$. With enough sampling, we can approximate the distribution of $\hat{v}$. Table S14.4 shows the resulting estimate, which is $\hat{v} = 0.43\times10^{-9}$ / bp /year (95% confidence interval of $0.38\text{-}0.49\times10^{-9}$ / bp /year).

Type	Mean	CI 95%
$\hat{\nu}\times\mathbf{10^{-9}}$	0.43	[0.38, 0.49]
$\hat{\nu}_{non-CpG}\times\mathbf{10^{-9}}$	0.35	[0.30, 0.39]
$\hat{\nu}_{CpG}\times\mathbf{10^{-9}}$	0.12	[0.10, 0.14]
$\hat{\nu}_{s-nCpG}\times\mathbf{10^{-9}}$	0.22	[0.20, 0.25]
$\hat{\nu}_{v}\times\mathbf{10^{-9}}$	0.16	[0.14, 0.20]

Table S14.4 Mutation rate estimates. *We present results and a 95% confidence interval in brackets. We report the estimate for all sites $\hat{\nu}$ and break down by mutation class. $\hat{\nu}_{non-CpG}$ denotes the mutation rate of non-CpG mutations, $\hat{\nu}_{CpG}$ the rate of CpG transitions, $\hat{\nu}_{s-nCpG}$ the rate of non-CpG transitions and $\hat{\nu}_{v}$ denotes the mutation rate of transversions.*

Breaking down by mutation class

We repeated the computation stratifying by mutation class, breaking the data down into transitions associated with CpG dinucleotides, which are known to have an inflated mutation rate, and other non-CpG mutations (Table S14.4).

Ideally, the sum of rates of different mutation classes should be equal to the total rate: $\hat{\nu} = \hat{\nu}_{v} + \hat{\nu}_{s-nCpG} + \hat{\nu}_{CpG}$. Empirically, however, we observe that the sum of the mutation rate estimates for the individual mutation classes is significantly greater than the total estimate. We are unable to find a satisfactory explanation to this inconsistency. This suggests that our mutation rate estimate should only be interpreted as tentative, and subject to error modalities as is the case for other previously reported mutation rate estimates. Table S14.5 shows the proportion of substitutions inferred to occur in different classes, comparing our mutation rate estimate in this note to other estimates. The numbers are in reasonable agreement

Table S14.5. Proportion of substitutions inferred to occur in different classes.

Method	*Transversions*	*Non-CpG transitions*	*CpG transitions*
PSMC	32%	44%	24%
Heterozygous sites in French_B	32%	45%	23%
De novo mutations (Kong et al.)	32%	51%	17%

Information required for using our mutation rate estimate in other analyses

A complication in interpreting any analysis of the human mutation rate is that any estimate is specific for the subset of the nucleotides in the genome that are analyzed. Since each study uses a different genome filtering strategy, and different filters restrict to segments of the genome with different underlying true mutation rates, it is therefore not obvious that the mutation estimates that emerge from different studies can be applied to other datasets.

As part of this paper, we therefore present two sets of data that aim to make it possible to use the mutation rate estimates obtained here for other studies.

First, in Online Table S14A we present a file in .bed format that gives the exact positions of the nucleotides in the genome that we filtered out obtain our mutation rate estimate. Other researchers who wish to make population genetic inferences that employ our mutation rate estimates can restrict their genomic analyses to this subset of nucleotides.

Second, in Table S14.6, we use our mutation rate ($\hat{\nu} = 0.43\times10^{-9}$ / bp /year) to estimate average genomic divergence times per year for the two chromosome of the French_B individual, the two chromosomes of the Yoruba_B individual, and human-chimpanzee divergence, over the subset of the nuclear genome analyzed for our PSMC analyses. These time estimates make the simplifying and possibly incorrect assumption that mutation rates per year have been constant over time. However, even if mutation rates have changed over time and the true genomic divergence times per year are wrong, the time estimates are not expected to be compromised for the purpose of calibrating inferences about history over the last tens of thousands of years since the date of the Ust'-Ishim bone in the history of non-Africans. Researchers who are using different subsets of the genome in their analyses may wish to use these times rather than our mutation rate to calibrate their results.

Table S14.6. Implied mean genomic divergence times in years.

Mean coalescent time	*Sequence divergence*	*Time (Mya)*
Two chromosomes of French_B	0.77×10^{-3}	0.90
Two chromosomes of Yoruba_B	1.02×10^{-3}	1.19
Human-chimpanzee	1.27×10^{-2}	14.77

Note: This table is based on the mutation rate of $\hat{\nu} = 0.43\times10^{-9}$ / bp /year for present-day humans.

Summary

We have derived a new method to estimate the mutation rate from ancient DNA, assuming the population size history of the ancient sample is identical to that of present humans prior to the death of the archaic individual. With this method we estimate that the human mutation rate is 0.43×10^{-9} per site per year, with a 95% confidence interval (0.38×10^{-9} - 0.49×10^{-9}).

References

1 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* **13**, 745-753, doi:10.1038/nrg3295 (2012).

3 Takahata, N. & Satta, Y. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4811-4815 (1997).

4 Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496, doi:10.1038/nature10231 (2011).

Supplementary Information 15

Mutation rate estimates from branch shortening

Qiaomei Fu*, Svante Pääbo, Janet Kelso, David Reich
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Summary

We estimate the human mutation rate per year (averaged over the last tens of thousands of years) based on missing evolution or "branch shortening". Specifically, we measure the reduced number of mutations on the lineage of the carbon-dated Ust'-Ishim sample and compare this to the number of mutations seen in a present-day human since they separated from a common ancestor. When we use different present-day human genomes in such a comparison, we observe that quality differences between the genomes lead to differences in the estimated mutation rate. It is notable, however, that our estimates for the present-day human individuals with the highest quality give similar estimates to those that we obtain from the PSMC analysis (SI14) – an approach which is not expected to be affected by quality differences. This suggests that reliable branch shortening estimates can be obtained with genome sequences of high quality.

Branch Shortening Estimates

Ancient genomes are expected to show a reduced number of derived mutations ("branch shortening") compared to genomes from living relatives since the ancient individual stopped accumulating mutations at the time of death. This phenomenon has previously been observed in the Neandertal [1] and Denisovan [2] high coverage genomes by comparing the number of lineage-specific mutations on leading to the extinct archaic individual to the number of lineage-specific mutations on the lineages to present-day humans.

Following the methodology described in Meyer et al. 2011 [2] and Prüfer et al. 2014 [1], we measured the extent to which the branch leading to the Ust'-Ishim individual is shortened relative to present-day humans (25 present-day humans from the "A and B-panel"), Denisova [2] and Neandertal [1]. We used the inferred ancestor of human and chimpanzee (Ensembl Compara v64 [3,4]) to determine the ancestral state. We excluded CpG sites, since deaminated

cytosines cannot be removed with UDG-treatment at methylated CpG sites, and thus these sites have a residual high rate of ancient DNA error.

We filtered the autosomal data in each individual as described in SI7, and sampled a random allele from the diploid genotype calls at each site in the autosomes. At sites meeting these criteria, we tabulated three types of substitutions:

individual#1-specific (i1) ≡ (individual#2 ≠ individual#1) and (individual#2 = anc.)
individual#2-specific (i2) ≡ (individual#2 ≠ individual#1) and (individual#1 = anc.)
common (c) ≡ (individual#2 = individual#1) and (individual#2 ≠ anc.)

We estimated branch shortening for individual#2 as the reduction of individual#2-specific counts compared to individual#1-specific counts (i1-i2)/i1. To express the shortening as a fraction of the divergence to chimpanzee, we normalized the estimate by the length of the extant lineage to the common ancestor of chimpanzee and human: ((i1-i2)/i1) × (i1/(i1+c)) = (i1-i2)/(i1+c).

The Ust'-Ishim branch is evidently shortened relative to all present-day humans, but the degree of shortening is less than for Denisova or Altai (Table S15.1).

Table S15.1: *Branch shortening of ancient samples relative to comparison individuals, expressed as a fraction of human-chimpanzee divergence.*

Individual#1		Individual#2		
Sample	**Alias**	**Ust'-Ishim**	**Denisova**	**Altai**
Ust'-Ishim	Ust'-Ishim		0.44%	0.65%
Denisova	Denisovan			0.22%
HGDP01029	San_A	0.65%	1.09%	1.31%
HGDP00456	Mbuti_A	0.59%	1.02%	1.24%
HGDP00927	Yoruba_A	0.63%	1.06%	1.28%
HGDP01284	Mandenka_A	0.62%	1.05%	1.27%
DNK02	Dinka_A	0.58%	1.01%	1.23%
HGDP00521	French_A	0.58%	1.01%	1.23%
HGDP00665	Sardinian_A	0.61%	1.05%	1.26%
HGDP00778	Han_A	0.60%	1.03%	1.25%
HGDP01307	Dai_A	0.62%	1.05%	1.27%
HGDP00998	Karitiana_A	0.61%	1.05%	1.26%
HGDP00542	Papuan_A	0.61%	1.04%	1.26%

HGDP01036	San_B	0.33%	0.77%	0.98%
HGDP00982	Mbuti_B	0.31%	0.74%	0.96%
HGDP01286	Mandenka_B	0.32%	0.76%	0.98%
HGDP00936	Yoruba_B	0.32%	0.75%	0.97%
DNK07	Dinka_B	0.36%	0.79%	1.01%
HGDP00533	French_B	0.37%	0.81%	1.03%
HGDP01076	Sardinian_B	0.37%	0.81%	1.02%
HGDP00775	Han_B	0.47%	0.90%	1.12%
HGDP01308	Dai_B	0.41%	0.85%	1.07%
HGDP01015	Karitiana_B	0.35%	0.79%	1.01%
MIXE0007	Mixe_B	0.36%	0.79%	1.01%
HGDP00546	Papuan_B	0.37%	0.81%	1.03%
WON,M	Australian_B1	0.33%	0.76%	0.98%
BUR,E	Australian_B2	0.40%	0.84%	1.05%

Mutation Rate Estimates

A calibrated carbon date of 46,880–43,210 cal BP (95.4% probability) has been determined for the Ust'-Ishim bone. With this as a calibration point, we can estimate the mutation rate as the observed reduction in substitutions over the compared bases divided by the carbon-dated age of Ust'-Ishim (Table S15.2). We compute standard errors (SE) using a Weighted Block Jackknife [5] by removing one chromosome at a time.

Table S15.2: *Autosomal mutation estimate per site per year $\times 10^{-9}$ based on the branch shortening in Ust'-Ishim relative to the A- and B-panel present-day humans (excluding CpG sites)*

			All autosomes		**European hg19 segments**		**African hg19 segments**	
Sample	**Alias**	**Panel**	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**
HGDP01029	San_A	A	0.76	0.03	0.74	0.05	0.79	0.07
HGDP00456	Mbuti_A	A	0.68	0.03	0.70	0.04	0.74	0.07
HGDP01284	Mandenka_A	A	0.71	0.02	0.74	0.04	0.78	0.06
HGDP00927	Yoruba_A	A	0.73	0.03	0.71	0.04	0.67	0.09
DNK02	Dinka_A	A	0.66	0.03	0.70	0.05	0.61	0.11
HGDP00521	French_A	A	0.69	0.02	0.71	0.05	0.65	0.07
HGDP00665	Sardinian_A	A	0.72	0.02	0.67	0.04	0.74	0.07
HGDP00778	Han_A	A	0.70	0.03	0.64	0.04	0.67	0.09
HGDP01307	Dai_A	A	0.70	0.03	0.70	0.05	0.62	0.08
HGDP00998	Karitiana_A	A	0.71	0.03	0.72	0.06	0.64	0.07
HGDP00542	Papuan_A	A	0.71	0.02	0.73	0.04	0.57	0.08
HGDP01036	San_B	B	0.36	0.02	0.35	0.05	0.38	0.08

HGDP00982	Mbuti_B	B	0.36	0.03	0.39	0.05	0.39	0.09
HGDP01286	Mandenka_B	B	0.40	0.03	0.42	0.05	0.42	0.08
HGDP00936	Yoruba_B	B	0.36	0.02	0.38	0.05	0.33	0.09
DNK07	Dinka_B	B	0.40	0.03	0.39	0.05	0.41	0.06
HGDP00533	French_B	B	0.44	0.03	0.44	0.05	0.43	0.07
HGDP01076	Sardinian_B	B	0.44	0.02	0.40	0.03	0.40	0.06
HGDP00775	Han_B	B	0.52	0.02	0.54	0.05	0.50	0.05
HGDP01308	Dai_ B	B	0.49	0.03	0.44	0.05	0.51	0.07
HGDP01015	Karitiana_B	B	0.43	0.02	0.42	0.04	0.35	0.07
MIXE0007	Mixe_B	B	0.43	0.02	0.44	0.04	0.36	0.06
HGDP00546	Papuan_B	B	0.44	0.02	0.45	0.05	0.36	0.07
WON,M	Australian_B2	B	0.39	0.02	0.37	0.04	0.40	0.07
BUR,E	Australian_B1	B	0.45	0.01	0.42	0.04	0.43	0.06

The autosomal mutation rate in the panel-B present-day humans after removing CpG sites is 0.66-0.76 $\times 10^{-9}$ / bp /year for panel-A and 0.36-0.52 $\times 10^{-9}$ / bp / year for panel-B (reporting the range over point estimates, which are themselves relatively precise with typical standard errors of ± 0.02-0.03). Based on the derived alleles that are only found in the lineage leading to the comparison individual (individual#1-specific i1), we estimate that CpG dinucleotides constitute on average 17% of all mutations analyzed. We therefore divide our mutation rate estimate by (1-0.17) to obtain corrected estimates of 0.80-0.91 $\times 10^{-9}$ / bp /year for panel-A and 0.44-0.63 $\times 10^{-9}$ / bp / year for the B-panel.

The panel-B mutation rate estimates overlap those from the analysis of the PSMC SI14 (0.40-0.49 $\times 10^{-9}$), while the mutation rate estimates from the lower-coverage panel-A genomes do not overlap those from the PSMC. As discussed previously [1], branch shortening is highly sensitive to the quality of the compared genome sequences and must be interpreted with caution in samples with differing qualities.

To determine whether alignment bias may be influencing the mutation rate computation, we repeated the computation restricting to segments of the hg19 reference human genome sequence that are known to derive from genomic segments of European and African ancestry, respectively (Table S15.2) [6]. We observe that the mutation rate estimates are not significantly different between genomic segments of African and European ancestry and conclude that alignment bias does not have a major effect on the branch shortening results.

Finally, we break down the estimates by substitution class. Table S15.3 estimates the transversion mutation rate and non-CpG transition mutation rate for B-panel individuals, pooling across samples to increase the precision of the estimate. Encouragingly, the resulting estimates are consistent with the PSMC (transversion: 0.14-0.21 × 10^{-9} / site / year; non-CpG transition: ~ 0.20-0.26 × 10^{-9} / site / year).

Table S15.3. *Summary of autosomal mutation rate estimates for panel-B individuals, expressed per site per year ×10^{-9} and broken down by substitution class*

Pool	All sites		Autosome transversion		Autosome non-CpG transition		Chromosme X non-CpG	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**
All B-panel	0.42	0.01	0.17	0.00	0.25	0.01	0.31	0.54
Non-African	0.45	0.01	0.19	0.01	0.26	0.01	0.38	0.90
African	0.37	0.01	0.14	0.01	0.23	0.01	0.22	0.59

Summary

From the Ust'-Ishim individual, which is directly carbon dated, we use branch shortening to estimate a nuclear mutation rate of 0.44-0.63 × 10^{-9} / site / year, which is lower than the value that has been widely used in the past (1 × 10^{-9}) [7]. This range is consistent with more recent estimates based on *de novo* substitution studies of 0.4-0.6 × 10^{-9} / site / year [8,9] and also with the estimate obtained using the PSMC method of SI14. However, we caution that the comparisons using lower quality A-panel individuals give significantly different results, indicating that this measure is sensitive to quality differences between the compared genomes.

References

1 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

3 Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* **18**, 1814-1828, doi:gr.076554.108 [pii] 10.1101/gr.076554.108 (2008).

4 Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* **18**, 1829-1843, doi:gr.076521.108 [pii] 10.1101/gr.076521.108 (2008).

5 Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* **17**, 1217-1241, doi:DOI 10.1214/aos/1176347265 (1989).

6 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

7 Takahata, N. & Satta, Y. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4811-4815 (1997).
8 Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471-475, doi:10.1038/nature11396 (2012).
9 Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* **13**, 745-753, doi:10.1038/nrg3295 (2012).

Supplementary Information 16

Neandertal ancestry in Ust'-Ishim

Qiaomei Fu*, Janet Kelso
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Summary

Here we analyse the high coverage genomes of the Ust'-Ishim individual together with the genomes of the Altai Neandertal, Denisovan and multiple present-day humans (the 25 individuals from Panels A and B [1]). The goal is to investigate whether the Ust'-Ishim individual shares the archaic admixture seen in present-day humans [2-4] and to estimate the proportion of the archaic admixture in the Ust'-Ishim individual.

We find that the Ust'-Ishim individual has an indistinguishable proportion of Neandertal admixture to that in non-Africans.

D-statistics

To determine whether the Ust'-Ishim individual has archaic admixture we used the *D*-statistic[2,4,5]. If *A, B, C, D* are 4 populations, then:

$$D(A,B;C,D) = \frac{E[n_{BABA}]-E[n_{ABBA}]}{E[n_{BABA}]+E[n_{ABBA}]} \quad (11.1)$$

where E[n_{BABA}] is the count of alleles agreeing in populations A, C and also in B, D (but differing in A, B) and E[n_{ABBA}] is the expected count of alleles agreeing in populations A, D and also in B, C (but different in A, B). A positive *D* is therefore an indication of more allele-sharing between either of the population pairs *(A, C)* or *(B, D)*. In contrast, a negative *D* indicates more shared alleles between either of the population pairs *(A, D)* or *(B, C)*.

Standard errors on all D-statistics reported here are computed using a Weighted Block Jackknife [6] with a block size of 5 million base pairs (5 Mb).

We tested whether Neandertals or Denisovans share alleles at a different rate with Ust'-Ishim than with diverse present-day non-Africans using the statistic *D(Ust'-Ishim, X; Archaic, Outgroup)*. Here, *X* is the tested individual, and the outgroup is the inferred common ancestral sequence of human and chimpanzee.

Table S16.1. D-statistics of the form *D(Ust'-Ishim, X; Archaic, Outgroup).*

X	Altai Neandertal		Denisovan	
Individual	**D stat**	**Z-Score**	**D stat**	**Z-Score**
Han_B	0.008	0.771	-0.005	0.955
Dai_B	0.018	1.827	0.002	0.286
$Karitiana_B$	0.015	1.474	0.004	0.579
$French_B$	0.031	3.384	0.010	1.788
$Sardinian_B$	0.019	1.892	0.009	1.479
$Papuan_B$	-0.007	-0.653	-0.067	-8.854

The Ust'-Ishim individual shares a similar proportion of alleles with the Neandertal and the Denisovan to that seen in present-day mainland Eurasians. There is a tendency, though not significant, for the Neandertal and Denisovan to share more alleles with Ust'-Ishim than with Western Eurasians, with the |Z|-scores for Neandertal comparisons to French and Sardinian being $3.4 > |Z| > 1.9$. The Papuan shares more alleles with the Denisovan than the individuals from other populations tested here do, which is consistent with the previously documented Denisovan gene flow into the ancestors of Papuans [7]

We also computed the statistic D(*X, Yoruba; Altai Neandertal, Outgroup*) (Table S16.2). The strongest signal is observed for Ust'-Ishim. However, the standard errors from the Block Jackknife overlap with those from the present-day non-Africans, so there is no significant evidence of more Neandertal ancestry in Ust'-Ishim than in eastern non-Africans, consistent with the findings in Table S16.1.

Table S16.2. D-statistics of the form *D(X, Yoruba; Altai, Outgroup).*

	Altai Neandertal		
Individual	D stat	Standard Error	Z-Score
Ust'-Ishim	0.068	0.008	8.53
Han_B	0.063	0.006	10.35
Dai_B	0.054	0.005	10.12
$French_B$	0.044	0.005	8.20
$Sardinian_B$	0.054	0.006	9.46

Estimates of Neandertal mixture proportion using f_4-ratio statistics

To determine the proportion of the genome of the Ust'-Ishim individual that was contributed by Neandertals, we use a ratio of f_4-statistics, which provides an unbiased estimate of the admixture proportion [7]. Briefly, we calculate a statistic that is proportional to the correlation in the allele frequency difference between Eurasians and the Altai Neandertal, and divide it by the same statistic calculated between another Neandertal (the Mezmaiskaya Neandertal) and the Altai Neandertal. Since genome coverage for the Mezmaiskaya individual is low (0.5×) we do not use genotypes, and instead randomly select reads with a map quality higher than 37. The ratio we use for the estimation is the same one that is used and justified in [1]. Specifically, we estimate Neandertal proportion as:

$$\hat{\alpha} = \frac{f_4(Den,\ Altai;\ Africa,\ X)}{f_4(Den,\ Altai;\ Africa,\ Mez)} \qquad (11.2)$$

Here, we represent sub-Saharan Africans by a pool of Mbuti, Yoruba and Dinka.

The proportion of Neandertal admixture in the Ust'-Ishim individual is estimated to be 2.26%, which is not significantly higher than any other present-day individual (Table S16.3).

Table S16.3. Estimates of Neandertal mixture proportion

	Proportion	**Std error**	**Z-Score**
Ust'-Ishim	2.26%	0.33%	6.86
French	1.62%	0.20%	7.92
Sardinian	1.82%	0.20%	8.97
Han	2.11%	0.23%	9.06
Dai	1.66%	0.22%	7.58
Mix	1.87%	0.24%	7.64
Karitiana	1.97%	0.24%	8.05

References

1 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

3 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

4 Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, doi:10.1126/science.1224344 (2012).
5 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, doi:10.1534/genetics.112.145037 (2012).
6 Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* **17**, 1217-1241, doi:DOI 10.1214/aos/1176347265 (1989).
7 Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* **89**, 516-528, doi:10.1016/j.ajhg.2011.09.005 (2011).

Supplementary Information 17

Denisovan ancestry in Ust'-Ishim

Qiaomei Fu*, Michael Lachmann, Svante Pääbo, Janet Kelso
* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de)

Summary

To address whether the Ust'-Ishim individual carries a similar amount of Denisovan admixture to that which has been reported for East Asian or European populations [1,2], we calculated an admixture statistic that takes advantage of local ancestry information and reflects the relative contribution of Neandertal and Denisovan to each present-day populations and Ust'-Ishim.

This analysis shows that while the Neandertal admixture in the Ust'-Ishim individual is comparable to that in all present-day non-African populations, the amount of Denisovan admixture is less than that in Asian populations and is comparable to that in Europeans.

We conclude that Ust'-Ishim probably does not share the small amount of Denisovan ancestry that has previously been reported for East Asians [3,4].

Neandertal admixture statistic

To identify sites in the genomes of modern humans that have Neandertal ancestry we use an "Enhancement" strategy that enriches for alleles that have a greatly increased likelihood of having arisen in Neandertals [5]. Specifically, we restrict to sites where:

- all 195 present-day Africans (YRI and LWK from the 1000 genomes plus the sub-Saharan Africans from the A are B panels) are ancestral (i.e.: match the human-chimpanzee ancestor) and the Neandertal is homozygous for the derived allele.

- all present day humans from panel A + B, the Ust'-Ishim, Denisovan, and Altai Neandertal pass the filters described in SI7.

We then identify sites in all present-day A and B panel non-Africans, the Ust'-Ishim individual, and the Denisovan that are either:

(i) homozygous for the derived state (scored as 2)

(ii) heterozygous for the derived state (scored as 1)

(iii) homozygous for the ancestral state (scored as 0)

Figure S17.1 illustrates a section of one autosome in all present-day non-Africans and Ust'-Ishim that we scored using these criteria. Blocks that appear yellow or blue correspond to positions that are either heterozygous or homozygous for the derived allele, respectively. Blocks of yellow and blue therefore indicate sites in the genomes of these individuals that share derived sites with the Altai Neandertal.

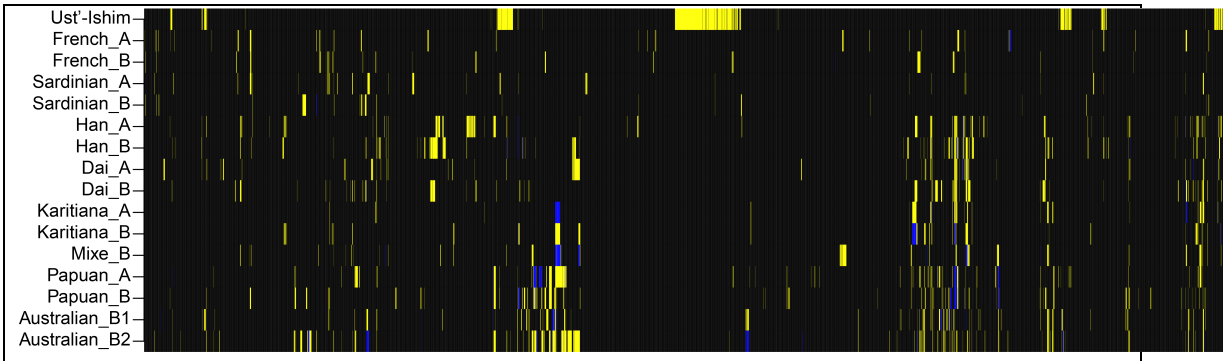


Figure S17.1. *A section of one autosome in present-day non-Africans (from the A and B panels) and Ust'-Ishim that is coloured according to sharing of derived alleles with Neandertals. Blocks that appear yellow or blue correspond to positions that are either heterozygous or homozygous for the derived (Neandertal) allele, respectively. Blocks of yellow and blue therefore indicate sites in the genomes of these individuals where they share derived sites with the Altai Neandertal.*

To convert this to a ratio, we tabulate the following:

- We count the total number of derived sites (either homozygous or heterozygous) in the Ust'-Ishim individual and in the Panel A and B present-day humans (sites that arose on the branch labeled **b** in Figure S17.2a).

- We condition on sites where the Denisovan is homozygous for the ancestral state, and count the total number of derived sites in Ust'-Ishim and in present-day non-Africans from Panel A and B (sites that arose on the branch labeled **a** in Figure S17.2a).

The sites being counted are illustrated in a schematic tree (Figure S17.2). If the present-day non-African populations and Ust'-Ishim experienced the same admixture events with Neandertals and Denisovans, then the number of derived sites in present-day humans and Ust'-Ishim from the **a** lineage (specific to Altai)**,** and the number of derived sites in present-day humans and Ust'-Ishim from the **b** lineage (shared between Denisova and Altai), are expected to be consistent.

We observe a similar **a/b** count ratio in Ust'-Ishim and the present-day humans outside of Oceania, indicating that these populations share admixture from similar archaic populations (Table S17.1). Based on previous studies in non-Africans outside of Oceania [2], it is likely that the primary archaic admixture in these populations that is driving these patterns is from Neandertals.

The pattern is different in Papuans and Australians (Oceanians) where we observe a lower ratio of **a/b**. This is driven by an increase in **b** in these populations (shared derived sites between the Neandertal and Denisovan contribute to **b** but not to **a**). This indicates that these populations experienced additional admixture not seen in other non-Africans, from an archaic population that did not have the same **a/b** ratio as the main archaic population that introgressed into other non-SAfricans. This is consistent with previous reports of Denisovan admixture into the ancestors of Oceanians [3].

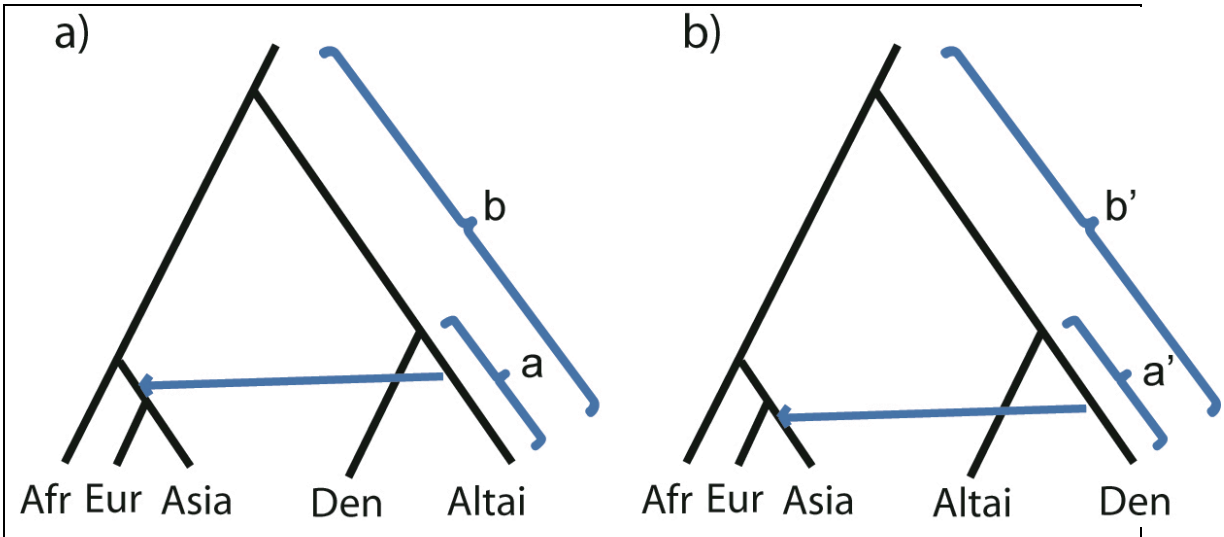


Figure S17.2. *Admixture model used for interpreting results*

Denisovan admixture statistic

Having demonstrated that this approach can distinguish between Neandertal and Denisovan admixture we switched the ascertainment scheme to identify derived sites in present-day humans and Ust'-Ishim that are shared with the Denisovan. Briefly:

- all 195 present-day Africans (YRI and LWK from the 1000 genomes plus the sub-Saharan Africans from the A are B panels) match the human-chimpanzee ancestor and the Denisovan is homozygous for the derived allele.

- all present day humans from panels A and B, Ust'-Ishim, the Denisovan, and the Altai Neandertal pass the filters described in SI7.

Figure S17.3 illustrates the section of one autosome in all present-day non-Africans and Ust'-Ishim scored using these criteria, restricting to sites where Neandertal is homozygous for the ancestral state. This analysis is similar to the conditioning of Pruefer et al. 2014 [2] (Supplementary Note 13), where a similar procedure was able to successfully enrich for genomics segments introgressed from Denisovans.

As in Figure S17.1, blocks that appear yellow or blue correspond to positions that are either heterozygous or homozygous for the derived allele, respectively. In other words, blocks of yellow and blue indicate places in the genomes of these individuals where they share derived sites with the Denisovan. From an inspection of the plot it is evident that Denisovan contribution to Papuans and Australians is highest.

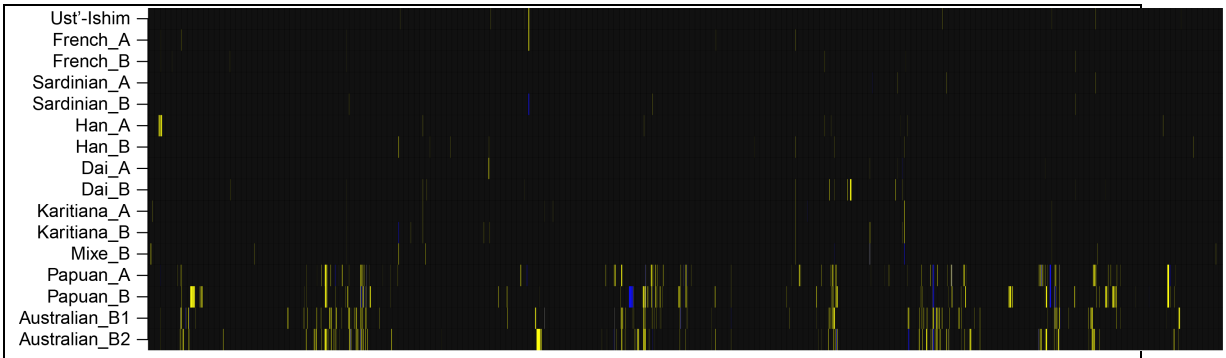


Figure S17.3. *A section of one autosome in all present-day non-Africans and Ust'-Ishim that is coloured according to Denisovan ancestry in these individuals. Blocks that appear yellow or blue correspond to positions that are either heterozygous or homozygous for the derived (Denisovan) allele, respectively. Blocks of yellow and blue therefore indicate sites in the genomes of these individuals where they share the derived sites with the Denisovan.*

For our ratio:

- We count the total number of derived positions (either homozygous or heterozygous) in the Ust'-Ishim individual and in the individuals from Panels A and B (sites that arose on the branch labeled **b'** in Figure S17.2b).

- We then use those sites where the Neandertal is homozygous for the ancestral state, and count the total number of derived positions observed in the Ust'-Ishim individual and in present-day non-Africans from Panels A and B (sites that arose on the branch labeled **a'** in Figure S17.2b).

We then compute the ratio **a'/b'**. If the present-day non-African populations and Ust'-Ishim experienced the same admixture with Neandertals and Denisovans, the ratio should be similar.

Empirically, we observe that the ratio is higher in East Asian and Native populations than in Europeans (Table S17.1). A Wilcoxon signed-rank test comparing the ratios for European populations and Ust'-Ishim to the ratios for the East Asian and Native American populations shows that the ratios for Europeans and Ust'-Ishim are significantly smaller than the ratios for East Asian + Native American populations ($P = 0.0028$). This likely reflects the small amount of admixture from Denisovans that has been reported for eastern non-Africans like East Asians and Native Americans [1,2]. The ratio for Ust'-Ishim is within that for Europeans, indicating that the Ust'-Ishim individual may not share the additional admixture with Denisovans that occurred in the population ancestral to East Asians and Native Americans.

Table S17.1. Ratio of sites in classes *a* and *b* across diverse present-day humans.

Individual	Neandertal lineage (a/b)	Denisovan lineage (a'/b')	Denisovan specific (a') / Neandertal specific (a)
Ust_Ishim	73%	11%	4%
French	76%	18%	6%
French2	71%	10%	4%
Sardinian	77%	14%	4%
Sardinian2	77%	17%	6%
Han	76%	22%	9%
Han2	74%	26%	12%
Dai	74%	24%	11%
Dai2	75%	25%	11%
Karitiana	73%	20%	9%
Karitiana2	74%	23%	10%
Mixe	75%	24%	10%
Papuan	58%	58%	99%
Papuan2	59%	60%	105%
Australian	59%	59%	104%
Australian1	62%	62%	101%

References

1 Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18301-18306, doi:10.1073/pnas.1108181108 (2011).

2 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

3 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

4 Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet* **89**, 516-528, doi:10.1016/j.ajhg.2011.09.005 (2011).

5 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

Supplementary Information 18

Dating Neandertal admixture in Ust'-Ishim

Qiaomei Fu+* Priya Moorjani+*, Michael Lachmann, Flora Jay, Montgomery Slatkin, Janet Kelso§, David Reich§

+ These authors contributed equally. § These authors co-mentored the work.

* To whom correspondence should be addressed (qiaomei_fu@eva.mpg.de, pm2730@columbia.edu)

Overview

A previous study used the mean tract lengths of Neandertal ancestry segments in present-day humans to infer that the average date of Neandertal gene flow into non-Africans was between 37,000 and 86,000 years before the present (BP) [1].

Since Ust'-Ishim, a 45,000-year-old anatomically modern human, shows evidence of a similar proportion of Neandertal ancestry as present-day non-Africans, it is clear that Neandertal gene flow into modern humans had already occurred by 45,000 years BP.

To estimate when the Neandertal introgression occurred into Ust'-Ishim, we use the insight that admixture between distinct populations induces correlation in ancestry across the genome of the admixed individual and that the extent of this correlation is informative about the time since mixture [2]. This implies that by studying the correlation in genotypes across pairs of SNPs that are most informative for the Neandertal admixture in Ust'-Ishim, we could estimate the time that has elapsed since admixture [3-5]. While this method is also applicable to present-day humans, the time that elapsed between the Neandertal admixture and the death of the Ust'-Ishim individual is much shorter than the time that has elapsed since the admixture into the ancestor of all present-day humans. Thus, the segments of Neandertal ancestry present in the Ust'-Ishim genome are expected to be longer.

To implement this idea, we use a SNP ascertainment scheme that is informative for Neandertal ancestry and compute the pairwise covariance for all pairs of SNPs that are at genetic distance d Morgans apart. We show via simulations that this statistic is expected to decay approximately exponentially and that the rate of decay of the fitted exponential is informative about the gene flow date (Table S18.2, Figure S18.2).

Analysis of real data

We ascertained 221,000 SNPs that are informative for Neandertal introgression based on the criteria that Africans (1000 genomes YRI and LWK [6]) are fixed for the ancestral allele and differ from the Altai Neandertal (which carries the derived allele) and at least one derived allele is observed at that position among the 1000 genomes individuals (excluding the LWK and YRI) to reduce the effect of error. The spatial distribution of stretches of the genome where a non-African sample carries the derived allele is qualitatively different for Ust'-Ishim than for present-day non-Africans, with the Ust'-Ishim containing fewer segments that match Neandertals but that these are larger in size compared to present-day non-Africans (Figure S18.1). This is consistent with Ust'-Ishim being closer in time to the date of the Neandertal gene flow, and motivates our analysis to utilize this signal to estimate the date of Neandertal introgression.

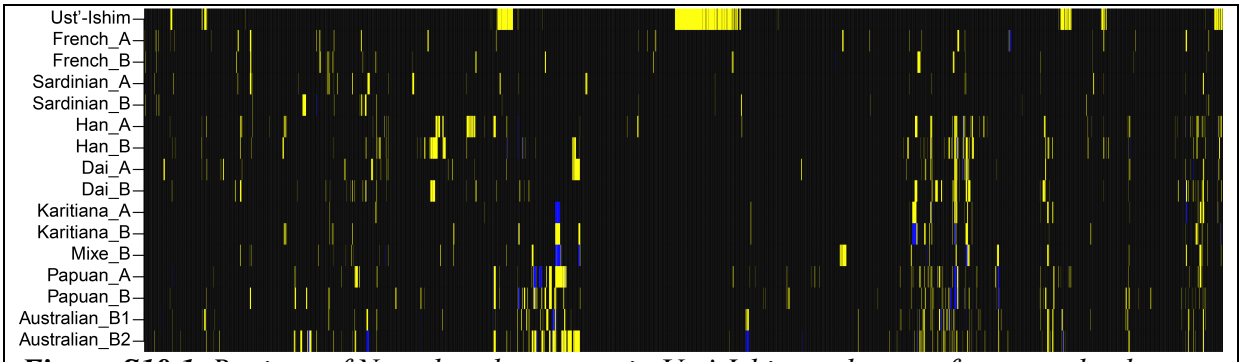


Figure S18.1. *Regions of Neandertal ancestry in Ust'-Ishim and a set of present-day humans based on an ascertainment scheme in which African genomes (1000 genomes YRI and LWK) carry the ancestral allele and Altai Neandertal carries the derived allele. Homozygous ancestral alleles are in black; heterozygous derived alleles in yellow, and homozygous derived alleles in blue. Ust'-Ishim has larger regions of likely Neandertal ancestry than present-day non-Africans, consistent with the signal of more recent Neandertal mixture that we study in this note.*

For all pairs of SNPs that match our ascertainment scheme, we use the Oxford combined genetic map and calculate the average covariance over all pairs of SNPs in 0.001 cM bins of genetic distance (from 0.02cM - 1cM). Assuming a single pulse of admixture, we fit an exponential function with an affine term ($y = Ae^{-nd} + c$), where n = number of generations since mixture and d = genetic distance (in Morgans) to infer the date of Neandertal introgression for present-day non-African populations in Panel B [7] and for the Ust'-Ishim individual. The covariance curves are shown in Figure S18.2. We compute standard errors using a Weighted Block Jackknife [8], removing one chromosome in each run and studying the

variability in the estimated dates of mixture. From these curves, we estimate that the time of Neandertal admixture in Ust'-Ishim occurred 331 (S.E.± 99) generations before the death of the Ust'-Ishim individual.

Assuming 29 years per generation [9], the date of admixture is 9,599 ± 2,871 years before the Ust'-Ishim individual lived. Since the Ust'-Ishim sample is 45,000 years old, this implies that the mixture occurred 51,728-57,740 years BP (considering mean ± SE). We caution, however, that this estimate does not take into account uncertainty in the generation interval, as well as uncertainty in the radiocarbon date.

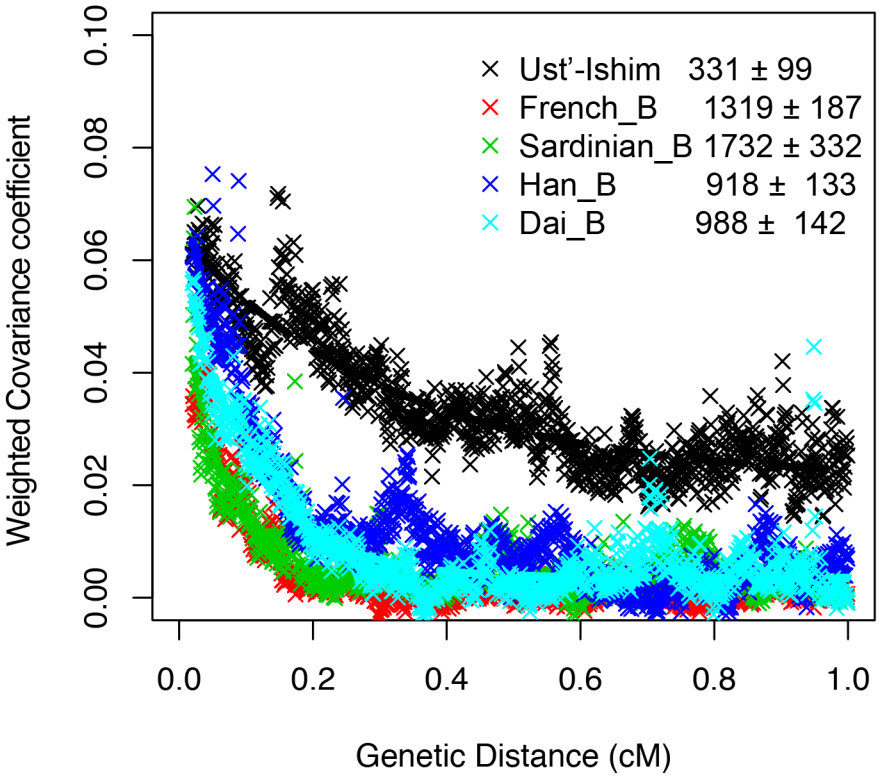


Figure S18.2. *Decay of pairwise covariance for SNPs that match the ascertainment scheme in which Africans carry the ancestral allele and Neandertal carries the derived allele. The data from the Ust'-Ishim individual shows a longer range covariance compared to that observed in present-day non-Africans.*

Robustness to uncertainties of the genetic map:

The analysis reported above assumed that the genetic map we are using is accurate. In fact, however, the genetic maps we use are limited in their resolution and also are known to vary across populations and over time [10]. While at broad scales (>1 Mb) recombination maps are

strongly correlated across human populations and even across species [10,11] at fine scales, differences in the maps may affect decorrelation rates.

The ideal genetic map for our application would be a pedigree-based map generated from a large number of Ust'-Ishim families. While such a map is unavailable, as an alternative, we repeated the computation using seven different genetic maps built in present-day humans based either on linkage disequilibrium patterns [6], observations of recombination events in parent-child transmissions [12] or breakpoints of admixture tracts [11]. These maps are built in different populations, are sensitive to different time periods of human history, and encompass more variability in population history than exists between the Ust'-Ishim individual and the combined Oxford genetic map [13] that we use for our main inferences. Thus, we can reasonably hypothesize that the variability in our inference depending on which map we use bounds the error due to not having access to an Ust'-Ishim genetic map.

Table S18.1 shows that the estimated date of Neandertal introgression into Ust'-Ishim is consistent (within two standard errors) regardless of the genetic map we use. This is likely because most maps are accurate at the scale of >100 kb, while the average length of the Neandertal introgressed segments in Ust'-Ishim is ~300 kb (Figure S18.1 and Figure S18.2). We conclude from this that map error is not likely to be a major problem for our inference.

Table 18.1 *Estimated dates of Neandertal admixture in Ust'-Ishim using different recombination maps.*

Recombination Map	Reference	Estimate	Std Error
Combined Oxford map	International HapMap Consortium [13]	331	99
African American	Hinch et al. 2011 [11]	361	132
deCODE	Kong et al. 2002 [12]	320	147
CEU	1000 genomes paper [6]	308	131
CHB	1000 genomes paper [6]	295	65
MXL	1000 genomes paper [6]	342	151
ASW	1000 genomes paper [6]	227	85
YRI	1000 genomes paper [6]	244	67

Simulation testing

To assess the utility of this statistic, we performed coalescent simulations using ms [14] under various demographic models. We generated data for three populations that we chose to have

demographic parameters roughly similar to what we expect for Neandertals, sub-Saharan Africans and non-Africans. For each simulation, we used an ascertainment scheme in which Africans are fixed for the ancestral allele and Neandertals have at least one derived allele. For all the SNPs that matched our ascertainment scheme, we performed analysis using the statistic described above to estimate the date of Neandertal mixture.

Simulation 1: We simulated data for 22 haploid genomes: 10 Non-Africans, 10 sub-Saharan Africans, and 2 Neandertals. We simulated 200 chromosomes that are each 10 Mb in length under a simple demographic model with gene flow from Neandertals into non-Africans that occurred at varying times of mixture (500-2,500 generations ago). We combined two haploid chromosomes to generate a diploid chromosome and performed the analysis for each non-African diploid individual separately. We set the following parameters:

- Mutation rate = $2x10^{-8}$ per bp/generation
- Recombination rate = $1x10^{-8}$ per bp/generation
- Effective population size (N_e) of modern humans = 10,000.
- Neandertals undergo a 120-generation bottleneck at time 3,120 generations in the past where the effective population size reduces to 200 individuals.
- Non-African split from Africans 2,500 generations before present.
- Sub-Saharan Africans split from Neandertal 10,000 generation before present.
- The proportion of Neandertal ancestry in non-Africans is 0.03.

Table S18.2 *Dates of Neandertal admixture for simulation 1.*

Ind	Expected	Estimate	Std Error
Non-African1	500	514	44
Non-African2	500	549	51
Non-African3	500	501	45
Non-African4	500	609	50
Non-African5	500	558	49
Non-African1	1000	1033	72
Non-African2	1000	976	75
Non-African3	1000	1112	85
Non-African4	1000	1111	71
Non-African5	1000	1124	82
Non-African1	1500	1552	98
Non-African2	1500	1342	101
Non-African3	1500	1469	131
Non-African4	1500	1487	123
Non-African5	1500	1430	133
Non-African1	2000	2084	160
Non-African2	2000	1644	122
Non-African3	2000	2007	202
Non-African4	2000	2158	186
Non-African5	2000	1987	224
Non-African1	2500	2433	277
Non-African2	2500	2137	226
Non-African3	2500	2617	197
Non-African4	2500	2212	183
Non-African5	2500	2400	215

The *ms* command line is as follows:

ms 22 200 -t 8000 -r 4000 10000000 -I 3 10 10 2 -em start-time 1 3 1200 -em end-time 1 3 0 -ej 0.0625 2 1 -en 0.075 3 0.02 -en 0.078 3 1 -ej 0.25 3 1

Here, pop1 = non-African, pop2 = African and pop3 = Neandertal. The start- and end-time correspond to the start and end of the simulated Neandertal gene flow into non-Africans. We assume a single pulse migration (single-generation migration) so end-time = start-time + 1.

The results are summarized in Table S18.2. For all dates of mixture tested here, we obtain accurate and unbiased estimates for the dates of Neandertal introgression into non-Africans.

Simulation 2: We simulated data for 48 haploid genomes from 5 populations: 2 introgressing Neandertals, 2 Altai Neandertals (representing an ancient genome), 20 sub-Saharan Africans, 20 present-day non-Africans and 4 ancient non-Africans (an ancient genome, meant to be similar to Ust'-Ishim) that are related to each other as shown in Figure S18.3. To simulate the two ancient samples (Ust'-Ishim and the Altai Neandertal), we set the population size to be 1 from time t = 0 to time t = 2,000 generations. This has the effect of stopping recombination during this interval. The simulated ancient non-African is inferred to have a shared demographic history with the other simulated non-Africans prior to 2,000 generations in the past. To sample an ancient diploid individual, we simulate two haploid individuals using the procedure, and then combined them.

We set the following parameters for the simulation:

- Mutation rate = 1.5×10^{-8} per bp/generation
- Recombination rate = 2×10^{-8} per bp/generation
- N_e of Neandertals = 2,500
- N_e of modern humans = 10,000.
- Non-Africans split from Africans 3,000 generations before present.
- The Altai Neandertal split from the introgressing Neandertal 4,000 generations ago.
- Africans split from Neandertals 12,000 generations before present
- The proportion of Neandertal ancestry in non-Africans is 0.03.
- We sample an ancient non-African (Ust'-Ishim) and Altai 2,000 generations before present.

The *ms* command line is as follows. Here, pop1= African, pop2=modern Europeans, pop3-6=Ust'-Ishim and pop7-10= Neandertal.

```
ms 48 1 -I 10 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 1 -en 0 3 1e-10 -en 0 4 1e-10 -en 0 5 1e-10 -en 0 6 1e-10 -en 0 7 1e-10 -en 0 8 1e-10 -en 0 9 1e-10 -en 0 10 1e-10 -ej 0.05 4 3 -ej 0.05 6 5 -ej 0.05 8 7 -ej 0.05 10 9 -en 0.05000025 3 1 -en 0.05000025 5 1 -en 0.05000025 7 0.25 -en 0.05000025 9 0.25 -ej 0.0500025 5 3 -ej 0.0500025 9 7 -en 0.050005 3 1 -en 0.050005 7 0.25 -ej 0.0500075 3 2 -en 0.05001 2 1 -es 0.0625 2 0.97 -en 0.0625025 11 0.25 -en 0.0625025 2 1 -ej 0.075 2 1 -en 0.0750025 1 1 -ej 0.1 11 7 -en 0.1000025 7 0.25 -ej 0.3 7 1 -en 0.3000025 1 1 -r 20000 50000000 -t 30000 -p 12 -seeds 45 46 47
```

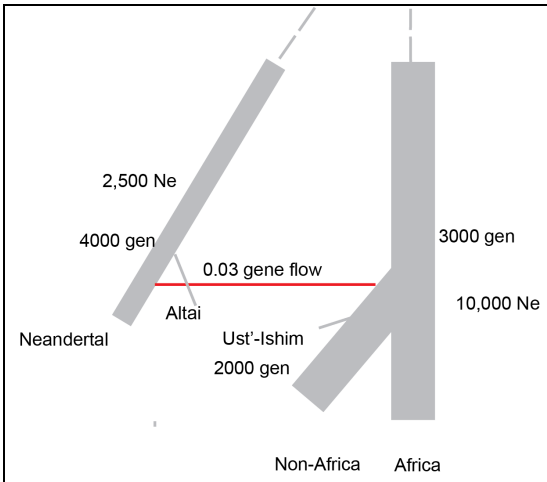


Figure S18.3 *Demographic model for simulation 2.*

The results are summarized in Table S18.3. For the ancient sample (Ust'-Ishim), we obtain accurate and unbiased dates of mixture (500 generations). However, the estimated dates are biased for present-day non-Africans where the mixture occurred ~2500 generations ago.

Table S18.3 *Dates of Neandertal admixture for simulation 2.*

Ind	Expected	Estimate	Std Error
Ust'-Ishim	500	552	52
Non-African1	2500	2101	288
Non-African2	2500	2272	265
Non-African3	2500	1902	198
Non-African4	2500	2083	117

Simulation 3: We introduced a more complex demographic history based loosely on the model fit by Gravel et al. 2011 [15] (Figure S18.4). We do not use the population sizes and the European-Asian population split time estimated by Gravel et al. (2011) (of 17,200–26,500 years before present), since these are incompatible with the recent observations that Tianyuan, an ancient sample dated to ~40,000 years before present, is already part of the lineage leading to present-day Asians [16]. We generated 20 African, 20 European, 20 Asian, 2 introgressing Neandertal, 2 Altai and 4 ancient non-African (Ust'-Ishim) haploid sequences. We set the parameters in Figure S18.4 as follows:

- Mutation rate = 1.5×10^{-8} per bp/generation
- Recombination rate = 2×10^{-8} per bp/generation
- Altai splits from the introgressing Neandertal 4,000 generations before present
- We sample Altai 2,400 generations before present
- Africans split from Neandertals 12,000 generations before present.
- N_e of Neandertals is 2,500
- We introduce a population expansion in the common ancestor of present-day humans at 6,000 generations ago with N_e increasing from 7,000 to 14,000
- Non-Africans split from Africans at 3,000 generations before present and undergo a bottleneck until 2,200 generations before present, reducing N_e to 1,860.
- Europeans and East Asians split 2,000 generations ago with a reduction in N_e of East Asians to 550, and N_e of Europeans to 1,032.
- Both European and East Asian populations subsequently undergo expansions until the present ($t = 0$) with the present-day East Asians have N_e of 45,300 and Europeans have N_e of 33,800.
- Gene flow from Neandertals into the ancestors of present-day non-Africans occurred at 2,500 generations before present
- The proportion of Neandertal ancestry in non-Africans is 0.03.
- We sample Ust'-Ishim 1,800 generations before present
- The split time between East Asians and Ust'-Ishim occurred 2,000 generations before present

The *ms* command line is as follows. Here, pop1= African, pop2=present-day Europeans, pop3=present-day Asian, pop4-7= Ust'-Ishim and pop8-11= Neandertal

ms 68 1 -I 11 20 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 2.41428571428571 -en 0 3 3.23571428571429 -eg 0 2 97.6909397920288 -eg 0 3 123.512032930849 -en 0 4 7.14285714285714e-11 -en 0 5 7.14285714285714e-11 -en 0 6 7.14285714285714e-11 -en 0 7 7.14285714285714e-11 -en 0 8 7.14285714285714e-11 -en 0 9 7.14285714285714e-11 -en 0 10 7.14285714285714e-11 -en 0 11 7.14285714285714e-11 -ej 0.0321428571428571 5 4 -ej 0.0321428571428571 7 6 -en 0.0321430357142857 4 0.714285714285714 -en 0.0321430357142857 6 0.714285714285714 -ej 0.0321446428571429 6 4 -en 0.0321464285714286 4 0.714285714285714 -ej 0.0357142857142857 2 3 -ej 0.0357142857142857 4 3 -en 0.0357160714285714 3 0.132857142857143 -en 0.0357160714285714 3 0.132857142857143 -ej 0.0428571428571429 9 8 -ej 0.0428571428571429 11 10 -en 0.0428573214285714 8 0.178571428571429 -en 0.0428573214285714 10 0.178571428571429 -ej 0.0428589285714286 10 8 -en 0.0428607142857143 8 0.178571428571429 -es 0.0446428571428571 3 0.97 -en 0.0446446428571429 12 0.178571428571429 -en 0.0446446428571429 3 0.132857142857143 -ej 0.0535714285714286 3 1 -en 0.0535732142857143 1 1 -ej 0.0714285714285714 12 8 -en 0.0714303571428571 8 0.178571428571429 -en 0.107142857142857 1 0.521428571428571 -ej 0.214285714285714 8 1 -en 0.2142875 1 0.714285714285714 -r 28000 50000000 -t 42000 -p 12 -seeds 67 68 69

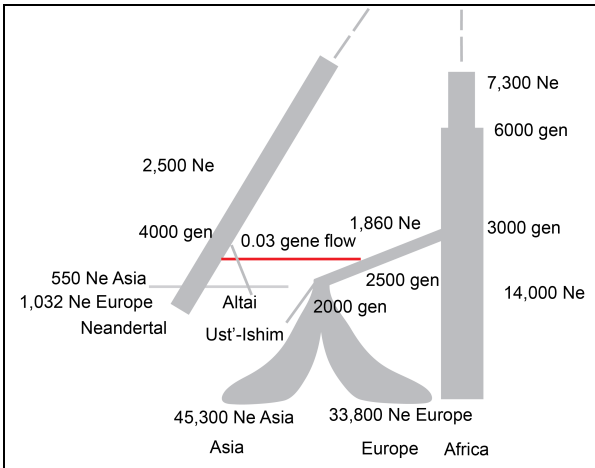


Figure S18.4 *Demographic model for simulation 3 based on the Gravel et al. model [15] with modifications described above.*

Table S18.4 shows the estimated dates of Neandertal introgression. The dates of the admixture in present-day non-Africans are downward biased, but this bias does not affect the dates in Ust'-Ishim, where the admixture is more recent and the inference is easier.

Table S18.4 *Dates of Neandertal admixture for simulation 3.*

Ind	Expected	Estimate	Std Error
Ust'-Ishim	700	651	57
Asia1	2500	1076	185
Asia2	2500	853	161
Asia3	2500	921	193
Asia4	2500	717	232
Europe1	2500	752	104
Europe2	2500	961	160
Europe3	2500	1157	143
Europe4	2500	776	230

Simulation 4: We modified simulation 3 based on the parameters inferred from real data. We changed the following parameters of simulation 3 as follows (see also Figure S18.5):

- Split time of Asian-European-Ust'-Ishim = 2,000 generations before present
- Gene flow from Neandertal to non-Africans = 2,200 generations before present

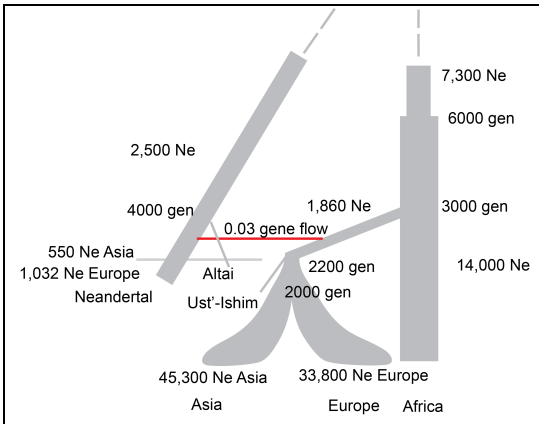


Figure S18.5 *Demographic model for simulation 4 is adapted from the model of Gravel et al. 2011, but is modified to be more realistic based on observations in real data.*

The *ms* command line is as follows. Here, pop1= African, pop2=present-day Europeans, pop3=present-day Asian, pop4-7= Ust'-Ishim, pop8-11= Neandertal.

ms 68 1 -I 11 20 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 2.41428571428571 -en 0 3 3.23571428571429 -eg 0 2 97.6909397920288 -eg 0 3 123.512032930849 -en 0 4 7.14285714285714e-11 -en 0 5 7.14285714285714e-11 -en 0 6 7.14285714285714e-11 -en 0 7 7.14285714285714e-11 -en 0 8 7.14285714285714e-11 -en 0 9 7.14285714285714e-11 -en 0 10 7.14285714285714e-11 -en 0 11 7.14285714285714e-11 -ej 0.0321428571428571 5 4 -ej 0.0321428571428571 7 6 -en 0.0321430357142857 4 0.714285714285714 -en 0.0321430357142857 6 0.714285714285714 -ej 0.0321446428571429 6 4 -en 0.0321464285714286 4 0.714285714285714 -ej 0.0357142857142857 2 3 -ej 0.0357142857142857 4 3 -en 0.0357160714285714 3 0.132857142857143 -en 0.0357160714285714 3 0.132857142857143 -es 0.0392857142857143 3 0.97 -en 0.0392875 12 0.178571428571429 -en 0.0392875 3 0.132857142857143 -ej 0.0428571428571429 9 8 -ej 0.0428571428571429 11 10 -en 0.0428573214285714 8 0.178571428571429 -en 0.0428573214285714 10 0.178571428571429 -ej 0.0428589285714286 10 8 -en 0.0428607142857143 8 0.178571428571429 -ej 0.0535714285714286 3 1 -en 0.0535732142857143 1 1 -ej 0.0714285714285714 12 8 -en 0.0714303571428571 8 0.178571428571429 -en 0.107142857142857 1 0.521428571428571 -ej 0.214285714285714 8 1 -en 0.2142875 1 0.714285714285714 -r 28000 50000000 -t 42000 -p 12 -seeds 46 47 48

Table S18.5 shows that the results of simulation 4 which are qualitatively similar to simulation 3. For more recent dates of admixture, such as in Ust'-Ishim, we continue to obtain unbiased dates of mixture.

Table S18.5 *Dates of Neandertal admixture for simulation 4.*

Ind	Expected	Estimate	Std Error
Ust'-Ishim	400	387	57
Asia1	2200	695	151
Asia2	2200	746	166
Asia3	2200	666	172
Asia4	2200	756	114
Europe1	2200	1035	86
Europe2	2200	1168	198
Europe3	2200	1004	166
Europe4	2200	1163	146

References

1 Sankararaman, S., Patterson, N., Li, H., Paabo, S. & Reich, D. The Date of Interbreeding between Neandertals and Modern Humans. *PLoS Genet* **8**, doi:DOI 10.1371/journal.pgen.1002947 (2012).

2 R Chakraborty, K. M. W. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci* **85**, 9119–9123 (1988).

3 Loh, P. R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233-1254, doi:10.1534/genetics.112.147330 (2013).

4 Moorjani P, P. N., Hirschhorn JN, Keinan A, Hao L, et al. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. . *PLoS Genet* **7**, e1001373, doi:doi:10.1371/journal.pgen.1001373 (2011).

5 Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747-751, doi:10.1126/science.1243518 (2014).

6 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

7 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

8 Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* **17**, 1217-1241, doi:DOI 10.1214/aos/1176347265 (1989).

9 Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology* **128**, 415-423, doi:10.1002/ajpa.20188 (2005).

10 Auton, A. *et al.* A fine-scale chimpanzee genetic map from population sequencing. *Science* **336**, 193-198, doi:10.1126/science.1216872 (2012).

11 Hinch, A. G. *et al.* The landscape of recombination in African Americans. *Nature* **476**, 170-175, doi:10.1038/nature10336 (2011).

12 Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat Genet* **31**, 241-247, doi:10.1038/ng917 (2002).

13 International HapMap, C. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861, doi:10.1038/nature06258 (2007).

14 Hudson, R. R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-338 (2002).

15 Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 11983-11988, doi:DOI 10.1073/pnas.1019276108 (2011).

16 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).