

# 1 **The genomic history of Australia**

2 **The human population history of Australia remains contentious, not least because of a lack of**  
3 **large extensive genomic data. We generated high-coverage genomes for 83 geographically diverse**  
4 **Aboriginal Australians (all speakers of Pama-Nyungan languages) and 25 Papuans from the New**  
5 **Guinea Highlands. We find that Papuan and Aboriginal Australian ancestors diversified from**  
6 **each other 25-40 thousand years ago (kya), suggesting early population structure in the ancient**  
7 **continent of Sahul (Australia, New Guinea and Tasmania). However, all contemporary**  
8 **Aboriginal Australian studied descend from a single founding population that differentiated**  
9 **around 10-32 kya. We find evidence for a population expansion in northeast Australia during the**  
10 **Holocene (past c.10 kya) associated with limited gene flow from this region to the rest of**  
11 **Australia. This is broadly consistent with the spread of the Pama-Nyungan languages and**  
12 **cultural changes taking place across the continent in the mid-Holocene. We find evidence for a**  
13 **single out of Africa dispersal for all contemporary humans and estimate that Aboriginal**  
14 **Australians and Papuans shared a common ancestor with other Eurasians 60-100 kya, with**  
15 **subsequent admixture with different archaic populations. Finally, we report evidence of selection**  
16 **in Aboriginal Australians potentially associated with living in the desert.**

17 During most of the last 100 ky, Australia, Tasmania and New Guinea formed a single continent, Sahul,  
18 which was separated from Sunda (the continental landmass including mainland and western island  
19 Southeast Asia) by a series of deep oceanic troughs never exposed by changes in sea level (the  
20 Wallacean region as defined by biogeographers). Colonisation of Sahul is thought to have required at  
21 least 8-10 separate sea crossings between islands<sup>1</sup>, potentially constraining the occupation of Australia  
22 and New Guinea by earlier hominins<sup>2</sup>. The age of the first occupation of Australia has been disputed.  
23 There are several archaeological sites in Australia dating to 40-45 kya (Figure 1), long argued to  
24 represent the age of first occupation<sup>3</sup> despite a few sites dating to  $\geq 50$  kya. However, recent studies  
25 support the earlier dates, suggesting that Sahul was first settled by 47.5-55 kya<sup>4-6</sup>. This is consistent  
26 with the earliest evidence for modern humans in Sunda at a similar time<sup>7</sup> (Figure 1). Moreover skeletal  
27 remains that share morphological similarities with the ancestors of Aboriginal Australians and Papuans  
28 are found in South East Asia up until about 3,5 kya<sup>8</sup>, suggesting that the ancestors of Aboriginal  
29 Australians and Papuans extended from Sahul to Sunda.

30 Historically, the morphological diversity among Aboriginal Australians was interpreted by some as  
31 indicating multiple ancestral migrations<sup>9-11</sup>, or descent from Javanese *Homo erectus*, with varying levels  
32 of gene flow from contemporaneous populations<sup>12</sup>. However, statistical analyses indicate that  
33 Australian crania show no evidence of *H. erectus* admixture<sup>13</sup>. Still, the distinctiveness of the  
34 Australian archaeological record has led to the suggestion that the ancestors of Aboriginal Australians  
35 and Papuans (hereafter referred to as Australo-Papuans), as well as a small number of other  
36 populations, left the African continent earlier than the ancestors of present-day Eurasians<sup>14</sup>. Although  
37 such multiple dispersals from Africa are supported by some genetic studies<sup>15,16</sup>, others have found  
38 support for only one out of Africa (OoA) event, with one<sup>17</sup> or two<sup>18</sup> independent founding waves into  
39 Asia, of which the earlier contributed to Australo-Papuan ancestry<sup>19,20</sup>. Recent genomic results have  
40 also shown that both Aboriginal Australian<sup>20</sup> and Papuan<sup>21</sup> ancestors likely admixed with Neanderthal  
41 and Denisovan archaic hominins after leaving Africa.

42 Once in Sahul, contact among groups would have been affected by rising sea-levels that separated the  
43 Australian continent from New Guinea and Tasmania 7-14.5 kya through the formation of the Arafura  
44 Sea and Bass Strait<sup>22,23</sup> (Figure 1). These events still appear to be part of the oral tradition of several  
45 Aboriginal Australian communities<sup>24</sup>. Similarly, environmental variation accentuated during the last  
46 glacial maximum (LGM) 19-26.5 kya, leading to greater desertification of Australia<sup>25</sup> and more  
47 challenging temperature gradients, appears to have had an impact on the number and density of human  
48 populations<sup>26,27</sup>. In the same context, morphological and physiological studies find that Aboriginal  
49 Australians living in the desert areas today have unique adaptations<sup>28-30</sup>, such as the absence of the  
50 increased metabolic rates observed in Europeans when exposed to the freezing night temperatures  
51 common in the desert<sup>31,32</sup>.

52 At the time of European contact, Aboriginal Australians spoke over 250 distinct languages<sup>33</sup>, two-thirds  
53 of which belong to the Pama-Nyungan family. The place of origin of this language family, which  
54 covers 90% of the Australian mainland, has been debated<sup>34</sup>, as has the effect of its extensive diffusion  
55 on its internal phylogenetic structure<sup>33</sup>. The pronounced similarity among Pama-Nyungan languages,  
56 together with shared socio-cultural patterns, have been interpreted as the result of a recent, mid-  
57 Holocene, expansion<sup>35</sup>. Other changes in the mid-late Holocene (~4 kya) include the efflorescence of  
58 backed blades (microliths<sup>36</sup>) and the introduction of the dingo<sup>37</sup>. The spatial distribution of microliths

59 roughly correlates with the Pama-Nyungan languages. It has even been suggested that Pama-Nyungan  
60 languages, dingoes and backed blades all reflect a recent migration into Australia<sup>38</sup>. Although an  
61 external origin for backed blades has been rejected<sup>36</sup>, dingoes were certainly introduced, most likely via  
62 island south-east Asia<sup>37</sup>. Rock art traditions also suggest contact between Sulawesi (Indonesia) and  
63 Australia<sup>38</sup>. Intriguingly, a recent genetic study found evidence of Indian gene flow into Australia at the  
64 approximate time of these Holocene changes<sup>39</sup>. Finally, substantial contact with Asians and Europeans  
65 is well documented in historical times<sup>40-43</sup>, suggesting potentially complex admixture among present-  
66 day Aboriginal Australians.

67 After a century of research, the origins and evolutionary history of Aboriginal Australians continue to  
68 be debated. To date, only three whole genome sequences have been described - one deriving from a  
69 historical tuft of hair from Western Desert Australia<sup>20</sup> and two others from cell lines with limited  
70 provenance information<sup>44</sup>. In this study we report the first extensive investigation of Aboriginal  
71 Australian genomic diversity by reporting and analysing the high-coverage genomes of 83 Pama-  
72 Nyungan-speaking Aboriginal Australians and 25 Highland Papuans.

## 73 **Dataset**

74 We collected saliva samples for DNA sequencing in collaboration with Aboriginal Australian  
75 communities and individuals in Australia (S01). We sequenced high-depth genomes (average depth of  
76 60X, range 20X-100X) from 83 Aboriginal Australian individuals representing a wide geographical  
77 distribution and a broad range of Pama-Nyungan languages (Figure 1, Extended Data Table 1, S02,  
78 S03, S04). Additionally, we sequenced 25 Highland Papuan genomes (38X-53X; S03) from five  
79 linguistic groups, and generated genotype data for 45 additional Papuans living or originating in the  
80 highlands (Figure 1). These datasets were combined with previously published genomes and SNP-chip  
81 genotype data, including Aboriginal Australian data from Arnhem Land and from a human diversity  
82 cell line panel from the European Collection of Cell Cultures<sup>44</sup> (ECCAC, Figure 1, S04).

83 We explored the extent of admixture in the Aboriginal Australian autosomal gene pool by estimating  
84 ancestry proportions with an approach based on sparse nonnegative matrix factorization (sNMF)<sup>45</sup>. We  
85 found that the genomic diversity of Aboriginal Australian populations is best modelled by a mixture of  
86 four main different genetic ancestries that can be assigned to four geographic regions based on their

87 relative frequencies: Europe, East Asia, New Guinea and Australia (Figure 2, Extended Data Figure 1,  
88 S05). The degree of admixture varies among groups (S05) with the Ngaanyatjarra speakers from  
89 central Australia (WCD) having a significantly higher “Aboriginal Australian component” (median  
90 value = 0.95) in their genomes compared to the median value of other Aboriginal Australian groups  
91 (median value = 0.64; Mann-Whitney rank sum test, one tail p-value = 3.55e-07). The “East Asian”  
92 and “Papuan” components are mostly present in northeastern Aboriginal Australian populations (Figure  
93 2b, Extended Data Figure 1, S05), while the “European component” is widely distributed across  
94 groups. In most of the subsequent analyses, we either selected specific samples or groups according to  
95 their level of Aboriginal Australian ancestry, or masked the data for the non-Aboriginal Australian  
96 ancestry genomic component (S06).

## 97 **Colonisation of Sahul and diversification of Australians and Papuans**

98 The origins of Aboriginal Australians is a source of much debate, as are the nature of the relationships  
99 among Aboriginal Australians and between Aboriginal Australians and Papuans. Using  $f_3$  statistics,  
100 estimates of genomic ancestry proportions and classical multi-dimensional scaling (MDS) analyses, we  
101 find that Aboriginal Australians and Papuans are closer to each other than to any other present-day  
102 worldwide population included in our study (Figure 2a, Figure 3a, S05). This is consistent with  
103 Aboriginal Australians and Papuans being derived from a common ancestral population, which initially  
104 colonised Sahul. Moreover, comparing outgroup  $f_3$  statistics we do not find significant differences  
105 between Papuan populations (highland Papuan groups and HGDP-Papuans) in their genetic affinities to  
106 Aboriginal Australians (Figure 3b), suggesting that the Papuan groups share a common ancestor after  
107 or at the same time as the divergence between Aboriginal Australians and Papuans.

108 To investigate the number of founding waves into Australia, we contrasted alternative models of  
109 settlement history through a composite likelihood method that compares the observed joint Site  
110 Frequency Spectrum (SFS) to that predicted under specific demographic models<sup>46,47</sup>(Figure 4a, S07).  
111 We compared the HGDP-Papuans to four Aboriginal Australian populations with low levels of  
112 European admixture (Extended Data Figure 1) from both northeastern (CAI and WPA) and  
113 southwestern (WON and WCD) Australia. We compared one and two-wave models where each  
114 Australian region was either colonized independently, or by descendants of a single Australian  
115 founding population after its divergence from Papuans. The one-wave model resulted in a better fit to

116 the observed SFS, suggesting that the ancestors of the sampled Aboriginal Australians diverged from a  
117 single ancestral population. This scenario is also supported by MDS analyses, even when masking  
118 Eurasian tracts, as well as by estimation of ancestry proportion analyses where all Aboriginal  
119 Australians form a cluster distinct from the Papuan populations (Figure 2, S05). Additionally, it is  
120 supported by  $f_3$  analyses where all Aboriginal Australians are largely equidistant from Papuans when  
121 adjusting for recent admixture (Figure 3c). Thus, our results based on 83 Pama-Nyungan speakers, do  
122 not support earlier claims of multiple ancestral migrations into Australia giving rise to contemporary  
123 Aboriginal Australian diversity<sup>9-11</sup>.

124 The SFS analysis suggests that there was a bottleneck in the ancestral Australo-Papuan population ~50  
125 kya (95% CI 35-54 kya, S07), which overlaps with archaeological evidence for the earliest occupation  
126 of both Sunda and Sahul, between 47.5-55 kya<sup>4,5,48</sup>. We further infer that the ancestors of Pama-  
127 Nyungan speakers and Highland Papuans diverged ~37 kya (95% CI 25-40 kya, Figure 4a, S07), which  
128 is in close agreement with results of an MSMC analysis (Figure 4b, S08), a method estimating cross  
129 coalescence rates between pairs of populations based on individuals' haplotypes<sup>49</sup>. It is also in  
130 agreement with previous estimates based on SNP array data<sup>39</sup> and the distribution of *Helicobacter*  
131 *pylori* strains<sup>50</sup>. These results imply that the divergence between sampled Papuans and Aboriginal  
132 Australians is older than the disappearance of the land bridge between New Guinea and Australia about  
133 8 kya, and suggest ancient genetic structure in Sahul. Such structure may be related to palaeo-  
134 environmental changes leading up to the onset of the LGM. Sedimentary studies show that the vast  
135 Lake Carpentaria (500 x 250 km, Figure 1) began to form ~40 kya, when sea-levels fell below the 53m-  
136 deep Arafura Sill<sup>51</sup>. Therefore, although Australia and New Guinea remained connected until the early  
137 Holocene, the flooding of the Carpentaria basin and its increasing salinity<sup>51</sup> may have promoted  
138 population isolation.

### 139 **Archaic admixture**

140 We characterised the number, timing and intensity of archaic gene flow events using three  
141 complementary approaches: SFS-based (Figure 4a, Figure 5c, S07), a goodness-of-fit analysis  
142 combining D-statistics (S09), and a method that directly infers putatively derived archaic 'haplotypes'  
143 (S11). Aboriginal Australians and Papuan genomes show an excess of putative Denisovan-derived

144 variants (Extended Data Figure 2d, S10), as well as substantially more putative Denisovan-derived  
145 haplotypes (PDH) than other non-Africans (Extended Data Figure 3). The number and total length of  
146 those putative haplotypes varied considerably across samples. However, the estimated number of PDH  
147 correlates almost perfectly ( $r^2 = 0.96$ ) with the estimated proportion of Australo-Papuan ancestry in  
148 each individual (Extended Data Figure 3). We also estimated that the values of  $F_{ST}$  between autosomal  
149 SNPs or PDHs assigned to WCD and Papuans were both around 0.12. Moreover, we found no  
150 significant difference in the distribution of the number of PDHs or the average length of PDHs between  
151 putatively unadmixed Australians and Papuans (Mann-Whitney U test,  $p > 0.05$ ). Taken together, these  
152 observations provide strong evidence for a single Denisovan admixture event that predates the  
153 population split between Australians and Papuans (see also<sup>52</sup>) and widespread recent Eurasian  
154 admixture in Aboriginal Australians (Figure 2, S05). Furthermore, using the SFS-based approach and  
155 constraining Denisovan admixture to have occurred before the Aboriginal Australian-Papuan  
156 divergence results in an admixture estimate of ~4% (95% CI 3-5%, Figure 5c, S07), similar to the  
157 estimates using D-statistics (~5%, S09). The SFS analyses further suggest that Denisovan/Australo-  
158 Papuan admixture took place ~44 kya (95% CI 31-50 kya, S07). We note that the point estimate for the  
159 age of the bottleneck overlaps with the confidence interval for the age of admixture, and that a  
160 bottleneck could have occurred anywhere along the dispersal route of Australo-Papuan populations  
161 from the ancestral source.

162 The SFS analysis also provides evidence for a primary Neanderthal admixture event (~2%, 95% CI 1-  
163 3%, Figure 5c, S07) taking place in the ancestral population of all non-Africans ~60 kya (95% CI 55-  
164 84 kya, Figure 5c, S07). Note that, although we cannot estimate absolute dates of archaic admixture  
165 from the lengths of PDHs and putative Neanderthal-derived haplotypes (PNHs), we can obtain a  
166 relative date. We found that for 20 putatively unadmixed Australians and 12 putatively unadmixed  
167 HGDP-Papuans, the average PNH length is 33.8 Kb and the average PDH length is 37.4 Kb. These are  
168 significantly different from each other ( $p = 9.65 * 10^{-6}$  using a conservative sign test), and suggest that  
169 the time since Neanderthal admixture was roughly 11% greater than the time since Denisovan  
170 admixture roughly in line with our SFS based estimates for Denisovan pulse (31-50 kya) versus the  
171 primary pulse of Neanderthal admixture (55-84 kya). The SFS analysis also suggests that the main  
172 Neanderthal pulse was followed by a further 1% (95% CI: 0.2-2.7%, Figure 5c, S07) pulse of

173 Neanderthal gene flow into the ancestors of Eurasians, and a smaller pulse into the ancestors of Asians  
174 (0.2%, 95% CI 0.1-1.0%, Figure 5c, S07), but there is little evidence for Neanderthal introgression  
175 private to Australo-Papuans, potentially limited to ~0.2% (95% CI 0.05-1.3%, Figure 5c, S07). In  
176 addition, the fact that the number of Neanderthal-specific introgressed sites increases from Europe to  
177 Australia (Extended Data Figure 2d, S10), and then decreases in Amerindians is consistent with  
178 recurrent Neanderthal (or Neanderthal-related archaic) gene flow during the waves of expansion into  
179 Eurasia. Our results are thus indicative of several pulses of Neanderthal gene flow into modern  
180 humans, as inferred previously<sup>53-55</sup>. Note however, the apparent high levels in Neanderthal-specific  
181 introgressed sites in Australo-Papuans can be explained by the expected number of misclassified  
182 Neanderthal introgressed sites resulting from the shared ancestry of these two archaic hominins (S10).  
183 Finally, using our SFS and haplotype based approaches, we explored additional models involving  
184 complex structure among the archaic populations. We found suggestive evidence that the archaic  
185 contribution could be more complex than a model involving discrete Denisovan and Neanderthal  
186 admixture pulses<sup>20,21</sup> (S07, S11), supporting the view that the archaic contribution in Australo-Papuans  
187 is likely more complex than was previously assumed<sup>20,21</sup> (S07).

## 188 **Out of Africa**

189 To investigate the relationship of Australo-Papuan ancestors to other world populations, we computed  
190 D-statistics<sup>56,57</sup> of the form ((H1=Aboriginal Australian,H2=Eurasian), H3=African) and  
191 ((H1=Aboriginal Australian,H2=Eurasian), H3=Ust'-Ishim). Several of these were significantly  
192 positive (S09), suggesting that Africans and Ust'-Ishim – a ~45 kya modern human from Asia<sup>58</sup> - are  
193 both closer to Eurasians than to Aboriginal Australians. These findings are in agreement with a model  
194 of Eurasians and Australo-Papuan ancestors dispersing from Africa in two independent waves.  
195 However, when correcting for a moderate amount of Denisovan admixture, Aboriginal Australians and  
196 Eurasians become equally close to Ust'-Ishim, as expected in a single OoA scenario (S09). Similarly,  
197 the D-statistics for ((H1=Aboriginal Australian, H2=Eurasian), H3=African) becomes much smaller  
198 after correcting for Denisovan admixture. Additionally, a goodness-of-fit approach combining D-  
199 statistics across worldwide populations indicates stronger support for two waves OoA, but when taking  
200 Denisovan admixture into account, a one-wave scenario fits the observed D-statistics equally well  
201 (Figure 5a, S09).

202 To further investigate the timing and number of OoA events giving rise to present-day Australo-Papuan  
203 and Eurasians (Sardinians and Han Chinese) we used the observed SFS in a model based composite  
204 likelihood framework. When considering only modern human genomes, we find evidence for two  
205 waves OoA, with a dispersal of Australo-Papuans ~14 ky before Eurasians (S07). However, when  
206 explicitly taking into account archaic Neanderthal and Denisovan introgression into modern  
207 humans<sup>44,59</sup>, the SFS analysis supports a single origin for the OoA populations marked by a bottleneck  
208 ~72 kya (95% CI 60-104 kya, S07). This scenario is reinforced by the observation that the ancestors of  
209 Australo-Papuan and Eurasians share a Neanderthal admixture event (95% CI 1.1-3.5%). Our analyses  
210 suggest that this single OoA ancestral population underwent two expansions at approximately the same  
211 time: one involving the ancestors of Australo-Papuan (51-72 kya) and the other, possibly slightly more  
212 recent, involving the ancestors of Eurasians (48-68 kya) (Figure 5c). Furthermore, modern humans have  
213 both an LD decay rate and a number of predicted deleterious homozygous mutations (recessive genetic  
214 load) that correlates with distance from Africa (S05, S10, and Extended Data Figure 2 a-c), again  
215 consistent with a single African origin. Aboriginal Australians also show levels of recessive load and  
216 LD that are intermediate between East Asians and Amerindians as expected if they all derive from the  
217 same OoA dispersal event.

218 The model estimated from the SFS analysis also suggests an early divergence of Australo-Papuans  
219 from the ancestors of all non-Africans, in agreement with two colonisation waves across Asia<sup>20,21,39</sup>.  
220 Under our best model, Australo-Papuans began to diverge from Eurasians ~58 kya (95% CI 51-72 kya,  
221 Figure 5c, S07), whereas Europeans and East Asians diverged from each other ~42 kya (95% CI 29-55  
222 kya, Figure 5c, S07) in agreement with previous estimates<sup>19,39,60,61</sup>. We find evidence for high levels of  
223 gene flow between the ancestors of Eurasians and Australo-Papuans, suggesting that, after the  
224 fragmentation of the OoA population (“Ghost” in Figure 5c) 57-58 kya, the groups remained in close  
225 geographical proximity for some time before Australo-Papuan ancestors dispersed eastwards.  
226 Furthermore, our results show multiple gene flow events between sub-Saharan Africans and Western  
227 Eurasians after ~42 kya. This supports previous findings of extensive contact between African and non-  
228 African populations<sup>60-62</sup>.

229 Our MSMC analyses suggest that the Yoruba/Australo-Papuans and the Yoruba/Eurasians cross-  
230 coalescence rates are distinct, implying that the Yoruba and Eurasian gene trees across the genome

231 have on average a more recent common ancestor (Figure 5b, S08). We show through simulations that  
232 these differences cannot be explained by archaic admixture. Moreover, the expected difference in  
233 phasing quality is not sufficient to fully explain this pattern either (see S08). While a similar separation  
234 in cross coalescence rate curves is obtained when comparing Eurasians or Australo-Papuans with  
235 Dinka, we find that, when comparing the Australo-Papuans or the Eurasians with the San, the cross  
236 coalescence curves are overlapping (S08). We also find that the change in effective population size  
237 through time of Aboriginal Australians, Papuans, and East Asians is very similar until around 50 kya,  
238 including a deep bottleneck around 60 kya (Extended Data Figure 7). Taken together, these MSMC  
239 results suggest complex population structure in Africa preceding a split of a single non-African  
240 ancestral population, combined with gene flow between the ancestors of Yoruba or Dinka (but not San)  
241 and the ancestors of Eurasians, which is not shared with Australo-Papuans. These results are  
242 qualitatively in line with the SFS-based analyses (see e.g., Figure 5b).

243

## 244 **Genetic structure of Aboriginal Australians**

245 Uniparental haplogroup diversity in this dataset (Extended Data Table 1, S12) is consistent with  
246 previous studies of mitochondrial DNA (mtDNA) and Y chromosome variation in Australia and  
247 Oceania, including the presence of typically European, Southeast and East Asian lineages<sup>63-68</sup>. The  
248 combined results provide important insights into the social structure of Aboriginal Australian societies.  
249 Aboriginal Australian groups exhibit greater between-group variation for mtDNA (16.8%) than for the  
250 Y chromosome (11.3%), in contrast to the pattern for most human populations<sup>69,70</sup>. This result suggests  
251 higher levels of male than female migration between Aboriginal Australian groups and may reflect the  
252 complex marriage and post-marital residence patterns among Pama-Nyungan Australian groups<sup>71</sup>.  
253 Moreover, the inferred European ancestry for the Y chromosome is much greater than that for mtDNA  
254 (31.8% vs. 2.4%), reflecting male-biased European gene flow into Aboriginal Australian groups during  
255 the colonial era.

256 Based on the genome sequences, we find genetic relationships within Australia that mirror geography,  
257 with a significant correlation ( $r_{\text{GEN,GEO}} = 0.59$ ,  $p\text{-value} < 0.0005$ ) when comparing the first two  
258 dimensions in an MDS analysis (S14). This correlation is higher when genomic regions of putative

259 recent European and East Asian (i.e., Han Chinese) origin are “masked” ( $r_{\text{GEN,GEO}} = 0.77$ , p-value <  
260 0.0005, Extended Data Figure 5). The main axis of genetic differentiation in the masked Aboriginal  
261 Australian genomes was determined using the Bearing correlogram approach. We found that an axis of  
262 angle =  $65^\circ$  compared to the equator (i.e., in the southwest to northeast direction) explains most of the  
263 genetic differentiation (S14).

264 Populations from the centre of the continent occupy positions genetically intermediate to this axis  
265 (Extended Data Figure 5). A similar result is observed with an  $F_{\text{ST}}$ -based tree for the masked data  
266 (Figure 6a, S05) as well as in analyses of genetic affinity based on the  $f_3$  statistic (Figure 3b),  
267 suggesting a population division between northeastern and southwestern groups. Such structure is  
268 further supported by the SFS analyses showing that populations from southwestern desert and  
269 northeastern regions diverged as early as ~31 kya (95% CI 10-32 kya), followed by limited gene flow  
270 (estimated  $2Nm < 0.01$ , 95% CI  $2 < Nm < 11.25$ ). The analysis of the major routes of gene flow within the  
271 continent supports the idea that the Australian interior has acted as a barrier to gene flow. Indeed, using  
272 a model inspired by principles of electrical engineering where gene flow is represented as a current  
273 flowing through the Australian continent and observed  $F_{\text{ST}}$  values are a measure of connectivity, we  
274 find that gene flow occurred preferentially along the coasts of Australia (Extended Data Figure 6, S14).  
275 These findings are consistent with a model of expansion followed by population fragmentation when  
276 and the extreme aridity in the interior of Australia<sup>25</sup> formed barriers to population movements during  
277 the LGM<sup>22</sup>.

278 We used MSMC based on autosomal data and mtDNA Bayesian Skyline Plots<sup>72</sup>(BSP) to estimate  
279 changes in effective population sizes within Australia. The MSMC analyses show evidence of a  
280 population expansion starting ~10 kya in the northeast, while both MSMC and BSP suggest a  
281 bottleneck in the southwestern desert populations taking place during the past ~10 kya (Extended Data  
282 Figure 7 , S08, S12). This is consistent with archaeological evidence for a population expansion  
283 associated with significant changes in socio-economic and subsistence strategies in the Holocene<sup>73,74</sup>.

284 European admixture almost certainly had not occurred before the late 18<sup>th</sup> century, but earlier East  
285 Asian and/or Papuan gene flow into Australia could have taken place. We characterized the mode and  
286 tempo of gene flow into Aboriginal Australians using three different approaches (S06, S07, S13). We  
287 used approximate Bayesian computation (ABC) to compare the observed mean and variance among

288 Aboriginal Australian individuals in the proportion of European, East Asian and Papuan admixture, to  
289 that computed from simulated datasets under various models of gene flow. We estimated the European  
290 and East Asian admixture to have occurred on the order of ten generations ago (S13), consistent with  
291 historical and ethnographic records. Consistent with this, the local ancestry approach based on RFMix  
292 suggests that the European and East Asian admixture is more recent than the Papuan admixture  
293 (Extended Data Figure 4a). In addition, both the ABC and SFS analyses suggest that the best fitting  
294 model for the Aboriginal Australian-Papuan data is one of continuous but modest gene flow, mostly  
295 unidirectional from Papuans to Aboriginal Australians, and geographically restricted to northeast  
296 Aboriginal Australians ( $2Nm=0.4$ , 95% CI 0.0-20.4, Figure 4a, S07).

297 To further investigate Papuan gene flow, we conducted follow-up analyses on the Papuan ancestry  
298 tracts obtained from the local ancestry analysis. We inferred local ancestry as the result of admixture  
299 between four components: European, East Asian, Papuans and Aboriginal Australian (S06). We chose  
300 WCD as the representative of Aboriginal Australian ancestry, because it is the least admixed  
301 population among our Australian samples (Figure 2, S05). Papuan tract length distribution show a clear  
302 geographic pattern, with “younger tracts” (higher median length and variance) in individuals closer to  
303 New Guinea and “older” (lower median length and variance) in individuals closer to WCD (Extended  
304 Data Figure 4b); there is a strong correlation of Papuan tract length variance with distance from WCD  
305 to other Aboriginal Australian groups ( $r=0.64$ ,  $p\text{-value}<0.0001$ ). The prevalence of short ancestry tracts  
306 of Papuan origin, compared to longer tracts of East Asian and European origin, suggests that a large  
307 fraction of the Papuan gene flow is much older than that from Europe and Asia, which is consistent  
308 with the ABC analysis (S13). We also investigated possible South Asian (Indian related) gene flow into  
309 Aboriginal Australian, as reported by a recent study<sup>39</sup>. However, we found no evidence of a component  
310 that can be uniquely assigned to Indian populations in the Aboriginal Australian gene pool using either  
311 admixture analyses or  $f_3$  and D-statistics (S05), even when including the original Aboriginal Australian  
312 genotype data from Arnhem Land. The different nature and size of the comparative datasets may  
313 account for the discrepancy in the results.

### 314 **Pama-Nyungan languages and genetic structure**

315 To investigate if linguistic relationships reflect genetic relationships among Aboriginal Australian  
316 populations, we built a Bayesian phylogenetic tree for the 28 different Pama-Nyungan languages

317 represented in this sample<sup>75</sup> (Figure 6b, S15). The linguistic and  $F_{ST}$ -based genetic trees obtained  
318 (Figure 6) share several well-supported partitions. For example, both trees indicate that the northeastern  
319 (CAI and WPA), and southwestern groups (ENY, NGA, WCD and WON) each form a cluster, while  
320 PIL, BDV and RIV are found between them. A distance matrix between pairs of languages, computed  
321 from the language-based tree, is significantly correlated with geographic distances ( $r_{GEO,LAN} = 0.83$ ,  
322 Mantel test two-tail p-value on 9,999 permutations = 0.0001). This suggests that differentiation among  
323 Pama-Nyungan languages in Australia follows geographic patterns, as observed in other language  
324 families elsewhere in the world<sup>15,76</sup>. Furthermore, we find a correlation between linguistics and genetics  
325 ( $r_{GEN,LAN} = 0.43$ , Mantel test p-value < 0.0005) that remains significant when controlling for geography  
326 ( $r_{GEN,LAN,GEO} = 0.26$ , Mantel test p-value < 0.0005). This is consistent with language differentiation after  
327 populations lose (genetic) contact with one another<sup>77</sup>. The correlation between the linguistic and genetic  
328 trees is all the more striking given the difference in time scales: the Pama-Nyungan family is generally  
329 accepted to have diversified within the last 6 ky<sup>78</sup>, while the genetic estimates are two to five times that  
330 age. The linguistic tree thus cannot simply reflect initial population dispersals, but rather reflects a  
331 genetic structure that has a complex history, with initial differentiation 10-32 kya, localised population  
332 expansions (northeast) and bottlenecks (southwest) ~10 kya, and subsequent limited gene flow from the  
333 northeast to the southwest. The latter may be the genetic signature that tracks the divergence of the  
334 Pama-Nyungan language family.

### 335 **Selection in Aboriginal Australians**

336 To identify any selection specific to Aboriginal Australians, we used two different methods based on  
337 the identification of SNPs with high allele frequency differences between Aboriginal Australians and  
338 other groups, similar to the often used Population-Branch Statistics<sup>79</sup> (PBS, S16). First, we scanned the  
339 Aboriginal Australian genomes for loci with an unusually large change in allele frequencies since the  
340 divergence from Papuans, taking recent admixture with Europeans and Asians into account. Among the  
341 top ranked genomic regions (Extended Data Table 2), we identified candidate loci that might be related  
342 to cold tolerance and dehydration resistance. One peak of high differentiation (the 7<sup>th</sup> highest peak) is  
343 located near the *NETO1* gene, which harbours alleles that have previously been shown to be associated  
344 with thyroid hormone levels. Interestingly, it has been suggested that thyroid hormone levels are  
345 associated with Aboriginal Australian specific adaptations to desert cold<sup>80</sup>. We investigated this

346 potential thermoregulatory adaptation further by identifying genomic regions showing high  
347 differentiation associated with different ecological regions in Australia (S16). The top candidate gene  
348 in this scan is *KCNJ2*, encoding a potassium channel protein harbouring alleles associated with  
349 thyrotoxic periodic paralysis<sup>81</sup>. This disease results from complications related to hyperthyroidism,  
350 providing additional support for the thyroid hormone system as a target of desert-related natural  
351 selection in Aboriginal Australians<sup>80</sup>.

352 Another locus of interest close to the 8<sup>th</sup> highest peak of differentiation, *SLC2A12*, is associated with  
353 serum urate levels<sup>82</sup>. The pathophysiology of dehydration includes elevated serum urate levels.  
354 Therefore, these results are suggestive of a locus that may be involved in tolerance to dehydration in  
355 Aboriginal Australians. Although further studies are needed to associate putative selected genetic  
356 variants in Aboriginal Australians with specific phenotypic effects, the current selection scan provides  
357 candidate genes for such future efforts.

## 358 **Discussion**

359 Our findings shed light, but also raise new questions, concerning on the population history of  
360 Aboriginal Australians. They suggest an early population structure in Sahul likely dating back ~37 kya  
361 (25-40 kya), when the ancestors of Highland Papuans and Pama-Nyungan Aboriginal Australians  
362 diversified. Intriguingly, despite this, our results also indicate that the population that diverged from  
363 Papuans was the ancestor of all the Aboriginal Australian groups sampled in this study; yet,  
364 archaeological evidence shows that by 40-45 kya, humans were widespread within Australia (Figure 1).  
365 Three non-exclusive demographic scenarios can account for this observation: 1) the Aboriginal  
366 Australian ancestral population prior to the divergence from Papuans was widespread, maintaining  
367 gene flow across the continent; 2) it was deeply structured, and only one group among the early settlers  
368 survived to give rise to Aboriginal Australians; and 3) other groups survived, but the descendants are  
369 not represented in our sample. Additional modern genomes, especially from Tasmania and the Non-  
370 Pama-Nyungan regions of the Northern Territory and Kimberley (both regions highly distinct  
371 linguistically<sup>83</sup> and not represented in our sample), as well as ancient genomes pre-dating European  
372 contact in Australia and other expansions across South East Asia<sup>38</sup>, should help resolve these questions  
373 in the future.

374 To add to this already complex picture, our estimates of ~44 kya (31-50 kya) for the time of admixture  
375 between the Australo-Papuan ancestors and an archaic hominin distantly related to Denisovans are very  
376 young. In the absence of paleontological evidence that archaic hominins crossed the Wallace Line,  
377 combined with evidence of much lower levels of Denisovan ancestry across East Asia and the  
378 Americas<sup>52,86</sup>, it is likely that the admixture occurred in Southeast Asia or even further to the west,  
379 constraining the age when the ancestors of living Australo-Papuan colonised Sahul and/or the actual  
380 timing of Denisovan admixture. In this context, it is noteworthy that our SFS based time estimates  
381 relies on the use of recently suggested molecular clock ( $1.25 \times 10^{-8}$ , see<sup>84</sup>) and generation time for  
382 humans (29 years<sup>85</sup>). Should any of these parameters change, our genetic-based time estimates will  
383 need revisions too.

384 Interestingly, our results also show that southwestern and northeastern Pama-Nyungan populations  
385 diverged 10-32kya. Together with the evidence for selection in genes that may have provided an  
386 advantage in extreme desert environments, such as those experienced in Western Desert populations  
387 during the LGM, these results point to a long-standing genetic structure among Pama-Nyungan  
388 Aboriginal Australians that survived post-glacial demographic changes. In other parts of the world,  
389 including South East Asia, Pleistocene demographic patterns were overlaid by post-glacial and  
390 Holocene expansions that left both genetic and linguistic regional signatures<sup>87</sup>. In Australia, the  
391 archaeological record also shows post-glacial expansions<sup>73,74</sup>, while the spread of Pama-Nyungan  
392 languages across the continent is generally accepted to be mid-to-late Holocene<sup>35</sup>. Our genetic findings  
393 indicate an early Holocene demographic expansion localized to northeast Aboriginal Australians, as  
394 well as gene flow spreading from the northeast across the continent. These observations are consistent  
395 with a possible origin and spread of the Pama-Nyungan languages from the northeast of Australia to the  
396 rest of the continent. Thus, evidence from genetics may add to the linguistic and cultural evidence -  
397 such as the spread of large ceremonial gatherings, trade and exchange intensification, broad alliance  
398 networks, cross-group male ritual induction, new plant foods, among several others<sup>35</sup> – that the  
399 dispersal of Pama-Nyungan languages has been driven by both cultural diffusion and demic expansion.

#### 400 **Data access**

401 The whole genome sequence data and SNP array data generated in this study are available upon request  
402 from E.W ([ewillerslev@snm.ku.dk](mailto:ewillerslev@snm.ku.dk)) and D.M.L. ([d.lambert@griffith.edu.au](mailto:d.lambert@griffith.edu.au)). The Papuan whole

403 genome sequence data generated in this study are also available under managed access through the  
404 EGA database (<https://www.ebi.ac.uk/ega>) under study accession number EGAS00001001247.

## 405 **References for the main text**

- 406 1. Birdsell, J. B. The recalibration of a paradigm for the first peopling of greater Australia. *Sunda Sahul Prehist.*  
407 *Stud. Southeast Asia Melanes. Aust.* 113–167 (1977).
- 408 2. Davidson, I. The colonization of Australia and its adjacent islands and the evolution of modern cognition.  
409 *Curr. Anthropol.* **51**, S177–S189 (2010).
- 410 3. O’Connell, J. F. & Allen, J. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of  
411 recent research. *J. Archaeol. Sci.* **31**, 835–853 (2004).
- 412 4. Summerhayes, G. R. *et al.* Human Adaptation and Plant Use in Highland New Guinea 49,000 to 44,000  
413 Years Ago. *Science* **330**, 78–81 (2010).
- 414 5. Clarkson, C. *et al.* The archaeology, chronology and stratigraphy of Madjedbebe (Malakunanja II): A site in  
415 northern Australia with early occupation. *J. Hum. Evol.* **83**, 46–64 (2015).
- 416 6. O’Connell, J. F. & Allen, J. The process, biotic impact, and global implications of the human colonization of  
417 Sahul about 47,000 years ago. *J. Archaeol. Sci.* **56**, 73–84 (2015).
- 418 7. Barker, G. *et al.* The ‘human revolution’ in tropical Southeast Asia: the antiquity of anatomically modern  
419 humans, and of behavioural modernity, at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52**, 243–261 (2007).
- 420 8. Matsumura, H. & Oxenham, M. F. Demographic transitions and migration in prehistoric East/Southeast  
421 Asia through the lens of nonmetric dental traits. *Am. J. Phys. Anthropol.* **155**, 45–65 (2014).
- 422 9. Topinard, P. *Etude sur les Tasmaniens.* (1869).
- 423 10. Birdsell, J. B. Preliminary data on the trihybrid origin of the Australian Aborigines. *Archaeol. Phys.*  
424 *Anthropol. Ocean.* 100–155 (1967).

- 425 11. Tbone, A. Morphological contrasts in Pleistocene Australians. *RL Kirk AG Tborne Eds Orig. Aust.* 95–1  
426 (1976).
- 427 12. Thorne, A. G. & Wolpoff, M. H. Regional continuity in Australasian Pleistocene hominid evolution. *Am. J.*  
428 *Phys. Anthropol.* **55**, 337–349 (1981).
- 429 13. Westaway, M. C. & Groves, C. P. The mark of Ancient Java is on none of them. *Archaeol. Ocean.* **44**, 84–95  
430 (2009).
- 431 14. Lahr, M. M. & Foley, R. Multiple dispersals and modern human origins. *Evol. Anthropol. Issues News Rev.* **3**,  
432 48–60 (1994).
- 433 15. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes*: (Princeton  
434 University Press, 1996).
- 435 16. Reyes-Centeno, H. *et al.* Genomic and cranial phenotype data support multiple modern human dispersals  
436 from Africa and a southern route into Asia. *Proc. Natl. Acad. Sci.* **111**, 7248–7253 (2014).
- 437 17. Consortium, T. H. P.-A. S. Mapping Human Genetic Diversity in Asia. *Science* **326**, 1541–1545 (2009).
- 438 18. Liu, H., Prugnolle, F., Manica, A. & Balloux, F. A Geographically Explicit Genetic Model of Worldwide  
439 Human-Settlement History. *Am. J. Hum. Genet.* **79**, 230–237 (2006).
- 440 19. Wollstein, A. *et al.* Demographic History of Oceania Inferred from Genome-wide Data. *Curr. Biol.* **20**, 1983–  
441 1992 (2010).
- 442 20. Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia.  
443 *Science* **334**, 94–98 (2011).
- 444 21. Reich, D. *et al.* Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and  
445 Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- 446 22. Clark, P. U. *et al.* The last glacial maximum. *science* **325**, 710–714 (2009).

- 447 23. Lewis, S. E., Sloss, C. R., Murray-Wallace, C. V., Woodroffe, C. D. & Smithers, S. G. Post-glacial sea-level  
448 changes around the Australian margin: a review. *Quat. Sci. Rev.* **74**, 115–138 (2013).
- 449 24. Nunn, P. D. & Reid, N. J. Aboriginal Memories of Inundation of the Australian Coast Dating from More than  
450 7000 Years Ago. *Aust. Geogr.* 1–37 (2015).
- 451 25. Reeves, J. M. *et al.* Climate variability over the last 35,000 years recorded in marine and terrestrial archives  
452 in the Australian region: an OZ-INTIMATE compilation. *Quat. Sci. Rev.* **74**, 21–34 (2013).
- 453 26. Veth, P. Islands in the Interior: A Model for the Colonization of Australia's Arid Zone. *Archaeol. Ocean.* **24**,  
454 81 (1989).
- 455 27. Hiscock, P. & Wallis, L. A. in *Desert Peoples* (eds. Veth, P., Smith, M. & Hiscock, P.) 34–57 (Blackwell  
456 Publishing Ltd, 2005). at <<http://onlinelibrary.wiley.com/doi/10.1002/9780470774632.ch3/summary>>
- 457 28. Abbie, A. A. & Australian Institute of Aboriginal Studies. *Studies in physical anthropology: volume II*.  
458 (Australian Institute of Aboriginal Studies, 1975). at <<http://catalog.hathitrust.org/Record/005995683>>
- 459 29. Kirk, R. L. *Aboriginal Man Adapting: The Human Biology of Australian Aborigines*. (Clarendon Press, 1981).
- 460 30. Birdsell, J. B. *Microevolutionary Patterns in Aboriginal Australia: A Gradient Analysis of Clines*. (Oxford  
461 University Press, 1993).
- 462 31. Scholander, P. F., Hammel, H. T., Hart, J. S., LeMessurier, D. H. & Steen, J. Cold Adaptation in Australian  
463 Aborigines. *J. Appl. Physiol.* **13**, 211–218 (1958).
- 464 32. Hammel, H. T., Elsner, R. W., Messurier, D. H. L., Andersen, H. T. & Milan, F. A. Thermal and metabolic  
465 responses of the Australian aborigine exposed to moderate cold in summer. *J. Appl. Physiol.* **14**, 605–615  
466 (1959).
- 467 33. Dixon, R. M. W. *Australian Languages: Their Nature and Development*. (Cambridge University Press, 2002).
- 468 34. Williams, A. N. *et al.* A continental narrative: Human settlement patterns and Australian climate change  
469 over the last 35,000 years. *Quat. Sci. Rev.* **123**, 91–112 (2015).

- 470 35. Evans, N. & McConvell, P. The enigma of Pama-Nyungan expansion in Australia. *Archaeol. Lang. II* 174–191  
471 (1997).
- 472 36. Hiscock, P. Review. *Archaeol. Ocean.* **43**, 44–47 (2008).
- 473 37. Savolainen, P., Leitner, T., Wilton, A. N., Matisoo-Smith, E. & Lundeberg, J. A detailed picture of the origin  
474 of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **101**,  
475 12387–12390 (2004).
- 476 38. Bellwood, P. *First Migrants: Ancient Migration in Global Perspective*. (Wiley-Blackwell, 2013).
- 477 39. Pugach, I., Delfin, F., Gunnarsdóttir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate  
478 Holocene gene flow from India to Australia. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1803–1808 (2013).
- 479 40. Haddon, A. C. (Alfred C. *et al. Reports of the Cambridge Anthropological Expedition to Torres Straits ..*  
480 (Cambridge [Eng.] : The University Press, 1901). at <<http://archive.org/details/reportsofcambrid02hadd>>
- 481 41. Macknight, C. C. Macassans and Aborigines. *Oceania* **42**, 283–321 (1972).
- 482 42. Chase, A. ‘All Kind of Nation’: Aborigines and Asians in Cape York Peninsula. *Aborig. Hist.* 7–19 (1981).
- 483 43. Macknight, C. C. Macassans and the Aboriginal Past. *Archaeol. Ocean.* **21**, 69–75 (1986).
- 484 44. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**,  
485 43–49 (2014).
- 486 45. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and Efficient Estimation of Individual  
487 Ancestry Coefficients. *Genetics* **196**, 973–983 (2014).
- 488 46. Nielsen, R. Estimation of Population Parameters and Recombination Rates From Single Nucleotide  
489 Polymorphisms. *Genetics* **154**, 931–942 (2000).
- 490 47. Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. Robust Demographic Inference from  
491 Genomic and SNP Data. *PLoS Genet* **9**, e1003905 (2013).

- 492 48. Allen, J. & O'Connell, J. F. Both half right: Updating the evidence for dating first human arrivals in Sahul.  
493 *Aust. Archaeol.* 86 (2014).
- 494 49. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome  
495 sequences. *Nat. Genet.* **46**, 919–925 (2014).
- 496 50. Moodley, Y. *et al.* The Peopling of the Pacific from a Bacterial Perspective. *Science* **323**, 527–530 (2009).
- 497 51. Holt, S. Palaeoenvironments of the Gulf of Carpentaria from the last glacial maximum to the present, as  
498 determined by foraminiferal assemblages. (2005).
- 499 52. Qin, P. & Stoneking, M. Denisovan Ancestry in East Eurasian and Native American Populations. *Mol. Biol.*  
500 *Evol.* msv141 (2015). doi:10.1093/molbev/msv141
- 501 53. Wall, J. D. *et al.* Higher Levels of Neanderthal Ancestry in East Asians than in Europeans. *Genetics* **194**, 199–  
502 209 (2013).
- 503 54. Vernot, B. & Akey, J. M. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes.  
504 *Science* **343**, 1017–1021 (2014).
- 505 55. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature advance*  
506 **online publication**, (2015).
- 507 56. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for Ancient Admixture between Closely Related  
508 Populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
- 509 57. Patterson, N. J. *et al.* Ancient Admixture in Human History. *Genetics* genetics.112.145037 (2012).  
510 doi:10.1534/genetics.112.145037
- 511 58. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**,  
512 445–449 (2014).
- 513 59. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**,  
514 222–226 (2012).

- 515 60. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the Joint Demographic  
516 History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genet* **5**, e1000695  
517 (2009).
- 518 61. Lukić, S. & Hey, J. Demographic Inference Using Spectral Methods on SNP Data, with an Analysis of the  
519 Human Out-of-Africa Expansion. *Genetics* **192**, 619–639 (2012).
- 520 62. Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci.* **111**,  
521 2632–2637 (2014).
- 522 63. gounder Palanichamy, M. *et al.* Phylogeny of Mitochondrial {DNA} Macrohaplogroup N in India, Based on  
523 Complete Sequencing: Implications for the Peopling of South Asia. *Am. J. Hum. Genet.* **75**, 966–978 (2004).
- 524 64. Kivisild, T. *et al.* The Role of Selection in the Evolution of Human Mitochondrial Genomes. *Genetics* **172**,  
525 373–387 (2006).
- 526 65. Hudjashov, G. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA  
527 analysis. *Proc. Natl. Acad. Sci.* **104**, 8726–8730 (2007).
- 528 66. van Holst Pellekaan, S. M., Ingman, M., Roberts-Thomson, J. & Harding, R. M. Mitochondrial genomics  
529 identifies major haplogroups in Aboriginal Australians. *Am. J. Phys. Anthropol.* **131**, 282–294 (2006).
- 530 67. Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and  
531 New Guinean aborigines. *Genome Res.* **13**, 1600–1606 (2003).
- 532 68. Friedlaender, J. *et al.* Expanding Southwest Pacific Mitochondrial Haplogroups P and Q. *Mol. Biol. Evol.* **22**,  
533 1506–1517 (2005).
- 534 69. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in  
535 humans. *Nat. Genet.* **20**, 278–280 (1998).
- 536 70. Lippold, S. *et al.* Human paternal and maternal demographic histories: insights from high-resolution Y  
537 chromosome and mtDNA sequences. *Investig. Genet.* **5**, 13 (2014).

- 538 71. Radcliffe-Brown, A. R. The Social Organization of Australian Tribes. *Oceania* **1**, 34–63 (1930).
- 539 72. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian Coalescent Inference of Past Population  
540 Dynamics from Molecular Sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
- 541 73. Haberle, S. G. & David, B. Climates of change: human dimensions of Holocene environmental change in low  
542 latitudes of the PEPH transect. *Quat. Int.* **118-119**, 165–179 (2004).
- 543 74. Lourandos, H & David, B. in *in Bridging Wallace's Line: the Environmental and Cultural History and*  
544 *Dynamics of the SE Asian-Australasian Region* (ed. A.P. Kershaw, B. David, N. Tapper, D. Penny & J. Brown.)  
545 (97-118).
- 546 75. Bowern, C. & Atkinson, Q. Computational phylogenetics and the internal structure of Pama-Nyungan.  
547 *Language* **88**, 817–845 (2012).
- 548 76. Excoffier, L., Harding, R. M., Sokal, R. R., Pellegrini, B. & Sanchez-Mazas, A. Spatial differentiation of RH and  
549 GM haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities. *Hum. Biol.* **63**, 273–  
550 307 (1991).
- 551 77. Bowern, C. & Evans, B. *The Routledge Handbook of Historical Linguistics*. (Routledge, 2015).
- 552 78. Evans, N. & Jones, R. in *Archaeology and linguistics: aboriginal Australia in global perspective* (Oxford  
553 University Press Australia, 1997).
- 554 79. Yi, X. *et al.* Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* **329**, 75–78  
555 (2010).
- 556 80. Qi, X., Chan, W. L., Read, R. J., Zhou, A. & Carrell, R. W. Temperature-responsive release of thyroxine and  
557 its environmental adaptation in Australians. *Proc. R. Soc. Lond. B Biol. Sci.* **281**, 20132747 (2014).
- 558 81. Cheung, C.-L. *et al.* Genome-wide association study identifies a susceptibility locus for thyrotoxic periodic  
559 paralysis at 17q24.3. *Nat. Genet.* **44**, 1026–1029 (2012).

- 560 82. Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout among African  
561 Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum. Mol. Genet.* **20**,  
562 4056–4068 (2011).
- 563 83. Evans, N. *The Non-Pama-Nyungan Languages of Northern Australia: Comparative Studies of the*  
564 *Continent’s Most Linguistically Complex Region.* (Pacific Linguistics, Research School of Pacific and Asian  
565 Studies, Australian National University, 2003).
- 566 84. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution.  
567 *Nat. Rev. Genet.* **13**, 745–753 (2012).
- 568 85. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based  
569 population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
- 570 86. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl. Acad. Sci.* **108**, 18301–18306  
571 (2011).
- 572 87. Bellwood, P. Early Agriculturalist Population Diasporas? Farming, Languages, and Genes. *Annu. Rev.*  
573 *Anthropol.* **30**, 181–207 (2001).
- 574 Supplementary Information (see annex)
- 575 S01 Ethical approvals in relation to sampling in Australia
- 576 S02 Ethnography and linguistics for the Aboriginal Australian individuals
- 577 S03 Sample collection, DNA extraction, array genotyping, whole-genome sequencing and processing
- 578 S04 Reference panels, relatedness and runs of homozygosity
- 579 S05 Linkage disequilibrium (LD) and population structure within Australia
- 580 S06 Local ancestry
- 581 S07 Demographic inferences
- 582 S08 MSMC analysis
- 583 S09 D-statistic based tests using sampled reads from sequencing data
- 584 S10 Mutation load analysis

- 585 S11 Archaic gene flow
- 586 S12 Uniparental markers
- 587 S13 ABC analysis to characterize recent European, East Asian and Papuan gene flow
- 588 S14 Spatial analyses
- 589 S15 Computational phylogenetics: Pama-Nyungan languages
- 590 S16 Scan for positive selection
- 591