Partitioning the genomic journey to becoming *Homo sapiens*

Luca Pagani[1,2], Riccardo Bertazzon[1], Vasili Pankratov[2,3], Leonardo Vallini[4], Davide Marnetto[5], Maurizio Esposito[6+], Ilaria Granata[6], En En Teo[7], Aswini Leela Loganathan[7], Pille Hallast[8], Charles Lee[8], Qasim Ayub[7], Massimo Mezzavilla[1]

Affiliations
1 Department of Biology, University of Padova, Via Ugo Bassi 58/B, 35131, Padova, Italy
2 Institute of Genomics, University of Tartu, Riia 23/b, 51010 Tartu, Estonia
3 Bioinformatics Research Centre, Aarhus University, Universitetsbyen 81, 3., Building 1872, DK-8000 Aarhus C Denmark
4 Institute of Organismic and Molecular Evolution, Palaeogenetics Group, Johannes Gutenberg University, Mainz 55128, Germany
5 Department of Neurosciences, University of Turin, 10126, Turin, Italy
6 Istituto di Calcolo e Reti ad alta prestazione, CNR, 80131, Naples, Italy
7 Monash University Malaysia Genomics Facility, School of Science, Monash University, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia
8 The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

Corresponding authors: luca.pagani@unipd.it, massimo.mezzavilla@unipd.it

Abstract
What makes us human? *Homo sapiens* diverged from its ancestors in fundamental ways, reflected in recent genomic acquisitions like the PAR2-Y chromosome translocation. Here we show that despite morphological and cultural differences between modern and archaic humans, these human groups share these recent acquisitions. Our modern lineage shows recent functional variants in only 56 genes, of which 24 are linked to brain functions and skull morphology. Nevertheless, these acquisitions failed to introgress into Neanderthals when archaic and modern populations admixed after 350 kya, suggesting their exclusive link to the modern human niche or that Neanderthal's small population size hindered their spread. Taken together, our results point to a scenario where Modern and Archaic should be regarded as populations of an otherwise common human species, which independently accumulated mutations and cultural innovations.

Introduction:

The acquisition of derived traits has punctuated the long evolutionary path followed by the ancestors of all humans, as witnessed by the fossil record. This constant accumulation of physical, cognitive and cultural innovations is what we collectively refer to when we think of us as "humans". The debate about the timing and dynamics that eventually led to the differentiation of modern humans from previous and parallel lineages remains unresolved (*1*). Recently, it has been proposed that around 900 thousand years ago (kya), a severe population bottleneck, perhaps promoted by a sympatric founder event, caused a profound reduction in the number of living ancestors of all present-day humans (*2*). Despite other methods failing to reproduce such a signal (*3*)(*4*), the 900 kya period may have coincided with a drastic change at the genomic level, such as the chromosomal fusion that yielded chromosome 2 in our current karyotype (*5*). In this light, because of meiotic incompatibility, the bottleneck could have been due solely to reproductive barriers with other members of the broader hominin population inhabiting Africa at the time. Still, it might also have been paralleled by novel mutations that increased the fitness of this new, yet small, population of hominids. In both scenarios, such events must have affected not only the ancestors of modern humans, but those of Neanderthals and Denisova alike, since these extinct groups separated from our lineage between 650 and 500 kya (*6*). It follows that this founding period cannot be seen as the moment that defined the birth of our species unless Neanderthals and Denisova were to be included within the same broader label.

With this work, we set out to exploit crucial events during the past 1 million years of our evolution to identify human-specific genomic loci (*7*), to ultimately shed light on the long-lasting question of what makes us human. The three events that enabled us to partition the genome are: Event 1, the aforementioned bottleneck (after 1000 kya (*2*)); Event 2, the separation between the ancestors of modern and archaic humans (650 kya (*6*)) and Event 3, the interaction between African human lineages and Neanderthals at around 350 kya, as demonstrated by autosomal and uniparental evidence (*8*, *9*)(*10*)(*11*)(*12*). This latter event is in the opposite direction and took place almost 300 thousand years before the more widely known human-Neanderthal introgression(*13*) that followed the Out of Africa expansion of the modern human population. We here leverage this deep admixture event, as it provides a time checkpoint to assess which genetic variants accumulated along the modern human lineage (e.g., after Event 2) were already present and displayed a high enough adaptive potential to enter the Neanderthal gene pool as a consequence of Event 3.

Results

Newly observed chromosomal rearrangement

We first focused our attention on a karyotype-level human-defining feature: the pseudo-autosomal region 2 (PAR2) of the human chromosome Y, which together with the chromosomal fusion that generated the human chromosome 2, is a genomic rearrangement (X to Y translocation) known to have emerged after the separation of our lineage from the chimpanzee lineage (*14*). Together with PAR1, PAR2 enables homologous pairing of the X and Y sex chromosomes during meiosis in human males. Firstly, we observed that the male Denisova genome displayed a sequencing depth on regions matching the X chromosome PAR2 higher than the rest of the chromosome (**Fig 1 A**). This suggests the presence of PAR2 in chromosome Y of the ancestors of modern and archaic humans, hence showing the translocation to be older than Event 2 unless recurrent acquisition is contemplated (*15*). Similar results obtained for a male Neanderthal genome (**Fig 1 A**) are provided as a control, given that Neanderthal Y chromosome replacement is among the known consequences of Event 3, and the presence of PAR2 in these archaic populations is therefore amply expected. Further positive control for our analysis is provided by the higher sequencing depth displayed by all modern and archaic humans at PAR1 (**Fig 1 A**) when compared to the rest of chromosome X, being the presence of PAR1 along all modern and archaic human Y chromosomes undisputable.

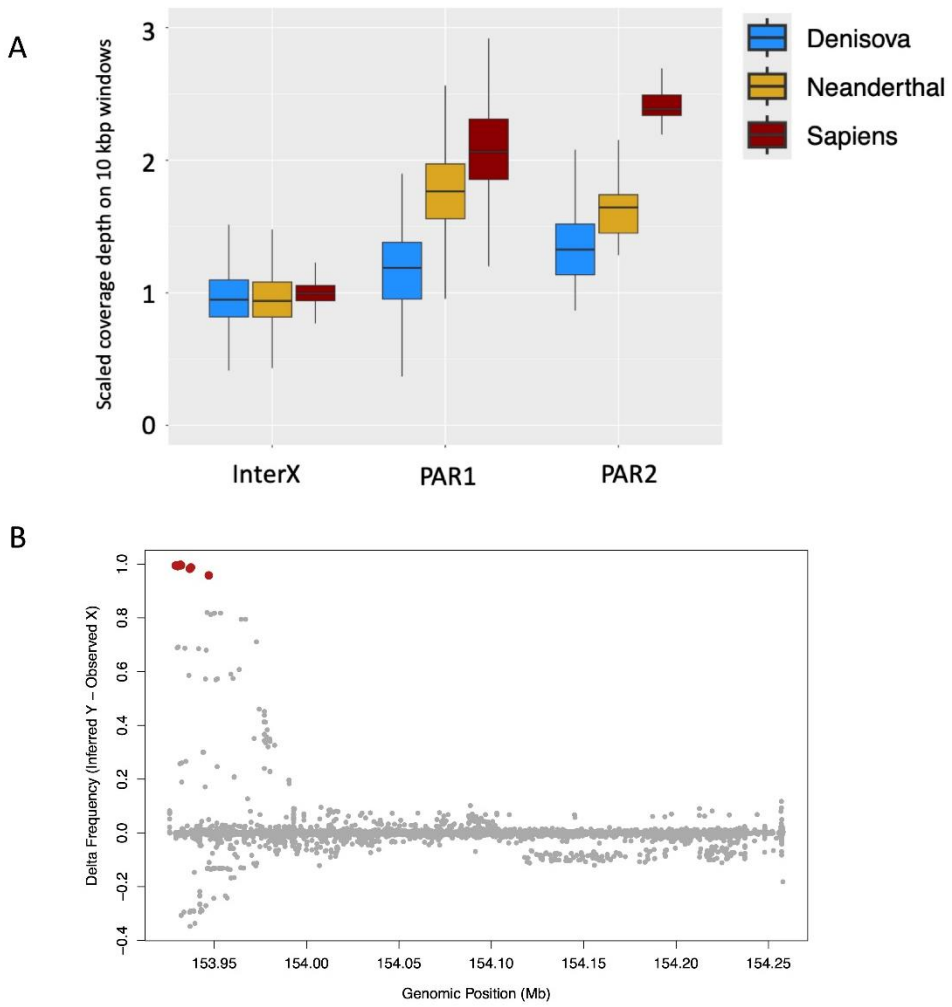


Fig 1. PAR2 coverage depth and Y-specific mutations. A) Spy Neanderthal, Denisova8 and Modern human (SS6004480 individual) sequencing depth averaged over windows of 10 kb from the central part of chromosome X, X-PAR1 and X-PAR2. The sequencing depths displayed are z-transformed based on the depth distribution of the internal part of the relevant chromosome X (InterX). In all modern and archaic genomes, both PAR regions show a statistically significant (Mann-Whitney p-value<1.6×10^{-6}) higher depth compared to their InterX, consistent with the presence of both a Y and an X chromosome PAR sequences in the sampled individuals. The Y reads are expected to map on the X chromosome because of the high sequence homology at PAR2 sites. **B**) Red dots mark the location of the eleven male-specific variants along the ~330 kb long PAR2 sequence when reads from both chromosomes of modern humans are collapsed together. (Red SNP positions: 153929157; 153929840; 153929878; 153930185; 153931728; 153931814; 153932007; 153932008; 153936764; 153937415; 153947049, mapped onto X-PAR2 with T2T-CHM13v2.0 coordinates, with the PAR2 starting point set at 153925834).

We then inspected the PAR2 mutational landscape within chromosome X of modern humans to characterize the substitutions that occurred since the translocation event onto chromosome Y, intending to provide molecular dating for said event. We found 11 single nucleotide variants (SNVs) present as heterozygous in male PAR2 chromosome X regions and virtually absent in females, hence likely attested as mismapped, male-specific mutations that occurred

after the common ancestor of all the Y-PAR2 sequences separated from the rest of the X-PAR2 gene pool **(Figure 1B, red dots)**. These were confirmed in Y chromosomes sequenced using long-read platforms (*16*). These mutations span a 21,215 bp region at the 5' end of the Y-PAR2. They are followed towards the 3' telomeric end of the region by additional mutations with a less polarized frequency difference between males and females, as expected under a scenario of increasing recombination events between the X and Y-PAR2 regions as a function of the genetic distance from the translocation point. As an upper limit for the PAR2 translocation, we estimated the time (T) since the ancestor of all Y-PAR2 sequences separated from the X-PAR2 gene pool being 518 kya, using the Y chromosome mutation rate of 3×10^{-8} mut/bp/gen (*17*). This estimate refers to the time elapsed between the translocation event and the most recent common ancestor of all modern human Y chromosomes, currently dated at 338 kya (*18*). Thus the X->Y PAR2 translocation event must have therefore occurred not earlier than 856 kya (95% CI: 633-1253 kya), coinciding with the putative population bottleneck (Even 1) described above and predating the separation of modern and archaic human lineages.

Partitioning the genomic landscape of modern and archaic

Next we focussed on the evolutionary events that occurred after the separation between the ancestors of modern and archaic humans that occurred 650 kya (Event 2). We characterized the genomic coalescence rates distribution of the genome, using 50 kb genomic windows (Figure S1, Table S1) and focussed on regions that showed coalescence events younger than 650 kya in all modern humans (Human650 regions, see Methods) after controlling for the local reliability of the molecular clock. These windows were chosen for their potential of having experienced selective events in modern humans after Event 2, hence providing a clue to what derived features best characterize our lineage. The advantage of using coalescence events within modern humans as a criterion for partitioning the genome is evident in Figure S1, where all the genomic windows displaying at least one missense variant with derived frequency >=98% in modern humans and fixed derived in both Altai Neanderthal and Denisova (*19*), only a few are indeed tagging recent events along the modern human lineage and can be deemed as highly informative in defining what makes us human. In addition, we were also interested in how many Human650 regions were transferred to the Neanderthal gene pool (here represented by the Altai Neanderthal) as a consequence of Event 3, hence displaying versatile and high adaptive potential also in this population of archaic humans. The date of Event 3 (~350 kya, **Fig. 2**) falls indeed within a remarkable period for the evolution of our species in Africa. A growing body of archaeological

(*20*)(*21*) and genetic (*6*, *22*, *23*) evidence points to ~300 kya as the time when our species experienced a sudden growth and subsequent separation into the main population branches that we still observe today. Although the first undisputable fossil remains attributed to modern humans date to at least 50 kya later (*24–26*), the window of time of 300-250 kya is currently seen as the defining moment for the emergence of our species in Africa. It follows that Event 3, taking place just before or around such a crucial time in our evolutionary history, can be instrumental in partitioning genomic regions (with functional variants therein) into the following three sets. 1) Regions with variants that experienced a sweep in modern humans between Event 2 and Event 3 and that did not enter the Neanderthal gene pool despite their potential to do so; 2) Regions with variants that experienced a sweep in modern humans between Event 2 and Event 3 and that entered the Neanderthal gene pool (**Fig. 2**) and 3) Regions with variants that experienced a sweep in modern humans after Event 3, and are characterizing the final evolutionary steps before our lineage split into the various present-day modern human population.

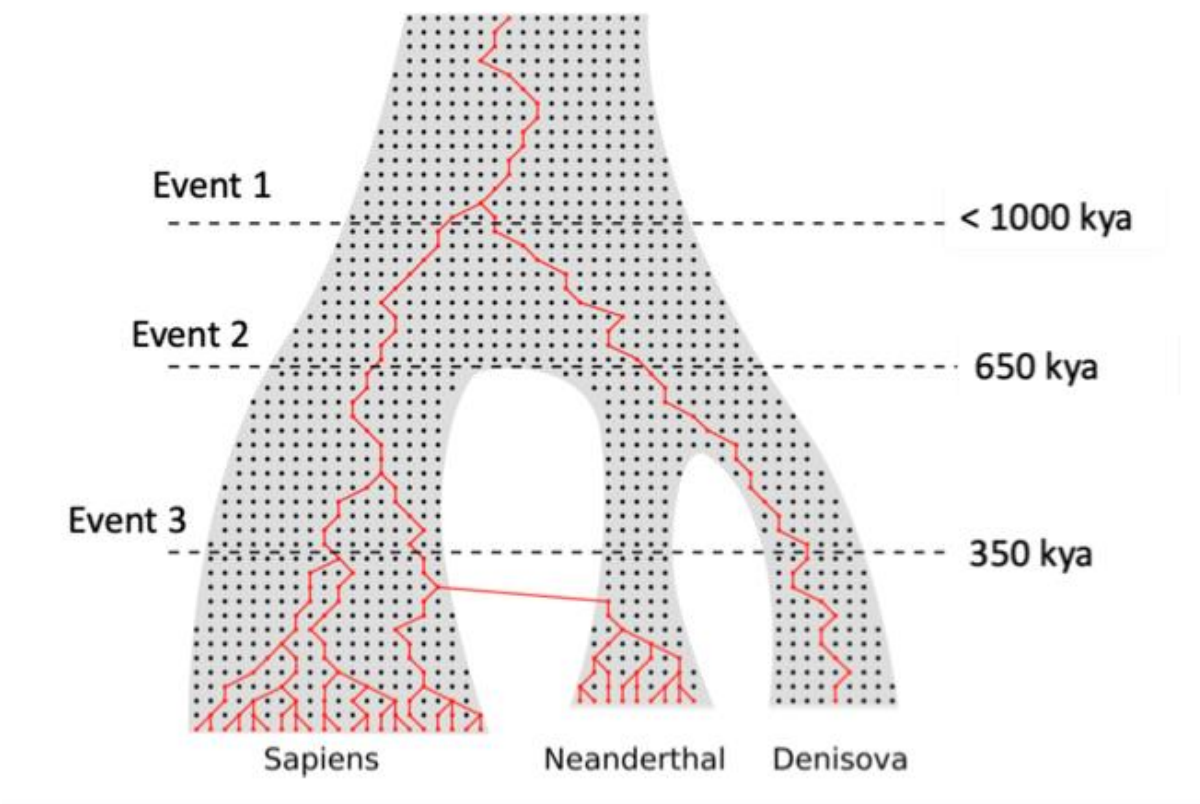


Fig 2. Events Partitioning. Defining Event 1 (the period preceding the separation between modern and archaic lineages, characterized by a putative bottleneck at around 1000 kya), Event 2 (the separation of modern and archaic lineages, at around 650 kya) and Event 3 (the contribution of modern human lineages to the gene pool of Neanderthals, after 350 kya), along the population tree of modern and archaic humans. The gene tree depicted in

red exemplifies a Human650 region that entered and was also fixed within the Neanderthal gene pool as a consequence of Event 3.

As a consequence of Event 3, we show that Altai Neanderthal displays more regions with a coalescence event <= 650 kya with the human reference genome than Denisova **(Figure S2,** blue line). However, this pattern may also be affected by the widespread introgression of Neanderthal alleles into Eurasian genomes (*13*), including into a non-negligible portion of the hg19 reference genome used for these analyses (as also visible in the top left quadrant of Figure S1). However, an even more striking difference, which can only be explained as a consequence of Event 3, is observed when looking only within Human650 regions (**Figure S2,** green line), maxing when the Altai Neanderthal-reference human genome coalescences are around the time of Event 3. Furthermore, within the Human650 regions, the number of regions with a Neanderthal-reference human genome coalescence <= 650 kya (Neanderthal Recent) (22 regions) is significantly higher (Fisher exact test, p-value 0.01) than the number of regions (7 regions) expected if Human650 and Neanderthal Recent were two independent sets of regions. This result points to Event 3 to have indeed caused a significant number of Human650 regions to have entered the Neanderthal gene pool.
We then tried to single out this genomic signal by sifting through all Human650 regions looking for the ones showing 1) a human reference-Neanderthal coalescence <=650kya, 2) at least one site with derived allele frequency >=98% in modern humans, homozygous derived in Altai Neanderthal and homozygous ancestral in Denisova and 3) as a further stringency filter, no sites with derived allele frequency >=98% in modern humans, homozygous derived in Denisova and homozygous ancestral in Altai Neanderthal. We dubbed these regions “what made them humans” (WMTH), being Human650 regions that contributed to their (Neanderthal’s) gene pool. We found ten WMTH regions (**Table 1**, first ten rows, **Figure S1**, orange dots) that matched all three criteria and an additional seven that matched just the first two criteria (**Table 1**, bottom three rows marked by an asterisk, **Figure S1**, purple diamonds). Many such regions land on chromosome X, since its smaller population size causes neutral coalescence to happen earlier than the ones of the autosomes. Further validation of the inferred TMRCA of the WMTH regions was performed with Relate (**Supplementary Information, Figure S3**).

Table 1. "What made them humans" (WMTH) regions. 17 Human650 regions that contributed to the gene pool of Neanderthals as a consequence of Event 3. The last seven regions, highlighted with an asterisk, display at least one derived site shared between modern humans and Denisova (D-Hder) but not Neanderthal and hence do not meet our full list of inclusion criteria. The abundance of regions landing on chromosome X is explained by the smaller effective population size of its gene pool, which in turn affects the distribution of expected neutral coalescences.

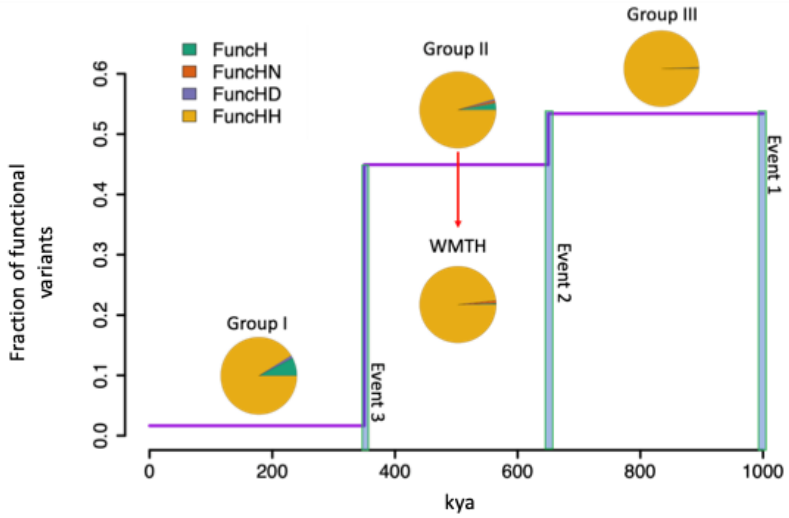
Chr	Start(kb)	End(kb)	NeaTMRCA	DenTMRCA	HumTMRCA	Genes	N-Hder(R,C)	D-Hder(R,C)	Hder(R,C)
5	45600	45650	551834	743851	436800	HCN1	1(0,0)	0(0,0)	4(0,0)
6	27900	27950	239976	3791166	484800	OR2W6P;RPLP2P1;OR2W4P;OR2B6	47(0,0)	0(0,0)	0(0,0)
7	30650	30700	143954	935888	484800	GARS;CRHR2	1(0,0)	0(0,0)	1(0,0)
10	103600	103650	623763	815804	460800	KCNIP2;C10orf76	5(**1**,0)	0(0,0)	0(0,0)
10	103650	103700	599820	791762	561600	C10orf76	8(0,0)	0(0,0)	4(0,0)
12	123550	123600	575896	1127774	643200	RP11-197N18.7;PITPNM2	4(0,0)	0(0,0)	1(0,0)
16	22000	22050	191946	1031670	532800	CTD-2649C14.1;PDZD9;C16orf52	12(0,0)	0(0,0)	0(0,0)
16	22050	22100	311919	1223706	504000	RP11-101E7.2;C16orf52	8(0,0)	0(0,0)	0(0,0)
X	25350	25400	479952	1007798	513600	-	4(0,0)	0(0,0)	7(0,0)
X	62950	63000	623925	671906	292800	ARHGEF9	1(0,0)	0(0,0)	1(0,0)
3*	51650	51700	647857	791842	552000	RAD54L2;TEX264	2(0,0)	**2(0,0)**	2(0,1)
3*	99850	99900	527873	1487613	600000	RP11-15N24.4;CMSS1	13(0,**1**)	**1(0,0)**	0(0,0)
11*	118300	118350	599796	863862	532800	RN7SL86P;ATP5L;RP11-770J1.4;KMT2A	1(0,0)	**1(0,0)**	0(0,0)
X*	81400	81450	623900	2135402	576000	-	26(0,0)	**2(0,0)**	1(0,0)
X*	81500	81550	503859	1511426	547200	-	10(0,0)	**1(0,0)**	7(0,0)
X*	96650	96700	647832	1319525	374400	DIAPH2	10(0,0)	**2(0,0)**	1(0,0)
X*	119650	119700	623888	1439741	600000	CUL4B	14(0,0)	**1(0,0)**	0(0,0)

Abbreviations: *chr: chromosome; NeaTMRCA: time to the most recent common ancestor (TMRCA) between the Altai Neanderthal sequence and the hg19 human reference; DenTMRCA: TMRCA between the Denisova sequence and the hg19 human reference; HumTMRCA: top 90% deepest TMRCA between a panel of 50 African human genomes and the hg19 human reference; N-Hder: sites that are homozygous derived in Altai Neanderthal and have a derived allele frequency of at least 98% in the 1000 Genomes Samples individuals and are homozygous ancestral in Denisova; R,C: of these, how many sites are annotated with Regulatory or Coding consequences (See Table S2 for a full list of these sites); D-Hder: sites that are homozygous derived in Denisova and have a derived allele frequency of at least 98% in the 1000 Genomes Samples individuals and are homozygous ancestral in Altai Neanderthal. All genomic coordinates are in build37 (hg19).*

Functional characterization

Finally we focused on the functional consequences of the derived sites included within the 50 kb regions we analyzed so far and that displayed TMRCA within modern humans <=1000 kya (e.g., after Event 1). Of all the derived alleles shared between Neanderthal and modern humans in the WMTH regions, only two(**Table 1**, in bold) were predicted to have regulatory or coding consequences, of which one fell in a region not meeting our 3rd inclusion criterion. We asked whether this

apparent scarcity of functional mutations is expected when looking at the broader set of Human650 regions with a human coalescence older than Event 3.
We considered, as a reference set, the functional (coding or regulatory) sites where at least 98% of modern humans included in Human650 regions have a human TMRCA older than Event 3 (Group II in **Fig 3**) and compared against the test set represented by the WMTH regions (**Table 1** and **Fig 3**), a strict subset of Group II. Among all sites derived in at least 98% of modern humans, WMTH regions include a similar proportion of functional sites (4.35%) compared to Group II regions (4.28%). In addition, we compared functional sites where modern humans and Neanderthals but not Denisovans share the derived allele (FuncHN in **Fig 3**) within WMTH windows (i.e., the functional sites that entered the Neanderthal pool during Event 3) against the functional sites where just modern humans, or modern humans and Neanderthals or modern humans and Denisovans share the derived allele (e.g., all the sites that were derived in modern humans and that could enter the Neanderthal pool during Event 3, FuncH, FuncHN and FuncHD in **Fig 3**). For these sites, the fraction of WMTH windows displaying at least one functional site is not enriched when compared with Group II (2 regions with at least 1 FuncHN out of 11 WMTH regions, 18.2%, compared to 60 out of 242, 24.7%, in Group II. Fisher exact test p-value=1). Conversely, looking at the total number of variants, the Neanderthal gene pool received significantly fewer modern human-derived functional sites than it could have, given the ones available in Group II (2 FuncHN out of 177 FuncHN, FuncHD, FuncH or FuncHH in WMTH regions, or 1.13%, compared to 112 FuncH, FuncHD or FuncHN out of 2465 FuncH, FuncHD, FuncHN or FuncHH in GroupII, or 4.5%. Fisher exact test p-value=0.03). It follows that, in contrast, WMTH regions showed an excess of derived sites shared by all archaic and modern humans (FuncHH, in gold in Figure 3), hence having occurred along lineage shared by all archaic and modern humans (here dubbed the *Homo heidelbergensis*, *HH* lineage). As a summary, WMTH compared to Group II regions are not enriched in regions with at least one functional modern human-derived site that entered the Neanderthal pool and, as a whole, are depleted of modern human-derived functional sites shared with Altai Neanderthal.



Groups	Sapiens	Neanderthal	Denisova	FuncH	FuncHN	FuncHD	FuncHH	Fraction Total Variants
I	<=350	-	>650	0.08	0.00	0.01	0.91	0.02
WMTH	>350 <=650	<=650	>650	0.01	0.01	0.00	0.98	0.03
II	>350 <=650	-	>650	0.03	0.01	0.01	0.95	0.45
III	>650 <=1000	>650 <=1000	>650 <=1000	0.00	0.00	0.00	0.99	0.53

Fig 3. Prevalence of lineage-specific functional variants according to the modern human TMRCA (shown on the x-axis) of the 50 kb window. The purple line represents the relative fraction of fixed-derived human functional variants exhibited by the genomic windows showing modern human TMRCA <= 1000 kya. Within each time period, a pie chart represents the breakdown of functional derived sites into FuncH (shared by >98% of modern humans but no archaics; FuncHN, shared by >=98% of modern humans, homozygous derived in Altai Neanderthal and homozygous ancestral in Denisova; FuncHD, shared by >=98% of modern humans, homozygous derived in Denisova and homozygous ancestral in Altai Neanderthal; FuncHH, shared by >=98% of modern humans and homozygous derived in both Denisova and Altai Neanderthal. Most functional variants that fall in a Human650 region with a TMRCA in modern humans <= Event 3 (Group I) are shared with all three lineages and only a small portion is unique to modern humans (FuncH). Among the variants within a Human650 with a TMRCA in modern humans between Event 3 and Event 2 (Group II), we observe that only a negligible fraction is FuncHN. This fraction is significantly reduced in the WMTH subset (i.e., variants that entered the Neanderthal gene pool during Event 3). This latter set (WMTH), in turn, shows an enrichment for FuncHH variants. As expected, regions showing a human TMRCA between Event 1 and Event 2 (Group III) are almost entirely constituted by FuncHH variants, given that fixation events in that time transect were, in the majority of cases, shared across archaic and modern humans.

Consistently with what was observed for human uniparental markers, which were shown to have permeated the gene pool of Neanderthals because of the accumulation of deleterious variants on the latter (*12*), our results based on autosomal and chromosome X data confirm that Event 3 consequences on the Neanderthal gene pool can be best described as a reintroduction of variants that were lost in Neanderthal due to the strong genetic drift caused by their reduced

effective population size, rather than as an introduction of novel, modern human-specific derived sites.

An interpretation of this emerging picture may imply that the vast majority of modern human-derived variants that experienced a human sweep between Event 2 and Event 3 encompassed an adaptive potential that contributed perhaps to the biological characterization of our lineage but that, on the other hand, was either not sufficiently overwhelming to swarm into the Neanderthal gene pool or its fitness was linked to specific environmental or social features of the modern human niche.

To better understand the biological significance of the modern human-derived variants enclosed within the whole Human650 regions, we generated protein-protein interaction networks using only the genes falling within a Human650 region that displayed a coding or regulatory human-derived variant, here defined as fixed derived in at least 98% of the 1000 Genomes Project individuals, present or not in the Neanderthal genome (**Fig 4A**, **Table S2**, **Fig S1** dark blue dots). The resulting list of 56 proteins, listed in **Fig 4**, as a whole, are enriched for a compelling list of phenotypes: the first three (Cognitive impairment, Intellectual disability, Abnormality of higher mental function) are related to higher-level brain function, while the next three (Abnormality of the face, Abnormality of the head, Neurodevelopmental abnormality) are related to developmental anomalies (**Fig 4 B**) (*27*). The 24 genes linked to the enrichment of one or more phenotypes listed in **Fig 4B** are highlighted in orange in the network above. Despite the phenotypes connected with the reported enrichment appearing to be pivotal aspects to becoming modern humans, only two of these genes (*CMSS1* and *C10orf76*, circled in blue) were passed on to Neanderthals during Event 3 (WMTH regions) and only *CMSS1* is involved in the enrichment and associated with a moderate CADD score (Combined annotated dependent depletion, a measure of the deleteriousness of a mutation) (*28*) (CADD=3.27, **Table S2**). On the other hand, seven genes (*ATRX*, *SPAG5*, *ARL13B*, *TMEM139*, *ALDOC*, *STARD7* and *ENOX2*) experienced a human sweep only after Event 3 (**Fig S1**, cyan dots), and their functional variants were potentially not available to Neanderthals during Event 3. Notably, SPAG5, ATRX and ARL13B interact and expand the network of proteins for which derived functional variants were already available at Event 3 (**Fig 4B**). A tempting explanation may involve the role of these latter three genes as the “detonator” to unlock the biological adaptive potential of other genes (namely *MAGT1*, *RAD54L*, *SPAG17*, *DNAH1*, *SPOP*, *HDAC*6 and *TCTN1*) that were not sufficient on their own to trigger a sweep in the Neanderthal gene pool after Event 3.

Importantly, *SPAG5* features in several published lists as the gene with the highest number of missense substitutions between modern humans and Neanderthal, and its function has been previously associated with the mitotic spindle (*29*)(*29*, *30*).

Additional evidence in support of the biological relevance of the modern human-derived allele for the reported SNPs (**Table S2**) is provided by the observation that the vast majority of the functional SNPs falling in the Human650 regions display 0% frequency of the ancestral and archaic (Neanderthal or Denisova) allele even in populations such as Papuan Highlanders (*31*, *32*), who experienced multiple introgressions from both populations of archaic humans (**Table S2**).

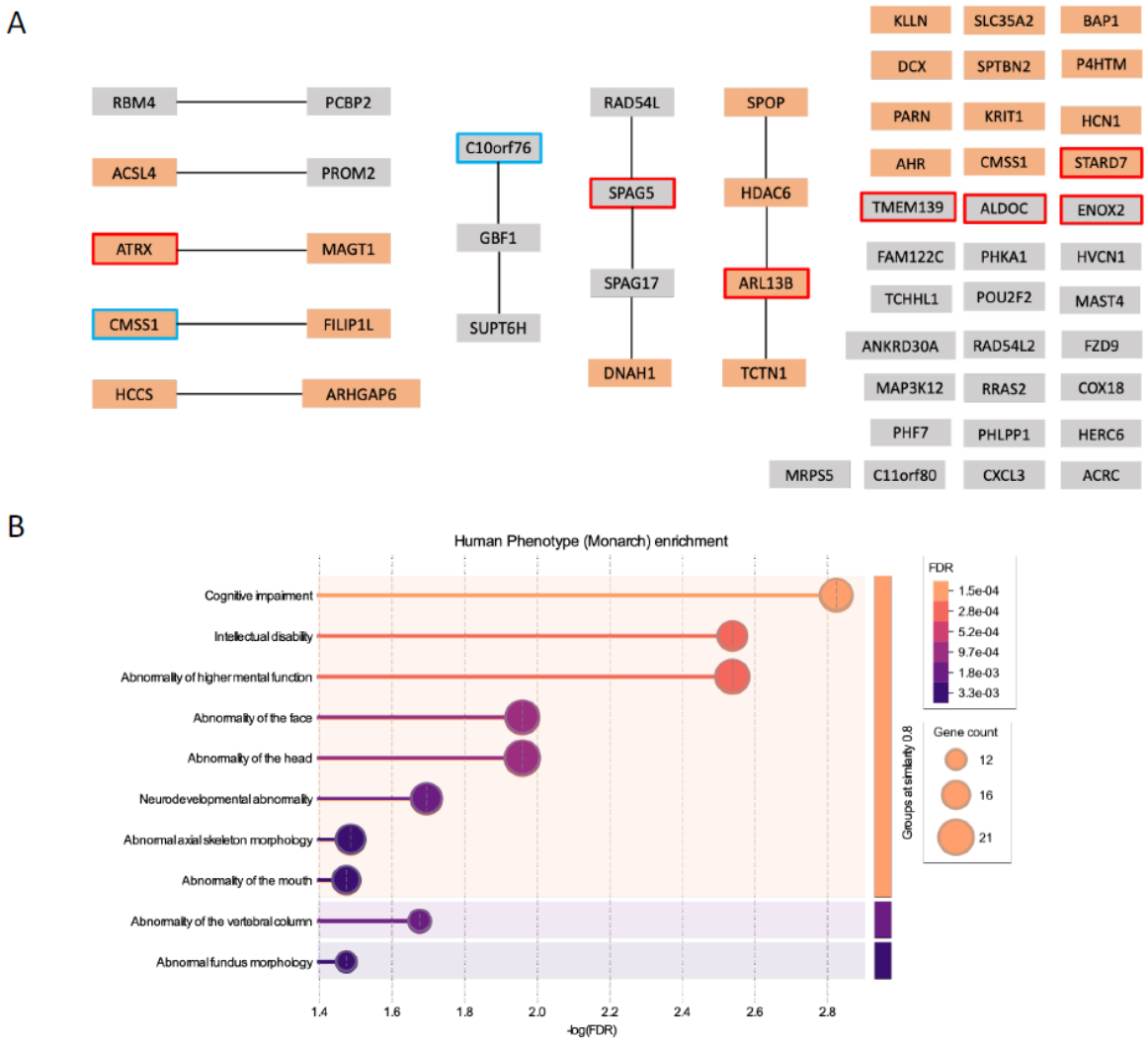


Fig 4. Functional enrichment analyses. A) Interaction network obtained from the STRING database using the 56 genes listed in Table S2 (*ACSL4, AHR, ALDOC, ANKRD30A, ARHGAP6, ARL13B, ARMH3, ATRX, BAP1, C11orf80, CMSS1, COX18, CXCL3, DCX, DNAH1, ENOX2, FAM122C, FILIP1L, FZD9, GBF1, GCNA, GLUD2, HCCS, HCN1, HDAC6, HERC6, HVCN1, KLLN, KRIT1, MAGT1, MAP3K12, MAST4, MRPS5, P4HTM, PARN, PCBP2, PHF7, PHKA1, PHLPP1, POU2F2, PROM2, RAD54L, RAD54L2, RBM4, RRAS2, SLC35A2, SPAG17, SPAG5, SPOP, SPTBN2, STARD7, SUPT6H, TCHH, TCHHL1, TCTN1, TMEM139*). Proteins circled in red represent genes with a recent coalescence in humans (after Event 3), and proteins circled in blue represent genes defined as WMTH in our manuscript. The 24 orange boxes represent proteins that contribute to at least one of the phenotype enrichments described in panel B.
B) Functional enrichments in our network, using the STRING database we selected only the phenotypes with an FDR<=0.05 present in the Monarch phenotypes initiative. The phenotype with the lowest FDR and highest enrichment is cognitive impairment.

Discussion

Putting together the various lines of genetic evidence accumulated so far, we can partition the derived features that concur in defining us as "modern humans" in a time- and evolutionarily-informed way. Of these features, the PAR2 region of human chromosome Y translocated from chromosome X at around 895 kya, during a period that was marked by other chromosomal rearrangements and demographic events experienced by the ancestors of all modern and archaic humans, after Event 1. After the separation of our lineage from archaic humans (i.e., between Event 2 and Event 3, or 650-350 kya) we observe a phase where putative modern human-specific sweeps (Human650 regions), encompassing coding and regulatory variants, occur in genes involved in higher-level brain function and morphological development. We also stress that no reproductive or fertility phenotype was associated with the reported genes, pointing to reproductive barriers not being the focus of this human evolutionary phase.

By leveraging on Event 3 (the human-Neanderthal admixture event that took place after 350 kya) we could further investigate whether the contribution to the human fitness of these derived variants was just relative to the social structure and ecological niche occupied by modern humans in Africa, or whether they could be selected in Neanderthals too, if given a chance to enter their gene pool.

On the one hand, the Human650 regions that entered the Neanderthal gene pool are not enriched for functional variants that emerged specifically along the modern human lineage and include only 1 of the 24 genes linked with the brain phenotypes mentioned above (circled in blue in Figure 4B). On the other hand, these genomic regions are enriched in functional variants that had already emerged before Event 2, when all humans still belonged to the *H. heidelbergensis* lineage, and that were putatively lost due to drift in Neanderthals after they left Africa (Table S3 for an inclusive list), showing that variants with sufficient adaptive potential could be picked up by Neanderthals during this important time checkpoint, despite their reduced effective population size(*9*). Finally, additional modern human-specific sweeps that occurred after Event 3 and encompassed *SPAG5*, *ATRX* and *ARL13*B genes may have further modified the biological pathways and progressively shaped our modern phenotype.

It is important to stress that more recent contributions of our lineage to the Neanderthal one (attested between 100 kya (*33*) and 200 kya(*9*)), if confirmed, would cause even these latter three genes to be added to the list of biological signatures derived along our lineage and that were not picked up by the archaic gene pool, given that their coalescences in modern humans all precede 200 kya and, in general, pre-date the deepest separations between present-day modern human populations (*9*, *23*). More generally, whether or not the human-derived variants present in *SPAG5*, *ATRX* and *ARL13B* were given a chance to show their adaptive potential in Neanderthals, the emerging scenario is one where the human-derived functional variants accumulated in the past 650 kya involved in higher brain functions were in all likelihood strictly linked to the ecological niche humans occupied in Africa.

We caution that our results build on the explicit assumption that modern human defining features must have experienced a sweep after Event 2. This strict experimental framework is bound to leave out additional signals that may have experienced a more intricate evolutionary history. One example is provided by the *TKL1* locus, which was shown to trigger considerable developmental differences in the brain when the modern or archaic human alleles are deployed (*34*), but which reached fixation of the derived allele in humans well before Event 2 (**Table S1**, chrX:153500000-153550000 HumanTMRCA ~1200 kya).

In conclusion, and given the impossibility to accurately dating derived substitutions that are fixed in all modern and archaic lineages, if we are searching for an instant that defined the human lineage we can state that the events that made all of us humans are represented by the chromosome 2 fusion and PAR2 translocation, and such events can be ascribed to the period that precedes 650 kya, which unites the ancestors of all Modern, Neanderthal and Denisova within the same *Homo sapiens* species.
After that moment, relaxation of selection promoted by a reduction in effective population size of the archaic populations, and accumulation of cultural innovations in both modern and archaics are in all likelihood the predominant causes that explain the reported differences at biological and behavioural levels, respectively, observed in the archaeological record between the various populations of the cosmopolitan species we all belong to.
Further functional studies are needed to elucidate the putative role of *ATRX*, *SPAG5* and *ARL13B* as recently acquired keystones of biological pathways that

had been selected by the ecological niche occupied by our human population in Africa since 650 kya, and that may play a role in the fine-tuning of the higher level brain functions, perhaps the most iconic asset of our modern human phenotype.

Material and Methods

Genomic Data

Publicly available Whole genome sequences (WGS) for modern humans were obtained from https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ , WGS for the archaic Neanderthal and Denisova and for the SS6004480 human Chromosome X were obtained from
http://cdna.eva.mpg.de/neandertal/altai/Denisovan/ ;
http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/ ;
http://cdna.eva.mpg.de/neandertal/Spy94a/;
http://cdna.eva.mpg.de/denisova/Den8/;
http://cdna.eva.mpg.de/neandertal/altai/ModernHumans/bam/ ;
http://cdna.eva.mpg.de/neandertal/Vindija/VCF/Vindija33.19/.
Genotype calls for the PAR2 were obtained from https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html?prefix=T2T/CHM13/assemblies/variants/1000_Genomes_Project/chm13v2.0/
PAR2 sites were mapped on the T2T-CHM13v2.0 reference. Genomic markers were instead mapped on build hg19 and analyses were run only at sites where the ancestral allele inferred from the 1000 Genomes project was available.

PAR 2 analyses

We obtained information about sequencing depth along the X chromosomes of a Neanderthal (Spy), a Denisova (Denisova 8) and a modern human male individual (SS6004480) by computing the average depth over non-overlapping 10 kb long windows along the whole chromosome X and displayed the distribution of sequencing depth for the PAR1 (X:0-2,394,410), PAR2 (chrX:153,925,834-154,259,566) (*35*) and the remaining portion (chrX:2,394,411-153,925,833) of the chromosome X. To graphically compare the coverage level between the reference modern human, Neanderthal and Denisova, for each region (Inter X chromosome, PAR1 and PAR2) we scaled the depth coverage using the *scale* function in R and

adding 1 to each value for each region (Inter X chromosome, PAR1 and PAR2) in the modern human, Neanderthal and Denisova separately.
As PAR2 behaves similarly to an autosomal region it is not possible to distinguish short reads coming from the X-linked from Y-linked PAR2 sequence in a given genome. We, therefore, retrieved all PAR2-derived reads in modern human male individuals mapping onto chromosome X (e.g., via masking the Y chromosome for its PAR2 portion), using the most up-to-date set of genotype calls made available by Rhie et al. (*35*) for the 1000 Genomes Project panel (*36*). We exploited the genotype calls performed by the T2T Y chromosome consortium on all the 1000 Genomes Project samples by mapping their reads onto the latest T2T-CHM13v2.0 reference sequence after masking out the chromosome Y-PAR2 region, therefore letting the X-PAR2 region (chrX:153925834-154259566, almost identical to the Y-PAR2 region chrY:62122809-62460029) in both males and females attracting the relevant reads from either chromosome (*35*).
In addition, we estimated the allele frequencies for all non-reference PAR2 sites in 1000 Genomes Project modern human males and females separately and used the observed allele frequencies in female and male X chromosomes to infer the allele frequency in male Y chromosomes:

$$Inferred\ Yfreq = 2 * (Observed\ Male\ X\ freq - Observed\ Female\ X\ freq)$$

and identified 11 sites that reached an allele frequency in male Y chromosomes close to 1 (hence distinguishing the Y-PAR2 from the X-PAR2 sequences). Other classes of notable mutations, namely the ones showing an X frequency of less than 0.5 in males (inferred to be less than 1 in Y-PAR2) and approximately 0 in females, and the ones showing an X frequency of approximately 0.5 in males (inferred to be approximately 1 in PAR2-Y chromosomes) and more than 0 in females, were annotated but then discarded as representing not yet fixed male-specific mutations or mutations present in the X-PAR2 at the time of the translocation, respectively, and were deemed outside of the scope of the present study.
The 11 mutations found to be fixed in all surveyed males and virtually absent from females spanned 21215 bps at the 5’ end of the Y-PAR2 sequence and were considered to have occurred since the divergence between the X-PAR2 sequences and the most recent common ancestor of all Y-PAR2 sequences and before the diversification of present-day Y chromosome lineages. We then dated the time that occurred for these mutations to accumulate by using the simple formula: T= g * ((accumulated mutations) / length) / mu), using 30 years as the generation time (g)

in humans and ruling out complex back-forth scenarios between the X and Y PAR2 pools, given the absence of these mutations within the X-PAR2 gene pool. When using the Y chromosome mutation rate of 3 x 10^{-8} mut/bp/gen (*17*) we obtained a T of 518 ky (95% CI: 295-915 ky), before modern human MRCA. We computed confidence intervals on our estimate by estimating the probability distribution of observing exactly 11 mutations in 21215 bp over different time intervals and taking the 2.5th and 97.5th quantiles of this distribution. The modern human MRCA has been attested at 338 kya, hence yielding a final 856 kya (95% CI: 633-1253 kya) kya as a final point estimate for the PAR2 translocation.

Computing window-based genomic divergences

We computed genetic divergences from the human reference sequence (hg19) by counting the number of differences between Altai Neanderthal or Denisova or 50 Sub-Saharan African individuals from the 1000 Genomes Project over a sliding window of 50 kb across all the autosomes and chromosome X, and by picking a phased haploid sequence in modern humans and a random allele at heterozygous sites in archaic humans. We then converted the obtained divergence ((number of differences/2) / sequence length) in a date of coalescence, or time since the most recent common ancestor (TMRCA) by simply dividing it by the average human mutation rate (1.25 x 10^{-8} mut/bp/gen) (*37*) and multiplying it by 30 years, here taken for the time of a human generation.

$$TMRCA = 30 * \frac{(0.5 * number\ of\ differences)/sequence\ length}{1.25 * 10^{-8}}$$

For each genomic window, we then obtained a point estimate for the divergence between each archaic human and the reference genome (here deemed as a proxy for a random modern human), and we took the top 90% most divergent estimate based on the 50 African individuals as the within- modern human divergence (humanTMRCA). We based our modern human estimates on African genomes to minimize the effect of the known Neanderthal introgression that occurred after the expansion of modern humans out of Africa 60 kya (*13*, *38*). Neanderthal chunks embedded within the reference genome are expected to yield deeper divergences between all tested Africans and the reference genome, but we deemed this to cause a marginal impact overall, and not relevant for our experimental design aimed at detecting regions with increased similarity between modern humans and Neanderthals. We filtered out from our downstream analyses regions with <50% of accessible sequence based on the 1000 Genomes Project strict accessibility masks

(**Table S1**) and <90% availability of archaic and ancestral alleles along all 50 kb sites.

Definition of Human650 and WMTH regions

Based on the data available from Table S1, we defined Human650 regions as regions with a) a HumanTMRCA <= 650 kya (e.g., after the separation of our lineage from the one of the archaic humans, Event 2) and b) a DenisovaTMRCA >=650 kya, as expected from the population tree, to exclude genomic regions that experienced apparent relaxation of the molecular clock due to purifying selection. Of these Human650 regions, we further identified What Made Them Human (WMTH) regions as the ones with i) a NeanderthalTMRCA <=650 kya (not expected by the population tree, unless human-Neanderthal admixture is involved), ii) at least one fixed derived mutation shared between Altai Neanderthal and modern humans (derived frequency >=0.98 in 1000 Genomes Project individuals) but not Denisovans and iii) no fixed derived mutations shared between Denisovans and modern humans but not Altai Neanderthal.

Enrichment analysis

We explored the distribution of archaic/modern and modern/modern TMRCAs computed in **Table S1**, looking for enrichments of Neanderthal/human TMRCA compared to Denisova/human TMRCAs within autosomal loci, hence excluding chromosome X from this particular analysis. We first examined regions displaying an archaic/human TMRCA falling within a moving cut-off (t) from 1000 to 100 kya, with a step of 10 ky and counted the number of 50 kb regions with a NeanderthalTMRCA less than t years ago and with DenisovaTMRCA>=650 kya, labelled as NRDA (Neanderthal Recent, Denisova Ancient), then we estimated the number of segments with DenisovaTMRCA less than t years ago and with NeanderthalTMRCA >=650 kya, these segments were labelled as DRNA (Denisova Recent - Neanderthal Ancient). We then tested whether the abundance of NRDA and DRNA regions was just following Incomplete Lineage Sorting by computing the normalized ratio

$$Archaic\ recent = {}^{NRDA}/_{(DRNA + NRDA)}$$

with

$$NRDA = N.segments(NeaTRMCA < T\ \ \&\ \ DenTMRCA \geq 650Kya)$$
$$DRNA = N.segments(DenTRMCA < T\ \ \&\ \ NeaTMRCA \geq 650Kya)$$

(Blue line in **Supplementary Figure 2)**. A value higher than 0.5 means an excess of NRDA regions compared to DRNA regions.
We then computed the same ratio by looking only within Human650 regions and obtained the Archaic Recent within Human650 enrichment curve (Green in **Supplementary Figure 2**). All these results show an excess that reaches saturation (equal to 1) after Event 3.

Relate
For each 50kb Autosomal WMTH region, we further dwelled in the analyses of TMRCA using Relate v1.1.9 (*39*). The trees were built only for the regions of interest using the entire 1000 Genomes dataset (3202 samples) plus three archaic genomes. With this, we inferred genomic regions corresponding to each tree (regions with no inferred historical recombination within the sample).
We then computed the TMRCA of each tree spanning at least 5kb using the procedure described above (Figure S1).

Variant Annotation and analyses
We used Ensembl VEP (Ensembl GRCh37 release 113) (*40*) to obtain the most severe consequence and the gene they encompassed for each of the sites that were fixed derived in at least 98% of all 1000 Genomes Project individuals (including all fixed sites) and falling in genomic windows with a HumanTMRCA <=1000 kya (e.g that experienced a human fixation event after Event 1). We categorized all obtained consequences into three main classes:
-Coding (missense_variant, splice_acceptor_variant, splice_donor_5th_base_variant, splice_donor_region_variant, splice_donor_variant, splice_polypyrimidine_tract_variant, splice_region_variant, start_lost, stop_gained, stop_lost)

-Regulatory (3_prime_UTR_variant, 5_prime_UTR_variant, mature_miRNA_variant, regulatory_region_variant, TF_binding_site_variant)

-Others (downstream_gene_variant, intergenic_variant, intron_variant, non_coding_transcript_exon_variant, synonymous_variant, upstream_gene_variant)

We referred to Coding and Regulatory consequences together as Functional and distinguished them into four categories: FuncH, >=98% derived allele frequency only in modern humans; FuncHD, >=98% derived allele frequency in modern humans and homozygous derived in Denisova but ancestral in Altai Neanderthal;

FuncHN, >=98% derived allele frequency in modern humans and homozygous derived Neanderthals but ancestral in Denisova, and FuncHH,>=98% derived allele frequency surveyed modern and homozygous derived in Altai Neanderthal and Denisova.

We then divided the Functional sites into non-overlapping sets based on the HumanTMRCA and ArchaicTMRCA displayed by the 50 kb windows containing them:
Group I: HumanTMRCAs <=350 kya, DenisovaTMRCA >650 kya;
Group II: 350 kya < HumanTMRCA <= 650 kya and DenisovaTMRCA > 650 kya;
Group III: 650 < (HumanTMRCA, DenisovaTMRCA, NeanderthalTMRCA) <=1000 kya;
A final Group, called WMTH, was formed by a subset of Group II regions displaying a NeanderthalTMRCA <= 650 kya.

To inspect the biological network encompassed by the genes showing functional modifications in the past 650 kya of modern human evolution, the list of genes affected by at least one FuncH or FuncHN variant of groups I and II (falling in Human650 regions, displaying a HumanTMRCA <= 650 kya) was analyzed using STRING v12.0 database (STRING database https://string-db.org/). The obtained network of protein-protein interactions was analysed for phenotype enrichment using the STRING enrichment tool against the Monarch Initiative database (*41*).

Acknowledgements
The authors would like to thank Prof. Gil Guastoni Rosenthal, Dr. Adeline Morez, Dr. Andra Meneganzin, Giacomo Moro Mauretto, Francesca Carollo and Hélios Delbrassine for the fruitful discussion during the drafting of the current manuscript, and Dr. Mathilde André for facilitating access to the Papuan allele frequencies. LP was funded by the Italian Ministry of University and Research PRIN N. 2022B27XYM.

Competing interests: C. L. is a scientific advisory board member of Nabsys and Genome Insight. The other authors declare that they have no competing interests.

Data and materials availability: All genomic data are publicly available at
https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/,
http://cdna.eva.mpg.de/neandertal/altai/Denisovan/;
http://cdna.eva.mpg.de/neandertal/altai/AltaiNeandertal/VCF/;
http://cdna.eva.mpg.de/neandertal/Spy94a/;
http://cdna.eva.mpg.de/denisova/Den8/;

http://cdna.eva.mpg.de/neandertal/Vindija/VCF/Vindija33.19/, TMRCA computations are available as supplementary data.

References

1. A. Meneganzin, M. Bernardi, Were Neanderthals and Homo sapiens “good species”? *Quat. Sci. Rev.* **303**, 107975 (2023).

2. W. Hu, Z. Hao, P. Du, F. Di Vincenzo, G. Manzi, J. Cui, Y.-X. Fu, Y.-H. Pan, H. Li, Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science* **381**, 979–984 (2023).

3. J. Terhorst, Accelerated Bayesian inference of population size history from recombining sequence data. *bioRxiv*, doi: 10.1101/2024.03.25.586640 (2024).

4. T. Cousins, A. Durvasula, Insufficient evidence for a severe bottleneck in humans during the Early to Middle Pleistocene transition, *bioRxiv* (2024)p. 2024.10.21.619456.

5. B. Poszewiecka, K. Gogolewski, P. Stankiewicz, A. Gambin, Revised time estimation of the ancestral human chromosome 2 fusion. *BMC Genomics* **23**, 616 (2022).

6. K. Prufer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Paabo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

7. C. V. Weiss, L. Harshman, F. Inoue, H. B. Fraser, D. A. Petrov, N. Ahituv, D. Gokhman, The cis-regulatory effects of modern human-specific variants. *eLife* **10** (2021).

8. H. Zeberg, M. Jakobsson, S. Pääbo, The genetic changes that shaped Neandertals, Denisovans, and modern humans. *Cell* **187**, 1047–1058 (2024).

9. L. Li, T. J. Comi, R. F. Bierman, J. M. Akey, Recurrent gene flow between Neanderthals and modern humans over the past 200,000 years. *Science*, doi: 10.1126/science.adi1768 (2024).

10. D. N. Harris, A. Platt, M. E. B. Hansen, S. Fan, M. A. McQuillan, T. Nyambo, S. W. Mpoloka, G. G. Mokone, G. Belay, C. Fokunang, A. K. Njamnshi, S. A. Tishkoff, Diverse African genomes reveal selection on ancient modern human introgressions in Neanderthals. *Curr. Biol.* **33**, 4905–4916.e5 (2023).

11. C. Posth, C. Wißing, K. Kitagawa, L. Pagani, L. van Holstein, F. Racimo, K. Wehrberger, N. J. Conard, C. J. Kind, H. Bocherens, J. Krause, Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat. Commun.* **8**, 16046 (2017).

12. M. Petr, M. Hajdinjak, Q. Fu, E. Essel, H. Rougier, I. Crevecoeur, P. Semal, L. V. Golovanova, V. B. Doronichev, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, M. V. Shunkov, M. B. Kozlikin, A. P. Derevianko, B. Vernot, M. Meyer, J. Kelso, The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science* **369**, 1653–1656 (2020).

13. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

14. S. P. Otto, J. R. Pannell, C. L. Peichel, T.-L. Ashman, D. Charlesworth, A. K. Chippindale, L. F. Delph, R. F. Guerrero, S. V. Scarpino, B. F. McAllister, About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* **27**, 358–367 (2011).

15. K. D. Makova, B. D. Pickett, R. S. Harris, G. A. Hartley, M. Cechova, K. Pal, S. Nurk, D. Yoo, Q. Li, P. Hebbar, B. C. McGrath, F. Antonacci, M. Aubel, A. Biddanda, M. Borchers, E. Bornberg-Bauer, G. G. Bouffard, S. Y. Brooks, L. Carbone, L. Carrel, A. Carroll, P.-C. Chang, C.-S. Chin, D. E. Cook, S. J. C. Craig, L. de Gennaro, M. Diekhans, A. Dutra, G. H. Garcia, P. G. S. Grady, R. E. Green, D. Haddad, P. Hallast, W. T. Harvey, G. Hickey, D. A. Hillis, S. J. Hoyt, H. Jeong, K. Kamali, S. L. K. Pond, T. M. LaPolice, C. Lee, A. P. Lewis, Y.-H. E. Loh, P. Masterson, K. M. McGarvey, R. C. McCoy, P. Medvedev, K. H. Miga, K. M. Munson, E. Pak, B. Paten, B. J. Pinto, T. Potapova, A. Rhie, J. L. Rocha, F. Ryabov, O. A. Ryder, S. Sacco, K. Shafin, V. A. Shepelev, V. Slon, S. J. Solar, J. M. Storer, P. H. Sudmant, Sweetalana, A. Sweeten, M. G. Tassia, F. Thibaud-Nissen, M. Ventura, M. A. Wilson, A. C. Young, H. Zeng, X. Zhang, Z. A. Szpiech, C. D. Huber, J. L. Gerton, S. V. Yi, M. C. Schatz, I. A. Alexandrov, S. Koren, R. J. O'Neill, E. E. Eichler, A. M. Phillippy, The complete sequence and comparative analysis of ape sex chromosomes. *Nature* **630**, 401–411 (2024).

16. P. Hallast, P. Ebert, M. Loftus, F. Yilmaz, P. A. Audano, G. A. Logsdon, M. J. Bonder, W. Zhou, W. Höps, K. Kim, C. Li, S. J. Hoyt, P. C. Dishuck, D. Porubsky, F. Tsetsos, J. Y. Kwon, Q. Zhu, K. M. Munson, P. Hasenfeld, W. T. Harvey, A. P. Lewis, J. Kordosky, K. Hoekzema, Human Genome Structural Variation Consortium (HGSVC), R. J. O'Neill, J. O. Korbel, C. Tyler-Smith, E. E. Eichler, X. Shi, C. R. Beck, T. Marschall, M. K. Konkel, C. Lee, Assembly of 43

diverse human Y chromosomes reveals extensive complexity and variation, *bioRxiv* (2023)p. 2022.12.01.518658.

17. A. Helgason, A. W. Einarsson, V. B. Guðmundsdóttir, Á. Sigurðsson, E. D. Gunnarsdóttir, A. Jagadeesan, S. S. Ebenesersdóttir, A. Kong, K. Stefánsson, The Y-chromosome point mutation rate in humans. *Nat. Genet.* **47**, 453–457 (2015).

18. F. L. Mendez, T. Krahn, B. Schrack, A.-M. Krahn, K. R. Veeramah, A. E. Woerner, F. L. M. Fomine, N. Bradman, M. G. Thomas, T. M. Karafet, M. F. Hammer, An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.* **92**, 454–459 (2013).

19. M. Kuhlwilm, C. Boeckx, A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Scientific Reports* **9**, 1–14 (2019).

20. A. Mounier, M. Mirazón Lahr, Deciphering African late middle Pleistocene hominin diversity and the origin of our species. *Nat. Commun.* **10**, 3406 (2019).

21. J.-J. Hublin, A. Ben-Ncer, S. E. Bailey, S. E. Freidline, S. Neubauer, M. M. Skinner, I. Bergmann, A. Le Cabec, S. Benazzi, K. Harvati, P. Gunz, New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature* **546**, 289–292 (2017).

22. A. P. Ragsdale, T. D. Weaver, E. G. Atkinson, E. G. Hoal, M. Möller, B. M. Henn, S. Gravel, A weakly structured stem for human origins in Africa. *Nature* **617**, 755–763 (2023).

23. C. M. Schlebusch, H. Malmström, T. Günther, P. Sjödin, A. Coutinho, H. Edlund, A. R. Munters, M. Vicente, M. Steyn, H. Soodyall, M. Lombard, M. Jakobsson, Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).

24. C. M. Vidal, C. S. Lane, A. Asrat, D. N. Barfod, D. F. Mark, E. L. Tomlinson, A. Z. Tadesse, G. Yirgu, A. Deino, W. Hutchison, A. Mounier, C. Oppenheimer, Age of the oldest known Homo sapiens from eastern Africa. *Nature* **601**, 579–583 (2022).

25. T. D. White, B. Asfaw, D. DeGusta, H. Gilbert, G. D. Richards, G. Suwa, F. C. Howell, Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* **423**, 742–747 (2003).

26. K. Harvati, C. Röding, A. M. Bosman, F. A. Karakostis, R. Grün, C. Stringer, P. Karkanas, N. C. Thompson, V. Koutoulidis, L. A. Moulopoulos, V. G. Gorgoulis, M. Kouloukoussa, Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia. *Nature* **571** (2019).

27. [No title]. https://academic.oup.com/nar/article/52/D1/D938/7449493.

28. M. Schubach, T. Maass, L. Nazaretyan, S. Röner, M. Kircher, CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to

improve genome-wide variant predictions. *Nucleic Acids Res.* **52**, D1143–D1154 (2024).

29. S. Peyrégne, J. Kelso, B. M. Peter, S. Pääbo, The evolutionary history of human spindle genes includes back-and-forth gene flow with Neandertals. doi: 10.7554/eLife.75464 (2022).

30. F. Mora-Bermúdez, P. Kanis, D. Macak, J. Peters, R. Naumann, L. Xing, M. Sarov, S. Winkler, C. E. Oegema, C. Haffner, P. Wimberger, S. Riesenberg, T. Maricic, W. B. Huttner, S. Pääbo, Longer metaphase and fewer chromosome segregation errors in modern human than Neanderthal brain development. *Science Advances*, doi: 10.1126/sciadv.abn7702 (2022).

31. N. Brucato, M. André, R. Tsang, L. Saag, J. Kariwiga, K. Sesuki, T. Beni, W. Pomat, J. Muke, V. Meyer, A. Boland, J. F. Deleuze, H. Sudoyo, M. Mondal, L. Pagani, R. I. Gallego, M. Metspalu, M. P. Cox, M. Leavesley, F. X. Ricaut, Papua New Guinean Genomes Reveal the Complex Settlement of North Sahul. *Molecular biology and evolution* **38** (2021).

32. M. André, N. Brucato, G. Hudjasov, V. Pankratov, D. Yermakovich, F. Montinaro, R. Kreevan, J. Kariwiga, J. Muke, A. Boland, J.-F. Deleuze, V. Meyer, N. Evans, M. P. Cox, M. Leavesley, M. Dannemann, T. Org, M. Metspalu, M. Mondal, F.-X. Ricaut, Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels. *Nat Commun* **15**, 3352 (2024).

33. M. Kuhlwilm, I. Gronau, M. J. Hubisz, C. de Filippo, J. Prado-Martinez, M. Kircher, Q. Fu, H. A. Burbano, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, T. Marques-Bonet, A. M. Andrés, B. Viola, S. Pääbo, M. Meyer, A. Siepel, S. Castellano, Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**, 429–433 (2016).

34. A. Pinson, L. Xing, T. Namba, N. Kalebic, J. Peters, C. E. Oegema, S. Traikov, K. Reppe, S. Riesenberg, T. Maricic, R. Derihaci, P. Wimberger, S. Pääbo, W. B. Huttner, Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science* **377**, eabl6422 (2022).

35. A. Rhie, S. Nurk, M. Cechova, S. J. Hoyt, D. J. Taylor, N. Altemose, P. W. Hook, S. Koren, M. Rautiainen, I. A. Alexandrov, J. Allen, M. Asri, A. V. Bzikadze, N.-C. Chen, C.-S. Chin, M. Diekhans, P. Flicek, G. Formenti, A. Fungtammasan, C. Garcia Giron, E. Garrison, A. Gershman, J. L. Gerton, P. G. S. Grady, A. Guarracino, L. Haggerty, R. Halabian, N. F. Hansen, R. Harris, G. A. Hartley, W. T. Harvey, M. Haukness, J. Heinz, T. Hourlier, R. M. Hubley, S. E. Hunt, S. Hwang, M. Jain, R. K. Kesharwani, A. P. Lewis, H. Li, G. A. Logsdon, J. K. Lucas, W. Makalowski, C. Markovic, F. J. Martin, A. M. Mc Cartney, R. C. McCoy, J. McDaniel, B. M. McNulty, P. Medvedev, A. Mikheenko, K. M. Munson, T. D. Murphy, H. E. Olsen, N. D. Olson, L. F. Paulin, D. Porubsky, T. Potapova, F. Ryabov, S. L. Salzberg, M. E. G. Sauria, F. J. Sedlazeck, K. Shafin, V. A. Shepelev, A. Shumate, J. M. Storer, L. Surapaneni, A. M. Taravella Oill, F. Thibaud-Nissen, W. Timp, M. Tomaszkiewicz, M. R. Vollger, B. P. Walenz, A. C. Watwood, M. H. Weissensteiner, A. M. Wenger, M. A. Wilson, S.

Zarate, Y. Zhu, J. M. Zook, E. E. Eichler, R. J. O'Neill, M. C. Schatz, K. H. Miga, K. D. Makova, A. M. Phillippy, The complete sequence of a human Y chromosome. *Nature* **621**, 344–354 (2023).

36. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

37. A. Scally, R. Durbin, Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012).

38. L. Vallini, C. Zampieri, M. J. Shoaee, E. Bortolini, G. Marciani, S. Aneli, T. Pievani, S. Benazzi, A. Barausse, M. Mezzavilla, M. D. Petraglia, L. Pagani, The Persian plateau served as hub for Homo sapiens after the main out of Africa dispersal. *Nat. Commun.* **15**, 1882 (2024).

39. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).

40. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

41. C. J. Mungall, J. A. McMurry, S. Köhler, J. P. Balhoff, C. Borromeo, M. Brush, S. Carbon, T. Conlin, N. Dunn, M. Engelstad, E. Foster, J. P. Gourdine, J. O. B. Jacobsen, D. Keith, B. Laraway, S. E. Lewis, J. NguyenXuan, K. Shefchek, N. Vasilevsky, Z. Yuan, N. Washington, H. Hochheiser, T. Groza, D. Smedley, P. N. Robinson, M. A. Haendel, The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res* **45**, D712–D722 (2017).