nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-021-01408-0

Widespread Denisovan ancestry in Island Southeast Asia but no evidence of substantial super-archaic hominin admixture

João C. Teixeira[1,2]✉, Guy S. Jacobs[3,4], Chris Stringer[5], Jonathan Tuke[6], Georgi Hudjashov[7], Gludhug A. Purnomo[1,8], Herawati Sudoyo[8,9,10], Murray P. Cox[7], Raymond Tobler[1,2,16], Chris S. M. Turney[11,12,16], Alan Cooper[13,14,16] and Kristofer M. Helgen[12,15,16]

The hominin fossil record of Island Southeast Asia (ISEA) indicates that at least two endemic 'super-archaic' species—*Homo luzonensis* and *H. floresiensis*—were present around the time anatomically modern humans arrived in the region >50,000 years ago. Intriguingly, contemporary human populations across ISEA carry distinct genomic traces of ancient interbreeding events with Denisovans—a separate hominin lineage that currently lacks a fossil record in ISEA. To query this apparent disparity between fossil and genetic evidence, we performed a comprehensive search for super-archaic introgression in >400 modern human genomes, including >200 from ISEA. Our results corroborate widespread Denisovan ancestry in ISEA populations, but fail to detect any substantial super-archaic admixture signals compatible with the endemic fossil record of ISEA. We discuss the implications of our findings for the understanding of hominin history in ISEA, including future research directions that might help to unlock more details about the prehistory of the enigmatic Denisovans.

Island Southeast Asia (ISEA) hosts a unique and diverse fossil record of hominin presence throughout the Pleistocene epoch[1]. The island of Java in modern Indonesia marks the southeastern extent of the range of *Homo erectus*, the first hominin species thought to have successfully dispersed outside Africa, where it maintained a presence from ~1.49 million years ago (Ma) until ~117–108 thousand years ago (ka)[2–4]. At least two additional endemic species lived in ISEA during the Pleistocene and are likely to have survived until the arrival of anatomically modern humans (AMH) > 50 ka[5–8]: *H. floresiensis* on Flores, in the Lesser Sunda Islands (also part of modern Indonesia)[9,10], and *H. luzonensis* on Luzon, in the northern Philippines[11]. The phylogenetic relationships of these two species to each other and to other hominins remains an area of debate. Recent interpretations suggest that *H. floresiensis* is either a close relative of *H. erectus*, or alternatively represents an even more archaic species of *Homo* that independently reached ISEA in a separate dispersal event out Africa[9,12,13]. The current classification of *H. luzonensis* is also uncertain; the available specimens share similarities in certain morphological traits with various hominin taxa including *Australopithecus*, Asian *H. erectus*, *H. floresiensis* and *H. sapiens*[11].

Genetic evidence preserved in modern human genomes suggests that at least one additional hominin group probably inhabited ISEA at the time of AMH arrival. Present-day human populations living in ISEA, New Guinea and Australia harbour substantial genetic ancestry from Denisovans, a sister lineage to Neanderthals with a fossil record that is limited to a few skeletal fragments from the eponymous cave in the Altai Mountains in Siberia[14,15] and a >160,000-year-old mandible found in the Tibetan Plateau[16], where Denisovan DNA has recently been recovered from cave sediments[17]. Despite this geographically circumscribed fossil record, the patterns of Denisovan ancestry in modern human populations suggest that they may have been present across ISEA at the time of AMH arrival[18]. While the complexities inherent to demographic and archaic ancestry inference make it hard to infer the precise number and geographical location(s) of the encounters between AMH and Denisovans, the discovery of multiple distinct pulses of Denisovan admixture in contemporary human populations[19–24] suggests that Denisovans had probably come to occupy several islands east of Wallace's (including Huxley's) Line by 50 ka. Stone tools found in Sulawesi dated to ~100–200 ka[25] are also suggestive of possible Denisovan presence east of Wallace's Line[26]; however, direct fossil evidence of Denisovans in ISEA remains conspicuously absent to date.

The disparity between the lack of a fossil record of Denisovans in ISEA and the mounting genetic evidence suggesting AMH–Denisovan mixing events in this region poses an important outstanding question in hominin prehistory. A parsimonious solution

[1]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia. [2]ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), The University of Adelaide, Adelaide, South Australia, Australia. [3]Complexity Institute, Nanyang Technological University, Singapore, Singapore. [4]Department of Archaeology, University of Cambridge, Cambridge, UK. [5]Centre for Human Evolution Research, Department of Earth Sciences, Natural History Museum, London, UK. [6]School of Mathematical Sciences, The University of Adelaide, Adelaide, South Australia, Australia. [7]Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North, New Zealand. [8]Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta, Indonesia. [9]Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia. [10]Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia. [11]Chronos 14Carbon-Cycle Facility, Earth and Sustainability Science Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia. [12]ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), University of New South Wales, Sydney, New South Wales, Australia. [13]South Australian Museum, Adelaide, South Australia, Australia. [14]BlueSky Genetics, Ashton, South Australia, Australia. [15]Australian Museum Research Institute, Australian Museum, Sydney, New South Wales, Australia. [16]These authors contributed equally: Raymond Tobler, Chris S. M. Turney, Alan Cooper, Kristofer M. Helgen. ✉e-mail: joao.teixeira@adelaide.edu.au

to this problem is that perhaps either *H. luzonensis* and/or *H. floresiensis* (or both) is the source of the Denisovan ancestry in modern human genomes in the region; however, the anatomical attributes of both of these extinct ISEA hominin species are not readily reconcilable with the few confirmed specimens of Denisovans from Altai and the Tibetan Plateau[9–13,27–29]. Moreover, morphological and archaeological data suggest that both *H. floresiensis* and *H. luzonensis* had an extensive history in the region that preceded the estimated emergence time of the Denisovans[9–13,27–29]; thus, they are interpreted as two distinct super-archaic hominin species that evolved in situ on their respective island locales. The other possible admixture source—Indonesian *H. erectus*—is precluded because of its last appearance date of ~117–108 ka[3]. Thus, the source(s) of Denisovan introgression into modern human genomes in ISEA remains elusive.

Alternatively, the possible survival of *H. floresiensis* and *H. luzonensis* in ISEA until the arrival of AMH in the region raises the possibility that they also admixed with the ancestors of modern populations now living in ISEA. Traces of super-archaic admixture have been detected in Altai Denisovans[30] and, potentially, in modern Andamanese populations[31–33], suggesting that interbreeding between super-archaic hominins and more derived hominin species has previously occurred and produced viable progeny. If such an event occurred between AMH and endemic ISEA hominins, evidence of this mixing may yet remain undetected in the genomes of present-day human populations now living in ISEA, and would indirectly confirm the past presence of one or more super-archaic species in ISEA.

Results

To address this question and provide further insights into the hominin prehistory of ISEA, we implemented the most comprehensive search for introgressed super-archaic regions in modern human genomes performed to date, to our knowledge. We searched a total of 426 human genomes from across the world, including 214 individuals from Papuan and ISEA populations[22] (Supplementary Table 1), for genomic signatures compatible with introgression from archaic hominins such as *H. floresiensis*, *H. luzonensis* or other hypothetical late-surviving super-archaic hominin species. To detect blocks of introgressed super-archaic DNA, we extended the analytical pipeline reported in ref. [22] by including a recently published hidden Markov model (HMM) detection method[34]—which we call $HMM_{archaic}$—along with the two methods used in ref. [22]: ChromoPainter (CP)[35] and an HMM[36,37]. Importantly, $HMM_{archaic}$ differs from CP and HMM in that it does not require a reference genome to guide the detection of introgressed DNA, making it suitable for identifying DNA from super-archaic groups for which no genome information currently exists. Accordingly, we were able to distinguish putative introgressed super-archaic blocks by running the three detection methods on all 426 genomes and retaining only those that did not overlap any of the Neanderthal and Denisovan blocks predicted by CP and/or HMM. We term the resulting set of putative super-archaic sequences as $residual_{archaic}$ blocks (see Methods). Importantly, to specifically focus on patterns of super-archaic ancestry in ISEA, our strategy purposely excludes genetic variation shared between African and non-African populations. Accordingly, any super-archaic admixture involving AMH in Africa (for example, with taxa such as *Homo naledi*[38]) would be excluded from our results.

No evidence for super-archaic introgression in AMH

Filtering the $HMM_{archaic}$ introgressed blocks overlapping Neanderthal- and Denisovan-introgressed tracts identified ~12.5 Mb of $residual_{archaic}$ sequence per individual (that is, sequences that are putatively introgressed from a super-archaic source; Fig. 1a). The amount of detected $residual_{archaic}$ sequence was consistent across worldwide populations, with a slightly higher amount found in east ISEA (~15 Mb), and Papuan and Australian populations (~18 Mb). In accordance with previous results, ISEA, Papuan and Australian populations also had the largest amounts of Denisovan ancestry (reaching ~60 Mb in Papuan and Australian genomes), meaning that these populations actually had the lowest proportion of $residual_{archaic}$ sequence relative to the total archaic ancestry observed across all analysed populations (Supplementary Fig. 1). Our results indicate that super-archaic ancestry could potentially comprise a small but consistent amount of the genomic ancestry of modern human populations outside Africa. However, the current lack of evidence for widespread super-archaic admixture in modern human populations suggests that this global $residual_{archaic}$ signal is more likely to be a methodological artefact, or a signal of ancient genetic structure in human populations that pre-dates the out-of-Africa migration, or segregation of highly divergent AMH-derived sequences that were not detected in our African reference samples that result from incomplete lineage sorting or balancing selection[39]. Similarly, the additional ~2.5 Mb to ~5 Mb of $residual_{archaic}$ sequence observed in Papuan and Australian populations may represent a small but meaningful amount of super-archaic ancestry specific to this region, or instead simply reflect inter-population variation in the power of the statistical methods to detect Denisovan fragments or some other methodological artefact.

To further discriminate if the $residual_{archaic}$ blocks were truly introgressed super-archaic DNA, we searched for concordant signatures by investigating genetically distinct mutation motifs (that is, allelic states) that are characteristic of introgressed super-archaic DNA within $residual_{archaic}$ blocks. Specifically, for each nucleotide position in a $residual_{archaic}$ block, we characterized the allelic state for the test individual (X), Denisovan (D), Neanderthal (N) and an African individual (H; see Methods). This resulted in a set of mutation motifs of the form [X, D, N, H], with patterns of the type [1000] and [0111] potentially indicative of super-archaic introgression signals. After enumerating these mutation motifs for all $residual_{archaic}$ blocks in each individual, we used generalized linear models to test if the proportion of motifs showed population-specific differences, and computed *P* values by contrasting the full model with a null model consisting of the intercept alone (see Methods).

The mutation motifs differed significantly between populations when considering a linear model (ANOVA *P* value 5.79×10^{-224}) but not when a multinomial logistic regression model was used (where motifs are not independent as is assumed for the linear model; Fig. 1b and Supplementary Fig. 2). However, these differences are extremely subtle and correlate strongly with known archaic ancestry, consistent with the presence of a confounding effect (Fig. 1c and Supplementary Figs. 3–6). For example, Papuan genomes show a slightly higher proportion of [1000] motifs (<2%) compared with other populations (Fig. 1b and Supplementary Fig. 2), but inter-individual variation is also high and we do not observe a similar increase in the proportion of the [0111] motif in the population, which is also expected under a scenario of super-archaic introgression (Supplementary Fig. 2 and Methods).

While precise accounting for all motif count differences is non-trivial, probable explanations include the misclassification of alleles as either ancestral or derived, complex demographic histories, and the persistence of Neanderthal and Denisovan archaic signals among the $residual_{archaic}$ blocks that were not removed during the filtering step. For instance, the 2.5–5 Mb extra $residual_{archaic}$ sequence observed in Papuans and Australians might have resulted from these populations having substantially more introgression from a Denisovan-like source that is highly divergent from the Altai Denisovan genome[22]. This may result in some of the more diverged blocks being detected by the reference-free $HMM_{archaic}$ scan, but not in the two methods that rely on reference genomes (that is, CP and HMM). Indeed, while Denisovan and Neanderthal ancestry is

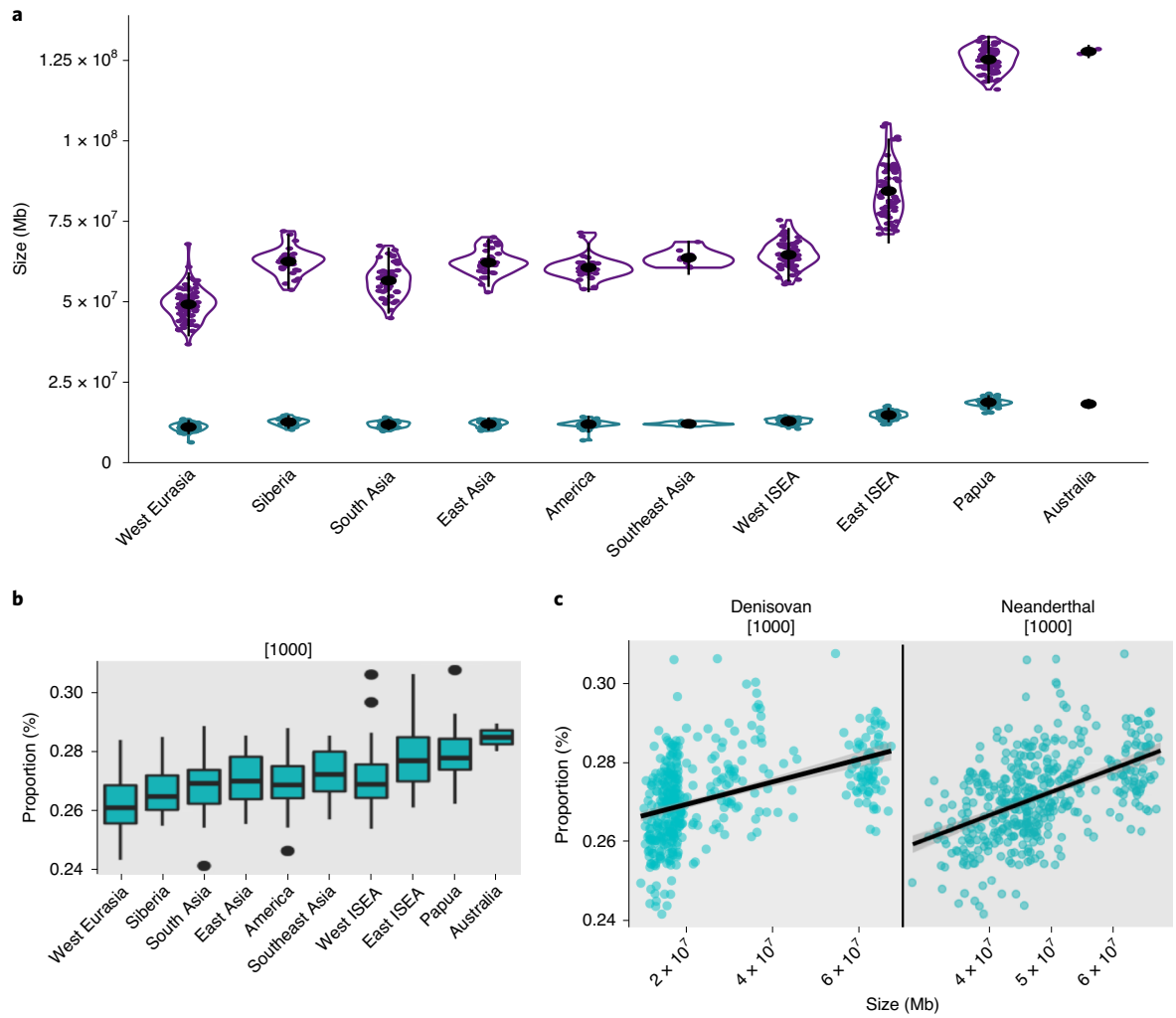

Fig. 1 | Introgression signals in extant populations across ISEA. a, Violin plots showing the cumulative amount (Mb) of Neanderthal and Denisovan ancestry (purple) estimated using HMM and residual$_{archaic}$ sequence (green) across different populations. Each dot represents a single sampled individual for a particular population. Within each violin plot, the population's mean and 95% values of the distribution are shown as a black dot and vertical line, respectively. **b**, The proportion of variants within residual$_{archaic}$ fragments that show mutation motifs compatible with super-archaic introgression [1000] per population. Each number on the string [1000] corresponds to the allelic states observed in [X, Denisovan, Neanderthal, Africa], where X is an individual from the test population (for example, Australia), and 1 and 0 define derived and ancestral allelic states, respectively. The lower and upper hinges of each boxplot correspond to the first and third quartiles, respectively, and the horizontal bar in the middle represents the median. The whiskers extend to the highest (upper whisker) or lowest (lower whisker) values of the distribution within 1.5 × IQR of each hinge, where IQR represents the interquartile range. Data values beyond the end of the whiskers are plotted as black dots. **c**, Scatter plot showing the association between the proportion of [1000] motifs within residual$_{archaic}$ fragments and the total amount of Denisovan (left) and Neanderthal (right) ancestry per individual.

positively correlated with the proportion of the [1000] motif across all populations, it is negatively correlated with the proportion of the [0111] motif (Supplementary Figs. 3 and 4, respectively), which strongly suggests that differences in the proportion of these motifs are caused by unassigned Neanderthal and Denisovan ancestry within residual$_{archaic}$ blocks.

Indirect introgression of super-archaic hominin DNA from Denisovans

A recent publication reported that modern human genomes carry traces (~4 Mb) of super-archaic ancestry that are embedded within introgressed Denisovan sequences (having previously been derived from ancient admixture events between Denisovans and an unknown super-archaic source)[40]. Importantly, the majority of these indirectly introgressed super-archaic segments were also detected in this study (20 out of 20, with 100% of the sequence length of each introgressed block being recovered; Supplementary Table 2); most of these were also included in our set of putative super-archaic blocks (17 out of 20, with an average of 47.5% of the sequence length of each block being recovered; Supplementary Table 3). Similar results were obtained by comparing HMM$_{archaic}$ and residual$_{archaic}$ blocks with super-archaic segments embedded within predicted Neanderthal introgressed regions (see Methods and Supplementary Tables 4 and 5).

Coalescent simulations support empirical observations

The accurate recovery of indirectly introgressed super-archaic fragments by our analytical pipeline suggests that it is sufficiently powerful to detect low levels of directly introgressed super-archaic ancestry, should it exist. Nonetheless, to rule out the possibility that the lack of evidence for super-archaic introgression into modern humans was due to a lack of statistical power, we used the coalescent

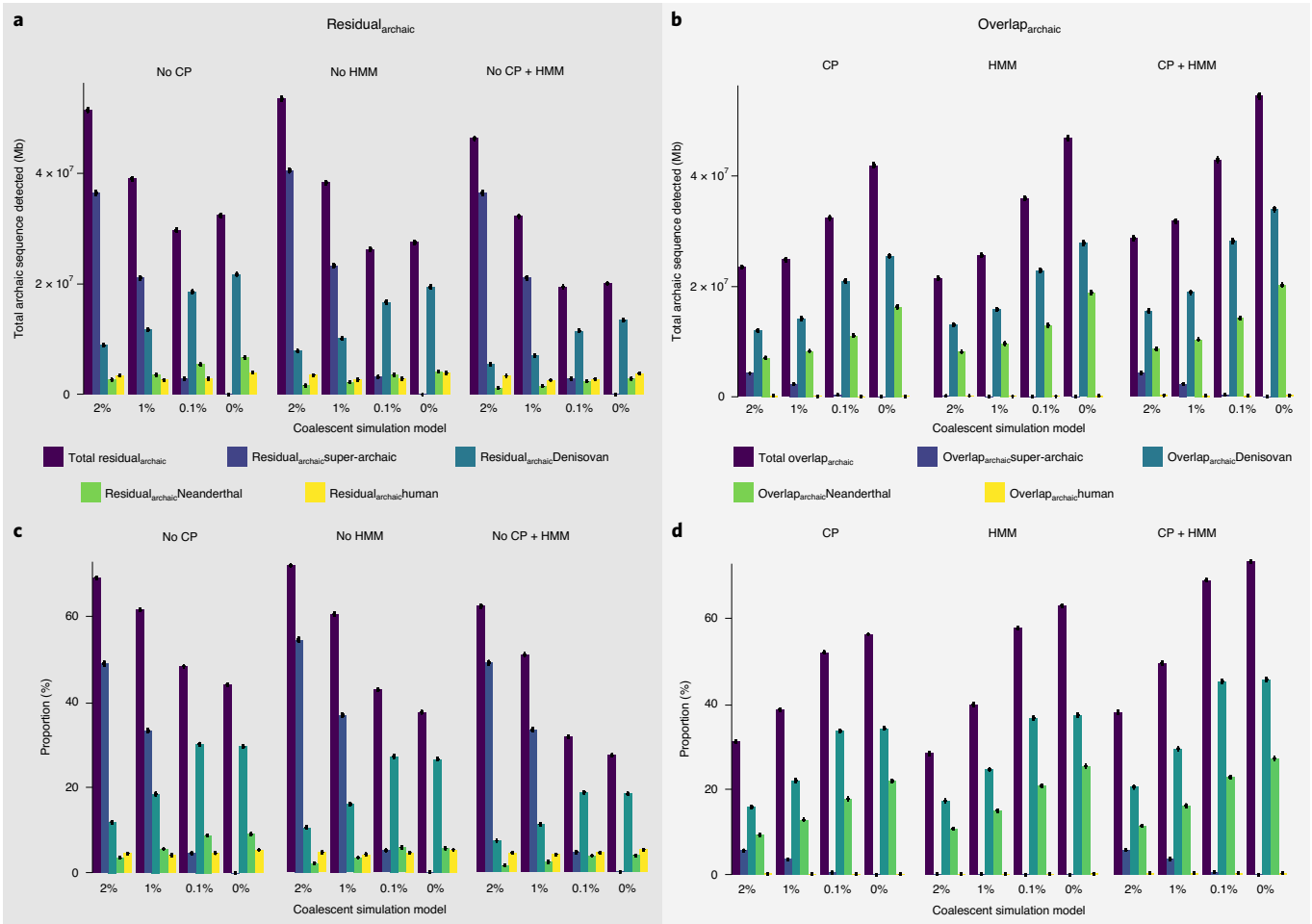

Fig. 2 | Results from coalescent simulations exploring the detection of archaic hominin introgressed sequences. a, Residual$_{archaic}$ after removing Neanderthal and Denisovan regions detected by CP, HMM and CP + HMM. The total residual$_{archaic}$ and the proportion of residual$_{archaic}$ that overlap simulated archaic regions for different species is shown from left to right, together with the amount of residual$_{archaic}$ that overlaps 'human' (non-archaic) regions. Different simulation models of super-archaic introgression are shown on the *x* axis from left to right. **b**, Overlap between regions inferred as archaic, showing the concordance between HMM$_{archaic}$ and the other two methods (overlap$_{archaic}$). **c**, The proportion of residual$_{archaic}$ sequence over the total amount of HMM$_{archaic}$ inferred to be archaic. **d**, The proportion of overlap$_{archaic}$ out of the total amount of HMM$_{archaic}$.

software msprime[41] to simulate Aboriginal Australian and Papuan histories under an empirically informed demographic model[42]. These simulations included separate Neanderthal and Denisovan admixture events along with differing amounts of super-archaic introgression (2%, 1%, 0.1% and 0%) in the common ancestral population of Australo–Papuans (see Methods). We then applied our full analytical pipeline to these simulated genomic datasets to detect super-archaic blocks and quantified the power and false discovery rate for the different levels of super-archaic introgression.

Our simulation results demonstrate that HMM$_{archaic}$ can confidently detect super-archaic blocks even in scenarios with extremely low levels of super-archaic ancestry—with true positive rates (TPRs) ranging from ~50% to ~95% for models with 0.1% and 2% super-archaic ancestry, respectively (Supplementary Fig. 9)—while maintaining extremely low false positive rates (FPRs; Supplementary Fig. 10).

The amount of residual$_{archaic}$ sequences detected per individual in the 0.1% and 0% super-archaic introgression models (~20 Mb; Fig. 2a) is strikingly close to that observed in the Papuan and Australian empirical data (~18 Mb; Fig. 1a). For these models, the majority of the residual$_{archaic}$ signal originates from Neanderthal and Denisovan introgression that went undetected by CP and HMM (Fig. 2a and Supplementary Fig. 12). In contrast, the 1% and 2% super-archaic introgression models detect at least two times more residual$_{archaic}$ sequence per individual than empirical estimates (~33 Mb and ~47 Mb, respectively; Fig. 2a), which was primarily caused by an inflation in the number of super-archaic blocks. Interestingly, the power to detect Neanderthal and Denisovan blocks using HMM$_{archaic}$ is negatively affected by increasing amounts of super-archaic ancestry, as the power of this method is proportionate to the divergence between the introgressing archaic population and the outgroup human population (see Supplementary Fig. 12 and Methods).

Similarly, the mutational motifs observed in the 0.1% and 0% super-archaic introgression models provide a closer fit to the empirical data than do higher levels of super-archaic introgression. For instance, the [1000] and [0111] mutational motifs comprise ~27% and ~6% on average in the empirical data, compared with ~26% and ~6.5% for the 0.1% model, and ~22.5% and ~4% for the 0% model (Supplementary Fig. 13). The close fit of the 0% and 0.1% models to our empirical observations provide strong support for there being little to no introgressed super-archaic sequences in non-African human genomes.

Discussion

The lack of any detectable super-archaic introgression in non-African modern human genomes in our analyses, beyond trace levels indirectly inherited via past admixture with Neanderthals

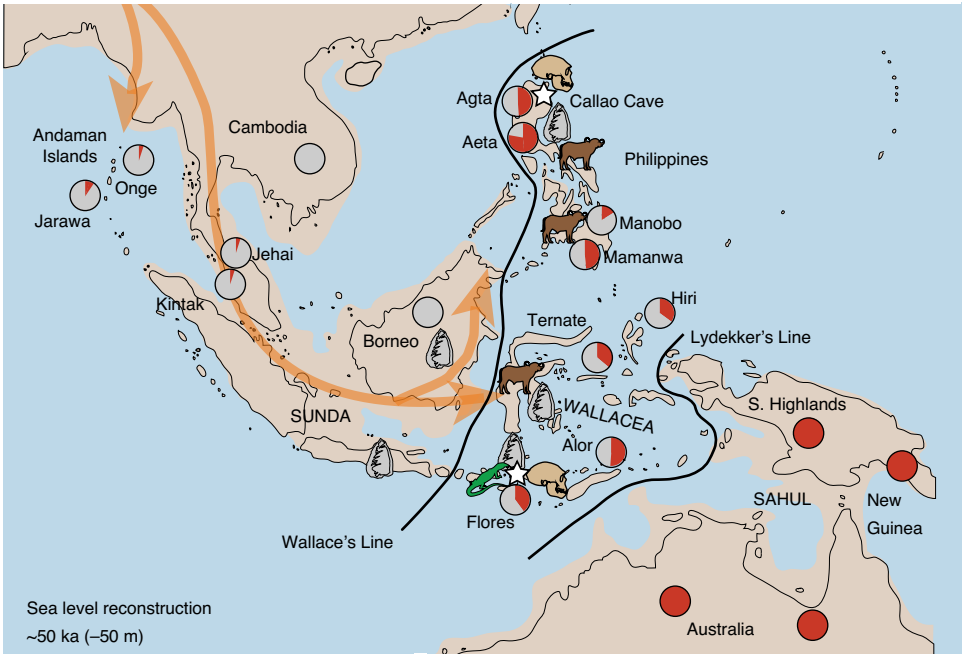

Fig. 3 | Hominin occupation and megafauna survival in ISEA at the time of modern human arrival. Confirmed presence of *H. floresiensis* and *H. luzonensis* is depicted by the skull and white star icons; regions with known artefacts associated with hominin presence are depicted by the stone tool icons; extant native megafauna east of Wallace's (including Huxley's) Line is depicted by the buffalo icon (representing mammals—*Bubalus*, *Rusa*, *Sus* and *Babirusa*) in the northern and southern Philippines and Sulawesi, and Komodo dragon icon on Flores and satellites. Inferred hominin presence covers the entry routes into Sahul, tentatively indicated by the orange arrows. The estimated Denisovan ancestry in modern populations is shown in red in the pie charts, relative to that observed in Australo–Papuan genomes. All populations containing large amounts of Denisovan ancestry are found east of Wallace's Line. Major biogeographic boundaries corresponding to Wallace's Line and Lyddeker's Line are shown as thick black lines and define Wallacea as the region separating the continental Sunda shelf from Sahul. Coastlines are defined as 50 m below present mean sea level, equivalent to the low sea level stand estimated ~50 ka.

and/or Denisovans, stands in stark contrast to the strong evidence of Denisovan admixture with the ancestors of present-day ISEA populations[18–23,43,44]. Based on current palaeoanthropological interpretations of *H. luzonenesis* and *H. floresiensis* as descendants of super-archaic hominin groups, our results indicate that interbreeding between these groups and AMH did not occur. However, we cannot outright reject the possibility that interbreeding did occur if these encounters either did not produce viable progeny or the offspring were viable and that these lineages have since died out. Evidence for super-archaic introgression into the ancestors of the Altai Denisovans[30] and, possibly, Andamanese populations[31–33] suggests that viable reproduction may actually have been possible, although further evaluation of these hypotheses is not possible at present given the available data.

An alternative explanation is that *H. luzonensis* and *H. floresiensis* belong to a hominin clade that is considerably less divergent from AMH than is currently accepted, possibly being the late-surviving descendants of an earlier radiation of a Denisovan-like lineage across ISEA. This would imply that hominin occupation of Flores (>1.01 Ma)[29,45] and the Philippines (from ~700 ka)[46] was not continuous and that the ubiquitous Denisovan ancestry across ISEA results from AMH admixture with one or both of these groups. Indeed, the patterning of Denisovan ancestry across ISEA is consistent with separate Denisovan introgression events in the Philippines[19] and, potentially, in Flores[23,43], the island homes of *H. luzonensis* and *H. floresiensis*, respectively. Further, it is possible that pronounced dwarfism and prolonged periods of endemic island evolution for *H. floresiensis* and *H. luzonensis* have complicated assessments of their morphology and possible phylogenetic relationships. While this explanation would provide a parsimonious answer to the identity of the 'southern' Denisovans, it does not align with the current consensus view that is based on interpretations of archaeological and fossil data[9,11–13,27–29].

A major complication in resolving these questions is the sparse Denisovan fossil record—currently consisting of one phalanx, a mandible, several teeth and some cranial fragments—which makes meaningful morphological comparisons very difficult. Potentially promising areas for further research include Sulawesi, where stone tool records are compatible with possible Denisovan occupation ~100–200 ka[25]. Intriguingly, Sulawesi is home to endemic dwarf buffalos (*Bubalus* spp.) and pigs (*Sus celebensis*, *Babirusa* spp.), which are among the few megafaunal species east of Wallace's Line known to have survived into the Holocene. Patterns of megafaunal survival in eastern ISEA coincide with known areas of pre-AMH hominin occupation, and include the living Komodo dragon (*Varanus komodoensis*) on Flores and its satellites, and surviving buffalo (*Bubalus mindorensis*), pigs (*Sus* spp.) and deer (*Rusa* spp.) in the oceanic Philippines (Fig. 3). This pattern suggests that long-term exposure to possible hunting pressures by archaic hominins might have facilitated the survival of megafaunal species in subsequent contacts with AMH. Therefore, such islands are good candidates for future research efforts to recover evidence of the elusive 'southern' Denisovans. Another intriguing (albeit unlikely) possibility is Australia, where the ~65 ka artefacts uncovered at Madjedbebe[6] might be associated with Denisovan presence.

Clearly, further resolution of hominin prehistory of ISEA will greatly benefit from direct fossil and archaeological evidence of Denisovan presence in the region, with the potential for proteomic studies to assist in resolving phylogenetic relationships where DNA is not recoverable. Nonetheless, the current fossil and archaeological

records, together with mounting genetic evidence from across ISEA, point to the widespread presence of archaic hominins east of Wallace's Line[26], and indicate that the first AMH populations to arrive in ISEA most probably encountered a variety of hominin populations, no matter which route they took to enter Sahul[47–52]. This hints that much of the Denisovan ancestry found in modern human populations in ISEA, New Guinea and Australia may have been locally acquired, emphasizing the need for more archaeological and genetic research across this understudied region in the future.

Methods

Samples. We examined 426 individuals from ten distinct populations (Supplementary Table 1), taking advantage of publicly available data from previous genomic studies and a recent effort to sequence hundreds of Indonesian genomes through the Indonesian Genome Diversity Project[22]. For a description of data preparation (single nucleotide polymorphism calling, quality control, phasing) see ref. [22].

Searching for signals of super-archaic admixture into modern humans. We searched for signals of super-archaic introgression in genomic sequences of AMH populations across the world, with a particular focus on ISEA and New Guinea (descendants from early AMH migrations into the region). These specific signatures are expected to include the existence of genetic variants that are not observed in Africa and that exhibit levels of linkage disequilibrium compatible with introgression events ~60–50 ka, similarly to observations for Neanderthal and Denisovan introgressed segments. However, we expect deep divergence times between extinct ISEA hominins (*H. luzonensis* and *H. floresiensis*) and *H. sapiens* if we consider the former are not part of the Denisovan/Neanderthal clade and are instead related to *H. erectus*, or represent additional *Homo* lineages that split from AMH ~2 Ma or earlier. Hence, the putative introgressed super-archaic regions are expected to be highly divergent to orthologous modern human genome sequences. Importantly, the absence of a genome sequence for the extinct ISEA hominin groups makes this inference far more complex than for Neanderthal or Denisovan introgression, for which reference genomes are available. Therefore, we searched for super-archaic introgression in the genomes of contemporary human populations around the world using a highly powerful hidden Markov chain model implemented in ref. [34] (termed here $HMM_{archaic}$), which is agnostic to the genome sequence of the putative archaic source. The rationale behind this strategy is that introgressed regions of the genome are enriched for genetic variants not seen in populations that have not admixed with the putative archaic source. In this case, we used African populations as an outgroup and assumed that these African populations have not interbred with Neanderthals, Denisovans, or any super-archaic source. It should be noted that the class of methods to which $HMM_{archaic}$ belongs are only indicative of archaic introgression. These methods might be prone to false positive detection of introgressed fragments due to incomplete lineage sorting or balancing selection maintaining old genetic diversity at specific selected loci. The $HMM_{archaic}$ method infers archaic admixture using a sliding window approach after controlling for genetic diversity existing in an outgroup (for example, African populations). We applied the method across all individuals from each of the ten sampled populations, using as an outgroup all individuals belonging to every African population contained in our dataset. After this, we further excluded positions where the Altai Neanderthal and Altai Denisovan individuals are heterozygous. We set the initial parameters to run $HMM_{archaic}$ following the author's implementation, specifically: states = ['Human', 'Archaic']; starting_probabilities = [0.98, 0.02]; transitions = [[0.9995,0.0005],[0.012,0.98]], emissions = [0.04, 0.1]. Importantly, the method can be applied to phased data, and hence extract putative introgressing haplotypes rather than unphased regions, allowing for downstream analysis that is more sensitive to the independent histories of homologous chromosomal regions. Hence, the model was trained and implemented on phased data, which was obtained as described in ref. [22]. We used a 1,000 base pair (bp) sliding window approach, as performed in the original implementation of the method[34], as the small size of the sliding windows across the genome allows a fine-scale resolution of even small introgressed fragments where other methods[35–37] are likely to fail.

The $HMM_{archaic}$ method outputs a posterior probability of introgression for each 1,000 bp window along each chromosome copy of each individual sample. These are called either 'Human' or 'Archaic' blocks, with each archaic block having posterior support >0.5; however, as we wish to focus on high-confidence introgressed blocks, we decided to drop archaic blocks with posterior probability support ≤0.95. Therefore, the archaic blocks we examined were all regions directly estimated from $HMM_{archaic}$ with posterior probability >0.95, with no further changes such as merging of the inferred archaic blocks.

Identifying Denisovan and Neanderthal introgressed fragments. We first sought to detect genomic signals of Neanderthal and Denisovan introgression using the CP[35] and HMM[36,37] introgression detection methods described in ref. [22]. These methods use phased data and seek to define haplotype blocks that are introgressed from an evolutionary relative of a sampled archaic genome, by detecting regions with a high density of variants that are shared with the archaic genome but not observed in an African outgroup sample. All parameters and details of the method implementations are given in ref. [22].

Obtaining $residual_{archaic}$ blocks. We then focused on regions inferred to be introgressed using $HMM_{archaic}$[34], which contain the introgressed fragments from Neanderthals and Denisovans and, potentially, additional introgressed signals not captured by CP or HMM. By subtracting the introgressed regions inferred to be of Neanderthal or Denisovan origin from CP and HMM, we produced a residual $HMM_{archaic}$ signal ($residual_{archaic}$) of blocks not overlapping Neanderthal or Denisovan fragments inferred with the other two methods. Specifically, for overlapping fragments, we subtract the overlapping $HMM_{archaic}$–CP/HMM regions, while still retaining the non-overlapping regions (refer to Supplementary Fig. 8 for an illustration). This approach is allied to the residual sequence obtained from the S^* statistic, as calculated in ref. [22], but differs in using more accurate phased archaic calls from $HMM_{archaic}$ and in the detail of the $residual_{archaic}$ block calling process. Note that identified $residual_{archaic}$ blocks may be in close proximity to Denisovan or Neanderthal introgressed regions (as is the case in Supplementary Fig. 8) and that these blocks are not suitable for some downstream analyses such as introgression time estimation based on introgressed block length, as they may correspond to subparts of larger introgressed blocks. We decided to adopt this strategy as there is potential for super-archaic blocks, in case they are present, to segregate close to, or overlap with, Neanderthal and Denisovan fragments, given the potential for non-random segregation of archaic blocks within the genome. While in the current work we do not present the results for an alternative strategy of completely removing $HMM_{archaic}$ blocks intersecting Neanderthal and Denisovan blocks to estimate $residual_{archaic}$, the findings are qualitatively similar to the ones presented here.

Looking at patterns of variation within $residual_{archaic}$ blocks. To further disentangle the patterns seen in $residual_{archaic}$ blocks, we looked at mutation motif patterns. We defined the mutation motifs as 0 (ancestral) and 1 (derived), and a combination of [X, D, N, H], where X represents the allelic state of a particular individual within an introgressed block (which can also be thought of as the test population—that is, Papuan, east ISEA, west ISEA and so on), D represents Denisova, N represents Neanderthal and H represents an individual from an African population (in our case Ju|'hoan—SS6004473). While all African variation was removed from the dataset prior to running $HMM_{archaic}$ (as Africans form the required outgroup), we reintroduced SS6004473 variation subsequently and only for this specific analysis. This means that, for example, the mutation motif [1001] is seen when X shares the derived allele with the African individual, and Neanderthal and Denisovan are ancestral; likewise, the mutation motif [1000] indicates regions where X carries a derived allele that is not observed in the African individual, Neanderthal or Denisovan. Hence, in the case of super-archaic introgression into modern humans, an enrichment in [1000] and [0111] motifs within introgressed blocks is to be expected.

Variation in motif proportion as a function of physical distance to introgressed regions. We investigated the proportion of different motifs as a function of physical distance to the putatively introgressed regions. In this case we divided the analyses into patterns seen within all $HMM_{archaic}$ introgressed fragments and those seen in $residual_{archaic}$ fragments (Supplementary Fig. 7). In this analysis, we define mutation motifs as [X, D, N, Af] where a single human outgroup is now represented by an indicator Af, 1 indicates that a variant is found in the derived state in one or more individuals in the African outgroup, and 0 indicates that the derived state is not observed. Thus, we are specifically focusing on whether variation is found at all in an African sample rather than a single African individual. When all $HMM_{archaic}$ fragments within the Papuan population are considered, we observe an excess of [1100] and [1010] motifs, compatible with introgression from Denisovan and Neanderthal into Papuan genomes, respectively, along with a sharp decrease of [1001] (where X shares a derived allele with Africa) motifs. These signatures consistently indicate Neanderthal and Denisovan introgression into Papuan genomes. When considering only $residual_{archaic}$ fragments, we observe a sharp increase in the [1000] motif (as expected) coupled with a reduction in the [1100] and the [1010] motifs (signals of Denisovan and Neanderthal introgression, respectively), suggesting that the remaining fragments do not show a clear signal of known archaic introgression. These Neanderthal and Denisovan signals increase in the regions around $residual_{archaic}$ blocks, indicating that they are often nested within introgressed Neanderthal and Denisovan sequences. This is an important observation, suggesting that much of the signal is contributed through known introgression, in support of the absolute increase in $residual_{archaic}$ in Papuan populations. Indeed, the definition of $residual_{archaic}$ does not exclude the detection of regions showing coalescent histories consistent with super-archaic introgression from within Denisovan and Neanderthal introgression (as would probably be the case for the blocks shown in example schematic Supplementary Fig. 8), and variation in the coalescent histories within blocks sharing the same introgression source is likely. While this suggests that $residual_{archaic}$ blocks may be retrieving super-archaic signals from within Denisovan and Neanderthal introgressing populations, we suggest that more data and more focused analysis, beyond the scope of this paper, are necessary to assess the significance of these patterns. The

sharp decrease in the [1001] motif observed in all $HMM_{archaic}$ blocks is replaced by a peak in $residual_{archaic}$ blocks, and a slight increase in the [0111] motif is now visible. In both cases, these indicate deep coalescence of $residual_{archaic}$ blocks not associated with the sampled Neanderthal or Denisovan sequences. While the [0111] signal is of particular interest in the context of super-archaic introgression, the lack of any global peaks in this motif (Supplementary Fig. 2) and elevated [1100] and [1010] signals surrounding $residual_{archaic}$ blocks argues that it more likely reflects deep coalescent histories within Denisovan and Neanderthal introgressed blocks than super-archaic introgression.

Motif proportion differences are correlated with known archaic ancestry. We explicitly test for a correlation between Neanderthal and Denisovan ancestry and motif proportions within $residual_{archaic}$ blocks between populations. Supplementary Figs. 3 and 4 show the correlation between inferred Denisovan and Neanderthal ancestry, respectively, and the proportion of different motifs, across all individuals. Interestingly, we find both positive and negative correlations between the proportion of different motifs and the detected amount of Denisovan and Neanderthal ancestry. In fact, these correlations are statistically significant for all but two motifs when regressing on Denisovan ancestry, [0100] (*P* value 0.289) and [1110] (*P* value 0.618), and for all but one motif when regressing on Neanderthal ancestry, [1110] (*P* value 0.221). These results are in agreement with the observations from simulations with no super-archaic introgression, which show that $residual_{archaic}$ sequence is essentially dominated by introgressed Neanderthal and Denisovan fragments that are undetected by both HMM and CP.

Comparing $HMM_{archaic}$ and $residual_{archaic}$ with predicted super-archaic regions. A recent study[40] proposes that a diminutive proportion of super-archaic ancestry survives in contemporary human populations due to introgression events between a highly divergent hominin and the ancestors of Neanderthals and Denisovans, who subsequently admixed with the ancestors of present-day people. To investigate whether our strategy allowed for the detection of rare super-archaic fragments, we contrasted the inferred $HMM_{archaic}$ and $residual_{archaic}$ blocks per individual with the putatively super-archaic fragments introgressed via Neanderthals and Denisovans proposed in ref. [40]. Specifically, we identified all the instances where an $HMM_{archaic}$ or a $residual_{archaic}$ block identified in each individual in our sample overlaps (even if only partially) a super-archaic fragment predicted in ref. [40]. We then counted the number of individuals containing at least one $HMM_{archaic}$ or $residual_{archaic}$ block overlapping each super-archaic fragment (that is, the overlap of each individual was counted only once per fragment, even in cases with multiple overlaps with the same fragment). After this, we combined all the overlaps across individuals and estimated the percentage of $HMM_{archaic}$ and $residual_{archaic}$ overlap per super-archaic fragment over the total length of the fragment. The results are reported in Supplementary Tables 2–5.

Simulating super-archaic introgression using msprime. To test the power of our experimental design to detect introgression from a highly diverged human lineage into the ancestors of ISEA populations/Australo–Papuans, we implemented a series of neutral coalescent simulations using the software msprime[41]. The simulations use demographic parameters derived from ref. [42], which models Aboriginal Australian history from full genome data from modern Australian and Papuan populations. The structure and parameters describing the standard demography (that is, excluding possible super-archaic introgression) followed a maximum likelihood model output. Briefly, we simulated a total of 35 African and 30 Australian individuals, and one Altai Denisovan individual that split from human populations 20,255 generations prior to the present, whereas African and Australian populations split from one another 3,916 generations ago. Additionally, we included one super-archaic individual that splits from the Human–Neanderthal–Denisova clade 70,000 generations in the past to mimic the deep split assumed for *H. floresiensis* and *H. luzonensis*, with haploid effective population size (Ne) = 13,249. Following ref. [42], Neanderthal (2.4%) and Denisovan (4.0%) introgression events were simulated 1,853 and 1,353 generations in the past, respectively, with the introgressing lineages being related to the Altai individuals, and additional minor Neanderthal contributions to the Eurasian clade (1.1%) and Australian clade (0.2%) 1,566 and 883 generations ago, respectively. For the super-archaic admixture, we assumed an introgression event occurring 1,353 generations ago. We set the mutation rate to 1.4×10^{-8} bp^{-1} per generation and the recombination rate to 1×10^{-8} bp^{-1} generation and simulated, per individual, a total of 300 chromosomes of 10 Mb in length each. This strategy allowed us to obtain a total simulated sequence that roughly matches the size of the human genome for each individual (~3 Gb of sequence), while ensuring sufficient independent replication. Importantly, after running the simulations, we sampled 65 human individuals (35 African and 30 Australian genomes), an Altai Neanderthal and an Altai Denisovan (related to, respectively, the introgressing Neanderthal and Denisovan populations), and one super-archaic individual.

A major advantage of using msprime to implement coalescent simulations is that the software allows the genealogy of each portion of simulated sequence to be traced back through time, including the migration of genomic regions between archaic and human populations (introgression). This means that, for each individual, we are able to know the exact amount and location of the introgressed segments, and are thus able to directly compute the strength of our approaches for detecting super-archaic introgression in the empirical data.

Models of super-archaic introgression. We initially implemented two models of super-archaic introgression: a model containing 2% introgression into the ancestors of Australians occurring at the same time as Denisovan introgression; and a second model without super-archaic introgression (0%). To estimate the power of our analytical framework to detected super-archaic introgression at low levels of admixture, for each simulated individual we created datasets with ~1% and ~0.1% super-archaic introgression by masking a specific proportion of super-archaic blocks in the 2% model. Specifically, this was achieved by (1) randomly sampling a proportion of introgressed super-archaic regions in each individual; and 2) merging all the regions sampled across all individuals and masking these merged super-archaic regions across all simulated individuals. This strategy ensured that the masked super-archaic regions were the same across all individuals. We were able to reduce the amount of super-archaic ancestry present in the simulated sequences to ~1% and ~0.1% by randomly sampling, per individual, ~10% and ~50% of introgressed super-archaic regions, respectively. Owing to the masking of the introgressed regions, the 1% and 0.1% models contained slightly less genetic sequence than the 0% and 1% models (~2.88 Gb and ~2.65 Gb simulated sequence, respectively); however, the masking did not alter the average proportion of introgressed sequences observed from either the Denisovan or Neanderthal lineages (Supplementary Fig. 12).

Power to uncover archaic introgression. We evaluated the performance of the analytical pipeline by comparing the results from our empirical data with four models of Australian super-archaic admixture at different introgression levels (2%, 1%, 0.1% and 0%). First, we estimated the power of each of the three detection methods used to compute archaic introgression in the empirical data; that is, CP, HMM and $HMM_{archaic}$. Analogous to the implementation in the empirical data, before running $HMM_{archaic}$, we excluded all variation present in the 35 simulated African genomes, along with positions for which the Altai Neanderthal and Denisovan individuals were heterozygous. Supplementary Fig. 9 shows the TPR of each method to detect archaic introgression. The TPRs were estimated as the length of detected regions that overlap the simulated introgressed regions over the total length of simulated introgressed regions (in base pairs). It was possible to estimate the TPR separately for introgression from the Neanderthal and Denisovan lineages for CP and HMM, although not for $HMM_{archaic}$ (which does not require a reference).

Both CP and HMM consistently detect Neanderthal introgression at a higher rate than Denisovan introgression, irrespective of the amount of super-archaic introgression present in the simulations (Supplementary Fig. 9). Considering that both CP and HMM rely on the availability of a reference sequence for the putatively introgressing archaic population, this observation is consistent with the fact that the simulated introgressing Neanderthal population is genetically closer to the reference Altai Neanderthal than the simulated introgressing Denisovan population is to the reference Altai Denisovan. Nevertheless, both methods seem to perform only slightly better in the absence of super-archaic introgression, presumably because, at least in the case of CP, a very small proportion of inferred Neanderthal and Denisovan introgression derives from super-archaic introgression (see 'Effects of super-archaic ancestry to detect Neanderthal/Denisovan introgression'). $HMM_{archaic}$ has extremely high power to detect super-archaic segments (Supplementary Fig. 9, top left) and, even though power decreases at lower levels of super-archaic introgression, it is always higher than the detection power for Neanderthal or Denisovan introgression across all four models (Supplementary Fig. 9).

FPR. We next examined the FPR of each method to detect archaic introgression. For the CP and HMM methods we define FPR as the proportion of sequence misassigned to a particular archaic population when that sequence is either introgressed from another hominin lineage or is from the human genealogy. For $HMM_{archaic}$ we simply estimated the proportion of sequence misassigned as archaic that overlaps simulated human regions. The results are shown in Supplementary Figs. 10 and 11. Both CP and HMM have relatively high FPRs when inferring Neanderthal introgression that actually results from Denisovan introgression (~40%), and vice versa (~35%; Supplementary Fig. 11). As expected, given the closer relationship of the introgressing Neanderthal population to the reference Altai Neanderthal compared with the introgressing Denisovan population to the reference Altai Denisovan, the FPR for CP and HMM is higher for Denisovan segments that were missasigned as Neanderthal (Supplementary Fig. 11, right panel) than vice versa (Supplementary Fig. 11, left panel). This pattern is also consistent with a close genetic relationship between Neanderthals and Denisovans, and the persistence of shared ancestral genetic diversity between the two species (incomplete lineage sorting). Importantly, however, the FPR of both methods is extremely low when inferring Neanderthal or Denisovan introgression when it either did not occur (Supplementary Fig. 11, middle columns, 'human') or the source was super-archaic (Supplementary Fig. 11, left columns, 'super-archaic'). Hence, our simulation results demonstrate that a negligible amount of introgressed super-archaic sequence will be mistaken for Neanderthal or Denisovan

introgression by CP and HMM. Finally, our stringent approach for detecting archaic introgression using $HMM_{Archaic}$ (posterior probability >0.95, see 'Searching for signals of super-archaic admixture into modern humans') results, as expected, in virtually no false positives in the simulations (Supplementary Fig. 10)—that is, a negligible portion of archaic $HMM_{archaic}$ overlaps with human genealogies.

Estimation of $residual_{archaic}$. We next investigated how this combination of TPRs and FPRs translated into the actual amount of recovered sequence. The results are shown in Supplementary Fig. 12, contrasting the total amount of simulated introgression versus the total amount detected for each archaic species using the different methods. Notably, the amount of Neanderthal and Denisovan introgression detected by $HMM_{archaic}$ consistently increases as the amount of super-archaic ancestry declines (see Effects of super-archaic ancestry to detect Neanderthal/Denisovan introgression). In contrast, the amount of Neanderthal and Denisovan detected by both CP and HMM is essentially independent from the amount of super-archaic ancestry present (as expected from the TPRs shown in Supplementary Fig. 9). As described above, the masking strategy adopted to reduce the amount of super-archaic in the simulations meant that models 1% and 0.1% contain a reduced amount of introgressed Neanderthal and Denisovan sequence overall (see explanation in Power to uncover archaic introgression).Therefore, we also present a corrected amount of simulated and detected archaic sequences by normalizing the total amounts to match the total amount of sequence considered in the empirical data (Supplementary Fig. 12b). This strategy also allowed us to compare the simulations directly with the results obtained for the empirical data, namely in terms of total $residual_{archaic}$ sequence present. After determining the total detected sequence in each method, we obtained the $residual_{archaic}$ regions by removing those regions that overlap with either the CP or HMM detected blocks ($residual_{archaic}$ in Fig. 2, overlapping blocks shown as $overlap_{archaic}$).

Effects of super-archaic ancestry to detect Neanderthal/Denisovan introgression. An interesting picture emerges when we consider the behaviour of $HMM_{archaic}$ in the presence of super-archaic introgression. The ability of $HMM_{archaic}$ to detect Neanderthal and Denisovan introgression is severely depleted at higher levels of super-archaic introgression, which appears to dominate the amount of detected archaic ancestry: less than 25% of truly introgressed Neanderthal and Denisovan sequences were detected when we simulate 2% super-archaic introgression, versus ~40–60% true rates for a model containing 0% super-archaic introgression (Supplementary Fig. 9, top panel). This pattern is consistent with the power of $HMM_{archaic}$ being proportionate to the divergence between the introgressing archaic population and the outgroup human population (that is, Africa). Importantly, we have simulated a super-archaic source whose divergence to modern humans is substantially higher than that of Neanderthals and Denisovans to mimic introgression from *H. floresiensis* and *H. luzonensis*, assuming that the latter are earlier diverging lineages of *Homo*. There is a considerably higher agreement between $HMM_{archaic}$ and both CP and HMM for a model with no super-archaic introgression compared with a model containing even 0.1% super-archaic introgression (Fig. 2). The most important signal for differentiating these scenarios, which have similar total simulated $residual_{archaic}$, is the concordance between $HMM_{archaic}$ and CP/HMM. Specifically, the excess divergence of super-archaic introgressed sequences means these blocks contain a higher amount of non-African variants and, therefore, are more efficiently detected by $HMM_{archaic}$. However, this process simultaneously impacts the internal optimization of $HMM_{archaic}$ emission parameters, causing the algorithm to seek more divergent introgressed blocks, which reduces the TPR for detecting known Denisovan and Neanderthal blocks. This is consistent with $HMM_{archaic}$ having a higher TPR for introgressed Neanderthal and Denisovan sequences when no super-archaic introgression is present, which in turn leads to a higher amount of Neanderthal and Denisovan sequence detected by all three methods (Fig. 2). This behaviour causes the concordance between methods to drop, and the $residual_{archaic}$ signal to increase as a proportion of total $HMM_{archaic}$, even when simulating minimal amounts of super-archaic introgression. The higher concordance between $HMM_{archaic}$, CP and HMM for the 0% model translates into a 27% proportion of $residual_{archaic}$ in this model (Fig. 2c), consistent with $residual_{archaic}$ regions computed in the empirical data (between ~15% in Papuan genomes and ~22% in West Eurasian genomes; Supplementary Fig. 1) and in contrast to ~33–60% for models with ≥0.1% super-archaic introgression. Importantly, in simulations containing higher proportions of super-archaic ancestry (1% and 2% models), we observe a much higher proportion of $residual_{archaic}$ sequence.

Investigating mutation motifs in $residual_{archaic}$ simulated models. To further investigate the nature of genetic diversity within $residual_{archaic}$ regions, we performed similar mutation motif analyses to those used in the empirical data (see 'Looking at patterns of variation within $residual_{archaic}$ blocks'). In particular, we investigated the amount of shared ancestral and derived alleles between individuals carrying the residual sequence (that is, test population), the simulated Altai Denisovan, the simulated Altai Neanderthal and a simulated African genome—again, while all African variation was excluded from $HMM_{archaic}$ analyses, we randomly sampled one individual and investigated allele sharing within $residual_{archaic}$ regions after running the method.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The genetic datasets analysed during the current study were downloaded from, and are available at, the European Genome-phenome Archive (accession number EGAS00001003054; https://www.ebi.ac.uk/ega/home) and the Estonian Biocentre data archive (http://evolbio.ut.ee).

Received: 24 July 2020; Accepted: 3 February 2021;

References

1. Kaifu, Y. Archaic hominin populations in Asia before the arrival of modern humans: their phylogeny and implications for the 'southern Denisovans'. *Curr. Anthropol.* **58**, S418–S433 (2017).
2. Morwood, M. J., O'Sullivan, P., Susanto, E. E. & Aziz, F. Revised age for Mojokerto 1, an early *Homo erectus* cranium from east Java, Indonesia. *Aust. Archaeol.* **57**, 1–4 (2003).
3. Rizal, Y. et al. Last appearance of *Homo erectus* at Ngandong, Java, 117,000–108,000 years ago. *Nature* **577**, 381–385 (2020).
4. Matsu'ura, S. et al. Age control of the first appearance datum for Javanese in the Sangiran area. *Science* **367**, 210–214 (2020).
5. Westaway, K. E. et al. An early modern human presence in Sumatra 73,000–63,000 years ago. *Nature* **548**, 322–325 (2017).
6. Clarkson, C. et al. Human occupation of northern Australia by 65,000 years ago. *Nature* **547**, 306–310 (2017).
7. Bowdler, S. 'Human occupation of northern Australia by 65,000 years ago' (Clarkson et al. 2017): a discussion. *Aust. Archaeol.* **83**, 162–163 (2017).
8. O'Connell, J. F. et al. When did *Homo sapiens* first reach Southeast Asia and Sahul? *Proc. Natl Acad. Sci. USA* **115**, 8482–8490 (2018).
9. Brown, P. et al. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
10. Sutikna, T. et al. Revised stratigraphy and chronology for *Homo floresiensis* at Liang Bua in Indonesia. *Nature* **532**, 366–369 (2016).
11. Détroit, F. et al. A new species of *Homo* from the Late Pleistocene of the Philippines. *Nature* **568**, 181–186 (2019).
12. Brown, P. & Maeda, T. Liang Bua *Homo floresiensis* mandibles and mandibular teeth: a contribution to the comparative morphology of a new hominin species. *J. Hum. Evol.* **57**, 571–596 (2009).
13. Argue, D., Groves, C. P., Lee, M. S. Y. & Jungers, W. L. The affinities of *Homo floresiensis* based on phylogenetic analyses of cranial, dental, and postcranial characters. *J. Hum. Evol.* **107**, 107–133 (2017).
14. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
15. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
16. Chen, F. et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature* **569**, 409–412 (2019).
17. Zhang, D. et al. Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. *Science* **370**, 584–587 (2020).
18. Reich, D. et al. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
19. Jinam, T. A. et al. Discerning the origins of the Negritos, First Sundaland People: deep divergence and archaic admixture. *Genome Biol. Evol.* **9**, 2013–2022 (2017).
20. Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S. & Akey, J. M. Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* **173**, 53–61.e9 (2018).
21. Mondal, M., Bertranpetit, J. & Lao, O. Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nat. Commun.* **10**, 246 (2019).
22. Jacobs, G. S. et al. Multiple deeply divergent Denisovan ancestries in Papuans. *Cell* **177**, 1010–1021.e32 (2019).
23. Teixeira, J. C. & Cooper, A. Using hominin introgression to trace modern human dispersals. *Proc. Natl Acad. Sci. USA* **116**, 15327–15332 (2019).
24. Bergström, A. et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
25. van den Bergh, G. D. et al. Earliest hominin occupation of Sulawesi, Indonesia. *Nature* **529**, 208–211 (2016).
26. Wallace, A. R. On the physical geography of the Malay Archipelago. *Proc. R. Geographical Soc. Lond.* **33**, 217–234 (1863).
27. Kaifu, Y. et al. Craniofacial morphology of *Homo floresiensis*: description, taxonomic affinities, and evolutionary implication. *J. Hum. Evol.* **61**, 644–682 (2011).
28. Kaifu, Y. et al. Unique dental morphology of *Homo floresiensis* and its evolutionary implications. *PLoS ONE* **10**, e0141614 (2015).
29. van den Bergh, G. D. et al. *Homo floresiensis*-like fossils from the early Middle Pleistocene of Flores. *Nature* **534**, 245–248 (2016).

30. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
31. Mondal, M. et al. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat. Genet.* **48**, 1066–1070 (2016).
32. Skoglund, P., Mallick, S., Patterson, N. & Reich, D. No evidence for unknown archaic ancestry in South Asia. *Nat. Genet.* **50**, 632–633 (2018).
33. Mondal, M., Casals, F., Majumder, P. P. & Bertranpetit, J. Reply to 'No evidence for unknown archaic ancestry in South Asia'. *Nat. Genet.* **50**, 1637–1639 (2018).
34. Skov, L. et al. Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet.* **14**, e1007641 (2018).
35. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
36. Seguin-Orlando, A. et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
37. Racimo, F. et al. Archaic adaptive introgression in *TBX15/WARS2*. *Mol. Biol. Evol.* **34**, 509–524 (2017).
38. Berger, L. R. et al. *Homo naledi*, a new species of the genus *Homo* from the Dinaledi Chamber, South Africa. *eLife* **4**, e09560 (2015).
39. Key, F. M., Teixeira, J. C., de Filippo, C. & Andrés, A. M. Advantageous diversity maintained by balancing selection in humans. *Curr. Opin. Genet. Dev.* **29**, 45–51 (2014).
40. Hubisz, M. J., Williams, A. L. & Siepel, A. Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. *PLoS Genet.* **16**, e1008895 (2020).
41. Kelleher, J., Barton, N. H. & Etheridge, A. M. Coalescent simulation in continuous space. *Bioinformatics* **29**, 955–956 (2013).
42. Malaspinas, A.-S. et al. A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016).
43. Tucci, S. et al. Evolutionary history and adaptation of a human pygmy population of Flores Island, Indonesia. *Science* **361**, 511–516 (2018).
44. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
45. Brumm, A. et al. Age and context of the oldest known hominin fossils from Flores. *Nature* **534**, 249–253 (2016).
46. Ingicco, T. et al. Earliest known hominin activity in the Philippines by 709 thousand years ago. *Nature* **557**, 233–237 (2018).
47. O'Connell, J. F. & Allen, J. The restaurant at the end of the universe: modelling the colonisation of Sahul. *Aust. Archaeol.* **74**, 5–31 (2012).
48. Kealy, S., Louys, J. & O'Connor, S. Islands under the sea: a review of early modern human dispersal routes and migration hypotheses through Wallacea. *J. Isl. Coast. Archaeol.* **11**, 364–384 (2016).
49. Kealy, S., Louys, J. & O'Connor, S. Reconstructing palaeogeography and inter-island visibility in the Wallacean Archipelago during the likely period of Sahul colonization, 65–45000 years ago. *Archaeol. Prospect.* **24**, 259–272 (2017).
50. Kealy, S., Louys, J. & O'Connor, S. Least-cost pathway models indicate northern human dispersal from Sunda to Sahul. *J. Hum. Evol.* **125**, 59–70 (2018).
51. Bird, M. I. et al. Palaeogeography and voyage modeling indicates early human colonization of Australia was likely from Timor-Roti. *Quat. Sci. Rev.* **191**, 431–439 (2018).
52. Bird, M. I. et al. Early human settlement of Sahul was not an accident. *Sci. Rep.* **9**, 8220 (2019).

Acknowledgements

We thank C. Huber and J. Schmidt for useful discussions on the genetic analyses, and K. Mitchell and F. Racimo for comments on the manuscript. This work was supported by ARC Indigenous Discovery Grant IN180100017 (J.C.T. and R.T.) and ARC Laureate Fellowships FL100100195 (C.S.M.T.) and FL140100260 (A.C.). C.S. acknowledges funding from the Calleva Foundation and The Human Origins Research Fund. G.S.J. acknowledges support from a Presidential Postdoctoral Fellowship from Nanyang Technological University.

Author contributions

J.C.T., K.M.H., G.S.J., M.P.C., G.H., G.A.P. and J.T. designed the methods and undertook the analyses. J.C.T., K.M.H., R.T., G.S.J., H.S., A.C., C.S., C.S.M.T. and M.P.C. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-021-01408-0.

Correspondence and requests for materials should be addressed to J.C.T.

Peer review information *Nature Ecology & Evolution* thanks Laurits Skov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): João C. Teixeira

Last updated by author(s): Nov 4, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software used to collect data
Data analysis	Software msprime used for coalescent simulations (Kelleher et al. 2016). Archaic introgression inferred using publicly available methods: ChromoPainter (CP - Lawson et al. 2012); a Hidden Markov Model (HMM - Jacobs et al. 2019); and HMMArchaic (Skov et al. 2018).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data analysed during this study are included in Jacobs et al. 2019 (https://doi.org/10.1016/j.cell.2019.02.035).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Analyses on publicly available genetic data to investigate super archaic introgression compatible with admixture between H. floresiensis and/or H. luzonensis and modern humans.
Research sample	Analysis of human genetic data from world-wide populations reported in previous studies, and available in Jacobs et al. 2019. The genetic dataset also comprises the Altai Neanderthal (Prüfer et al. 2016) and Denisovan (Meyer et al. 2012) genomes.
Sampling strategy	We examined 426 individuals from 10 distinct world-wide human populations.
Data collection	No data was collected in this study.
Timing and spatial scale	World-wide human populations were included in our study to ensure the accurate identification of archaic segments and to distinguish signatures of introgression between different populations. The main geographical focus of our study is Island Southeast Asia and New Guinea.
Data exclusions	No data were excluded.
Reproducibility	We searched for genomic signatures compatible with introgression from archaic hominins using three different publicly available methods: ChromoPainter (CP - Lawson et al. 2012); a Hidden Markov Model (HMM - Jacobs et al. 2019); and HMMArchaic (Skov et al. 2018). Each of these methods was applied independently in different human populations (Jacobs et al. 2019). We validated our experimental approach with coalescent simulations using msprime (Kelleher et al. 2016). A detailed explanation of the methods, including all the parameters used in both empirical data analyses and coalescent simulations is provided in the Methods section.
Randomization	Genetic samples were grouped according to geographical location.
Blinding	We undertook a range of methods to ensure blind analysis including SNP calling and phasing.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging