Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	A brief, numerical description of file contents. i.e.: Supplementary Figures 1-4, Supplementary Discussion, and Supplementary
Supplementary Information	Yes.	Supplementary_Figures_Teixeira_et_al_2021.pdf	Supplementary Figures S1- S15.
Reporting Summary	Yes.	Teixeira_nr-reporting-summary_revised.pdf	

1

Туре	Number If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: Smith_ Supplementary Video 1.mov	Legend or Descriptive Caption Describe the contents of the file
Supplementary Table	1-5	Supplementary_Tables_Teixeira_et_al_2021.xlsx	Supplementary Tables 1-5.

2

3 Widespread Denisovan ancestry in Island Southeast Asia but no evidence of

4 significant super-archaic hominin admixture

- 5 João C. Teixeira^{1,2*}, Guy S. Jacobs³, Chris Stringer⁴, Jonathan Tuke⁵, Georgi Hudjashov⁶,
- 6 Gludhug A. Purnomo^{1,7}, Herawati Sudoyo^{7,8,9}, Murray P. Cox⁶, Ray Tobler^{1,2#}, Chris S.M.
- 7 Turney^{10,11#}, Alan Cooper^{12,13#} and Kristofer M. Helgen^{11,14#}

- 8 ¹Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, SA 5005, 9 Australia
- 10 ²ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), The University of Adelaide, SA
- 11 5005, Australia
- 12 ³Complexity Institute, Nanyang Technological University, Singapore 637723, Singapore
- 13 ⁴Centre for Human Evolution Research, Department of Earth Sciences, Natural History Museum, London SW7
- 14 5BD, UK
- 15 ⁵School of Mathematical Sciences, The University of Adelaide, SA 5005, Australia
- 16 ⁶Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North 4442,
- 17 New Zealand
- 18 ⁷Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia
- 19 ⁸Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta 10430, Indonesia
- 20 ⁹Sydney Medical School, University of Sydney, NSW 2006, Australia
- 21 ¹⁰Chronos 14Carbon-Cycle Facility, Palaeontology, Geobiology and Earth Archives Research Centre, School of
- 22 Biological, Earth and Environmental Sciences, University of New South Wales, NSW 2052, Australia
- 23 ¹¹ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), University of New South Wales,
- 24 NSW 2052, Australia
- 25 ¹²South Australian Museum, Adelaide, SA 5000, Australia
- 26 ¹³BlueSky Genetics, PO Box 287, SA 5137, Australia
- ¹⁴Australian Museum Research Institute, Australian Museum, 1 William St, Sydney, NSW 2010, Australia 27
- 28 *Corresponding author: joao.teixeira@adelaide.edu.au
- 29

These authors contributed equally.

30 Abstract

31 The hominin fossil record of Island Southeast Asia (ISEA) indicates that at least two endemic 32 'super-archaic' species - Homo luzonensis and H. floresiensis - were present around the time 33 anatomically modern humans (AMH) arrived in the region >50,000 years ago. Intriguingly, 34 contemporary human populations across ISEA carry distinct genomic traces of ancient 35 interbreeding events with Denisovans - a separate hominin lineage that currently lacks a fossil 36 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 37 38 genomes, including over 200 from ISEA. Our results corroborate widespread Denisovan ancestry 39 in ISEA populations but fail to detect any significant super-archaic admixture signals compatible 40 with the endemic fossil record of ISEA. We discuss the implications of our findings for the 41 understanding of hominin history in ISEA, including future research directions that might help to 42 unlock more details about the prehistory of the enigmatic Denisovans.

43 Main Text

44 Island Southeast Asia (ISEA) hosts a unique and diverse fossil record of hominin presence throughout the Pleistocene¹. The island of Java in modern Indonesia marks the south-eastern 45 46 extent of the range of *Homo erectus* – the first hominin species thought to have successfully dispersed outside of Africa – where it maintained a presence from ~ 1.49 Ma until $\sim 117-108$ ka²⁻⁴. 47 48 At least two additional endemic species lived in ISEA during the Pleistocene and are likely to 49 have survived until the arrival of anatomically modern humans (AMH) >50 thousand years ago $(ka)^{5-8}$: *H. floresiensis* on Flores, in the Lesser Sunda Islands (also part of modern Indonesia)^{9,10}, 50 and *H. luzonensis* on Luzon, in the northern Philippines¹¹. The phylogenetic relationships of 51 52 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 53 54 represents an even more archaic species of *Homo* that independently reached ISEA in a separate dispersal event out of Africa^{9,12,13}. The current classification of *H. luzonensis* is also uncertain; 55 the available specimens share similarities in certain morphological traits with various hominin 56 taxa including Australopithecus, Asian H. erectus, H. floresiensis and H. sapiens¹¹. 57

58

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 significant 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-yearold mandible found in the Tibetan Plateau¹⁶, where Denisovan DNA has been recently recovered from cave sediments¹⁷. Despite this geographically circumscribed fossil record, the patterns of 66 Denisovan ancestry in modern human populations suggest that they may have been present across ISEA at the time of AMH arrival¹⁸. While the complexities inherent to demographic and 67 archaic ancestry inference make it hard to infer the precise number and geographical location(s) 68 69 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 70 71 probably come to occupy several islands east of Wallace's Line by 50ka. Stone tools found in Sulawesi dated to $\sim 100-200$ ka²⁵ are also suggestive of possible Denisovan presence east of 72 Wallace's Line²⁶; however, direct fossil evidence of Denisovans in ISEA remains conspicuously 73 74 absent to date.

75

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

79 solution to this problem is that perhaps either H. luzonensis and/or H. floresiensis (or both) are 80 the sources of the Denisovan ancestry in modern human genomes in the region; however, the 81 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. 82 83 Moreover, morphological and archaeological data suggest that both *H. floresiensis* and *H.* 84 *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 85 that evolved in-situ on their respective island locales. The other possible admixture source -86 Indonesian *H. erectus* – is precluded because of its last appearance date of $\sim 117-108$ ka³. Thus, 87 88 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 90 91 of AMH in the region raises the possibility that they also admixed with the ancestors of modern 92 populations now living in ISEA. Traces of super-archaic admixture have been detected in Altai Denisovans³⁰ and, potentially, in modern Andamanese populations^{31–33}, suggesting that 93 interbreeding between super-archaic hominins and more derived hominin species has previously 94 95 occurred and produced viable progeny. If such an event occurred between AMH and endemic 96 ISEA hominins, evidence of this mixing may yet remain undetected in the genomes of present-97 day human populations now living in ISEA, and would indirectly confirm the past presence of

98 one or more super-archaic species in ISEA.

99 **Results**

100 To address this question and provide further insights into the hominin prehistory of ISEA, we 101 implemented the most comprehensive search for introgressed super-archaic regions in modern 102 human genomes performed to date. We searched a total of 426 human genomes from across the world, including 214 individuals from Papuan and ISEA populations²² (Supplementary Table 1), 103 104 for genomic signatures compatible with introgression from archaic hominins such as H. 105 floresiensis, H. luzonensis or other hypothetical late-surviving super-archaic hominin species. To 106 detect blocks of introgressed super-archaic DNA, we extended the analytical pipeline reported by Jacobs et al.²² by including a recently published HMM detection method³⁴ – which we call 107 108 HMM_{Archaic} - along with the two methods used by Jacobs and colleagues; i.e. ChromoPainter (CP)³⁵ and a Hidden Markov Model (HMM)^{36,37}. Importantly, HMM_{Archaic} differs from CP and 109 110 HMM in that it does not require a reference genome to guide the detection of introgressed DNA, 111 making it suitable for identifying DNA from super-archaic groups for which no genome 112 information currently exists. Accordingly, we were able to distinguish putative introgressed 113 super-archaic blocks by running the three detection methods on all 426 genomes and only 114 retaining those that did not overlap any of the Neanderthal and Denisovan blocks predicted by 115 CP and/or HMM. We term the resulting set putative super-archaic sequences as residual_{Archaic} 116 blocks (see Methods). Importantly, to specifically focus on patterns of super-archaic ancestry in 117 ISEA, our strategy purposely excludes genetic variation shared between African and non-African 118 populations. Accordingly, any super-archaic admixture involving AMH in Africa (e.g. with taxa such as *Homo naledi*³⁸) would be excluded from our results. 119

120

121 No evidence for super-archaic introgression in AMH

122 Filtering the HMM_{Archaic} introgressed blocks overlapping Neanderthal- and Denisovan-123 introgressed tracts identified ~12.5Mb of residual_{Archaic} sequence per individual (i.e. sequences 124 that are putatively introgressed from a super-archaic source; Figure 1a). The amount of detected 125 residual_{Archaic} sequence was consistent across worldwide populations, with a slightly higher 126 amount found in East ISEA (~15Mb), and Papuan and Australian populations (~18Mb). In 127 accordance with previous results, ISEA, Papuan, and Australian populations also had the largest 128 amounts of Denisovan ancestry (reaching ~60Mb in Papuan and Australian genomes), meaning 129 that these populations actually had the lowest proportion of residual_{Archaic} sequence relative to the 130 total archaic ancestry observed across all analysed populations (Supplementary Figure 1). Our 131 results indicate that super-archaic ancestry could potentially comprise a small but consistent 132 amount of the genomic ancestry of modern human populations outside of Africa. However, the 133 current lack of evidence for widespread super-archaic admixture in modern human populations 134 suggests that this global residual_{Archaic} signal is more likely a methodological artefact, or a signal 135 of ancient genetic structure in human populations that predates the out-of-Africa migration, or 136 segregation of highly divergent AMH-derived sequences that were not detected in our African reference samples that result from incomplete lineage sorting or balancing selection³⁹. Similarly, 137 138 the additional ~2.5 to ~5Mb of residual_{Archaic} sequence observed in Papuan and Australian 139 populations may represent a small but meaningful amount of super-archaic ancestry specific to 140 this region, or instead simply reflect inter-population variation in the power of the statistical 141 methods to detect Denisovan fragments or some other methodological artefact.

142

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 (i.e.

145	allelic states) that are characteristic of introgressed super-archaic DNA within residual _{Archaic}
146	blocks. Specifically, for each nucleotide position in a residual _{Archaic} block, we characterized the
147	allelic state for the test individual (X), Denisovan (D), Neanderthal (N), and an African
148	individual (H) (see Methods). This resulted in a set of mutation motifs of the form [X, D, N, H],
149	with patterns of the type [1000] and [0111] potentially indicative of super-archaic introgression
150	signals. After enumerating these mutation motifs for all residual _{Archaic} blocks in each individual,
151	we used generalised linear models to test if the proportion of motifs showed population-specific
152	differences, and computed <i>p</i> -values by contrasting the full model to a null model consisting of
153	the intercept alone (see Methods).

154

155 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 156 157 (where motifs are not independent as is assumed for the linear model; Figure 1b and 158 Supplementary Figure 2). However, these differences are extremely subtle and correlate strongly 159 with known archaic ancestry, consistent with the presence of a confounding effect (Figure 1c and 160 Supplementary Figures 3-6). For example, Papuan genomes show a slightly higher proportion of 161 [1000] motifs (<2%) compared to other populations (Figure 1b and Supplementary Figure 2), but 162 inter-individual variation is also high and we do not observe a similar increase in the proportion 163 of the [0111] motif in the population, which is also expected under a scenario of super-archaic 164 introgression (Supplementary Figure 2 and Methods).

165

166 While precise accounting for all motif count differences is non-trivial, likely explanations

167 include the misclassification of alleles as either ancestral or derived, complex demographic

168 histories, and the persistence of Neanderthal and Denisovan archaic signals amongst the 169 residual_{Archaic} blocks that were not removed during the filtering step. For instance, the 2.5-5Mb 170 extra residual_{Archaic} sequence observed in Papuans and Australians might have resulted from 171 these populations having substantially more introgression from a Denisovan-like source that is highly divergent from the Altai Denisovan genome²². This may result in some of the more 172 diverged blocks being detected by the reference-free HMM_{Archaic} scan, but not in the two 173 174 methods that rely on reference genomes (i.e. CP and HMM). Indeed, while Denisovan and 175 Neanderthal ancestry is positively correlated with the proportion of the [1000] motif across all 176 populations, it is negatively correlated with the proportion of the [0111] motif (Supplementary 177 Figures 3 and 4, respectively), which strongly suggests that differences in the proportion of these 178 motifs are caused by unassigned Neanderthal and Denisovan ancestry within residual_{Archaic} 179 blocks.

180

181 Indirect introgression of super-archaic hominin DNA from Denisovans

182 A recent publication reported that modern human genomes carry traces (~4Mb) of super-archaic 183 ancestry that are embedded within introgressed Denisovan sequences (having previously been 184 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 185 186 also detected in this study (20 out of 20, with 100% of the sequence length of each introgressed 187 block being recovered; Supplementary Table 2), most of these were also included in our set of 188 putative super-archaic blocks (17 out of 20, with an average of 47.5% of the sequence length of 189 each block being recovered; Supplementary Table 3). Similar results were obtained by

comparing HMM_{Archaic} and residual_{Archaic} blocks to super-archaic segments embedded within
 predicted Neanderthal introgressed regions (see Methods and Supplementary Tables 4 and 5).

193 Coalescent simulations support empirical observations

194 The accurate recovery of indirectly introgressed super-archaic fragments by our analytical 195 pipeline suggests that it is sufficiently powerful to detect low levels of directly introgressed 196 super-archaic ancestry, should it exist. Nonetheless, to rule out the possibility that the lack of 197 evidence for super-archaic introgression into modern humans was due to a lack of statistical power, we used the coalescent software $msprime^{41}$ to simulate Aboriginal Australian and Papuan 198 histories under an empirically-informed demographic model⁴². These simulations included 199 200 separate Neanderthal and Denisovan admixture events along with differing amounts of super-201 archaic introgression (2%, 1%, 0.1% and 0%) in the common ancestral population of Australo-202 Papuans (see Methods). We then applied our full analytical pipeline to these simulated genomic 203 datasets to detect super-archaic blocks and quantified the power and false discovery rate for the 204 different levels of super-archaic introgression.

205

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
(TPR) ranging from ~50% to ~95% for models with 0.1% and 2% super-archaic ancestry,
respectively (Supplementary Figure 9) – while maintaining extremely low false positive rates
(Supplementary Figure 10).

The amount of residual_{Archaic} sequences detected per individual in the 0.1% and 0% super-archaic
introgression models (~20Mb – Figure 2a) is strikingly close to that observed in the Papuan and

213 Australian empirical data ($\sim 18Mb - Figure 1a$). For these models, the majority of the 214 residual_{Archaic} signal originates from Neanderthal and Denisovan introgression that went 215 undetected by CP and HMM (Figure 2a and Supplementary Figure 12). In contrast, the 1% and 216 2% super-archaic introgression models detect at least 2 times more residual_{Archaic} sequence per 217 individual than empirical estimates (~33Mb and ~47Mb, respectively – Figure 2a), which was 218 primarily caused by an inflation in the number of super-archaic blocks. Interestingly, the power 219 to detect Neanderthal and Denisovan blocks using HMM_{Archaic} is negatively affected by 220 increasing amounts of super-archaic ancestry, as the power of this method is proportionate to the 221 divergence between the introgressing archaic population and the outgroup human population (see 222 Supplementary Figure 12 and Methods).

223

Similarly, the mutational motifs observed in 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 to ~26% and ~6.5% for the 0.1% model, and ~22.5% and ~4% for the 0% model (Supplementary Figure 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.

231 Discussion

232 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 233 234 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 paleoanthropological 235 236 interpretations of H. luzonenesis and H. floresiensis as descendants of super-archaic hominin 237 groups, our results indicate that interbreeding between these groups and AMH did not occur. 238 However, we cannot outright reject the possibility that interbreeding did occur if these 239 encounters either did not produce viable progeny or the offspring were viable and that these 240 lineages have since died out. Evidence for super-archaic introgression into the ancestors of the Altai Denisovans³⁰ and, possibly, Andamanese populations^{31–33}, suggests that viable 241 242 reproduction may actually have been possible, though further evaluation of these hypotheses at 243 present is not possible given the available data.

244

245 An alternative explanation is that H. luzonensis and H. floresiensis belong to a hominin clade 246 that is considerably less divergent from AMH than is currently accepted, possibly being the late-247 surviving descendants of an earlier radiation of a Denisovan-like lineage across ISEA. This would imply that hominin occupation of Flores (>1.01Ma)^{29,45} and the Philippines (from 248 ~700ka)⁴⁶ was not continuous and that the ubiquitous Denisovan ancestry across ISEA results 249 from AMH admixture with one or both of these groups. Indeed, the patterning of Denisovan 250 251 ancestry across ISEA is consistent with separate Denisovan introgression events in the Philippines¹⁹ and, potentially, in $\text{Flores}^{23,43}$, the island homes of *H. luzonensis* and *H.* 252 253 floresiensis, respectively. Further, it is possible that pronounced dwarfism and prolonged periods

of endemic island evolution for *H. floresie*nsis 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 upon interpretations of archaeological and fossil
data^{9,11-13,27-29}.

259

260 A major complication in resolving these questions is the sparse Denisovan fossil record – 261 currently consisting of one phalanx, a mandible, several teeth and some cranial fragments -262 which makes meaningful morphological comparisons very difficult. Potentially promising areas 263 for further research include Sulawesi, where stone tool records are compatible with possible Denisovan occupation ~100-200ka²⁵. Intriguingly, Sulawesi is home to endemic dwarf buffalos 264 265 (Bubalus spp.) and pigs (Sus celebensis, Babirusa spp.), which are among the few megafaunal 266 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 267 268 include the living Komodo dragon (Varanus komodoensis) on Flores and its satellites, and 269 surviving buffalo (Bubalus mindorensis), pigs (Sus spp.) and deer (Rusa spp.) in the oceanic 270 Philippines (Figure 3). This pattern suggests that long-term exposure to possible hunting 271 pressures by archaic hominins might have facilitated the survival of megafaunal species in 272 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 273 unlikely) possibility is Australia, where the ~65ka artefacts uncovered at Madjedbebe⁶ might be 274 275 associated with Denisovan presence.

277 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 278 279 proteomic studies to assist in resolving phylogenetic relationships where DNA is not 280 recoverable. Nonetheless, the current fossil and archaeological records, together mounting 281 genetic evidence from across ISEA, point to the widespread presence of archaic hominins east of Wallace's Line²⁶, and indicate that the first AMH populations to arrive in ISEA most likely 282 encountered a variety of hominin populations, no matter which route they took to enter Sahul^{47–} 283 284 ⁵². This hints that much of the Denisovan ancestry found in modern human populations in ISEA, 285 New Guinea, and Australia may have been locally acquired, emphasizing the need for more 286 archaeological and genetic research across this understudied region in the future.

287 Methods

Samples. We examined 426 individuals from 10 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
 (IGDP)²². For a description of data preparation (SNP calling, QC, phasing) see Jacobs et al.²².

293 Searching for signals of super-archaic admixture into modern humans. We searched for signals 294 of super-archaic introgression in genomic sequences of AMH populations across the world, with 295 a particular focus on ISEA and New Guinea (descendants from early AMH migrations into the 296 region). These specific signatures are expected to include the existence of genetic variants that 297 are not observed in Africa, and which exhibit levels of linkage disequilibrium compatible with 298 introgression events ~60-50 ka, similarly to observations for Neanderthal and Denisovan 299 introgressed segments. However, we expect deep divergence times between extinct ISEA 300 hominins (*H. luzonensis* and *H. floresiensis*) and *H. sapiens* if we consider the former are not 301 part of the Denisovan/Neanderthal clade and are instead related to H. erectus, or represent 302 additional *Homo* lineages that split from AMH ~2 Ma or earlier. Hence, the putative introgressed 303 super-archaic regions are expected to be highly divergent to orthologous modern human genome 304 sequences. Importantly, the absence of a genome sequence for the extinct ISEA hominin groups 305 makes this inference far more complex than for Neanderthal or Denisovan introgression, for 306 which reference genomes are available. Therefore, we searched for super-archaic introgression in 307 the genomes of contemporary human populations around the world using a highly powerful Hidden Markov Chain model implemented by Skov et al.³⁴ (termed here HMM_{Archaic}), which is 308 agnostic to the genome sequence of the putative archaic source. The rationale behind this 309

310 strategy is that introgressed regions of the genome are enriched for genetic variants not seen in 311 populations which have not admixed with the putative archaic source. In this case, we used 312 African populations as an outgroup and assumed that these African populations have not 313 interbred with Neanderthals, Denisovans, or any super-archaic source. It should be noted that the 314 class of methods to which HMM_{Archaic} belongs are only indicative of archaic introgression. These 315 methods might be prone to false positive detection of introgressed fragments due to incomplete 316 lineage sorting or balancing selection maintaining old genetic diversity at specific selected loci. 317 The HMM_{Archaic} method infers archaic admixture using a sliding-window approach after 318 controlling for genetic diversity existing in an outgroup (e.g. African populations). We applied 319 the method across all individuals from each of the ten sampled populations, using as an outgroup 320 all individuals belonging to every African population contained in our dataset. After this, we 321 further excluded positions where the Altai Neanderthal and Altai Denisovan individuals are 322 heterozygous. We set the initial parameters to run HMM_{Archaic} following the author's 323 implementation, specifically: states =['Human', 'Archaic']; starting_probabilities = [0.98, 0.02]; 324 transitions = [[0.9995, 0.0005], [0.012, 0.98]], emissions = [0.04, 0.1]. Importantly, the method can 325 be applied to phased data, and hence extract putative introgressing haplotypes rather than 326 unphased regions, allowing for downstream analysis that is more sensitive to the independent 327 histories of homologous chromosomal regions. Hence, the model was trained and implemented on phased data, which was obtained as described in Jacobs et al.²². We used a 1,000bp sliding-328 window approach, as performed in the original implementation of the method³⁴, as the small size 329 330 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. 331

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.

339

340 Identifying Denisovan and Neanderthal introgressed fragments. We first sought to detect

341 genomic signals of Neanderthal and Denisovan introgression using the CP³⁵ and

HMM^{36,37} introgression-detection methods described in Jacobs et al²². 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 Jacobs et al²².

347

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

355 non-overlapping regions (refer to Supplementary Figure 8 for an illustration). This approach is allied to the residual S^* calculated in Jacobs et al²², but differs in using more accurate phased 356 archaic calls from HMM_{Archaic} and in the detail of the residual_{Archaic} block calling process. Note 357 358 that identified residual_{Archaic} blocks may be in close proximity to Denisovan or Neanderthal 359 introgressed regions (as is the case in Supplementary Figure 8) and that these blocks are not 360 suitable for some downstream analyses such as introgression time estimation based on 361 introgressed block length, as they may correspond to subparts of larger introgressed blocks. We 362 decided to adopt this strategy as there is potential for super archaic blocks, in case they are 363 present, to segregate close to, or overlap with, Neanderthal and Denisovan fragments, given the 364 potential for non-random segregation of archaic blocks within the genome. While in the current 365 work we do not present the results for an alternative strategy of completely removing 366 Neanderthal and Denisovan blocks to estimate residual_{Archaic}, the findings are qualitatively 367 similar to the ones presented here.

368

369 Looking at patterns of variation within residual_{Archaic} blocks. In order to further disentangle the 370 patterns seen in residual_{Archaic} blocks, we looked at mutation-motif patterns. We defined the 371 mutation motifs as 0 (ancestral) and 1 (derived), and a combination of [X, D, N, H], where 'X' 372 represents the allelic state of a particular individual within an introgressed block (which can also 373 be thought of as the test population - i.e., Papuan, East ISEA, West ISEA, etc), 'D' represents Denisova, 'N' represents Neanderthal, and 'H' represents an individual from an African 374 375 population (in our case Ju'hoan - SS6004473). While all African variation was removed from the 376 dataset prior to running HMM_{Archaic} (as Africans form the required outgroup), we reintroduced 377 SS6004473 variation subsequently and for this specific analysis only. 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.

384 Variation in motif proportion as a function of physical distance to introgressed regions. We 385 investigated the proportion of different motifs as a function of physical distance to the putatively 386 introgressed regions. In this case we divided the analyses into patterns seen within all 387 HMM_{Archaic} introgressed fragments and those seen residual_{Archaic} fragments (Supplementary 388 Figure 7). In this analysis, we define mutation motifs as [X, D, N, Af] where a single human 389 outgroup is now represented by an indicator Af, 1 indicates that a variant is found in the derived 390 state in one or more individuals in the African outgroup, and 0 indicates that the derived state is 391 not observed. Thus, we are specifically focusing on whether variation is found at all in an 392 African sample rather than a single African individual. When all HMM_{Archaic} fragments within 393 the Papuan population are considered, we observe an excess of [1100] and [1010] motifs, 394 compatible with introgression from Denisovan and Neanderthal into Papuan genomes, 395 respectively, along with a sharp decrease of [1001] (where X shares a derived allele with Africa) 396 motifs. These signatures consistently indicate Neanderthal and Denisovan introgression into 397 Papuan genomes. When considering residual_{Archaic} fragments only, we observe a sharp increase 398 in the [1000] motif (as expected) coupled with a reduction in the [1100] and the [1010] motifs 399 (signals of Denisovan and Neanderthal introgression, respectively), suggesting that remaining 400 fragments do not show a clear signal of known archaic introgression. These Neanderthal and

401 Denisovan signals increase in the regions around residual_{Archaic} blocks, indicating that they are 402 often nested within introgressed Neanderthal and Denisovan sequences. This is an important 403 observation, suggesting that much of the signal is contributed through known introgression, in 404 support of the absolute increase in residual_{Archaic} in Papuan populations. Indeed, the definition of 405 residual_{Archaic} does not exclude the detection of regions showing coalescent histories consistent 406 with super-archaic introgression from within Denisovan and Neanderthal introgression (as would 407 likely be the case for the blocks shown in example schematic Supplementary Figure 8), and 408 variation in the coalescent histories within blocks sharing the same introgression source is likely. 409 While this suggests that residual_{Archaic} blocks may be retrieving super-archaic signals from within 410 Denisovan and Neanderthal introgressing populations, we suggest that more data and more 411 focused analysis, beyond the scope of this paper, are necessary to assess the significance of these 412 patterns. The sharp decrease in the [1001] motif observed in all HMM_{Archaic} blocks is replaced by 413 a peak in residual_{Archaic} blocks, and a slight increase in the [0111] motif is now visible. In both 414 cases, these indicate deep coalescence of residual_{Archaic} blocks not associated with the sampled 415 Neanderthal or Denisovan sequences. While the [0111] signal is of particular interest in the 416 context of super-archaic introgression, the lack of any global peaks in this motif (Supplementary 417 Figure 2) and elevated [1100] and [1010] signals surrounding residual_{Archaic} blocks argues that it 418 more likely reflects deep coalescent histories within Denisovan and Neanderthal introgressed 419 blocks than super-archaic introgression.

420

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 Figures 3 and 4 show the correlation

424 between inferred Denisovan, and Neanderthal ancestry, respectively, and the proportion of 425 different motifs, across all individuals. Interestingly, we find both positive and negative 426 correlations between the proportion of different motifs and the detected amount of Denisovan 427 and Neanderthal ancestry. In fact, these correlations are statistically significant for all but two 428 motifs when regressing on Denisovan ancestry, [0100] (P-value 0.289) and [1110] (P-value 429 0.618), and for all but one motif when regressing on Neanderthal ancestry, [1110] (P-value 430 (0.221). These results are in agreement with the observations from simulations with no super-431 archaic introgression, which show that residual_{Archaic} sequence is essentially dominated by 432 introgressed Neanderthal and Denisovan fragments that are undetected by both HMM and CP. 433

434 *Comparing HMM*_{Archaic} and residual_{Archaic} to predicted super-archaic regions

A recent study by Hubisz et al.⁴⁰ proposes that a diminutive proportion of super-archaic ancestry 435 436 survives in contemporary human populations due to introgression events between a highly 437 divergent hominin and the ancestors of Neanderthals and Denisovans, who subsequently 438 admixed with the ancestors of present-day people. In order to investigate whether our strategy 439 allowed for the detection of rare super-archaic fragments, we contrasted the inferred HMM_{Archaic} 440 and residual_{Archaic} blocks per individual to the putatively super-archaic fragments introgressed via Neanderthals and Denisovans proposed by Hubisz et al⁴⁰. Specifically, we identified all the 441 442 instances where an HMM_{Archaic} or a residual_{Archaic} block identified in each individual in our 443 sample overlaps (even if only partially) a super-archaic fragment predicted by Hubisz et al. We 444 then counted the number of individuals containing at least one HMM_{Archaic} or residual_{Archaic} block 445 overlapping each super-archaic fragment (i.e. the overlap of each individual was only counted 446 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.

450

451 *Simulating super-archaic introgression using msprime*

452 In order to test the power of our experimental design to detect introgression from a highly 453 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 454 simulations use demographic parameters derived from Malaspinas et al.⁴², which models 455 456 Aboriginal Australian history from full genome data from modern Australian and Papuan 457 populations. The structure and parameters describing the standard demography (i.e. excluding 458 possible super-archaic introgression) followed a maximum likelihood model output. Briefly, we 459 simulated a total of 35 African and 30 Australian individuals, and one Altai Denisovan 460 individual that split from human populations 20,255 generations prior to the present, while 461 African and Australian populations split from one another 3,916 generations ago. Additionally, 462 we included one super-archaic individual, that splits from the Human-Neanderthal-Denisova 463 clade 70,000 generations in the past, to mimic the deep split assumed for *H. floresiensis* and *H. luzonensis*, with haploid Ne = 13,249. Following Malaspinas et al.⁴², Neanderthal (2.4%) and 464 465 Denisovan (4.0%) introgression events were simulated at, respectively, 1,853 and 1,353 466 generations in the past, with the introgressing lineages being related to the Altai individuals, and 467 additional minor Neanderthal contributions to the Eurasian clade (1.1%) and Australian clade (0.2%) at 1566 and 883 generations ago, respectively. For the super-archaic admixture, we 468 469 assumed an introgression event occurring 1,353 generations ago. We set the mutation rate to

470 1.4e-8/bp/generation and the recombination rate to 1e-8/bp/generation and simulated, per 471 individual, a total of 300 chromosomes of 10Mb in length each. This strategy allowed us to 472 obtain a total simulated sequence that roughly matches the size of the human genome for each 473 individual (~3Gb of sequence), while ensuring sufficient independent replication. Importantly, 474 after running the simulations, we sampled 65 human individuals (35 African and 30 Australian 475 genomes), an Altai Neanderthal and an Altai Denisovan (related to, respectively, the 476 introgressing Neanderthal and Denisovan populations), and one super-archaic individual. 477 A major advantage of using *msprime* to implement coalescent simulations is that the software 478 allows the genealogy of each portion of simulated sequence to be traced back through time, 479 including the migration of genomic regions between archaic and human populations (i.e. 480 introgression). This means that, for each individual, we are able to know the exact amount and 481 location of the introgressed segments, and are thus able to directly compute the strength of our 482 approaches for detecting super-archaic introgression in the empirical data.

483

484 Models of super-archaic introgression

485 We initially implemented two models of super-archaic introgression: a model containing 2% 486 introgression into the ancestors of Australians occurring at the same time as Denisovan 487 introgression, and a second model without super-archaic introgression (0%). To estimate the 488 power of our analytical framework to detected super-archaic introgression at low levels of 489 admixture, for each simulated individual we created datasets with ~1% and ~0.1% super-archaic 490 introgression by masking a specific proportion of super-archaic blocks in the 2% model. 491 Specifically, this was achieved by 1) randomly sampling a proportion of introgressed super-492 archaic regions in each individual; and 2) merging all the regions sampled across all individuals

493 and masking these merged super-archaic regions across all simulated individuals. This strategy 494 ensured that the masked super-archaic regions were the same across all individuals. We were 495 able to reduce the amount of super-archaic ancestry present in the simulated sequences to $\sim 1\%$ 496 and ~0.1% by randomly sampling, per individual, ~10% and ~50% of introgressed super-archaic 497 regions, respectively. Due to the masking of the introgressed regions, the 1% and 0.1% models 498 contained slightly less genetic sequence than the 0% an 1% models (~2.88Gb and ~2.65Gb 499 simulated sequence, respectively); however the masking did not alter the average proportion of 500 introgressed sequences observed from either the Denisovan of Neanderthal lineages 501 (Supplementary Figure 12).

502

503 *Power to uncover archaic introgression*

504 We evaluated the performance of the analytical pipeline by comparing the results from our 505 empirical data to four models of Australian-super-archaic admixture at different introgression 506 levels (i.e. 2%, 1%, 0.1% and 0%). First, we estimated the power of each of the three detection 507 methods utilized to compute archaic introgression in the empirical data; i.e. CP, HMM and 508 HMM_{Archaic}. Analogous to the implementation in the empirical data, before running HMM_{Archaic}, 509 we excluded all variation present in the 35 simulated African genomes, along with positions for 510 which the Altai Neanderthal and Denisovan individuals were heterozygous. Supplementary 511 Figure 9 shows the True Positive Rate (TPR) of each method to detect archaic introgression. The 512 TPRs were estimated as the length of detected regions that overlap the simulated introgressed 513 regions over the total length of simulated introgressed regions (in base-pairs). It was possible to 514 estimate the TPR separately for introgression from the Neanderthal and Denisovan lineages for 515 CP and HMM, though not for HMM_{Archaic} (which does not require a reference).

516 Both CP and HMM consistently detect Neanderthal introgression at a higher rate than Denisovan 517 introgression, irrespective of the amount of super-archaic introgression present in the simulations 518 (Supplementary Figure 9). Considering that both CP and HMM rely on the availability of a 519 reference sequence for the putatively introgressing archaic population, this observation is 520 consistent with the fact that the simulated introgressing Neanderthal population is genetically 521 closer to the reference Altai Neanderthal than the simulated introgressing Denisovan population 522 is to the reference Altai Denisovan. Nevertheless, both methods seem to perform only slightly 523 better in the absence of super-archaic introgression, presumably because, at least in the case of 524 CP, a very small proportion of inferred Neanderthal and Denisovan introgression derives from 525 super-archaic introgression (see below). HMM_{Archaic} has extremely high power to detect super-526 archaic segments (Supplementary Figure 9, top left) and, even though power decreases at lower 527 levels of super-archaic introgression, it is always higher than the detection power for Neanderthal 528 or Denisovan introgression across all four models (Supplementary Figure 9).

529

530 *False positive rate*

531 We next examined the False Positive Rate (FPR) of each method to detect archaic introgression. 532 For the CP and HMM methods we define FPR as the proportion of sequence misassigned to a 533 particular archaic population when that sequence is either introgressed from another hominin 534 lineage or is from the human genealogy. For HMM_{Archaic} we simply estimated the proportion of 535 sequence misassigned as archaic that overlaps simulated human regions. The results are shown in 536 Supplementary Figures 10 and 11. Both CP and HMM have relatively high FPRs when inferring 537 Neanderthal introgression that actually results from Denisovan introgression ($\sim 40\%$), and vice-538 versa (~35%) (Supplementary Figure 11). As expected, given the closer relationship of the

539 introgressing Neanderthal population to the reference Altai Neanderthal compared to the 540 introgressing Denisovan population to the reference Altai Denisovan, the FPR for CP and HMM 541 is higher for Denisovan segments that were missasigned as Neanderthal (Supplementary Figure 542 11, right panel) than vice-versa (Supplementary Figure 11, left panel). This pattern is also 543 consistent with a close genetic relationship between Neanderthals and Denisovans, and the 544 persistence of shared ancestral genetic diversity between the two species (incomplete lineage 545 sorting). Importantly, however, the FPR of both methods is extremely low when inferring 546 Neanderthal or Denisovan introgression when it either did not occur (Supplementary Figure 11, 547 middle columns - 'Human') or the source was super-archaic (Supplementary Figure 11, left 548 columns - 'super-archaic'). Hence, our simulation results demonstrate that a negligible amount 549 of introgressed super-archaic sequence will be mistaken for Neanderthal or Denisovan 550 introgression by CP and HMM. Finally, our stringent approach for detecting archaic 551 introgression using HMM_{Archaic} (posterior probability >0.95, see above) results, as expected, in 552 virtually no false positives in the simulations (Supplementary Figure 10) – i.e. a negligible 553 portion of archaic HMM_{Archaic} overlaps with human genealogies.

554

555 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 Figure 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 below – *Effects of super-archaic ancestry to detect Neanderthal/Denisovan introgression*). 562 In contrast, the amount of Neanderthal and Denisovan detected by both CP and HMM is 563 essentially independent from the amount of super-archaic ancestry present (as expected from the 564 TPRs shown in Supplementary Figure 9). As described above, the masking strategy adopted to 565 reduce the amount of super-archaic in the simulations meant that models 1% and 0.1% contain a 566 reduced amount of introgressed Neanderthal and Denisovan sequence overall (see explanation in 567 *Power to uncover archaic introgression*). Therefore, we also present a corrected amount of 568 simulated and detected archaic sequences by normalizing the total amounts to match the total 569 amount of sequence considered in the empirical data (Supplementary Figure 12, panel b). This 570 strategy also allowed us to compare the simulations directly to the results obtained for the 571 empirical data, namely in terms of total residual_{Archaic} sequence present. After determining the 572 total detected sequence in each method, we obtained the residual_{Archaic} regions by removing those 573 regions that overlap with either the CP or HMM detected blocks (residual_{Archaic} in Figure 2, 574 overlapping blocks shown as overlap_{Archaic}).

575

576 Effects of super-archaic ancestry to detect Neanderthal/Denisovan introgression

577 An interesting picture emerges when we consider the behaviour of HMM_{Archaic} in the presence of 578 super-archaic introgression. The ability of HMM_{Archaic} to detect Neanderthal and Denisovan 579 introgression is severely depleted at higher levels of super-archaic introgression, which appears 580 to dominate the amount of detected archaic ancestry: less than 25% of truly introgressed 581 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 582 (Supplementary Figure 9, top panel). This pattern is consistent with the power of HMM_{Archaic} 583 584 being proportionate to the divergence between the introgressing archaic population and the

585 outgroup human population (i.e. Africa). Importantly, we have simulated a super-archaic source 586 whose divergence to modern humans is significantly higher than that of Neanderthals and 587 Denisovans to mimic introgression from H. floresiensis and H. luzonensis, assuming that the 588 latter are earlier diverging lineages of *Homo*. There is a considerably higher agreement between 589 HMM_{Archaic} and both CP and HMM for a model with no super-archaic introgression compared to 590 a model containing even 0.1% super-archaic introgression (Figure 2). The most important signal 591 for differentiating these scenarios, which have similar total simulated residual_{Archaic}, is the 592 concordance between HMM_{Archaic} and CP/HMM. Specifically, the excess divergence of super-593 archaic introgressed sequences means these blocks contain a higher amount of non-African 594 variants and, therefore, are more efficiently detected by HMM_{Archaic}. However, this process 595 simultaneously impacts the internal optimisation of HMM_{Archaic} emission parameters, causing the 596 algorithm to seek more divergent introgressed blocks, which reduces the TPR for detecting 597 known Denisovan and Neanderthal blocks. This is consistent with HMM_{Archaic} having a higher 598 TPR for introgressed Neanderthal and Denisovan sequences when no super-archaic introgression 599 is present, which in turn leads to a higher amount of Neanderthal and Denisovan sequence 600 detected by all three methods (Figure 2). This behaviour causes the concordance between 601 methods to drop, and the residual_{Archaic} signal to increase as a proportion of total HMM_{Archaic}, 602 even when simulating minimal amounts of super-archaic introgression. The higher concordance 603 between HMM_{Archaic}, CP, and HMM for the 0% model translates into a 27% proportion of 604 residual_{Archaic} in this model (Figure 2c) – consistent with residual_{Archaic} regions computed in the empirical data (between ~15% in Papuan genomes and ~22% in West Eurasian genomes -605 606 Supplementary Figure 1) – and in contrast to $\sim 33\%$ to 60% for models with $\geq 0.1\%$ super-

607	archaic introgression. Importantly, in simulations containing higher proportions of super-archaic
608	ancestry (1% and 2% models), we observe a much higher proportion of residual _{Archaic} sequence.
609	

610 Investigating mutation motifs within residual_{Archaic} simulated models

611 In order to further investigate the nature of genetic diversity within residual_{Archaic} regions, we

612 performed similar mutation motif analyses to those used in the empirical data (see above). In

613 particular, we investigated the amount of shared ancestral and derived alleles between

614 individuals carrying the residual sequence (i.e. test population), the simulated Altai Denisovan,

615 the simulated Altai Neanderthal, and a simulated African genome – again, while all African

616 variation was excluded from HMM_{Archaic} analyses, we randomly sampled one individual and

617 investigated allele sharing within residual_{Archaic} regions after running the method.

618 Data availability

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

622 **References**

623	1.	Kaifu,	Y. Archaic	Hominin	Population	ns in As	ia before	the A	Arrival	of Moder	n Humans:
-----	----	--------	------------	---------	------------	----------	-----------	-------	---------	----------	-----------

- 624 Their Phylogeny and Implications for the 'Southern Denisovans'. *Curr Anthropol* 58,
- 625 S418–S433 (2017).
- 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).
- 628 3. Rizal, Y. *et al.* Last appearance of Homo erectus at Ngandong, Java, 117,000–108,000 years
 629 ago. *Nature* 577, 381–385 (2020).
- 630 4. Matsu'ura, S. *et al.* Age control of the first appearance datum for Javanese in the Sangiran
 631 area. *Science* 367, 210–214 (2020).
- 632 5. Westaway, K. E. *et al.* An early modern human presence in Sumatra 73,000-63,000 years
 633 ago. *Nature* 548, 322–325 (2017).
- 634 6. Clarkson, C. *et al.* Human occupation of northern Australia by 65,000 years ago. *Nature*635 547, 306–310 (2017).
- 636 7. Bowdler, S. 'Human occupation of northern Australia by 65,000 years ago' (Clarkson et al.
 637 2017): a discussion. *Aust. Archaeol.* 83, 162–163 (2017).
- 638 8. O'Connell, J. F. *et al.* When did Homo sapiens first reach Southeast Asia and Sahul? *Proc.*639 *Natl. Acad. Sci. U.S.A.* 115, 8482–8490 (2018).
- 640 9. Brown, P. et al. A new small-bodied hominin from the Late Pleistocene of Flores,
- 641 Indonesia. *Nature* **431**, 1055–1061 (2004).
- 642 10. Sutikna, T. *et al.* Revised stratigraphy and chronology for Homo floresiensis at Liang Bua
 643 in Indonesia. *Nature* 532, 366–369 (2016).
- 11. Détroit, F. *et al.* A new species of Homo from the Late Pleistocene of the Philippines.

- 645 *Nature* **568**, 181–186 (2019).
- 646 12. Brown, P. & Maeda, T. Liang Bua Homo floresiensis mandibles and mandibular teeth: a
- 647 contribution to the comparative morphology of a new hominin species. J. Hum. Evol. 57,
- 648 571–596 (2009).
- 649 13. Argue, D., Groves, C. P., Lee, M. S. Y. & Jungers, W. L. The affinities of Homo
- 650 floresiensis based on phylogenetic analyses of cranial, dental, and postcranial characters. *J.*651 *Hum. Evol.* 107, 107–133 (2017).
- Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060 (2010).
- Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226 (2012).
- 656 16. Chen, F. *et al.* A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau.
 657 *Nature* 569, 409–412 (2019).
- 658 17. Zhang, D. *et al.* Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave
 659 on the Tibetan Plateau. *Science* 370, 584–587 (2020).
- 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).
- Inam, 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
- 665 Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* 173, 53–61.e9
 666 (2018).
- 667 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).
- 670 22. Jacobs, G. S. *et al.* Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell* 177, 1010–1021.e32 (2019).
- 672 23. Teixeira, J. C. & Cooper, A. Using hominin introgression to trace modern human dispersals.
 673 *Proc. Natl. Acad. Sci. U. S. A.* 116, 15327–15332 (2019).
- 674 24. Bergström, A. *et al.* Insights into human genetic variation and population history from 929
 675 diverse genomes. *Science* 367, (2020).
- 676 25. van den Bergh, G. D. *et al.* Earliest hominin occupation of Sulawesi, Indonesia. *Nature* 529,
 677 208–211 (2016).
- 678 26. Wallace, A. R. On the Physical Geography of the Malay Archipelago. *Proceedings of the*679 *Royal Geographical Society of London* 7, 206 (1862).
- 680 27. Kaifu, Y. *et al.* Craniofacial morphology of Homo floresiensis: description, taxonomic
- affinities, and evolutionary implication. J. Hum. Evol. **61**, 644–682 (2011).
- 682 28. Kaifu, Y. et al. Unique Dental Morphology of Homo floresiensis and Its Evolutionary
- 683 Implications. *PLoS One* **10**, e0141614 (2015).
- van den Bergh, G. D. *et al.* Homo floresiensis-like fossils from the early Middle Pleistocene
 of Flores. *Nature* 534, 245–248 (2016).
- 686 30. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains.
- 687 *Nature* **505**, 43–49 (2014).
- Mondal, M. *et al.* Genomic analysis of Andamanese provides insights into ancient human
 migration into Asia and adaptation. *Nat. Genet.* 48, 1066–1070 (2016).
- 690 32. Skoglund, P., Mallick, S., Patterson, N. & Reich, D. No evidence for unknown archaic

- 691 ancestry in South Asia. *Nature genetics* vol. 50 632–633 (2018).
- 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).
- 694 34. Skov, L. *et al.* Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet.*695 14, e1007641 (2018).
- 696 35. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure
 697 using dense haplotype data. *PLoS Genet.* 8, e1002453 (2012).
- 698 36. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years.
 699 *Science* 346, 1113–1118 (2014).
- 700 37. Racimo, F. *et al.* Archaic Adaptive Introgression in TBX15/WARS2. *Mol. Biol. Evol.* 34,
 701 509–524 (2017).
- 38. Berger, L. R. *et al.* Homo naledi, a new species of the genus Homo from the Dinaledi
 Chamber, South Africa. *Elife* 4, (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
- 707 through demography-aware inference of the ancestral recombination graph. *PLoS Genet*.
 708 16, e1008895 (2020).
- 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).
- 713 43. Tucci, S. *et al.* Evolutionary history and adaptation of a human pygmy population of Flores

- 714 Island, Indonesia. *Science* **361**, 511–516 (2018).
- Vernot, B. *et al.* Excavating Neandertal and Denisovan DNA from the genomes of
 Melanesian individuals. *Science* 352, 235–239 (2016).
- 717 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. Forum 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. Island Coast.
 Archaeol. 11, 364–384 (2016).
- 49. Kealy, S., Louys, J. & O'Connor, S. Reconstructing Palaeogeography and Inter-island
- 727 Visibility in the Wallacean Archipelago During the Likely Period of Sahul Colonization,
- 728 65-45 000 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, (2019).

Acknowledgements: We thank Christian Huber and Joshua Schmidt for useful discussions on
the genetic analyses, and Kieren Mitchell and Fernando Racimo for comments on the
manuscript. This work was supported by ARC Indigenous Discovery Grant IN180100017 (JCT
and RT) and ARC Laureate Fellowships FL100100195 (CMST) and FL140100260 (AC). CS
acknowledges funding from the Calleva Foundation and The Human Origins Research Fund.

740 Author contributions: JCT, KMH, GSJ, MPC, GH, GAP and JT designed the methods and

value of the analyses. JCT, KMH, RT, GSJ, HS, AC, CS, CMST and MPC wrote the

742 manuscript with input from all authors.

743

744 **Competing interests:** The authors declare no competing interests.

745 Figure Legends:

746

747 Figure 1. Introgression signals in extant populations across Island Southeast Asia. (a) Violin plots showing the 748 cumulative amount (Mb) of Neanderthal and Denisovan ancestry (purple) estimated using HMM and residual_{Archaic} 749 sequence (green) across different populations. Each dot represents a single sampled individual for a particular 750 population. Within each violin plot, the population's mean and 95% values of the distribution are shown as a black 751 dot and vertical line, respectively. (b) The proportion of variants within residual_{Archaic} fragments that show mutation 752 motifs compatible with super-archaic introgression [1000] per population. Each number on the string [1000] 753 corresponds to the allelic states observed in [X, Denisovan, Neanderthal, Africa], where X is an individual from the 754 test population (e.g. Australia), and 1 and 0 define derived and ancestral allelic states, respectively. (c) Scatter plot 755 showing the association between the proportion of [1000] motifs within residual_{Archaic} fragments and the total 756 amount of Denisovan (left) and Neanderthal (right) ancestry per individual. 757 758 Figure 2. Results from coalescent simulations exploring the detection of archaic hominin introgressed sequences. a) 759 residual_{Archaic} after removing Neanderthal and Denisovan regions detected by CP, HMM and CP+HMM. The total 760 residual_{Archaic} and the proportion of residual_{Archaic} that overlap simulated archaic regions for different species is 761 shown from left to right, together with the amount of residual_{Archaic} that overlaps 'Human' (i.e. non-archaic) regions. 762 Different simulation models of super-archaic introgression are shown in the x-axis from left to right. b) Overlap 763 between regions inferred as 'archaic', showing the concordance between HMM_{Archaic} and the other two methods 764 (overlap_{Archaic}). c) The proportion of residual_{Archaic} sequence over the total amount of HMM_{Archaic} inferred to be 765 'archaic'. d) The proportion of overlap_{Archaic} out of the total amount of HMM_{Archaic}. 766 767 Figure 3. Hominin occupation and megafauna survival in Island Southeast Asia at the time of modern human 768 arrival. Confirmed presence of H. floresiensis and H. luzonensis depicted by skull icons; regions with known 769 artefacts associated with hominin presence are depicted by the stone tool icons; extant native megafauna east of the 770 Wallace Line is depicted by the buffalo icon (representing mammals-Bubalus, Rusa, Sus, and Babirusa) in the 771 northern and southern Philippines and Sulawesi, and Komodo dragon icon on Flores and satellites. Inferred hominin

772 presence covers the entry routes into Sahul, indicated by the orange arrows. The estimated Denisovan ancestry in

- 773 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 and Lyddeker's Lines are shown as thick black lines and define Wallacea as
- the region separating the continental Sunda shelf from Sahul. Coastlines are defined as -50 metres below present
- 777 mean sea level, equivalent to the low sea level stand estimated at ~50ka.
- 778







а



