ANCIENT PROTEINS

The proteome of the late Middle Pleistocene Harbin individual

Qiaomei Fu[1,2]*, Fan Bai[1], Huiyun Rao[1]*, Shaokun Chen[3], Yannan Ji[4], Aoran Liu[3], E. Andrew Bennett[1], Feng Liu[1], Qiang Ji[3]*

Denisovans are a hominin group primarily known through genomes or proteins, but their precise morphological features remain elusive because of the fragmentary nature of the discovered fossils. Here, we report 95 endogenous proteins retrieved from a nearly complete cranium from Harbin, China, dating to at least 146,000 years ago and previously assigned to a new species, *Homo longi*. This individual has three Denisovan-derived amino acid variants and clusters with Denisova 3, suggesting that the Harbin individual belongs to a Denisovan population. This study fills the gap between morphological and molecular evidence, enhancing our understanding of Denisovans' spatiotemporal dispersal and evolutionary history.

Denisovans represent an ancient hominin group distributed in Asia, and two genetically distinct populations have been identified using ancient DNA from Denisovans in Siberia and introgressed regions in modern humans (*1*–*4*). The currently known physical remains of Denisovans encompass a sparse assemblage of teeth, bone fragments, a phalanx, and a partial cranial fragment from a Denisovan cave in Siberia (*1*–*6*) that have limited diagnostic features for identifying different hominin groups. Thus, Denisovans to date have primarily been identified by DNA evidence. Recent advances in mass spectrometry technology have enabled the recovery of ancient hominin proteins from Early Pleistocene contexts (*7*–*11*) (Fig. 1A), expanding studies of Denisovans from solely genomic to proteomic as well. One benefit of proteomic analysis is that proteins are better preserved than DNA, thus expanding molecular studies into time periods thought to be out of reach for ancient DNA survival. Proteomic analysis has allowed the identification of more Denisovan-related specimens outside of Siberia, including a partial mandible and a rib from Baishiya Cave in China (*9*, *10*) and a Denisovan mandible from the Penghu Channel in Taiwan (*12*), as well as the collection of proteomic data from a possible Denisovan molar from the Tam Ngu Hao 2 (Cobra Cave) limestone cave in Laos (*11*). Identifying Denisovan-like specimens in China dating as far back as 160,000 years has broadened our knowledge of the geological range and time span of this extinct population. However, all fossil hominin specimens from which Denisovan-like genomic or proteomic material has been retrieved are fragmentary. Thus, despite Denisovans having been identified as a distinct archaic human lineage relative to Neanderthals and modern humans in 2010, we still do not know what they looked like.

Previous studies have inferred that Denisovans may have been widely distributed in East Asia (*13*, *14*), where substantial Middle Pleistocene hominin specimens have been unearthed. Some of these fossils [e.g., Xujiayao (*15*) and Harbin (*16*, *17*)], with nearly complete crania and informative morphological characteristics, have been suggested as possible Denisovans given their age, geography, and archaic features. However, whether a connection exists between these hominin fossils and Denisovans remains unclear. Retrieving ancient DNA or proteins from any of these fossils would help to identify whether they were Denisovans and, if so, to finally associate informative morphological characteristics with Denisovan genomic and proteomic material. Clarifying Denisovan morphology would greatly expand the identification of Denisovans across Asia and assist in calibrating Asia's Middle Pleistocene hominin systematics, further improving our understanding of the spatiotemporal dispersal of hominin populations in Asia and the relationship of Denisovans to recent Asian populations.

The Harbin fossil is an undistorted and nearly complete cranium discovered in the Harbin area of Heilongjiang Province, China (*16*, *18*) (Fig. 1A). Direct uranium-series dating of the cranium yielded a minimum age of ~146 thousand years ago (ka), whereas regional stratigraphic correlations suggest a possible origin from the upper part of the Upper Huangshan Formation (138 to 309 ka) (*19*). In one paper, it was made the type specimen of a new hominin species, *Homo longi* ("Dragon Man") (*16*). Since the first report of this cranium in 2021, intense debate has arisen regarding its taxonomic affiliation. Using morphology and phylogenetic analyses, it was grouped with Xiahe1 from Baishiya Karst Cave on the Tibetan Plateau (*18*), which was previously identified as a Denisovan-related population through proteomic analysis (*9*). Furthermore, a range of other Middle Pleistocene Chinese fossils, such as Dali, Hualongdong, and Jinniushan, were grouped with Harbin in these morphological analyses (*18*). The work on Harbin has since been extended to include the early Pleistocene Yunxian 2 cranium (often classified as *H. erectus*), the Penghu mandible, Xujiayao, and fossils from Denisova Cave, which in this work were grouped within the same clade as Harbin and were determined to be more closely related to *H. sapiens* than to Neanderthals (*20*). However, other recent morphological studies (*15*, *20*, *21*) have suggested that Harbin and Xiahe are not closely related. These studies have argued for three distinct groups: (i) the Julurens/Denisovans group (e.g., Xujiayao, Xiahe, and Penghu) characterized by larger cranial capacity and some shared Neanderthal features; (ii) the "Dragon Man" group (Harbin, Dali, and Jinniushan); and (iii) a third group consisting of Narmada, Maba, and Hualongdong.

Across these studies, the lack of genomic or proteomic material from a Middle Pleistocene specimen with informative morphology has impeded definitive phylogenetic assignments, particularly for Denisovans. Given the nearly complete morphology of the Harbin cranium, any ancient DNA or protein recovered from this individual would provide crucial evidence to connect molecular material to morphology, thereby improving our understanding of human evolutionary genealogy during the Middle Pleistocene in Asia.

Results

To understand the phylogenetic placement of the Harbin cranium, we first attempted to retrieve ancient DNA from six samples from the petrous bone and teeth (4.8 to 9.5 mg each) of the Harbin individual. No sample showed evidence of ancient human DNA (see the materials and methods and table S1). Subsequently, to assess the preservation of ancient proteins, we extracted proteins from two petrous samples (64.3 and 35.1 mg). Different fractions were collected to maximize amino acid sequence coverage, including an ammonium bicarbonate fraction, three acid-soluble fractions (>10, 3 to 10, and >3 kD), and two acid-insoluble fractions (see the materials and methods and data S2).

Recovery of the endogenous proteome

To gain more comprehensive information about the Harbin individual, a proteome was recovered by searching against a database composed of the modern human proteome and translated common bone and dentine protein sequences from archaic hominin genomes (see the supplementary text, section S2). A total of 308,458 peptide-spectrum matches (PSMs) and 20,455 peptides were identified by PEAKS Online (*22*). The peptides matching multiple genes were considered as nonunique and were excluded from subsequent analysis. We then assessed whether ancient hominin proteins were present in the Harbin individual by

[1]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China. [2]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China. [3]Hebei International Joint Research Center for Paleoanthropology, Hebei GEO University, Shijiazhuang, China. [4]Chinese Academy of Geological Sciences, Beijing, China. ***Corresponding author. Email: jiqiang@hgu.edu.cn (Q.J.); fuqiaomei@ivpp.ac.cn (Q.F.) raohuiyun@ivpp.ac.cn (H.R.)**

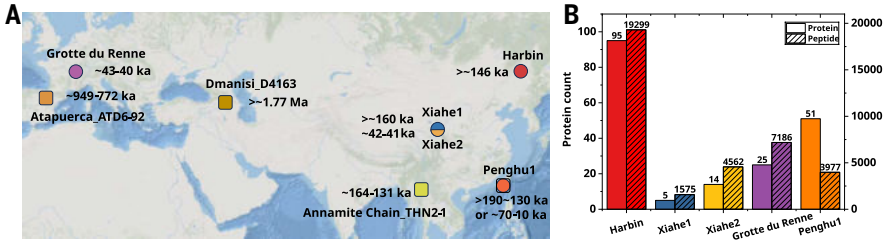

Fig. 1. Geographic locations and proteomic profiles for the Pleistocene hominin individuals with palaeoproteomic data. (**A**) Map of the geographic locations and hominin specimens studied by palaeoproteomic analysis. Circles with different colors indicate different bone or dentine specimens; squares with different colors indicate different enamel specimens. Because Xiahe1 and Xiahe2 were discovered at the same site, each specimen is represented by half of the same circle. Because both bone and enamel of Penghu1 have been analyzed, this is indicated as a circle within a square. Detailed information is provided in data S1. (**B**) Protein and peptide counts identified from five specimens. The data from Penghu1 were obtained from a related study (*12*). Because no accurate, nonredundant peptide counts were reported for the remaining three specimens, their data were obtained by processing corresponding raw proteomic data files using the same strategy as that used for Harbin.

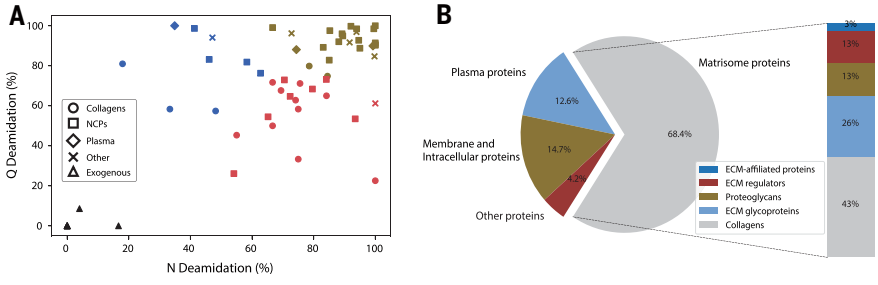

Fig. 2. Proteomic profile of the Harbin cranium. (**A**) Deamidation values of the identified proteins used in cluster analysis: 0% indicates no deamidation and 100% indicates complete deamidation. There is a clear separation between contaminating proteins (group 1: black) and endogenous proteins (group 2: red, group 3: blue, and group 4: brown) based on K-means cluster analysis. Cluster details are described in the supplementary text, section S2. Different shapes represent different types of proteins. "NCPs" refers to noncollagen matrisome proteins. (**B**) Composition across the 95 proteins retrieved for the Harbin individual. "Plasma proteins" here refers to nonmatrisome plasma proteins.

calculating the deamination values of glutamine (Q) and asparagine (N) in the detected proteins based on peptide intensity, which required at least two Q/N-containing peptides. Higher deamination values of Q or N suggest a more extensive degradation of related proteins and thus can be used to authenticate the retrieved ancient proteome (*7*–*9*). There are 59 proteins with both reliable Q and N deamidation values, and 48 proteins with either reliable Q or reliable N deamidation values. Proteins with both reliable Q and N deamidation values were included for cluster analysis to discriminate endogenous from contaminating proteins (Fig. 2A). Eight proteins (including keratins and trypsin) with low deamination values of Q and N (Q deamidation average: 1.07 ± 3.04%; N deamidation average: 2.58 ± 5.86%) were excluded as contaminants (group 1 in Fig. 2A), which is consistent with previously published common contaminants (*8*, *23*). The remaining 51 proteins showed elevated deamidation values (Q deamidation range: 22.56 to 100%; N deamidation range: 17.97 to 100%) (groups 2, 3, and 4 in Fig. 2A) and were considered authentic ancient proteins. In addition, 44 of 48 proteins showed reliable elevated Q or N deamidation values (Q deamidation range: 30 to 100%; N deamidation range: 44.77 to 100%). In total, 95 proteins with elevated deamidation values were detected and were considered to belong to the Harbin individual's endogenous proteome (data S3). The main composition of these 95 proteins included 68.4% matrisome proteins (54 core matrisome proteins and 11 matrisome-associated proteins) and 12.6% nonmatrisome plasma proteins (Fig. 2B), which is comparable to those described from well-preserved bone specimens (*24*) (see the supplementary text, section S3).

By retrieving 95 proteins with as many as 19,299 peptides from the Harbin cranium (>146 ka) (Fig. 1B), we not only increased the limited pool of proteomes retrieved from hominins dating to >100 ka, we also sequenced a much larger portion of the human proteome than from all other ancient individuals (protein count: 5 to 51; peptide count: 1575 to 7186). Comparing with previously published proteomic data, there was a significant 11.25-fold increase in peptide count over the dentine proteome of the contemporaneous Xiahe1 mandible (>160 ka) (*9*). There was also a higher peptide quantity relative to the younger Xiahe2 rib (*10*) and the Grotte du Renne bone fragments (*8*) (Fig. 1B). These results demonstrate the recovery of a highly informative ancient hominin proteome from the Harbin petrous bone.

Species confirmation

To determine the population assignment of the Harbin individual at a proteomic scale, we calculated peptide counts covering amino acid substitutions unique to modern humans, Neanderthals, Denisovans, chimpanzees, gorillas, and orangutans within all 95 endogenous proteins obtained. A minimum of two peptides was required for each position. In total, 122 single–amino acid polymorphisms (SAPs) were recovered from this endogenous proteome, showing a clear connection between the Harbin cranium and the Denisovans (Fig. 3 and data S4). First, positions derived in the gorillas, chimpanzees, and orangutans were all ancestral in the Harbin individual, with a match frequency of 100% (2025/2025 peptides, 102 positions). The Harbin cranium also showed 15 derived amino acid variants assigned to the *Homo* genus with a match frequency of 100% (74/74 peptides). These SAPs confidently assigned this individual to the *Homo* genus, which is consistent with the preliminary species confirmation based on matrix-assisted laser desorption/ionization (MALDI) spectra and a collagen database search (see the supplementary text, section S1).

Population assignment and possible physiological insight

We then evaluated all of the SAPs within the *Homo* genus and only retained positions with enough peptides and reliable PSMs (see the materials and methods and data S5). No reliable SAPs derived in Neanderthals were detected, whereas one SAP derived in modern humans showed the ancestral amino acid variant for the Harbin individual with a match frequency of 100% (2/2 peptides, 1 position). Finally, we identified four positions unique to Denisovans, of which three showed derived Denisovan amino acid variants with a match frequency of 86.57% (4/4, 109/126, and 3/4 peptides; data S4), and the remaining position retained the ancestral amino acid variant covered by two peptides and 13 PSMs.

We further investigated what amino acids these four Denisovan-derived positions showed in the Harbin individual, two Xiahe individuals (*9*, *10*), Penghu1, and Denisova 3 (*3*) (Table 1 and table S2). One Denisovan-derived allele (COL1A2 R996K) in the Harbin individual was already identified in the two Xiahe individuals and Penghu1, as well as in the Denisova 3 genome, which showed a homozygous derived allele at the corresponding genomic position. Even though COL1A2 996R was observed in Harbin at a low percentage of PSMs (4.5%) and related intensity (0.6%), the possibility of heterozygosity at this position cannot be excluded. For the three remaining positions not covered in the Xiahe individuals and Penghu1, Denisova 3 was

heterozygous at all corresponding genomic positions, whereas the Harbin cranium was homozygous derived for one position (COL18A1 G1033R), heterozygous for a second (COL26A1 R196G), and homozygous ancestral for a third (COL4A3 R368H). However, these amino acid alleles are not of high quality (Table 1). Because current proteomic search engines may not work so well on the accurate identification of PTM locations (table S2), particularly those in collagen, and because none of these three positions was detected in pFind, future studies could potentially uncover additional insights from the data by using complementary approaches for validation reliability [e.g., Orthrus (*25*)]. However, the identification of one highly confident and two potential Denisovan-derived amino acid variants supports the conclusion that the Harbin individual was from a Denisovan population.

Pathogenicity predictions for related protein variants were sourced from the AlphaMissense dataset (*26*). Of all the SAPs, only a heterozygous COL2A1 G1170S variant in the Harbin cranium is predicted to be pathogenic, being previously reported to possibly cause skeletal abnormalities such as avascular necrosis of the femoral head, Legg-Calvé-Perthes disease, premature hip osteoarthritis, and Stickler syndrome type 1 (*27*–*30*). The specific impact of the heterozygous state on the morphology and health of this individual is uncertain and requires further analysis using ancient phenotypic data and mutation expressivity studies.

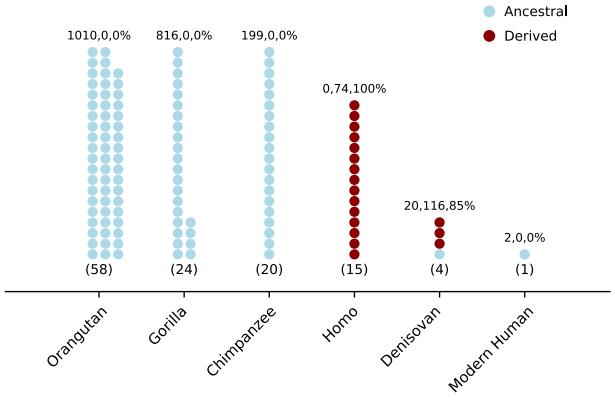

Fig. 3. SAP assignments for the Harbin cranium. Filled circles represent the observed genotype for each amino acid in the Harbin individual (blue: ancestral amino acids; red: derived amino acids). The number of ancestral and derived peptide counts, respectively, followed by the frequency of derived peptides, are shown at the top of the bar separated by commas. The number of SAPs is shown at the bottom of the bar in parentheses.

Phylogenetic reconstruction

To determine a more specific taxonomic attribution for the Harbin cranium, we reconstructed the consensus sequences of its endogenous proteins for phylogenetic analysis. For each endogenous protein of the Harbin cranium, reliable alleles were retained if they met the following requirements: PSM count ≥2, PSM ratio ≥10%, and intensity ratio ≥10% (*31*). Three confident heterozygous positions in which multiple amino acid alleles were retained with reliable spectra were identified in the Harbin cranium (Table 1). Consensus protein sequences were generated, with one variant randomly selected for heterozygous positions, and positions with single peptide coverage or indistinguishable, unreliable, or less-reliable SAPs were marked as "X." Less-confident sites (COL18A 1033, COL26A1 196, and COL4A3 368) related to Denisovans were also assigned to "X." We obtained 9422 amino acid positions from 88 proteins for the Harbin cranium, which is 11.37% of the total concatenated protein sequence alignment. The reference dataset for phylogenetic analysis was built using translated protein sequences from previously published high-coverage genomes, including Denisova 3 (*3*), two Neanderthals (Altai Neanderthal and Vindija19) (*32*, *33*), one early modern human (Ust'-Ishim) (*34*), present-day human individuals (*35*), and a chimpanzee (*36*, *37*). Two different sets of modern human individuals were used, one with 58 global present-day humans and the other with 64 pan-African individuals.

The phylogenetic trees for these two datasets are highly congruent with a commonly accepted primary topology also found for nuclear genomes, i.e., *H. sapiens* is a sister group to a clade containing Neanderthals and Denisovans (Fig. 4 and fig. S4). This contrasts with morphology-based phylogenetic analyses (*18*, *20*) suggesting that Harbin was in the sister group to the *H. sapiens* clade rather than to Neanderthals. The Harbin cranium and Denisovan 3 form a monophyletic group with high support (posterior probability 100% in the pan-African dataset and 98%

Table 1. Details of SAPs within *Homo* sp. identified in the Harbin cranium*

SAP assignment	Protein	Position	Accession no.	No. of peptides	No. of PSMs	Harbin	Xiahe1	Xiahe2	Penghu1	Denisova3	Altai Neanderthal	Vindija19	Ust'-Ishim	HGDP frequency	Great apes
Denisovan	**COL1A2**	**996**	**P08123**	**109/17**	**2277/107**	**K/R†**	**K**	**K**	**K**	**K**	**R**	**R**	**R**	**Fixed R**	**R**
Denisovan	COL18A1	1033	P39060	4	4	R‡	X	X	X	R,G	G	G	G	Fixed G	G
Denisovan	COL26A1	196	Q96A83	3/1	3/1	G/R‡	X	X	X	G,R	R	R	R	Fixed R	R
Denisovan	**COL4A3**	368	**Q01955**	2	13	R‡	**X**	**X**	X	**H,R**	**R**	**R**	**R**	**Fixed R**	R
Xiahe1	**COL2A1**	583	**P02458**	10/7	15/10	E/G†	**G**	**E**	E	**E**	**E**	**E**	**E**	**Fixed E**	E
Harbin	COL2A1	1022	P02458	17/1	28/1	M/V†	V	V	X	V	V	V	V	Fixed V	V
Harbin	SCARA3	490	Q6AZY7	2	2	A‡	X	X	X	G	G	G	G	Fixed G	G
Harbin heterozygous	COL1A2	549	P08123	234/50	1647/189	A/P	X	A	A	A	A	A	A	P (23.7%), A (76.3%)	A
Harbin heterozygous	COL2A1	601	P02458	12/13	68/21	IL/V	X	V	V	V	V	V	V	Fixed V	V
Harbin heterozygous	COL2A1	1170	P02458	38/33	207/158	G/S	G	G	G	G	G	G	G	Fixed G	G
Modern human	KNG1	178	P01042	2	5	T	X	X	X	T	T	T	M	M (47.8%), T (52.2%)	T

*Positions are based on the Uniprot accession number for the related human proteins. The listed amino acid calls for Xiahe1 (*9*), Xiahe2 (*39*), and Penghu1 (*12*) were obtained from the published protein sequences. The amino acid calls for Denisova3 (*3*), Altai Neanderthal (*32*), Vindija19 (*33*), and Ust'-Ishim (*34*) were translated from genomic data. The frequency was calculated from 929 high-coverage genomes of the Human Genome Diversity Project (*35*). †Considered less-reliable identifications because these amino acid alleles do not fulfill the requirements of PSM count ≥2, PSM ratio ≥10%, and intensity ratio ≥10% (*31*), and therefore the possibility of heterozygosity is low for these sites. Some derived alleles, such as COL1A2 996K and COL2A1 1022 M, appear to be better supported than the ancestral ones. ‡Considered less-reliable identifications because of limited spectra support (COL26A1 196R and SCARA3 490A), relatively low-quality spectra (COL18A1 1033R and COL26A1 196G), or possible contaminant origins (COL4A3 368R). For COL18A1 1033R, none of the supported spectra show a predominance of y ions. For COL26A1 196G, all supported PSMs have this SAP at the N terminus without surrounding y and b ions, which is consistent with the fragmentation patterns for HCD, where b1 and y_{N-1} ions are usually missing in the MS/MS spectrum (*40*). For COL4A3 368R, the supported peptides may also match possible contaminants such as mouse or bacteria.

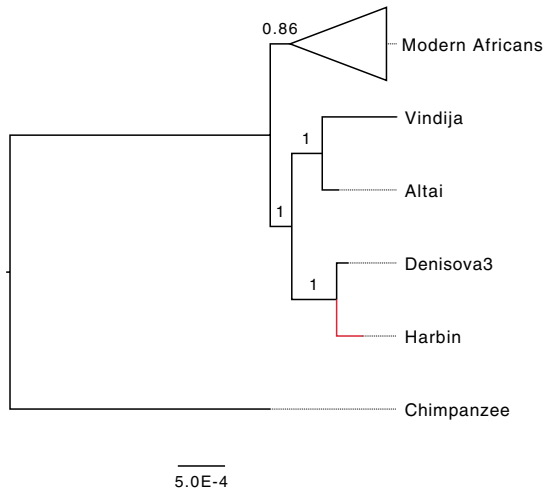

Fig. 4. Phylogenetic position of the Harbin cranium through Bayesian analysis using the dataset with 64 pan-Africans. Chimpanzee (*Pan troglodytes*) was used as an outgroup.

in the global dataset). This result is consistent with the population assignment using SAPs described above, in which we assigned the Harbin individual to a Denisovan population.

Discussion and conclusions

Recent studies have demonstrated the high potential of paleoproteomic analysis in tracking human evolutionary history, extending insights into geographical areas and time periods where substantial genetic information is absent (*7*, *9*, *38*). In this study, despite the lack of preserved ancient human DNA in the petrous bone and teeth of the Harbin cranium, abundant proteins were found, resulting in a high-quality archaic hominin proteome. Through proteomic analysis, we were able to definitively identify this individual as Denisovan.

By using proteomic analysis to connect the Harbin individual's nearly complete cranium with Denisovans, we now have the first comprehensive morphological blueprint for Denisovan populations, helping to address an unresolved question that has persisted over the past decade about what Denisovans looked like. In addition, the findings from our ancient proteomic analysis provide solid evidence that Denisovans were widely distributed across East Asia, as has been inferred in previous studies (*12*–*14*, *18*, *20*). The Harbin cranium, along with Xiahe1, Xiahe2, and Penghu1, have all been grouped through proteomic analysis with Denisovans, with a geographical range extending from Northeast China to the Tibetan Plateau to Taiwan and a chronological range spanning from the late Middle Pleistocene to the late Pleistocene. This study finally connects more complete Denisovan morphology to the genomic and proteomic data previously studied, further improving our understanding of Denisovan spatiotemporal dispersal and evolutionary history.

Additionally, proteomic identification of the Harbin cranium prompts a reconsideration of Asia's Middle Pleistocene hominin systematics. Denisovan populations in East Asia inhabited a variety of environments, including low-lying plains, high plateaus, and subtropical woodlands. Furthermore, many hominin fossils across East Asia, such as Yunxian 2, Dali, Jinniushan, and Hualongdong, show morphological features similar to those found in the Harbin cranium (*18*, *20*). Our findings strongly suggest that despite studies suggesting that this group may be distinguished from fossils more similar morphologically to the Denisovan-linked Xiahe individuals (e.g., Xujiayao and Penghu), these hominins potentially also belong to Denisovan populations. Additional molecular sampling of Middle and Late Pleistocene hominin fossils in East Asia, especially any with ancient DNA, will be necessary for determining the differentiation and dispersal patterns of Denisovan populations in East Asia and how these East Asian archaic hominins interacted with other archaic human populations and affected more recent Asian populations.

REFERENCES AND NOTES

1. J. Krause *et al.*, *Nature* **464**, 894–897 (2010).
2. D. Reich *et al.*, *Nature* **468**, 1053–1060 (2010).
3. M. Meyer *et al.*, *Science* **338**, 222–226 (2012).
4. V. Slon *et al.*, *Sci. Adv.* **3**, e1700186 (2017).
5. B. T. Viola, P. Gunz, S. Neubauer, "A parietal fragment from Denisova cave," paper presented at the 88th Annual Meeting of the American Association of Physical Anthropologists, Cleveland, OH, 27 to 30 March 2019.
6. S. Brown *et al.*, *Nat. Ecol. Evol.* **6**, 28–35 (2022).
7. F. Welker *et al.*, *Nature* **580**, 235–238 (2020).
8. F. Welker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11162–11167 (2016).
9. F. Chen *et al.*, *Nature* **569**, 409–412 (2019).
10. H. Xia *et al.*, *Nature* **632**, 108–113 (2024).
11. F. Demeter *et al.*, *Nat. Commun.* **13**, 2557 (2022).
12. T. Tsutaya *et al.*, *Science* **388**, 176–180 (2025).
13. J. Ruan *et al.*, *Science* **381**, 699–704 (2023).
14. D. Reich *et al.*, *Am. J. Hum. Genet.* **89**, 516–528 (2011).
15. X. Wu, C. J. Bae, *Paleoanthropology* (2024).
16. Q. Ji, W. Wu, Y. Ji, Q. Li, X. Ni, *Innovation (Camb.)* **2**, 100132 (2021).
17. É. Harney, N. Patterson, D. Reich, J. Wakeley, *Genetics* **217**, iyaa045 (2021).
18. X. Ni *et al.*, *Innovation (Camb.)* **2**, 100130 (2021).
19. Q. Shao *et al.*, *Innovation (Camb.)* **2**, 100131 (2021).
20. X. Feng *et al.*, The phylogenetic position of the Yunxian cranium elucidates the origin of Dragon Man and the Denisovans. bioRxiv 594603 [Preprint] (2024); https://doi.org/10.1101/2024.05.16.594603.
21. C. J. Bae, W. Liu, X. Wu, Y. Zhang, X. Ni, *Innovation (Camb.)* **4**, 100527 (2023).
22. L. Xin *et al.*, *Nat. Commun.* **13**, 3108 (2022).
23. H. Rao *et al.*, *Sci. Rep.* **10**, 16674 (2020).
24. E. Cappellini *et al.*, *J. Proteome Res.* **11**, 917–926 (2012).
25. Y. Chiang, M. J. Collins, Orthrus: an AI-powered, cloud-ready, and open-source hybrid approach for metaproteomics. bioRxiv 623814 [Preprint] (2024); https://doi.org/10.1101/2024.11.15.623814.
26. J. Cheng *et al.*, *Science* **381**, eadg7492 (2023).
27. Y. F. Liu *et al.*, *N. Engl. J. Med.* **352**, 2294–2301 (2005).
28. Y. Miyamoto *et al.*, *Hum. Genet.* **121**, 625–629 (2007).
29. P. Su *et al.*, *Arthritis Rheum.* **58**, 1701–1706 (2008).
30. S. Savasta *et al.*, *Eur. J. Pediatr.* **174**, 697–701 (2015).
31. P. P. Madupe *et al.*, *Science* **388**, 969–973 (2025).
32. K. Prüfer *et al.*, *Nature* **505**, 43–49 (2014).
33. K. Prüfer *et al.*, *Science* **358**, 655–658 (2017).
34. Q. Fu *et al.*, *Nature* **514**, 445–449 (2014).
35. A. Bergström *et al.*, *Science* **367**, eaay5012 (2020).
36. Z. N. Kronenberg *et al.*, *Science* **360**, eaar6343 (2018).
37. Y. Mao *et al.*, *Cell* **187**, 1547–1562.e13 (2024).
38. F. Welker *et al.*, *Nature* **576**, 262–265 (2019).
39. H. Xia, Y. Li, D. Zhang, F. Chen, *Kexue Tongbao* **69**, 5155–5160 (2024).
40. C. Shao, Y. Zhang, W. Sun, *J. Proteomics* **109**, 26–37 (2014).
41. Y. Perez-Riverol *et al.*, *Nucleic Acids Res.* **53**, D543–D553 (2025).

ACKNOWLEDGMENTS

We thank R. Yang, S. Yang and M. Shi for help in the laboratory; F. Yang and J. Wang for conducting some of the LC MS/MS runs; S. Tang for visualizing the figure; Z. Wu and Z. Zhou for initial investigations; M. A. Yang, W. Ping, and S. Pääbo for helpful comments; and the BSI team for making special improvements for our PEAKS Online software. **Funding:** This work was supported by the Chinese Academy of Sciences (CAS) (grant YSBR-019) and the National Natural Science Foundation of China (grant 41842039). **Author contributions:** Q.F. designed and supervised the research project. Q.J. prepared the archaeological samples and materials. Q.F., H.R., and F.B. analyzed data. E.A.B. and F.B. prepared the database. H.R, Q.F., and F.L. performed or supervised the wet laboratory work. Q.F. wrote the manuscript. Q.F., H.R., E.A.B., and Q.J. revised the manuscript. Q.F., H.R., E.A.B., and F.B. wrote and prepared the supplementary information. All authors discussed, critically revised, and approved the final version of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All proteomic mass spectrometry data have been deposited to the ProteomeXchange Consortium through the PRIDE (*41*) partner repository with dataset identifier PXD058447. Protein consensus sequences for the Harbin cranium used for phylogenetic analysis are available in data S6.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adu9677
Materials and Methods; Supplementary Text; Figs. S1 to S4; Tables S1 and S2; References (*42*–*101*); MDAR Reproducibility Checklist; Data S1 to S7

Submitted 29 November 2024; accepted 9 June 2025; published online 18 June 2025
10.1126/science.adu9677

Article

The Innovation

Massive cranium from Harbin in northeastern China establishes a new Middle Pleistocene human lineage

Xijun Ni,[1,2,3,4,*] Qiang Ji,[1,*] Wensheng Wu,[1] Qingfeng Shao,[5] Yannan Ji,[6] Chi Zhang,[2,4] Lei Liang,[1] Junyi Ge,[2,4] Zhen Guo,[1] Jinhua Li,[7] Qiang Li,[2,4] Rainer Grün,[8,9] and Chris Stringer[10,*]

[1]Hebei GEO University, Shijiazhuang 050031, China

[2]CAS Center for Excellence in Life and Paleoenvironment, Chinese Academy of Science, Beijing 100044, China

[3]CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Science, Beijing 100104, China

[4]University of Chinese Academy of Sciences, Beijing 100049, China

[5]Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China

[6]China Geo-Environmental Monitoring Institute, Beijing 100081, China

[7]Key Laboratory of Earth and Planetary Physics, Innovation Academy for Earth Science, Chinese Academy of Sciences, Beijing 100029, China

[8]Australian Research Centre for Human Evolution, Griffith University, Nathan, QLD, Australia

[9]Research School of Earth Sciences, The Australian National University, Canberra, ACT, Australia

[10]Centre for Human Evolution Research, Department of Earth Sciences, Natural History Museum, London, UK

*Correspondence: jiqiang@hgu.edu.cn (Q.J.); nixijun@hgu.edu.cn (X.N.); c.stringer@nhm.ac.uk (C.S.)

Received: May 10, 2021; Accepted: June 4, 2021; Published Online: June 25, 2021; https://doi.org/10.1016/j.xinn.2021.100130

Citation: Ni X., Ji Q., Wu W., et al., (2021). Massive cranium from Harbin in northeastern China establishes a new Middle Pleistocene human lineage. The Innovation **2(3)**, 100130.

It has recently become clear that several human lineages coexisted with *Homo sapiens* during the late Middle and Late Pleistocene. Here, we report an archaic human fossil that throws new light on debates concerning the diversification of the *Homo* genus and the origin of *H. sapiens*. The fossil was recovered in Harbin city in northeastern China, with a minimum uranium-series age of 146 ka. This cranium is one of the best preserved Middle Pleistocene human fossils. Its massive size, with a large cranial capacity (~1,420 mL) falling in the range of modern humans, is combined with a mosaic of primitive and derived characters. It differs from all the other named *Homo* species by presenting a combination of features, such as long and low cranial vault, a wide and low face, large and almost square orbits, gently curved but massively developed supraorbital torus, flat and low cheekbones with a shallow canine fossa, and a shallow palate with thick alveolar bone supporting very large molars. The excellent preservation of the Harbin cranium advances our understanding of several less-complete late Middle Pleistocene fossils from China, which have been interpreted as local evolutionary intermediates between the earlier species *Homo erectus* and later *H. sapiens*. Phylogenetic analyses based on parsimony criteria and Bayesian tip-dating suggest that the Harbin cranium and some other Middle Pleistocene human fossils from China, such as those from Dali and Xiahe, form a third East Asian lineage, which is a part of the sister group of the *H. sapiens* lineage. Our analyses of such morphologically distinctive archaic human lineages from Asia, Europe, and Africa suggest that the diversification of the *Homo* genus may have had a much deeper timescale than previously presumed. Sympatric isolation of small populations combined with stochastic long-distance dispersals is the best fitting biogeographical model for interpreting the evolution of the *Homo* genus.

Keywords: human phylogeny; human cranium fossil; human dispersal; human diversification

INTRODUCTION

The origin of modern humans (*Homo sapiens*, our own species) has long been a controversial topic. During the late Middle and Late Pleistocene, several human lineages, evidently at species level, coexisted with *H. sapiens* across Africa and Eurasia. These extinct hominins include *H. heidelbergensis*/*H. rhodesiensis*, *Homo naledi*, *Homo floresiensis*, *H. luzonensis*, Denisovans, Neanderthals (*Homo neanderthalensis*), and *Homo erectus*.[1–5] The phylogenetic relationship between these coexisting hominins and *H. sapiens* has long been debated. Before the appearance of undoubted modern humans in Asia, some archaic fossils, such as those from Narmada, Maba, Dali, Jinniushan, Xuchang, and Hualongdong show mosaic combinations of features present in *H. erectus*, *H. heidelbergensis*/*H. rhodesiensis*, Neanderthals, and *H. sapiens*. Therefore, it is widely believed that these Asian hominins are critical for studying the later evolution of the genus *Homo* and the origin of *H. sapiens*. The incomplete preservation of these fossils and the fact that they have largely been described by advocates of regional continuity have made it difficult to integrate them into the wider picture of human evolution. For example, Xuchang, Dali, and Hualongdong have all been described as transitional forms between Chinese *H. erectus* and *H. sapiens*, whose affinities can be understood in the context of a braided stream network model of gene flow.[6–9] Here, we report a fossil human cranium that is characterized by a combination of large cranial capacity, short face, and small check bones as in *H. sapiens*, but also a low vault, strong browridges, large molars, and alveolar prognathism as in most archaic humans. Through phylogenetic and biogeographic analyses, we argue that this fossil is the most complete representative of a distinct Middle Pleistocene lineage, with a separate evolutionary history in East Asia.

The Harbin human fossil is represented by a single cranium (HBSM2018-000018(A), housed in the Geoscience Museum of Hebei GEO University, Shijiazhuang, Hebei Province, China), which was reportedly discovered in 1933 during construction work when a bridge (Dongjiang Bridge) was built over the Songhua River in Harbin city (Figure 1). Because of a long and confused history since the discovery (see the supplemental information), the exact site of the find is uncertain. We tested the concentrations of rare earth elements (REEs) and the Sr isotopic composition of the human fossil and a range of mammalian fossils collected from deposits of the Songhua River near the supposed locality (Dongjiang Bridge), and used non-destructive X-ray fluorescence analyses to examine the element distributions of these human and mammalian fossils. The results of our experiments show that element distributions and REE concentrations of the Harbin cranium and the mammalian fossils found near Dongjiang Bridge have similar distribution patterns.[10] The Sr isotopic composition of the Harbin cranium falls in the range of the local Middle Pleistocene-Early Holocene human and mammalian fossils.[10] We also directly dated the Harbin fossil cranium by the uranium-series disequilibrium (U-series) method. The results suggest a minimum age for the cranium of ~146 ka.[10] While these results cannot pin the Harbin cranium to an exact site and layer, they are consistent with the conclusion that the cranium is from the late Middle Pleistocene of the Harbin area.[10]

Article

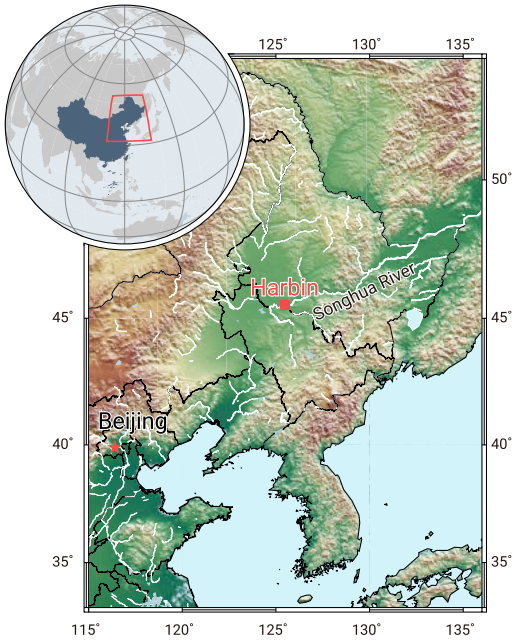

Figure 1. Geographic location of the Harbin cranium The red square indicates the Dongjiang Bridge in Harbin city.

RESULTS AND DISCUSSION

Morphology

The Harbin cranium is undistorted and almost intact, with the main losses being all but one tooth (the left M^2), and slight damage to the left zygomatic arch (Figure 2). It is massive in size, showing the largest values in our comparative fossil database (see the supplemental information) for measurements, such as maximum cranial length, nasio-occipital length, and supraorbital torus breadth, and the second largest values for measurements, such as biauricular breadth, frontal chord, zygomatic breadth, and biorbital breadth. Detailed morphological descriptions and comparisons of the cranium are given in the supplemental information, and are summarized below.

The cranial vault is voluminous (~1,420 mL capacity, measured using high-resolution computed tomography [CT] scanning and three-dimensional reconstruction of the endocranial cast). However, the braincase is clearly archaic, with a very wide supraorbital torus, base and palate, and a long and low shape in lateral view, with a receding frontal and evenly curved parietal contour. Nevertheless, it lacks both the angulated occipital with a strong transverse torus found in *H. erectus* and *H. heidelbergensis*/*H. rhodesiensis* crania, and the protruding occipital region with a central suprainiac fossa typical of Neanderthals. In posterior view the unkeeled cranium is widest in the supramastoid area, below which the well-developed mastoid processes slope inward. The temporals and parietals do not converge strongly as in *H. erectus* fossils, but there is no upper parietal expansion, as found in recent *H. sapiens*, nor the "en bombe" shape typical of Neanderthals. In lateral view the face is relatively low in height and retracted under the cranial vault, lacking the total anterior projection typical of *H. erectus* and *H. heidelbergensis*/*H. rhodesiensis*. The upper face and nasal aperture are very wide, but the zygomaxillary region is transversely flat and faces anteriorly, with a morphology like that of *H. sapiens*.

The combination of an archaic but large-brained cranial vault and a wide but *H. sapiens*-like face is striking, and is also found in the less-complete Middle Pleistocene Chinese fossils from Dali and Jinniushan, although they differ in details of morphology (see the supplemental information and Videos S1–S3). The less-complete Hualongdong cranium resembles Dali more closely in several respects, and some of its differences may be due to its immaturity, while the Xuchang and Maba partial crania appear more distinct (see the supplemental information for more details and comparative data).

Overall, the Harbin cranium shows an individual combination of traits, and probably represents a distinct species of *Homo* from other designated Middle-Late Pleistocene human taxa, such as *H. sapiens*, *H. neanderthalensis*, and *H. heidelbergensis*/*H. rhodesiensis*. Its enormous overall size sets it apart from nearly every other fossil but, in terms of cranial vault proportions, the braincase clearly overlaps in shape with those of other large-sized late archaic *Homo* species. However, the face, despite its enormous breadth dimensions, is relatively low in height and has an *H. sapiens* and *H. antecessor*-like zygomaxillary shape that is also found in the Middle Pleistocene Chinese fossils from Dali and Jinniushan. It is also hafted onto the braincase with reduced prognathism, as in recent humans. In its combination of traits, the Harbin cranium is more like fossils attributed to early *H. sapiens*, such as Jebel Irhoud 1 and Eliye Springs, than to later members of our lineage. Finally, and perhaps significantly, the morphology and large size of the surviving Harbin M^2 (Figure S1, mesiodistal length 13.6 mm and buccolingual width 16.6 mm) are matched most closely in the Late Pleistocene record by the permanent molars from Denisova Cave (Denisovan 4: $M^{2/3}$, mesiodistal length 13.1 mm, and buccolingual width 14.7 mm; Denisovan 8: M^3, mesiodistal length 14.3 mm, and buccolingual width 14.65 mm).[11,12]

Life reconstruction

The overall size, robustness, thick and strong supraorbital tori, large mastoid processes, and salient temporal lines of the Harbin cranium suggest that

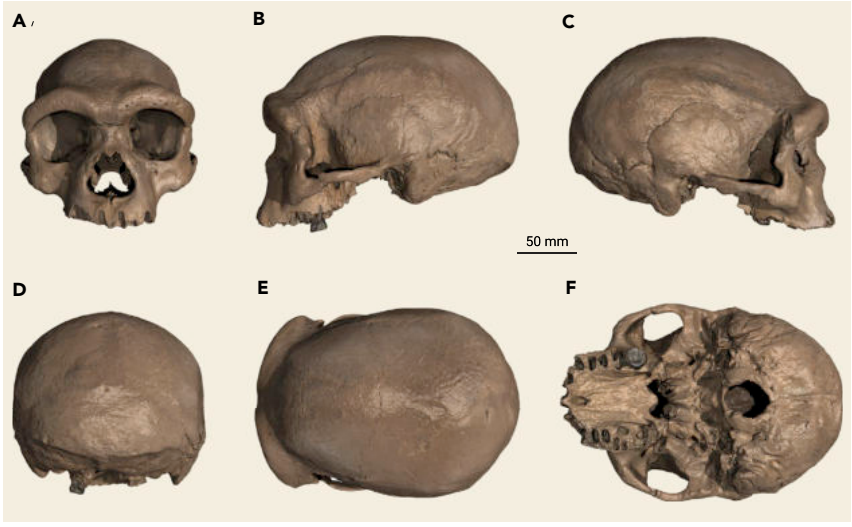

Figure 2. The Harbin cranium in standard views (A) Anterior view.
(B) Lateral view, left side.
(C) Lateral view, right side.
(D) Posterior view.
(E) Superior view.
(F) Inferior view. Scale bar, 50 mm.

it probably represents a male individual. The ectocranial sutures at the midlambdoid, lambdoid, obelion, anterior sagittal, superior sphenotemporal, incisive, anterior and posterior median palatine, and transverse palatine are all completely obliterated. The ectocranial sutures at bregma, midcoronal, pterion, sphenofrontal, and inferior sphenotemporal show significant closure. For the standard of *H. sapiens*, the ectocranial suture composite scores would suggest an old adult around 50 years old.[11,12] However, the tooth wear seems to suggest a younger age. The only preserved M^2 still has much enamel present, and dentine exposure is present on the protocone and paracone. The relatively complete ectocranial suture closure may be related to the robustness of the Harbin cranium. The large square eye sockets with strong supraorbital tori indicate deep eyes. The large and wide piriform aperture indicates a large and bulbous nose. The expanded paranasal region and relatively projecting middle face are matched with flat and short modern human-like cheek regions. Large incisor and canine tooth sockets indicate that the man probably had quite large front teeth and a broad mouth. The mandible of this individual is not known, but the phylogenetic analyses suggest that the Harbin cranium and the Xiahe mandible from Gansu Province of China form a sister group. The M^2 size of the Harbin cranium matches the tooth size of the Xiahe mandible. It is reasonable to deduce that the Harbin cranium probably matches a mandible as robust as the Xiahe mandible and without a chin. It is hard to reconstruct the skin tone and hair color of the Harbin individual without genetic information, but available genetic data suggest that Neanderthals, Denisovans, and early *H. sapiens* generally had relatively dark skin, hair, and eye color. Considering the high latitude of the provenance of the Harbin cranium, we have chosen to give the reconstruction only a medium-dark skin color (Figure 3).

Phylogenetic position of the Harbin cranium

Our extensive phylogenetic analyses based on parsimony criteria[13] and Bayesian inference[14–17] firstly support the monophyly of Neanderthals and the monophyly of *H. sapiens* (Figures 4 and S19–S23). The Irhoud fossils from Morocco form the most basal operational taxonomic unit (OTU) of the *H. sapiens* clade, and the Sima de los Huesos crania from Spain form the most basal OTU of the Neanderthal clade, in line with other current interpretations.[18–20] The Harbin cranium and Xiahe mandible form a sister group, and they, plus the Dali, Hualongdong, Jinniushan specimens, the European *H. antecessor* partial cranium, the African Eliye Springs cranium, and Rabat palate, form a monophyletic group. This clade forms the sister group of the similarly monophyletic *H. sapiens* clade. The specimens traditionally grouped in *H. heidelbergensis*/*H. rhodesiensis* do not constitute a monophyletic group and the Asian and African *H. erectus* specimens similarly form a paraphyletic group. When backbone constraints are used to reflect the results from palaeoproteomic and ancient DNA research by forcing the Xiahe mandible as the sister group of Neanderthals[21] and *H. antecessor* outside of the *H. sapiens*-Neanderthal clade,[22] Chinese late Middle Pleistocene humans, including the Harbin cranium, form a monophyletic clade as the sister group of Neanderthals (Figure S20). Both most parsimonious and backbone partially constrained phylogenetic trees support the monophyly of the group, including Dali, Jinniushan, Hualongdong, Xiahe, and Harbin.

Some researchers have proposed that all Middle Pleistocene hominins belong to a single lineage leading to modern humans, with Asian Middle Pleistocene hominins, such as Dali and Hualongdong, suggested as transitional forms between Asian *H. erectus* and Asian *H. sapiens* specimens.[6,9,24] Some other researchers have recognized these Asian hominins as part of the *H. heidelbergensis*/*H. rhodesiensis* hypodigm.[25–27] A previous analysis based on overall similarity showed differences between Dali-Maba and the *H. heidelbergensis* hypodigm, and the potential connection between Dali-Maba and African early *H. sapiens*.[28] Our analyses suggest that the Harbin cranium, together with Dali, Jinniushan, Hualongdong, and Xiahe, is not a part of the African and European *H. heidelbergensis*/*H. rhodesiensis* clade, but is the sister group of *H. sapiens* (see also the backbone partially constrained parsimony analysis in the supplemental information). The sister relationship between Harbin and Xiahe, as identified by Bayesian inference (but not parsimony analysis, see the supplemental information), is particularly interesting. The Xiahe mandible shows some proteomic features of the Denisovans,[21] who were informally called "*Homo sapiens altaiensis*" or "*Homo altaiensis*,"[12,29] and sediments from Baishiya Cave have yielded Denisovan mtDNA.[30] The Harbin M^2 also matches the known permanent Denisovan molars in size and root morphology, and, ever since the discovery of Denisovans, Asian Middle Pleistocene hominins, such as Dali, Jinniushan, and Xujiayao, have been suspected to represent an East Asian population of the Denisovans.[31] More mandibular specimens for the Harbin population or cranial specimens corresponding to the Xiahe mandible will test how close the Harbin and Xiahe humans are morphologically, while new genetic material will test the relationship of these populations to each other and to the Denisovans.

The results of the Bayesian tip-dating analyses suggest that the Harbin and Xiahe fossils shared a common ancestor ~188 ka (397–155 ka), and the clade, including the Harbin cranium and *H. sapiens* shared a common ancestor at ~949 ka (1,041.41–875.25 ka). The Neanderthal-*H. sapiens* divergence time in our analysis was ~1,007 ka (1,114–919 ka). This estimation falls in the range based on mtDNAs for the split between the basal Neanderthal (Sima de los Huesos) and the *H. sapiens* lineage,[20] but is much older than the estimation based on nuclear DNAs for the splits between the Neanderthal and *H. sapiens* lineages.[32–34] However, it is possible that this younger estimated divergence date is an artifact of statistical averaging between "super-archaic" and "recent gene flow" events.[35] The common ancestor of the *H. sapiens* OTUs included in our analysis is as old as ~770 ka (922–622 ka), suggesting that the *H. sapiens* clade has a much deeper origin time than previously estimated. The Eurasian *H. sapiens* OTUs share a common ancestor ~416 ka (534–305 ka) old. Outside of Africa, however, the earliest known *H. sapiens* fossil is only ~210 ka.[36]

There is a large time gap between the hypothetical common ancestor of Eurasian *H. sapiens* and the actual fossil record, from the Bayesian tip-dating analysis. One plausible hypothesis is that the ancestral population of Eurasian *H. sapiens* may have diversified in Africa for many millennia before they dispersed into Eurasia. Genetic studies on ancient DNA suggest that the initial genetic exchanges between Neanderthals and *H. sapiens* occurred between 468 and 219 ka,[33] or between ~370 and 100 ka,[34] and the introgression may

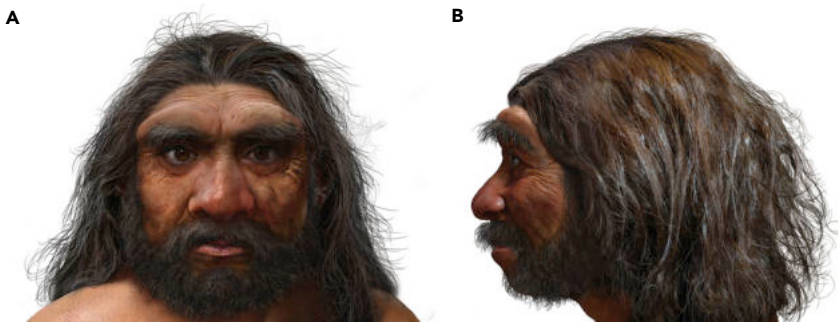

Figure 3. Life reconstruction of the Harbin cranium
(A) Anterior view.
(B) Lateral view, left side.