ARTICLE

doi:10.1038/nature18299

A genomic history of Aboriginal Australia

Anna-Sapfo Malaspinas[1,2,3]*, Michael C. Westaway[4]*, Craig Muller[1]*, Vitor C. Sousa[2,3]*, Oscar Lao[5,6]*, Isabel Alves[2,3,7]*, Anders Bergström[8]*, Georgios Athanasiadis[9], Jade Y. Cheng[9,10], Jacob E. Crawford[10,11], Tim H. Heupink[4], Enrico Macholdt[12], Stephan Peischl[3,13], Simon Rasmussen[14], Stephan Schiffels[15], Sankar Subramanian[4], Joanne L. Wright[4], Anders Albrechtsen[16], Chiara Barbieri[12,17], Isabelle Dupanloup[2,3], Anders Eriksson[18,19], Ashot Margaryan[1], Ida Moltke[16], Irina Pugach[12], Thorfinn S. Korneliussen[1], Ivan P. Levkivskyi[20], J. Víctor Moreno-Mayar[1], Shengyu Ni[12], Fernando Racimo[10], Martin Sikora[1], Yali Xue[8], Farhang A. Aghakhanian[21], Nicolas Brucato[22], Søren Brunak[23], Paula F. Campos[1,24], Warren Clark[25], Sturla Ellingvåg[26], Gudjugudju Fourmile[27], Pascale Gerbault[28,29], Darren Injie[30], George Koki[31], Matthew Leavesley[32], Betty Logan[33], Aubrey Lynch[34], Elizabeth A. Matisoo-Smith[35], Peter J. McAllister[36], Alexander J. Mentzer[37], Mait Metspalu[38], Andrea B. Migliano[29], Les Murgha[39], Maude E. Phipps[21], William Pomat[31], Doc Reynolds[40], Francois-Xavier Ricaut[22], Peter Siba[31], Mark G. Thomas[28], Thomas Wales[41], Colleen Ma'run Wall[42], Stephen J. Oppenheimer[43], Chris Tyler-Smith[8], Richard Durbin[8], Joe Dortch[44], Andrea Manica[18], Mikkel H. Schierup[9], Robert A. Foley[1,45], Marta Mirazón Lahr[1,45], Claire Bowern[46], Jeffrey D. Wall[47], Thomas Mailund[9], Mark Stoneking[12], Rasmus Nielsen[1,48], Manjinder S. Sandhu[8], Laurent Excoffier[2,3], David M. Lambert[4] & Eske Willerslev[1,8,18]

The population history of Aboriginal Australians remains largely uncharacterized. Here we generate high-coverage genomes for 83 Aboriginal Australians (speakers of Pama–Nyungan languages) and 25 Papuans from the New Guinea Highlands. We find that Papuan and Aboriginal Australian ancestors diversified 25–40 thousand years ago (kya), suggesting pre-Holocene population structure in the ancient continent of Sahul (Australia, New Guinea and Tasmania). However, all of the studied Aboriginal Australians descend from a single founding population that differentiated ~10–32 kya. We infer a population expansion in northeast Australia during the Holocene epoch (past 10,000 years) associated with limited gene flow from this region to the rest of Australia, consistent with the spread of the Pama–Nyungan languages. We estimate that Aboriginal Australians and Papuans diverged from Eurasians 51–72 kya, following a single out-of-Africa dispersal, and subsequently admixed with archaic populations. Finally, we report evidence of selection in Aboriginal Australians potentially associated with living in the desert.

During most of the last 100,000 years, Australia, Tasmania and New Guinea formed a single continent, Sahul, which was separated from Sunda (the continental landmass including mainland and western island Southeast Asia) by a series of deep oceanic troughs never exposed by changes in sea level. Colonization of Sahul is thought to have required at least 8–10 sea crossings between islands, potentially constraining the occupation of Australia and New Guinea by earlier hominins[1]. Recent assessments suggest that Sahul was settled by 47–55 kya[2,3] (Fig. 1). These dates align with those for the earliest evidence for modern humans in Sunda[4].

The distinctiveness of the Australian archaeological and fossil record has led to the suggestion that the ancestors of Aboriginal Australians and Papuans ('Australo-Papuans' hereafter) left the African continent earlier than the ancestors of present-day Eurasians[5]. Although some

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark. [2]Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland. [3]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. [4]Research Centre for Human Evolution, Environmental Futures Research Institute, Griffith University, Nathan, Queensland 4111, Australia. [5]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain. [6]Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. [7]Population and Conservation Genetics Group, Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal. [8]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [9]Bioinformatics Research Centre, Aarhus University, 8000 Aarhus, Denmark. [10]Department of Integrative Biology, University of California, Berkeley, California 94720, USA. [11]Verily Life Sciences, 2425 Garcia Ave, Mountain View, California 94043, USA. [12]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. [13]Interfaculty Bioinformatics Unit University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland. [14]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Building 208, 2800 Kongens Lyngby, Denmark. [15]Department for Archaeogenetics, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, D-07745 Jena, Germany. [16]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. [17]Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, D-07745 Jena, Germany. [18]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [19]Integrative Systems Biology Laboratory, Division of Biological and Environmental Sciences & Engineering, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Saudi Arabia. [20]Institute for Theoretical Physics, ETH Zürich, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland. [21]Jeffrey Cheah School of Medicine & Health Sciences, Monash University Malaysia, Jalan Lagoon Selatan, Sunway City, 46150 Selangor, Malaysia. [22]Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, UMR 5288, Centre National de la Recherche Scientifique, Université de Toulouse 3, 31073 Toulouse, France. [23]Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen N, Denmark. [24]CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal. [25]National Parks and Wildlife, Sturt Highway, Buronga, New South Wales 2739, Australia. [26]Explico Foundation, Vågavegen 16, 6900 Florø, Norway. [27]Giriwandi, Gimuy Yidinji Country, Queensland 4868, Australia. [28]Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK. [29]UCL Department of Anthropology, 14 Taviton Street, London WC1H 0BW, UK. [30]Yinhawangka elder, Perth, Western Australia 6062, Australia. [31]Papua New Guinea Institute of Medical Research, PO Box 60, Goroka, Papua New Guinea. [32]Archaeology, School of Humanities & Social Sciences, University PO Box 320, University of Papua New Guinea & College of Arts, Society & Education, James Cook University, Cairns, Queensland 4811, Australia. [33]Ngadju elder, Coolgardie, Western Australia 6429, Australia. [34]Wongatha elder, Kurrawang, Western Australia 6430, Australia. [35]Department of Anatomy, University of Otago, Dunedin 9054, New Zealand. [36]2209 Springbrook Road, Springbrook, Queensland 4213, Australia. [37]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [38]Estonian Biocentre, Riia 23b, Tartu 51010, Estonia. [39]86 Workshop Road, Yarrabah, Queensland 4871, Australia. [40]Esperance Nyungar elder, Esperance, Western Australia 6450, Australia. [41]Atakani Street, Napranum, Queensland 4874, Australia. [42]2 Wynnum North Road, Wynnum, Queensland 4178, Australia. [43]School of Anthropology and Museum Ethnography, Oxford University, Oxford OX2 6PE, UK. [44]Centre for Rock Art Research and Management, M257, University of Western Australia, Perth, Western Australia 6009, Australia. [45]Leverhulme Centre for Human Evolutionary Studies, Department of Archaeology and Anthropology, University of Cambridge, Fitzwilliam Street, Cambridge CB2 1QH, UK. [46]Department of Linguistics, Yale University, 370 Temple Street, New Haven, Connecticut 06520, USA. [47]Institute for Human Genetics, University of California, San Francisco, California 94143, USA. [48]Departments of Integrative Biology and Statistics, University of California, Berkeley, California 94720, USA.
*These authors contributed equally to this work.

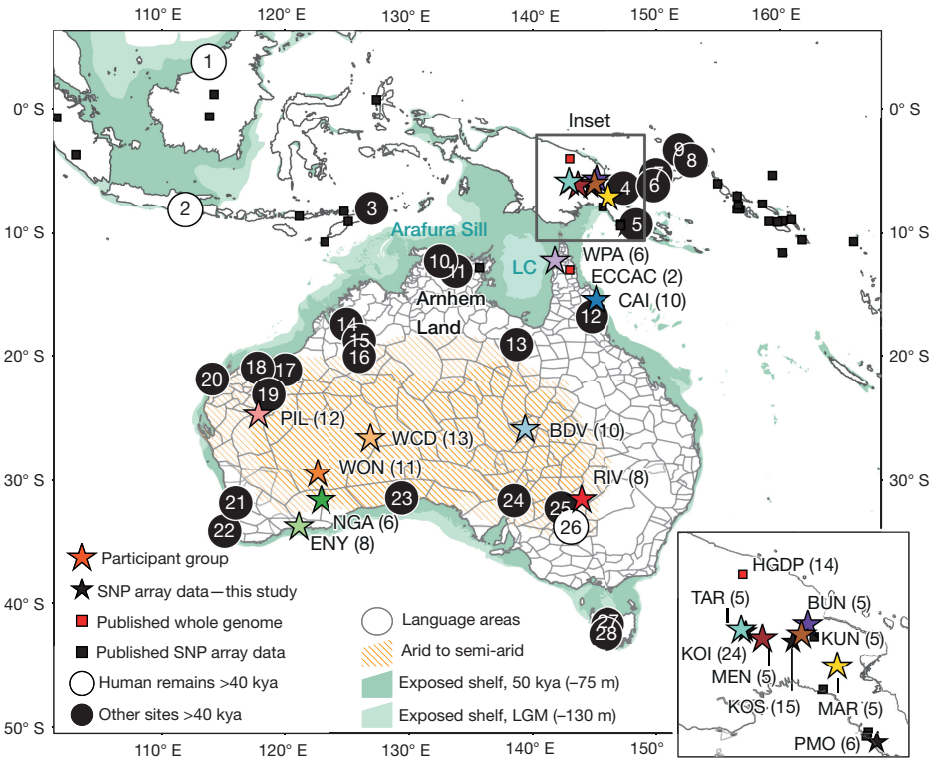


Figure 1 | Aboriginal Australian and Papuan samples used in this study, as well as archaeological sites and human remains dated to ~40 kya or older in southern Sunda and Sahul. The stars indicate the centroid location for each sampling group (sample size in parentheses). Publicly available genetic data (see Supplementary Information section S04) used as a reference panel in this study are shown as squares. Sites with dated human remains are shown as white circles and the archaeological sites as black circles. The associated dates can be found in Supplementary Information section S03. Grey boundaries correspond to territories defined by the language groups provided by the Australian Institute of Aboriginal and Torres Strait Islander Studies[45]. Sampled Aboriginal Australians self-identify primarily as: Yidindji and Gungandji from the Cairns region (CAI, $n = 10$, see also Supplementary Information section S02); Yupangati and Thanakwithi from northwest Cape York (WPA, $n = 6$), Wangkangurru and Yarluyandi from the Birdsville region (BDV, $n = 10$, 9 sequenced at high depth), Barkindji from southeast (RIV, $n = 8$); Pilbara area Yinhawangka and Banjima (PIL, $n = 12$), Ngaanyatjarra from western central desert (WCD, $n = 13$), Wongatha from Western Australia's northern Goldfields (WON, $n = 11$), Ngadju from Western Australia's southern Goldfields (NGA, $n = 6$); and Nyungar from southwest Australia (ENY, $n = 8$). Papuans include samples from the locations Bundi (BUN, $n = 5$), Kundiawa (KUN, $n = 5$), Mendi (MEN, $n = 5$), Marawaka (MAR, $n = 5$) and Tari (TAR, $n = 5$). We generated SNP array data (black stars) for 45 Papuan samples including 24 Koinambe (KOI) and 15 Kosipe (KOS)—described previously[46]—and 6 individuals with Highland ancestry sampled in Port Moresby (PMO). Lake Carpentaria (LC), which covered a significant portion of the land bridge between Australia and New Guinea 11.5–40 kya and thus potentially acted as a barrier to gene flow, is also indicated. Map data were sourced from the Australian Government, http://www.naturalearthdata.com/ and our research.

genetic studies support such multiple dispersals from Africa[6], others favour only one out-of-Africa (OoA) event, with one or two independent founding waves into Asia, of which the earlier contributed to Australo-Papuan ancestry[7,8]. In addition, recent genomic studies have shown that both Aboriginal Australian[8] and Papuan[9] ancestors admixed with Neanderthal and Denisovan archaic hominins after leaving Africa.

Increased desertification of Australia[10] during the last glacial maximum (LGM) 19–26.5 kya affected the number and density of human populations[11]. In this context, unique morphological and physiological adaptations have been identified in Aboriginal Australians living in desert areas today[12]. In particular, desert groups were hypothesized to withstand sub-zero night temperatures without showing the increase in metabolic rates observed in Europeans under the same conditions.

At the time of European contact, Aboriginal Australians spoke over 250 distinct languages, two-thirds of which belong to the Pama–Nyungan family and cover 90% of the Australian mainland[13]. The place of origin of this language family and the effect of its extensive diffusion on its internal phylogenetic structure have been debated[14], but the pronounced similarity among Pama–Nyungan languages, together with shared socio-cultural patterns, have been interpreted as resulting from a mid-Holocene expansion[15]. Other changes in the mid-late Holocene (~4 kya) include the proliferation of backed blades and the introduction of the dingo[16]. It has been suggested that Pama–Nyungan languages, dingoes and backed blades all reflect the same recent migration into Australia[17]. Although an external origin for backed blades has been rejected, dingoes were certainly introduced, most likely via island Southeast Asia[16]. A recent genetic study found evidence of Indian gene flow into Australia at the approximate time of these Holocene changes[18], suggesting a possible association, while substantial admixture with Asians and Europeans is well documented in historical times[19].

To date, only three Aboriginal Australian whole genome sequences have been described—one deriving from a historical tuft of hair from Australia's Western Desert[8] and two others from cell lines with limited provenance information[20]. In this study, we report the first extensive investigation of Aboriginal Australian genomic diversity by analysing the high-coverage genomes of 83 Pama–Nyungan-speaking Aboriginal Australians and 25 Highland Papuans.

Dataset

We collected saliva samples for DNA sequencing in collaboration with Aboriginal Australian communities and individuals in Australia (Supplementary Information section S01). We sequenced genomes at high-depth (average of 60×, range 20–100×) from 83 Aboriginal Australian individuals widely distributed geographically and linguistically (see Fig. 1 and associated legend for the location and label for each group as well as Extended Data Table 1, Supplementary Information

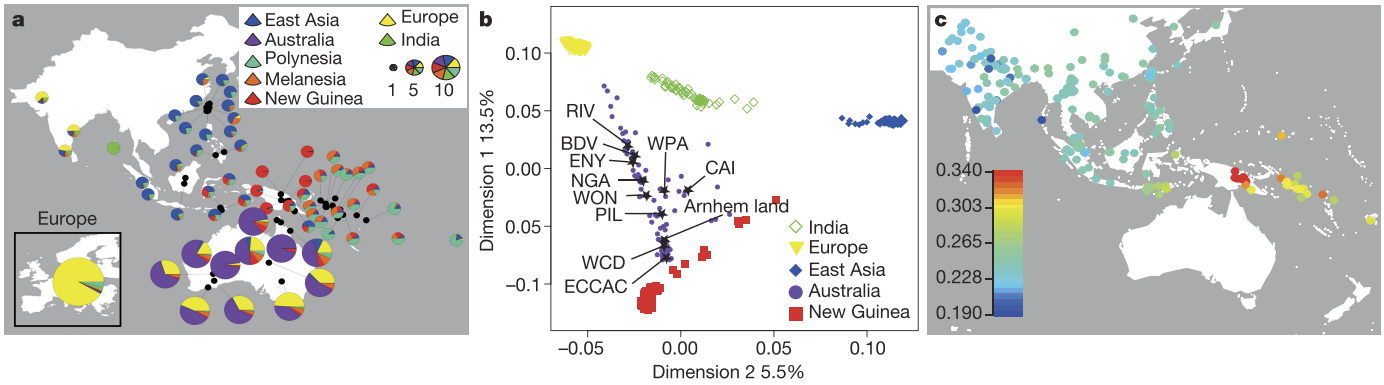


Figure 2 | Genetic ancestry of Aboriginal Australians in a worldwide context. a, Estimation of genomic ancestry proportions for the best number of ancestral components ($K = 7$) based on Aboriginal Australian and Papuan whole-genome sequence and SNP array data from this study (see Fig. 1), and publicly available SNP array data[18,26,47,57] (Supplementary Information section S05). Each ancestry component has been labelled according to the geographic region showing the corresponding highest frequency. The area of each pie chart is proportional to the sample size (as depicted in the legend). The genomes of Aboriginal Australian populations are mostly a mixture of European and Aboriginal Australian ancestry components. Northern Aboriginal Australian groups (Arnhem Land, CAI, ECCAC, PIL and WPA) are also assigned to components mainly present in East Asian populations, while northeastern Aboriginal Australian groups (CAI and WPA) also show components mainly present in New Guinean populations. A background of 5% 'Melanesian' component is observed in all the Aboriginal Australian populations; however, this component is widely spread over the geographic area shown in this figure, being present from Taiwan to India. We detected on average 1.5% 'Indian' component and 1.4% 'Polynesian' component across the Aboriginal Australian samples, but we attribute these residual ancestry components to statistical noise as they are present in other Southeast Asian populations and are not supported by other analyses (Supplementary Information section S05). **b**, Classical Multidimensional scaling (MDS) plot of first two dimensions based on an identity-by-state (IBS) distance matrix (based on 54,971 SNPs) between individuals from this study and worldwide populations, including publicly available data[18,26,47,57]. The first two dimensions explain 19% of the variance in the IBS distance matrix. Individuals are colour-coded according to sampling location, grouped into Australia (Arnhem Land, ECCAC, BDV, CAI, ENY, NGA, PIL, RIV, WCD, WON, WPA); East Asia (Cambodian, Dai, Han, Japanese, Naxi); Europe (English, French, Sardinian, Scottish, Spanish); India (Vishwabrahmin, Dravidian, Punjabi, Guaharati); and New Guinea (HGDP-Papuan, Central Province, Eastern Highlands, Gulf Province, Highlands, PMO, KOI, KOS, BUN, KUN, MEN, TAR, MAR). Stars indicate the centroid for each Aboriginal Australian group. Aboriginal Australians from this study as well as from previous studies are closest to Papuans and also show signals of admixture with Eurasians (see Supplementary Information section S05 for details). **c**, A heat map displaying outgroup f_3 statistics of the form f_3(Mbuti; WCD02, *X*), quantifying genetic drift shared between the putatively unadmixed individual WCD02 chosen to represent the Aboriginal Australian population, and various populations throughout the broader region for which either array genotypes or whole-genome sequencing data were publicly available or generated in this study. We used 760,116 SNPs for which WCD02 had non-missing array genotypes that overlapped with any other datasets. Standard errors, as estimated from block jack-knife resampling across the genome, were in the range 0.002–0.007.

sections S02–S04 for more information). Additionally, we sequenced 25 Highland Papuan genomes (38–53×; Supplementary Information sections S03, S04) from individuals representative of five linguistic groups, and generated genotype data for 45 additional Papuans living or originating in the Highlands (Fig. 1). These datasets were combined with previously published genomes and SNP array genotype data, including Aboriginal Australian data from Arnhem Land, and from a human diversity cell line panel from the European Collection of Cell Cultures[20] (ECCAC, Fig. 1, Supplementary Information section S04).

We explored the extent of admixture in the Aboriginal Australian autosomal gene pool by estimating ancestry proportions with an approach based on sparse non-negative matrix factorization (sNMF)[21]. We found that the genomic diversity of Aboriginal Australian populations is best modelled as a mixture of four main genetic ancestries that can be assigned to four geographic regions based on their relative frequencies: Europe, East Asia, New Guinea and Australia (Fig. 2a, Extended Data Fig. 1, Supplementary Information section S05). The degree of admixture varies among groups (Supplementary Information section S05), with the Ngaanyatjarra speakers from central Australia (WCD) having a significantly higher 'Aboriginal Australian component' (median value = 0.95) in their genomes than the other groups sampled (median value = 0.64; Mann–Whitney rank sum test, one-tail $P = 3.55 \times 10^{-7}$). The East Asian and New Guinean components are mostly present in northeastern Aboriginal Australian populations, while the European component is widely distributed across groups (Fig. 2a, Extended Data Fig. 1, Supplementary Information section S05). In most of the subsequent analyses, we either selected specific samples or groups according to their level of Aboriginal Australian ancestry, or masked the data for the non-Aboriginal Australian ancestry genomic components (Supplementary Information section S06).

Colonization of Sahul

The origin of Aboriginal Australians is a source of much debate, as is the nature of the relationships among Aboriginal Australians, and between Aboriginal Australians and Papuans. Using f_3 statistics[22], estimates of genomic ancestry proportions and classical multidimensional scaling (MDS) analyses, we find that Aboriginal Australians and Papuans are closer to each other than to any other present-day worldwide population considered in our study (Fig. 2b, c, Supplementary Information section S05). This is consistent with Aboriginal Australians and Papuans originating from a common ancestral population which initially colonized Sahul. Moreover, outgroup f_3 statistics do not reveal any significant differences between Papuan populations (Highland Papuan groups sampled in this study and the Human Genome Diversity Project (HGDP-Papuans)) in their genetic affinities to Aboriginal Australians (Extended Data Fig. 2a), suggesting that Papuan populations diverged from one another after or at the same time as they diverged from Aboriginal Australians.

To investigate the number of founding waves into Australia, we contrasted alternative models of settlement history through a composite likelihood method that compares the observed joint site frequency spectrum (SFS) to that predicted under specific demographic models[23] (Fig. 3, Supplementary Information section S07). We compared HGDP-Papuans to four Aboriginal Australian population samples with low levels of European admixture (Extended Data Fig. 1) from both northeastern (CAI and WPA) and southwestern desert (WON and WCD) Australia. We compared one- and two-wave models, where each Australian region was either colonized independently, or by descendants of a single Australian founding population after its divergence from Papuans. The one-wave model provides a better fit to the

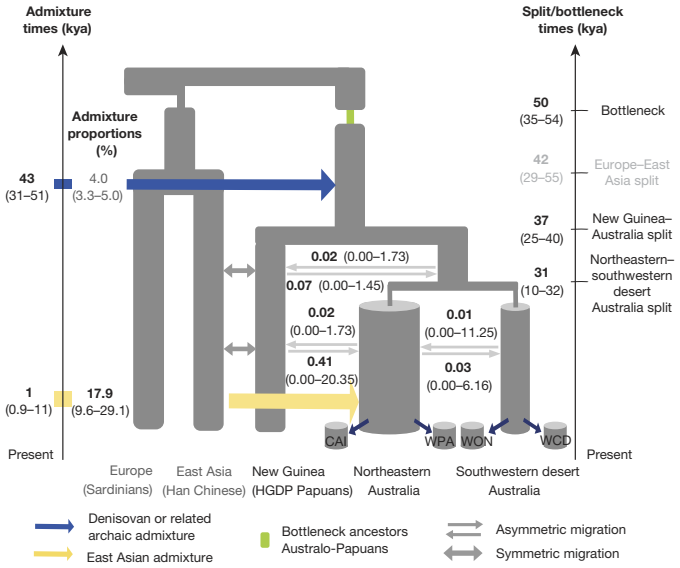


Figure 3 | Settlement of Australia. Best supported demographic model of the colonization of Australia and New Guinea. The demographic history of Aboriginal Australian populations was modelled by considering that sampled individuals are from sub-populations ('islands') that are part of two larger regions ('continents'), which geographically match the northeast and the southwestern desert regions of Australia. Maximum likelihood parameter estimates were obtained from the joint SFS of Han Chinese, HGDP-Papuans, CAI, WPA, WON and WCD. The 95% CI, obtained by non-parametric block bootstrap, are shown within parentheses. Estimated migration rates scaled by the effective population size ($2Nm$) are shown above/below the corresponding arrows. Only Aboriginal Australian individuals with low European ancestry were included in this analysis. In this model, we estimated parameters specific to the settlement of Australia and New Guinea (numerical values shown in black); keeping all the other demographic parameters set to the point estimates shown in Fig. 4 (numerical value shown in grey here). Only admixture events involving proportions >0.5% are shown. The inferred parameters were scaled using a mutation rate of 1.25×10^{-8} per generation per site[41] and a generation time of 29 years corresponding to the average hunter–gatherer generation interval for males and females[42]. See Supplementary Information section S07 for further details.

observed SFS, suggesting that the ancestors of the sampled Aboriginal Australians diverged from a single ancestral population. This conclusion is also supported by MDS analyses (Fig. 2b), as well as by estimation of ancestry proportion[58] where all Aboriginal Australians form a cluster distinct from the Papuan populations (Extended Data Fig. 1, Supplementary Information section S05). Additionally, it is supported by outgroup f_3 analyses, where all Aboriginal Australians are largely equidistant from Papuans when adjusting for recent admixture (Extended Data Fig. 2b). Thus, our results, based on 83 Pama–Nyungan speakers, do not support earlier claims of multiple ancestral migrations into Australia giving rise to contemporary Aboriginal Australian diversity[24].

The SFS analysis indicates that there was a bottleneck in the ancestral Australo-Papuan population ~50 kya (95% confidence intervals (CI) 35–54 kya, Supplementary Information section S07), which overlaps with archaeological evidence for the earliest occupation of both Sunda and Sahul 47–55 kya[2–4]. We further infer that the ancestors of Pama–Nyungan speakers and Highland Papuans diverged ~37 kya (95% CI 25–40 kya, Fig. 3, Supplementary Information section S07), which is in close agreement with results of multiple sequentially Markovian coalescent (MSMC) analyses (Extended Data Fig. 2c, Supplementary Information section S08), a method estimating cross coalescence rates between pairs of populations based on individuals' haplotypes[25]. This result is also in agreement with previous estimates, for example, based on SNP array data[18].

Archaic admixture

We characterized the number, timing and intensity of archaic gene-flow events using three complementary approaches: SFS-based (Supplementary Information section S07), a goodness-of-fit analysis combining D-statistics (Supplementary Information section S09), and a method that infers putatively derived archaic 'haplotypes' (Supplementary Information section S10). Aboriginal Australian and Papuan genomes show an excess of putative Denisovan introgressed sites (Extended Data Fig. 3a, Supplementary Information section S11), as well as substantially more putative Denisovan-derived haplotypes (PDHs) than other non-Africans (Extended Data Fig. 3b, Supplementary Information section S10). The number and total length of those putative haplotypes vary considerably across samples. However, the estimated number of PDHs correlates almost perfectly ($r^2 = 0.96$) with the estimated proportion of Australo-Papuan ancestry in each individual (Extended Data Fig. 3c). We found no significant difference in the distribution of the number of PDHs or the average length of PDHs between putatively unadmixed Aboriginal Australians and Papuans (Mann–Whitney U-test, $P > 0.05$). Moreover, the genetic differentiation between WCD and Papuans was also similar for both autosomal SNPs and PDHs, with F_{ST} values around 0.12. Taken together, these analyses provide evidence for Denisovan admixture predating the population split between Aboriginal Australians and Papuans (see also refs 26, 53) and widespread recent Eurasian admixture in Aboriginal Australians (Fig. 2a, b, Supplementary Information section S05). By constraining Denisovan admixture as having occurred before the Aboriginal Australian–Papuan divergence, the SFS-based approach results in an admixture estimate of ~4.0% (95% CI 3.3–5.0%, Fig. 4, Supplementary Information section S07), similar to that obtained by D-statistics (~5%, Supplementary Information section S09). The SFS analyses further suggest that Denisovan/Australo-Papuan admixture took place ~44 kya (95% CI 31–50 kya, Supplementary Information section S07), a date that overlaps with an estimate from a more recent study[54].

The SFS analysis also provides evidence for a primary Neanderthal admixture event (~2.3%, 95% CI 1.1–3.5%) taking place in the ancestral population of all non-Africans ~60 kya (95% CI 55–84 kya, Fig. 4, Supplementary Information section S07). Although we cannot estimate absolute dates of archaic admixture from the lengths of PDHs and putative Neanderthal-derived haplotypes (PNHs) in our samples, we can obtain a relative date. We found that, for putatively unadmixed Aboriginal Australians and HGDP-Papuans, the average PNH and PDH lengths are 33.8 kb and 37.4 kb, respectively (Extended Data Fig. 3b). These are significantly different from each other ($P = 9.65 \times 10^{-6}$ using a conservative sign test), and suggest that the time since Neanderthal admixture was about 11% greater than the time since Denisovan admixture, roughly in line with our SFS-based estimates for the Denisovan pulse (31–50 kya, Fig. 4) versus the primary pulse of Neanderthal admixture (55–84 kya). The SFS analysis also indicates that the main Neanderthal pulse was followed by a further 1.1% (95% CI 0.2–2.7%, Fig. 4, Supplementary Information section S07) pulse of Neanderthal gene flow into the ancestors of Eurasians. Finally, using our SFS- and haplotype-based approaches, we explored additional models involving complex structure among the archaic populations. We found suggestive evidence that the archaic contribution could be more complex than the model involving the discrete Denisovan and Neanderthal admixture pulses[8,9] shown in Fig. 4 (Supplementary Information sections S07, S10).

Out of Africa

To investigate the relationship of Australo-Papuan ancestors with other world populations, we computed D-statistics[22] of the form ((H1 = Aboriginal Australian, H2 = Eurasian), H3 = African) and ((H1 = Aboriginal Australian, H2 = Eurasian), H3 = Ust'-Ishim). Several of these were significantly positive (Supplementary Information section S09), suggesting that Africans and Ust'-Ishim—the 45,000-year-old remains of a modern human from Asia[27]—are both closer to

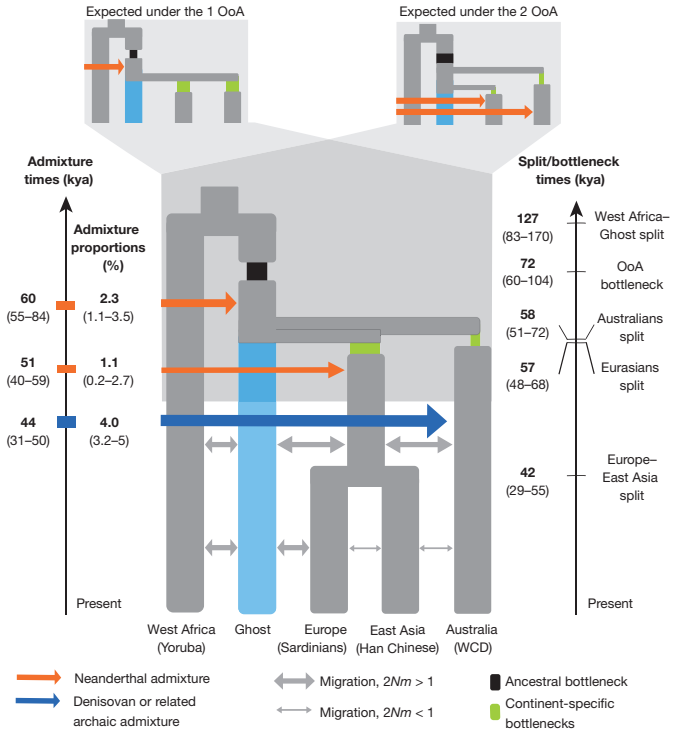


Figure 4 | Out of Africa. We used a likelihood-based approach to investigate whether the joint SFS supports the one-wave (1 OoA) or two-wave (2 OoA) scenarios. The maximum likelihood estimates are indicative of which scenario is best supported. As shown on the top left inset, under the 1 OoA scenario we expect (i) the presence of an ancestral bottleneck (in black); (ii) a relatively large Neanderthal admixture pulse shared by the ancestors of all non-Africans; and (iii) overlapping divergence times of the ancestors of Aboriginal Australians and Eurasians. In contrast, the top right inset shows parameters expected under a 2 OoA scenario: (i) a limited/absent ancestral bottleneck (in black) in the ancestors of all non-Africans; (ii) no shared Neanderthal admixture in the ancestors of all non-Africans; (iii) distinct divergence times for Aboriginal Australians and Eurasians. The main population tree shows the best fitting topology, which supports the 1 OoA scenario, and maximum likelihood estimates (MLEs) for the divergence and admixture times and the admixture proportions (with 95% CI obtained by non-parametric block bootstrap shown within parentheses). We assume that the OoA event is associated with the ancestral bottleneck. The 'Ghost' population represents an unsampled population related to Yoruba that is the source of the out-of-Africa event(s). Our results suggest that these two African populations split significantly earlier (~125 kya) than the estimated time of dispersals into Eurasia. Note that under a 1 OoA scenario, this ghost population becomes, after the ancestral bottleneck, the ancestral population of all non-Africans that admixed with Neanderthals. Arrow thicknesses are proportional to the intensity of gene flow and the admixture proportions, and only admixture events involving proportions >0.5% are displayed. The inferred parameters were scaled as for Fig. 3. See Supplementary Information section S07 for further details.

Eurasians than to Aboriginal Australians. These findings are in agreement with a model of Eurasians and Australo-Papuan ancestors dispersing from Africa in two independent waves. However, when correcting for a moderate amount of Denisovan admixture, Aboriginal Australians and Eurasians become equally close to Ust'-Ishim, as expected in a single OoA scenario (Supplementary Information section S09). Similarly, the D-statistics for ((H1 = Aboriginal Australian, H2 = Eurasian), H3 = African) became much smaller after correcting for Denisovan admixture. Additionally, a goodness-of-fit approach combining D-statistics across worldwide populations indicates stronger support for two waves OoA, but when taking Denisovan admixture into account, a one-wave scenario fits the observed D-statistics equally well (Extended Data Fig. 4a, b, Supplementary Information section S09).

To investigate the timing and number of OoA events giving rise to present-day Australo-Papuans and Eurasians further, we used the observed SFS in a model-based composite likelihood framework[23]. When considering only modern human genomes, we find evidence for two waves OoA, with a dispersal of Australo-Papuans ~14,000 years before Eurasians (Supplementary Information section S07). However, when explicitly taking into account Neanderthal and Denisovan introgression into modern humans[9,20], the SFS analysis supports a single origin for the OoA populations marked by a bottleneck ~72 kya (95% CI 60–104 kya, Fig. 4, Supplementary Information section S07). This scenario is reinforced by the observation that the ancestors of Australo-Papuans and Eurasians share a 2.3% (95% CI 1.1–3.5%) Neanderthal admixture pulse. Furthermore, modern humans have both a linkage disequilibrium decay rate and a number of predicted deleterious homozygous mutations (recessive genetic load) that correlate with distance from Africa (Supplementary Information sections S05, S11, Extended Data Fig. 5), again consistent with a single African origin.

The model best supported by the SFS analysis also suggests an early divergence of Australo-Papuans from the ancestors of all non-Africans, in agreement with two colonization waves across Asia[8,18]. Under our best model, Australo-Papuans began to diverge from Eurasians ~58 kya (95% CI 51–72 kya, Fig. 4, Supplementary Information section S07), whereas Europeans and East Asians diverged from each other ~42 kya (95% CI 29–55 kya, Fig. 4, Supplementary Information section S07), in agreement with previous estimates[7,18,28]. We find evidence for high levels of gene flow between the ancestors of Eurasians and Australo-Papuans, suggesting that, after the fragmentation of the OoA population ('Ghost' in Fig. 4) 57–58 kya, the groups remained in close geographical proximity for some time before Australo-Papuan ancestors dispersed eastwards. Furthermore, we find evidence for gene flow between sub-Saharan Africans and Western Eurasians after ~42 kya, in agreement with previous findings[28].

MSMC analyses suggest that the Yoruba/Australo-Papuans and the Yoruba/Eurasians cross-coalescence rates are distinct, implying that the Yoruba and Eurasian gene trees across the genome have, on average, more recent common ancestors (Extended Data Fig. 4c, Supplementary Information section S08). We show through simulations that these differences cannot be explained by typical amounts of archaic admixture (<20%, Extended Data Fig. 4d). Moreover, the expected difference in phasing quality among genomes is not sufficient to explain this pattern fully (Supplementary Information section S08). While a similar separation in cross coalescence rate curves is obtained when comparing Eurasians and Australo-Papuans with Dinka, we find that, when comparing Australo-Papuans and Eurasians with San, the cross coalescence curves overlap (Extended Data Fig. 4c). We also find that the inferred changes in effective population size through time of Aboriginal Australians, Papuans, and East Asians are very similar until around 50 kya, including a deep bottleneck around 60 kya (Extended Data Fig. 6a). Taken together, these MSMC results are consistent with a split of both Australo-Papuans and Eurasians from a single African ancestral population, combined with gene flow between the ancestors of Yoruba or Dinka (but not San) and the ancestors of Eurasians that is not shared with Australo-Papuans. These results are qualitatively in line with the SFS-based analyses (see Fig. 4). While our results do not exclude the possibility of an earlier OoA expansion, they do indicate that any such event left little trace in the genomes of modern Australo-Papuans, in line with conclusions from related work appearing alongside this study[55,56].

Genetic structure of Aboriginal Australians

Uniparental haplogroup diversity in this dataset (Extended Data Table 1, Supplementary Information section S12) is consistent with previous studies of mitochondrial DNA (mtDNA) and Y chromosome variation in Australia and Oceania[29], including the presence of typically European, Southeast and East Asian lineages[30]. The combined results

provide important insights into the social structure of Aboriginal Australian societies. Aboriginal Australians exhibit greater between-group variation for mtDNA (16.8%) than for the Y chromosome (11.3%), in contrast to the pattern for most human populations[31]. This result suggests higher levels of male- than female-mediated migration, and may reflect the complex marriage and post-marital residence patterns among Pama–Nyungan Australian groups[32]. As expected (Supplementary Information section S02), the inferred European ancestry for the Y chromosome is much greater than that for mtDNA (31.8% versus 2.4%), reflecting male-biased European gene flow into Aboriginal Australian groups during the colonial era.

On an autosomal level, we find that genetic relationships within Australia reflect geography, with a significant correlation ($r_{GEN,GEO} = 0.77$, $P < 0.0005$, Extended Data Fig. 7b) between the first two dimensions of an MDS analysis on masked genomes and geographical location (Supplementary Information section S13). Populations from the centre of the continent occupy genetically intermediate positions (Extended Data Fig. 7a, b). A similar result is observed with an F_{ST}-based tree for the masked genomic data (Extended Data Fig. 7c, Supplementary Information section S05) as well as in analyses of genetic affinity based on f_3 statistics (Extended Data Fig. 2a), suggesting a population division between northeastern and southwestern groups. This structure is further supported by SFS analyses showing that populations from southwestern desert and northeastern regions diverged as early as ~31 kya (95% CI 10–32 kya, Fig. 3), followed by limited gene flow (estimated scaled migration rate ($2Nm$) ~0.01, 95% CI 0.00–11.25). An analysis of the major routes of gene flow within the continent supports a model in which the Australian interior acted as a barrier to migration. Using a model inspired by principles of electrical engineering where gene flow is represented as a current flowing through the Australian continent and using observed F_{ST} values as a proxy for resistance, we infer that gene flow occurred preferentially along the coasts of Australia (Extended Data Fig. 7e–g, Supplementary Information section S13). These findings are consistent with a model of expansion followed by population fragmentation when the extreme aridity in the interior of Australia formed barriers to population movements during the LGM[33].

We used MSMC on autosomal data and mtDNA Bayesian skyline plots[34] (BSP) to estimate changes in effective population size within Australia. The MSMC analyses provide evidence of a population expansion starting ~10 kya in the northeast, while both MSMC and BSP indicate a bottleneck in the southwestern desert populations taking place during the past ~10,000 years (Extended Data Fig. 6, Supplementary Information sections S08, S12). This is consistent with archaeological evidence for a population expansion associated with significant changes in socio-economic and subsistence strategies in Holocene Australia[35].

European admixture almost certainly had not occurred before the late 18th century, but earlier East Asian and/or New Guinean gene flow into Australia could have taken place. We characterized the mode and tempo of gene flow into Aboriginal Australians using three different approaches (Supplementary Information sections S06, S07, S14). We used approximate Bayesian computation (ABC) to compare the observed mean and variance in the proportion of European, East Asian and Papuan admixture among Aboriginal Australian individuals to that computed from simulated datasets under various models of gene flow. We estimated European and East Asian admixture to have occurred approximately ten generations ago (Supplementary Information section S14), consistent with historical and ethnographic records. Consistent with this, a local ancestry approach suggests that European and East Asian admixture is more recent than Papuan admixture (Extended Data Fig. 8, Supplementary Information section S06). In addition, both ABC and SFS analyses indicate that the best-fitting model for the Aboriginal Australian-Papuan data is one of continuous but modest gene flow, mostly unidirectional from Papuans to Aboriginal Australians, and geographically restricted to northeast Aboriginal Australians ($2Nm = 0.41$, 95% CI 0.00–20.35, Fig. 3, Supplementary Information section S07).

To investigate gene flow from New Guinea further, we conducted analyses on the Papuan ancestry tracts obtained from the local ancestry analysis. We inferred local ancestry as the result of admixture between four components: European, East Asian, Papuan and Aboriginal Australian (Supplementary Information section S06). The Papuan tract length distribution shows a clear geographic pattern (Extended Data Fig. 8b); we find a significant correlation of Papuan tract length variance with distance from WCD to other Aboriginal Australian groups ($r = 0.64$, $P < 0.0001$). The prevalence of short ancestry tracts of Papuan origin, compared to longer tracts of East Asian and European origin, suggests that a large fraction of the Papuan gene flow is much older than that from Europe and Asia, consistent with the ABC analysis (Supplementary Information section S14). We also investigated possible South Asian (Indian-related) gene flow into Aboriginal Australians, as reported recently[18]. However, we found no evidence of a component that can be uniquely assigned to Indian populations in the Aboriginal Australian gene pool using either admixture analyses or f_3 and D-statistics (Supplementary Information section S05), even when including the original Aboriginal Australian genotype data from Arnhem Land. The different size and nature of the comparative datasets may account for this discrepancy.

Pama–Nyungan languages and genetic structure

To investigate whether linguistic relationships reflect genetic relationships among Aboriginal Australian populations, we inferred a Bayesian phylogenetic tree for the 28 different Pama–Nyungan languages represented in this sample[13] (Extended Data Table 1, Supplementary Information section S15). The resulting linguistic and F_{ST}-based genetic trees (Extended Data Fig. 7c, d) share several well-supported partitions. For example, both trees indicate that the northeastern (CAI and WPA) and southwestern groups (ENY, NGA, WCD and WON) form two distinct clusters, while PIL, BDV and RIV are intermediate. A distance matrix between pairs of languages, computed from the language-based tree, is significantly correlated with geographic distances ($r_{GEO,LAN} = 0.83$, Mantel test two-tail P on 9,999 permutations = 0.0001, Supplementary Information section S13). This suggests that differentiation among Pama–Nyungan languages in Australia follows geographic patterns, as observed in other language families elsewhere in the world[36]. Furthermore, we find a correlation between linguistics and genetics ($r_{GEN,LAN} = 0.43$, Mantel test $P < 0.0005$, Supplementary Information section S13) that remains significant when controlling for geography ($r_{GEN,LAN.GEO} = 0.26$, partial Mantel test $P < 0.0005$, Supplementary Information section S13). This is consistent with language differentiation after populations lost (genetic) contact with one another. The correlation between the linguistic and genetic trees is all the more notable given the difference in time scales: the Pama–Nyungan family is generally accepted to have diversified within the last 6,000 years[37], while the genetic estimates are two to five times that age. The linguistic tree thus cannot simply reflect initial population dispersals, but rather reflects a genetic structure that has a complex history, with initial differentiation 10–32 kya, localized population expansions (northeast) and bottlenecks (southwest) ~10 kya, and subsequent limited gene flow from the northeast to the southwest. The latter may be the genetic signature that tracks the divergence of the Pama–Nyungan language family.

Selection in Aboriginal Australians

To identify selection signatures specific to Aboriginal Australians, we used two different methods based on the identification of SNPs with high allele-frequency differences between Aboriginal Australians and other groups, similar to the population-branch statistics[38] (PBS, Supplementary Information section S16). First, we scanned the Aboriginal Australian genomes for loci with unusually large changes in allele frequency since divergence from Papuans, taking recent

admixture with Europeans and Asians into account ('global scan'). Second, we identified genomic regions showing high differentiation associated with different ecological regions within Australia ('local scan', Supplementary Information section S16). Among the top ranked peaks (Extended Data Table 2) we found genes associated with the thyroid system (*NETO1*, seventh peak in the global scan, and *KCNJ2*, first peak in the local scan) and serum urate levels (eighth peak in the global scan). Thyroid hormone levels are associated with Aboriginal-Australian-specific adaptations to desert cold[39] and elevated serum urate levels with dehydration[40]. These genes are therefore candidates for potential adaptation to life in the desert. However, further studies are needed to associate putative selected genetic variants with specific phenotypic adaptations in Aboriginal Australians.

Discussion

Australia has one of the longest histories of continuous human occupation outside Africa, raising questions of origins, relatedness to other populations, differentiation and adaptation. Our large-scale genomic data and analyses provide some answers but also raise new questions. We find that Aboriginal Australians and Eurasians share genomic signatures of an OoA dispersal—a common African ancestor, a bottleneck and a primary pulse of Neanderthal admixture. However, Aboriginal Australian population history diverged from that of other Eurasians shortly after the OoA event, and included private admixture with another archaic hominin.

Our genetic-based time estimates are relative, and to obtain absolute dates we relied on two rescaling parameters: the human mutation rate and generation time (assumed to be 1.25×10^{-8} per generation per site and 29 years, respectively, based on recent estimates[41,42]). Although the absolute estimates we report would need to be revised if these parameters were to change, the current values can be the starting point of future research and should be contextualized.

We find a relatively old divergence between the ancestors of Pama–Nyungan speakers and Highland Papuans, only ~10% younger than the European–East Asian split time. With the assumed rescaling parameters this corresponds to ~37 kya (95% CI 25–40 kya), implying that the divergence between sampled Papuans and Aboriginal Australians is older than the disappearance of the land bridge between New Guinea and Australia ~7–14.5 kya, and thus suggests ancient genetic structure in Sahul. Such structure may be related to palaeo-environmental changes leading up to the LGM. Sedimentary studies show that the large Lake Carpentaria (500 × 250 km, Fig. 1) formed ~40 kya, when sea levels fell below the 53-m-deep Arafura Sill[43]. Although Australia and New Guinea remained connected until the early Holocene, the flooding of the Carpentaria basin and its increasing salinity[43] may have thus promoted population isolation.

Our results imply that Aboriginal Australian groups are the descendants of the ancestral population that first colonized Australia[8,44]. They also indicate that the population that diverged from Papuans ~37 kya was ancestral to all Aboriginal Australian groups sampled in this study; yet, archaeological evidence shows that by 40–45 kya, humans were widespread within Australia (Fig. 1). Three non-exclusive scenarios could account for this observation: (1) the Aboriginal Australian ancestral population was widespread before the divergence from Papuans, maintaining gene flow across the continent; (2) it was deeply structured, and only one group survived to give rise to modern Aboriginal Australians; and (3) other groups survived, but the descendants are not represented in our sample. Additional genomes, especially from Tasmania and the non-Pama–Nyungan regions of the Northern Territory and Kimberley, as well as ancient genomes predating European contact in Australia and other expansions across Southeast Asia[17], may help resolve these questions in the future.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 October 2015; accepted 4 May 2016.
Published online 21 September 2016.

1. Davidson, I. The colonization of Australia and its adjacent islands and the evolution of modern cognition. *Curr. Anthropol.* **51,** S177–S189 (2010).
2. Clarkson, C. *et al.* The archaeology, chronology and stratigraphy of Madjedbebe (Malakunanja II): A site in northern Australia with early occupation. *J. Hum. Evol.* **83,** 46–64 (2015).
3. O'Connell, J. F. & Allen, J. The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. *J. Archaeol. Sci.* **56,** 73–84 (2015).
4. Barker, G.*et al.* The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behaviour of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52,** 243–261 (2007).
5. Lahr, M. M. & Foley, R. Multiple dispersals and modern human origins. *Evol. Anthropol. Issues News Rev.* **3,** 48–60 (1994).
6. Reyes-Centeno, H. *et al.* Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc. Natl Acad. Sci. USA* **111,** 7248–7253 (2014).
7. Wollstein, A. *et al.* Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20,** 1983–1992 (2010).
8. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–98 (2011).
9. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
10. Reeves, J. M. *et al.* Climate variability over the last 35,000 years recorded in marine and terrestrial archives in the Australian region: an OZ-INTIMATE compilation. *Quat. Sci. Rev.* **74,** 21–34 (2013).
11. Hiscock, P. & Wallis, L. A. in *Desert Peoples* (eds Veth, P., Smith, M. & Hiscock, P.) 34–57 (Blackwell Publishing Ltd, 2005).
12. Birdsell, J. B. *Microevolutionary Patterns in Aboriginal Australia: A Gradient Analysis of Clines.* (Oxford University Press, 1993).
13. Bowern, C. & Atkinson, Q. Computational phylogenetics and the internal structure of Pama–Nyungan. *Language* **88,** 817–845 (2012).
14. Dixon, R. M. W. *Australian Languages: Their Nature and Development.* (Cambridge University Press, 2002).
15. Evans, N. & McConvell, P. in *Archaeology and Language II: Archaeological Data and Linguistic Hypotheses* (eds Blench, R. & Spriggs, M.) Ch. 7 (Routledge, 1999).
16. Hiscock, P. *Archaeology of ancient Australia.* (Routledge, 2008).
17. Bellwood, P. *First Migrants: Ancient Migration in Global Perspective.* (Wiley-Blackwell, 2013).
18. Pugach, I., Delfin, F., Gunnarsdóttir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proc. Natl Acad. Sci. USA* **110,** 1803–1808 (2013).
19. Ellinghaus, K. Absorbing the 'Aboriginal problem': controlling interracial marriage in Australia in the late 19th and early 20th centuries. *Aborig. Hist.* **27,** 183–207 (2003).
20. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
21. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196,** 973–983 (2014).
22. Patterson, N. J. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
23. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9,** e1003905 (2013).
24. Thorne, A. G. in *The Origin of the Australians* (eds Kirk, R. L. & Thorne, A. G.) 95–112 (Canberra: Australian Institute of Aboriginal Studies, 1976).
25. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46,** 919–925 (2014).
26. Qin, P. & Stoneking, M. Denisovan Ancestry in East Eurasian and Native American Populations. *Mol. Biol. Evol.* **32,** 2665–2674 (2015).
27. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
28. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5,** e1000695 (2009).
29. Bergström, A. *et al.* Deep roots for Aboriginal Australian Y chromosomes. *Curr. Biol.* **26,** 809–813 (2016).
30. Hudjashov, G. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl Acad. Sci. USA* **104,** 8726–8730 (2007).
31. Lippold, S. *et al.* Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig. Genet.* **5,** 13 (2014).
32. Radcliffe-Brown, A. R. The social organization of Australian tribes. *Oceania* **1,** 34–63 (1930).
33. Veth, P. Islands in the interior: a model for the colonization of Australia's arid zone. *Archaeol. Ocean.* **24,** 81–92 (1989).
34. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22,** 1185–1192 (2005).

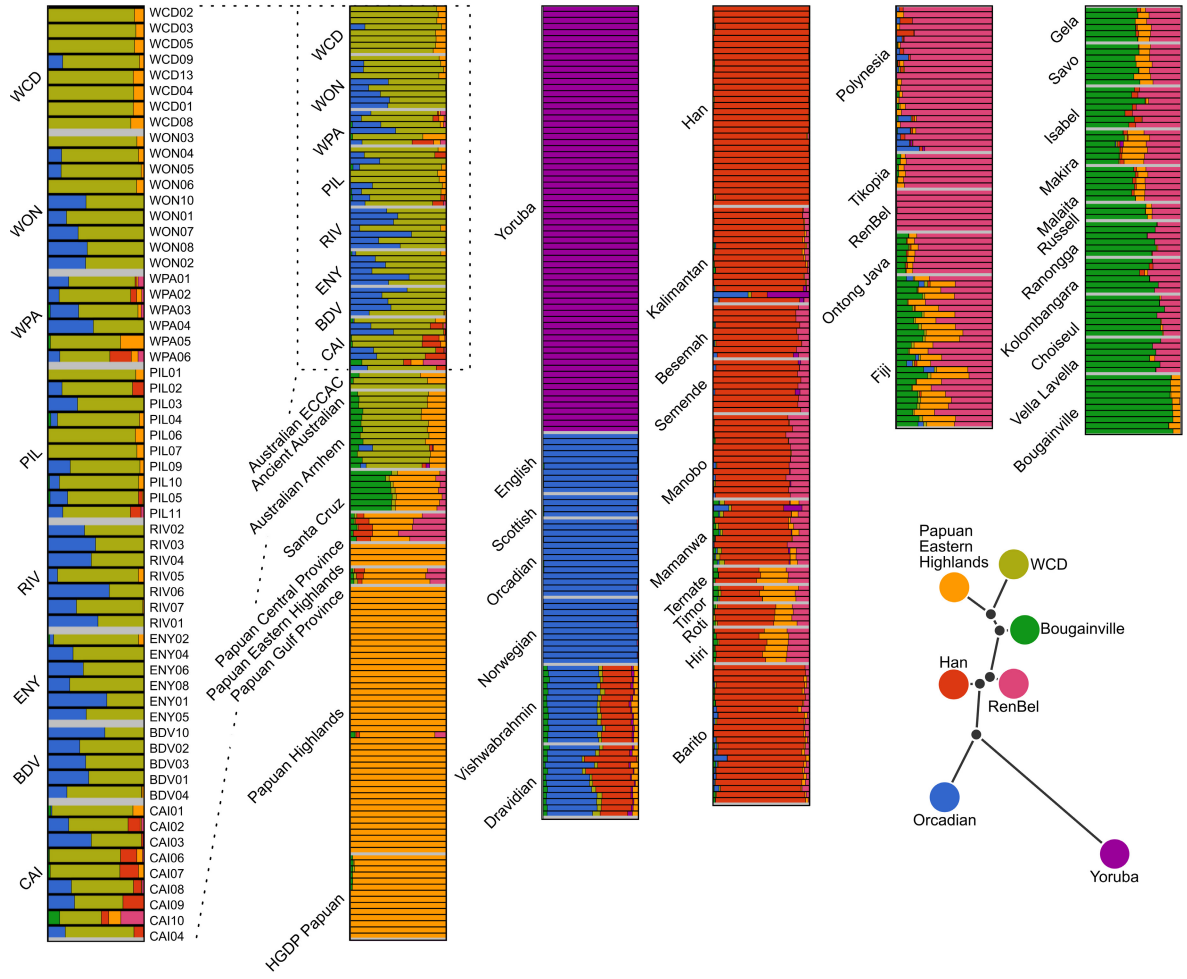
35. Lourandos, H. & David, B. in *Bridging Wallace's Line: the Environmental and Cultural History and Dynamics of the SE Asian-Australasian Region* (eds Kershaw, A. P., David, B., Tapper, N., Penny, D. & Brown, J.) *Advances in GeoEcology* **34,** 97–118 (2002).
36. Cavalli-Sforza, L. L. Genes, peoples and languages. *Proc. Natl Acad. Sci. USA* **91,** 7719–7724 (1997).
37. Evans, N. & Jones, R. in *Archaeology and linguistics: Aboriginal Australia in global perspective* (Oxford University Press Australia, 1997).
38. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329,** 75–78 (2010).
39. Qi, X., Chan, W. L., Read, R. J., Zhou, A. & Carrell, R. W. Temperature-responsive release of thyroxine and its environmental adaptation in Australians. *Proc. Biol. Sci.* **281,** 20132747 (2014).
40. Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum. Mol. Genet.* **20,** 4056–4068 (2011).
41. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13,** 745–753 (2012).
42. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128,** 415–423 (2005).
43. Holt, S. *Palaeoenvironments of the Gulf of Carpentaria from the Last Glacial Maximum to the Present, as Determined by Foraminiferal Assemblages.* PhD thesis, Univ. Wollongong (2005).
44. Heupink, T. H. *et al.* Ancient mtDNA sequences from the First Australians revisited. *Proc. Natl Acad. Sci. USA* **113,** 6892–6897 (2016).
45. Horton, D. (ed) *The Encyclopaedia of Aboriginal Australia.* (Aboriginal Studies Press, 1994).
46. Migliano, A. B. *et al.* Evolution of the pygmy phenotype: evidence of positive selection from genome-wide scans in African, Asian, and Melanesian pygmies. *Hum. Biol.* **85,** 251–284 (2013).
47. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
48. Wall, J. D. *et al.* Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics* **194,** 199–209 (2013).
49. Vernot, B. & Akey, J. M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343,** 1017–1021 (2014).
50. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
51. Wang, C. *et al.* Comparing spatial maps of human population-genetic variation using Procrustes analysis. *Stat. Appl. Genet. Mol. Biol.* **9,** 13 (2010).
52. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).
53. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl. Acad. Sci. USA* **108,** 18301–18306 (2011).
54. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr Biol.* **26,** 1241–1247 (2016).
55. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* http://dx.doi.org/10.1038/nature18964 (this issue).
56. Pagani, L. *et al.* Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* http://dx.doi.org/10.1038/nature19792 (this issue).
57. 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).
58. Cheng, J. Y., Mailund, T., & Nielsen, R. Ohana, a tool set for population genetic analyses of admixture components. *bioRxiv* doi:10.1101/071233 (2016).

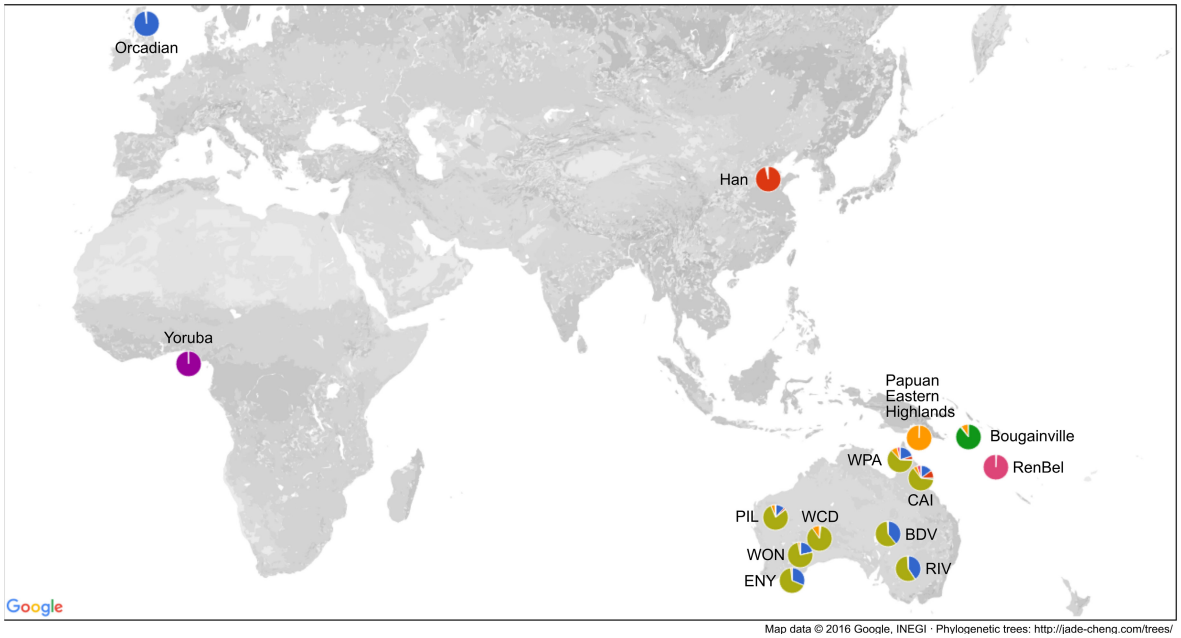
Supplementary Information is available in the online version of the paper.

Acknowledgements We thank all sample donors for contributing to this study. We thank Macrogen (http://www.macrogen.com/) for sequencing of the Aboriginal Australian genomes, M. Rasmussen, C. Der Sarkissian, M. Allentoft, D. Cooper, R. Gray, S. Greenhill, A. Seguin-Orlando, T. Carstensen, M. Przeworski, J. D. Jensen and L. Orlando for helpful discussions. We thank E. Thorsby for sample collection and contributing the DNA extract for the P2077 genome, I. Lissimore for support with data storage and distribution. We thank T. Parks, K. Auckland, K. Robson, A. V. Hill, J. B. Clegg, D. Higgs, D. J. Weatherall and M. Alpers for assistance in sample collection and discussion. L.E., V.C.S., I.A., I.D. and S.P. are grateful to the High Performance Computation platform of the University of Bern for providing access to the UBELIX cluster. This work was supported by the Danish National Research Foundation, the Lundbeck Foundation, the KU2016 grant and the Australian Research Council. A.-S.M. was supported by an ambizione grant with reference PZ00P3_154717 from the Swiss National Science Foundation (SNSF). M.C.W. was supported by the Australian Research Council (ARC) Discovery grants DP110102635 and DP140101405 and by a Linkage grant LP140100387. V.C.S., I.D. and S.P. were supported by SNSF grants to L.E. with references 31003A-143393 and CRSII3_141940. O.L. was supported by a Ramón y Cajal grant from the Spanish Ministerio de Economia y Competitividad (MINECO) with reference RYC-2013-14797 and by a BFU2015-68759-P (MINECO/FEDER) grant. I.A. was supported by a grant with reference SFRH/BD/73150/2010 from the Portuguese Foundation for Science and Technology (FCT). A.B., S.Sc., Y.X., C.T.-S. and R.D. were supported by a Wellcome Trust grant with reference WT098051. E.M., C.Ba., I.P., S.N. and M.St. acknowledge the Max Planck Society. S.Su. was supported by an ARC Discovery grant with reference DP140101405. J.L.W. was supported by a PhD scholarship from Griffith University. A.A. acknowledges the Villum foundation. I.M. was supported by a grant from the Danish Council for Independent Research with reference DFF–4090-00244. J.V.M.-M. acknowledges the Consejo Nacional de Ciencia y Tecnología (Mexico) for funding. N.B. and F.-X.R. were supported by the French Ministry of Foreign and European Affairs and French ANR with the grant ANR14-CE31-0013-01. S.B. was supported by a Novo Nordisk Foundation grant with reference NNF14CC0001. P.G. and A.B.M. were supported by a Leverhulme Programme grant number RP2011-R-045 to A.B.M. at UCL Department of Anthropology and M.G.T. at UCL Department of Genetics, Evolution and Environment. A.J.M. was supported by a Wellcome Trust grant with reference 106289/Z/14/Z. M.M. acknowledges the EU European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre; Estonian Institutional Research grant IUT24-1. M.G.T. was supported by a Wellcome Trust Senior Investigator Award with grant number 100719/Z/12/Z. S.J.O. was supported by a Wellcome Trust Core Award Grant Number 090532/Z/09/Z. A.Man. was supported by an ERC Consolidator Grant 647787 'LocalAdaptation'. M.E.P. would like to acknowledge the cardio-metabolic research cluster at Jeffrey Cheah School of Medicine & Health Sciences, Monash University Malaysia and Ministry of Science, Technology & Innovation, Malaysia for research grant 100-RM1/BIOTEK 16/6/2B. M.H.S. was supported by a grant from the Danish Independence Research Council with reference FNU 12-125062. R.A.F. was supported by the Leverhulme Trust. M.M.L. is supported by an ERC Advanced Grant 295907 'In-Africa'. C.Bo. was supported by USA National Science Foundation (NSF) grants BCS-0844550 and BCS-1423711, awarded to C.Bo. and Yale University. T.M. was supported by a grant from the Danish Independence Research Council with reference FNU 1323-00749. M.S.S. was supported by a Wellcome Trust grant with reference WT098051. L.E. was supported by Swiss NSF grant number 31003A-143393, D.M.L. was supported by ARC Discovery Grants DP110102635 and DP140101405 and Linkage grants LP140100387, LP120200144 and LP150100583. E.W. is grateful to St John's College in Cambridge for help and support.

Author Contributions G.A., J.Y.C., J.E.C., T.H.H., E.M., S.P., S.R., S.Sc., S.Su. and J.L.W. contributed equally and are listed alphabetically in the author list; A.A., C.Ba., I.D., A.E., A.Mar., I.M. and I.P. contributed equally and are listed alphabetically in the author list; T.S.K., I.P.L., J.V.M.-M., S.N., F.R., M.Si. and Y.X. contributed equally and are listed alphabetically in the author list. E.W. and D.M.L. initially conceived and headed the project. L.E. led the genetic load and the SFS-based demographic analyses. M.S.S. headed the research at the Wellcome Trust Sanger Institute. A.-S.M. planned and coordinated the genetic analyses and the sequencing of the Aboriginal Australian genomes. C.M., J.L.W., T.H.H., P.F.C., W.C., G.F., D.I., B.L., A.L., P.J.M., L.M., D.R., T.W., C.W., J.D., M.C.W. and E.W. collaborated with local groups to collect Aboriginal Australian samples. N.B., P.G., G.K., M.L., A.J.M., A.B.M., W.P., F.-X.R., P.S., M.G.T. and S.J.O. collaborated with local groups to collect Papuan samples. S.E. collaborated with local groups to collect the Rapanui sample. A.Mar. extracted DNA for the Aboriginal Australian genomes. M.S.S., A.B. and C.T.-S. coordinated the design and sequencing of the Papuan genomes. O.L., V.C.S., I.A., A.-S.M., A.B., G.A., J.Y.C., J.E.C., T.H.H., E.M., S.P., S.R., S.Sc., S.Su., J.L.W., A.A., C.Ba., I.D., A.E., A.Man., I.M., I.P., T.S.K., I.P.L., J.V.M.-M., S.N., F.R., M.Si., F.A., S.B., L.E., J.D.W. and T.M. analysed genetic data. C.Bo. collected and analysed linguistic data. L.E., E.W., D.M.L., Y.X., M.E.P., C.T.-S., R.D., M.S.S., A.Man., M.H.S., T.M., M.St. and R.N. supervised genetic analyses. M.C.W., C.M., W.C., G.F., D.I., B.L., A.L., P.J.M., L.M., D.R., T.W., C.W., E.A.M.-S., M.M., M.E.P., S.J.O., J.D., A.B.M., R.A.F. and M.M.L. provided archaeological, anthropological and historical context. A.-S.M., V.C.S., O.L., I.A., A.B., M.M.L., R.N., L.E., D.M.L. and E.W. wrote the manuscript with critical input from G.A., T.H.H., E.M., S.Sc., S.Su., J.L.W., C.Ba., A.E., I.P., E.A.M.-S., M.S.S., S.J.O., C.T.-S., R.D., M.G.T., J.D., A.Man., M.H.S., R.A.F., C.Bo., J.D.W., T.M., M.St. and all other coauthors. A.-S.M., V.C.S., O.L., I.A. and A.B. revised and compiled the Supplementary Information.

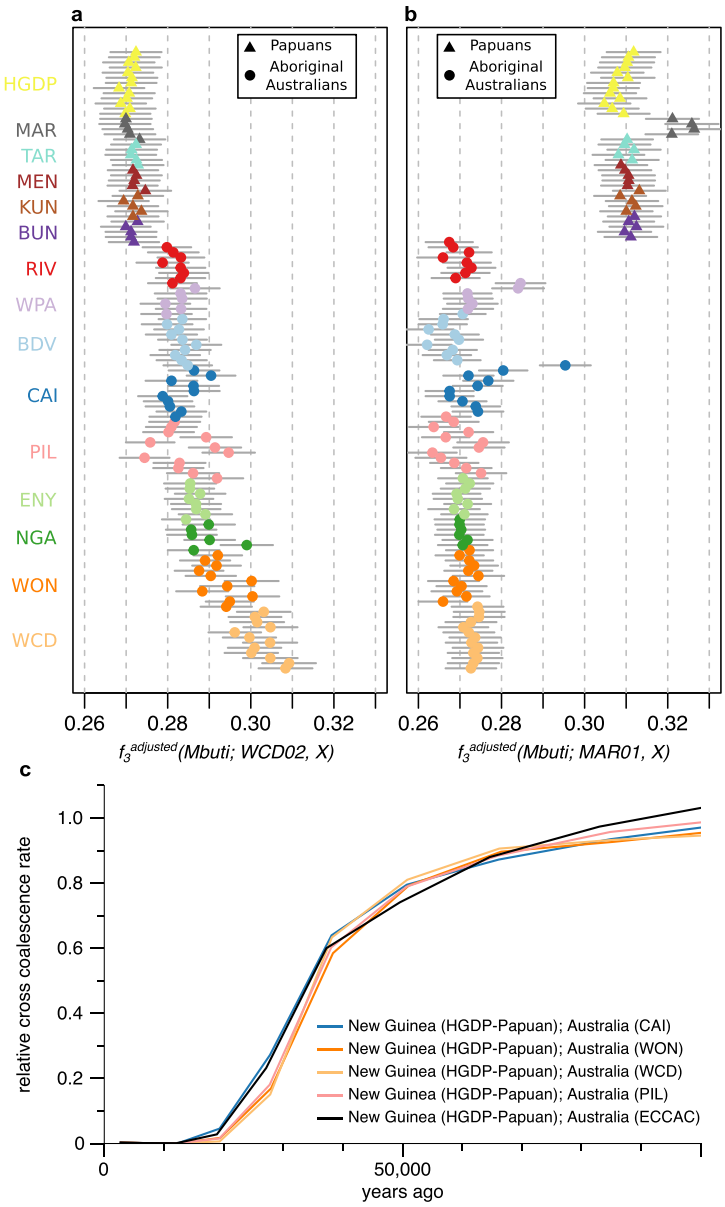
Author Information The Aboriginal Australian and Papuan whole genome sequence data generated in this study have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the EBI, under the accession numbers EGAS00001001766 and EGAS00001001247, respectively. The Papuan SNP array data generated in this study can be found under http://geogenetics.ku.dk/latest-news/alle_nyheder/2016/data. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. (ewillerslev@snm.ku.dk), D.M.L. (d.lambert@griffith.edu.au), L.E. (laurent.excoffier@iee.unibe.ch) and M.S.S. (ms23@sanger.ac.uk).



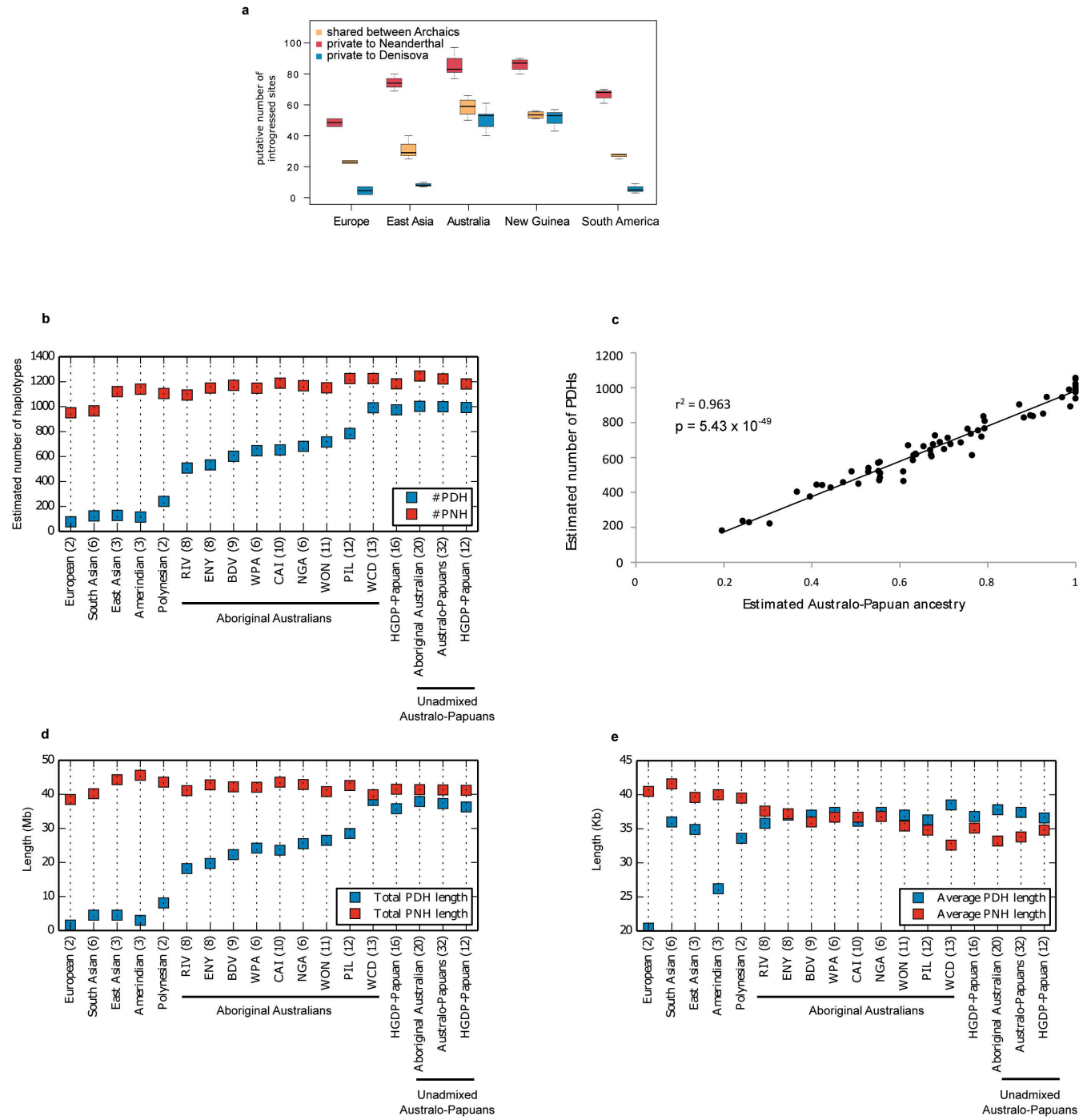


Extended Data Figure 1 | See next page for caption.

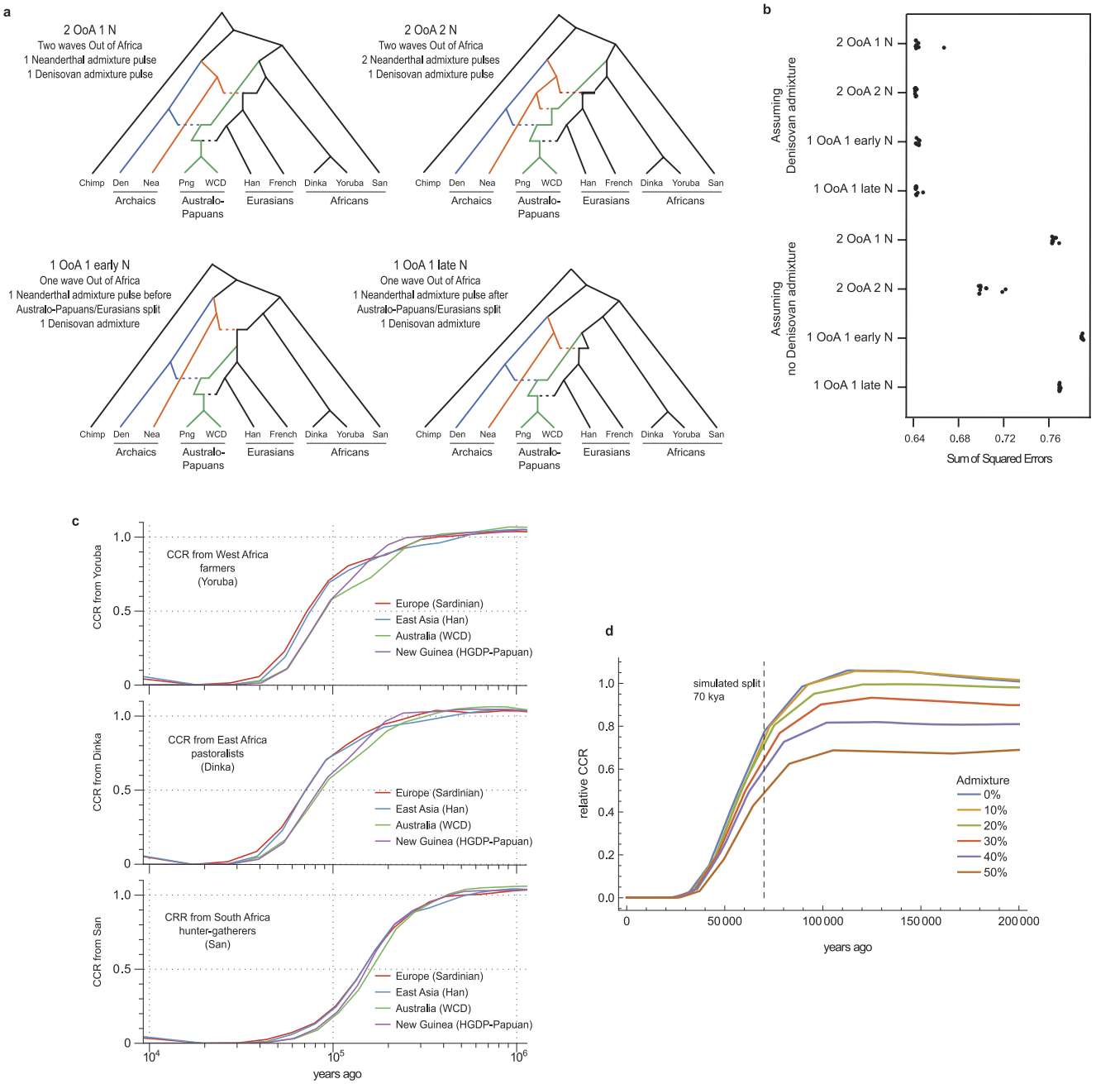
Extended Data Figure 1 | Per-individual admixture proportions of $K = 7$ ancestral components including Aboriginal Australians, New Guineans, Europeans, Africans, Melanesians and Polynesians. The genome of each individual is depicted as a bar and is coloured according to the estimated genome-wide proportions of ancestry components. An unrooted tree showing the relationships between the identified ancestral components is also estimated by our method. Each ancestry has been labelled with the name of the population (see also map) showing the highest fraction of that ancestral component. The cross-validation error is minimized for this value of K for fivefold cross-validation. The rooted tree supports the shared genetic origin of Aboriginal Australians, Papuans and Bougainvilleans. Note that only individuals with more than 50% of Aboriginal Australian ancestry in their genomes (defined in Supplementary Information section S06) were included in the analyses. Refer to ref. 58 and Supplementary Information section S05 for details about the method and the analysis. Map data ©2016 Google, INEGI. Tree constructed with http://jade-cheng.com/trees/.



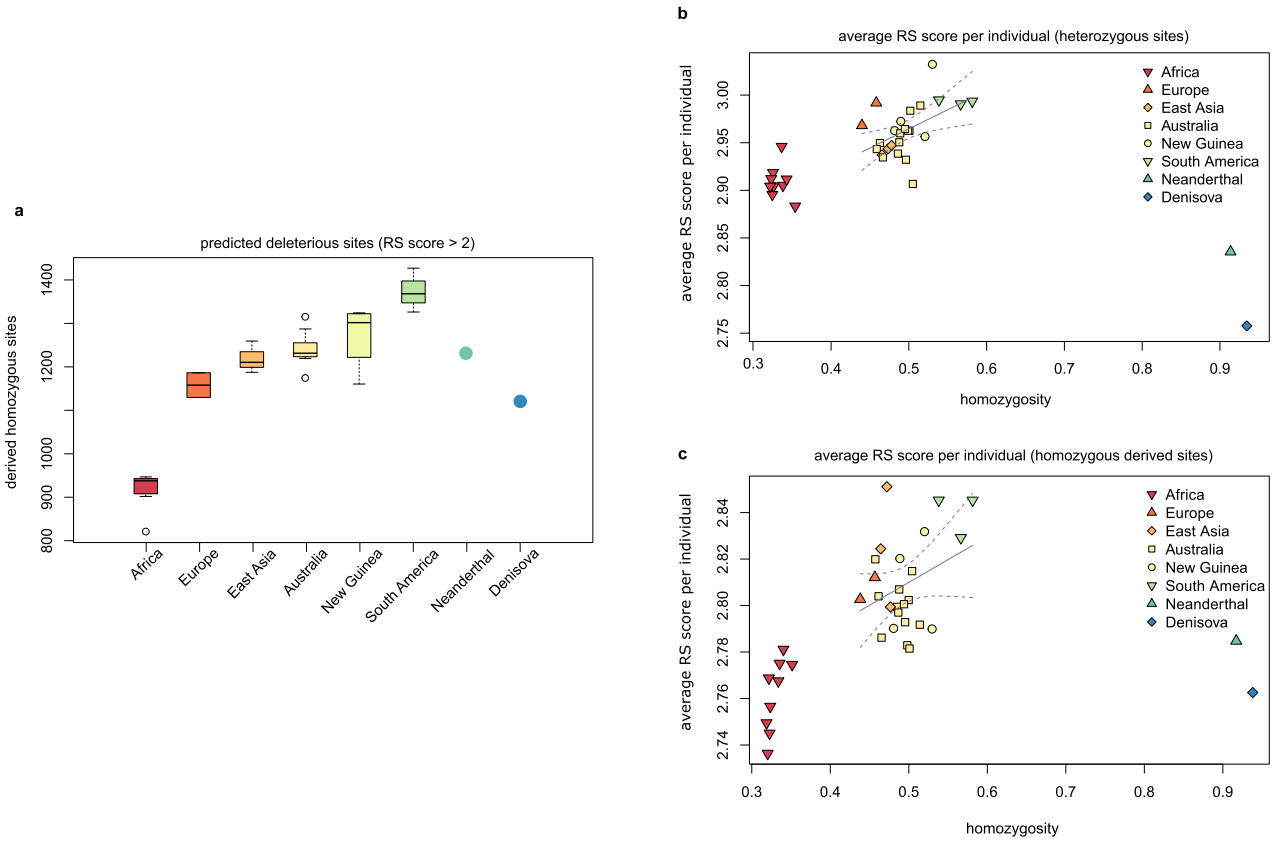
Extended Data Figure 2 | Genetic relationships of Aboriginal Australians and Papuans. a, Genetic affinities between a western central desert (WCD02) genome and Aboriginal Australians and Papuans. Outgroup f_3 statistics between WCD02 and all other Aboriginal Australians and Highland Papuan individuals that were whole-genome sequenced for this study, using the genotypes called from the sequencing data. Because the widespread recent admixture in Aboriginal Australians has large confounding effects on the f_3 statistics, the values were adjusted using the slope coefficient from a simple linear regression model fitted to the relationship between f_3 and the fraction of non-indigenous (that is, neither Aboriginal Australian nor Papuan) ancestry in each individual genome. The adjusted f_3 statistics display a genetic gradient that separates western and eastern Aboriginal Australian populations. However, we find no differences between Papuan population samples in their level of Aboriginal Australian affinity (Kruskal–Wallis test, $P = 0.083$). Horizontal lines correspond to ±1 standard error. **b**, Genetic affinities between a Papuan highlander genome and Aboriginal Australians and Papuans. The Papuan highlander sample MAR01 from the Marawaka area was arbitrarily chosen as a reference point for this analysis. f_3 values were adjusted for recent admixture as in **a**. All Aboriginal Australian groups display a similar level of Highland Papuan affinity (with the exception of three outlier individuals from the north-eastern WPA and CAI populations: WPA06, WPA05 and CAI10, the latter two of which are known to have at least one parent with origins in Papua New Guinea or the Torres Strait Islands). While some differences between groups are actually statistically significant (Kruskal–Wallis test, $P = 0.0002$, after removing the three outliers), which could be consistent with, for example, low levels of Papuan gene flow into some Aboriginal Australian groups (see Supplementary Information sections S06 and S07), we caution that some of these differences are probably due to imperfect adjustment for Eurasian admixture (the adjusted f_3 is highest in the WCD population, which has the least Eurasian admixture). Horizontal lines correspond to ±1 standard error. **c**, MSMC analyses. Linear interpolation through the midpoints of the time intervals of the relative cross coalescence rate estimates from MSMC[25] using pairs of individuals including one HGDP-Papuan and one other individual as indicated. We used CAI01, PIL06, WCD01, WON03 and an ECCAC sample for this analysis (see Supplementary Information section S08 for details). The MSMC results were scaled using a mutation rate of 1.25×10^{-8} per generation per site as suggested in ref. 41 and a generation time of 29 years, corresponding to the average hunter–gatherer generation interval for males and females[42].



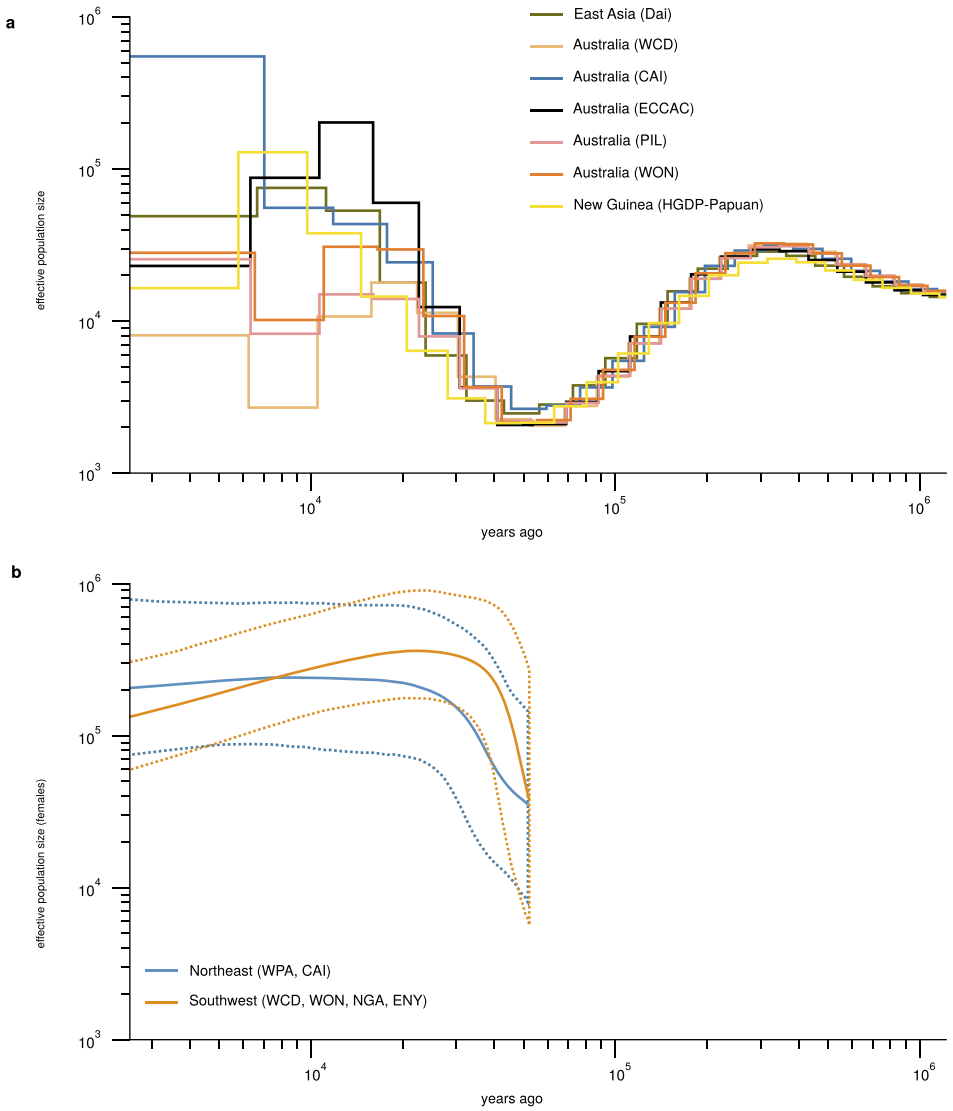
Extended Data Figure 3 | Introgressed archaic sites and putative Denisovan and Neanderthal haplotypes. a, Distribution of number of putative introgressed sites per individual from archaic humans. The number of Neanderthal-specific introgressed sites per individual increases from Europe to Australia, and then decreases in Amerindians, which is consistent with recurrent Neanderthal (or Neanderthal-related archaic) gene flow during the expansion into Eurasia. Our results are thus indicative of several pulses of Neanderthal gene flow into modern humans, as inferred previously[48–50]. We note, however, that the apparent high levels of Neanderthal-specific introgressed sites in Australo-Papuans can be explained by the expected number of misclassified Neanderthal introgressed sites resulting from the shared ancestry with Denisovans (see Supplementary Information section S11 for details). **b**–**e**, Putative Denisovan (PDH) and Neanderthal haplotypes (PNH). The putative haplotypes correspond to clusters (four or more SNPs spanning at least 4 kb) of heterozygous or homozygous genotypes in complete linkage disequilibrium ('diplotypes') that are potentially the result of Neanderthal or Denisovan admixture. Those diplotypes are homozygous ancestral in 10 Africans, homozygous derived in the Denisovan for the PDH (respectively Neanderthal for the PNH), homozygous ancestral in the Neanderthal for the PDH (respectively Denisovan for the PNH), and with the derived allele segregating in all other contemporary non-African humans (see Supplementary Information section S10 for details). We report the average number of PDHs and PNHs (**b**), the correlation between the estimated amount of Australo-Papuan ancestry (see Fig. 2a, Extended Data Fig. 1, Supplementary Information section S05) and the number of identified PDHs for each Australian sample (**c**), the sum of the lengths (**d**) and the average length (**e**) of the PDHs and PNHs per individual for worldwide populations included in our reference panel (see Supplementary Information section S04).



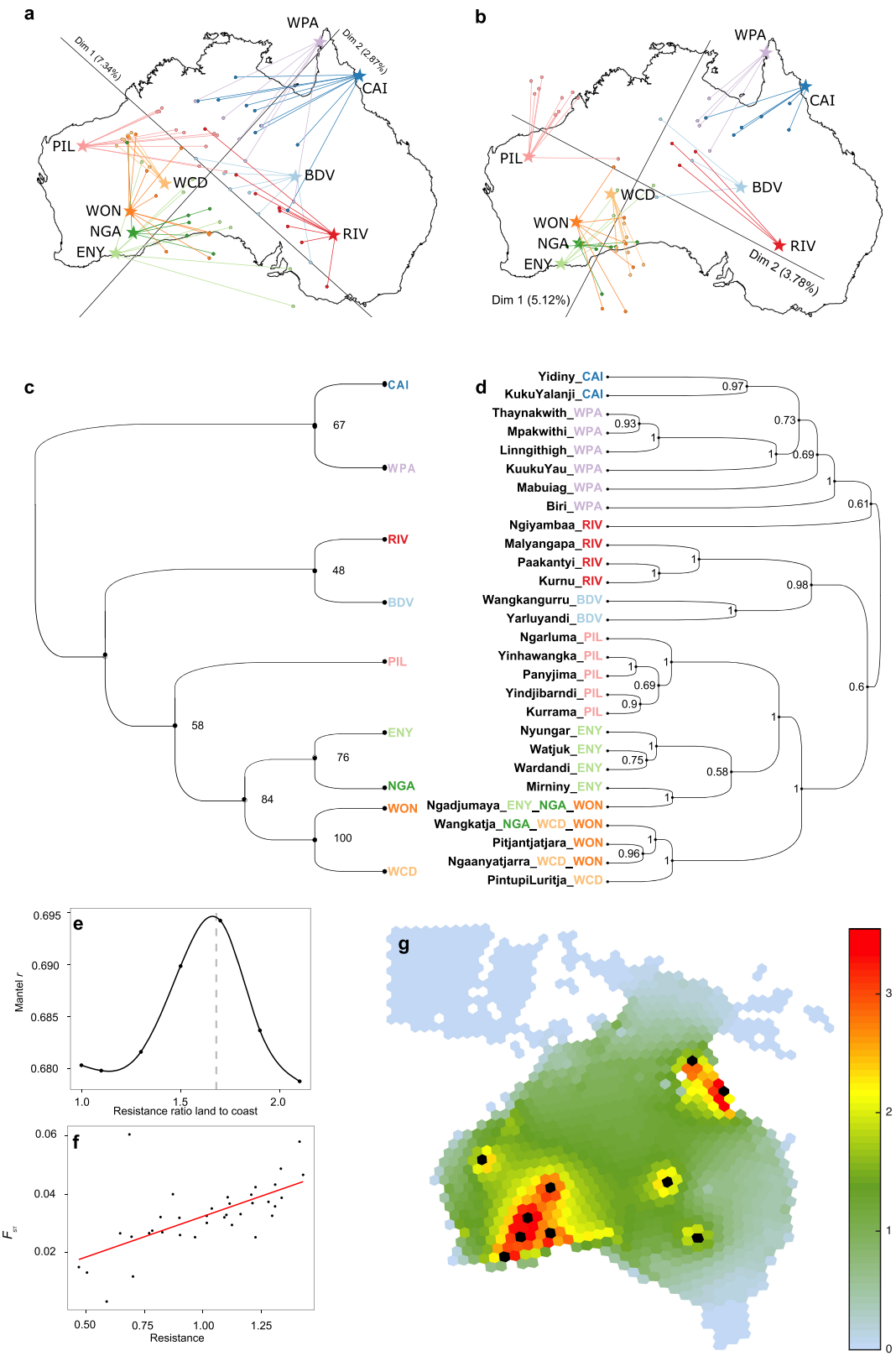
Extended Data Figure 4 | Out of Africa: admixture graphs based on D-statistics and MSMC analyses. a, Admixture graphs representing some of the topologies considered for the two waves and one wave Out of Africa models assuming Denisovan admixture. All topologies are identical except for the coloured lineages representing Australo-Papuans (green), Neanderthal (Nea, orange) and Denisovan (Den, blue). The graphs differ in (1) the number of OoA events, and (2) the number of Neanderthal admixture pulses. Png, HGDP-Papuan. **b**, Sum of squared errors between the observed D-statistics and the expectations for each quartet in the graph involving the chimpanzee as an outgroup for each of the admixture graphs shown in **a** and the corresponding four without Denisovan admixture. Each point is the result of the optimization procedure with a different starting point. See Supplementary Information section S09 for details. **c**, Relative cross coalescence rate (CCR) estimates from MSMC[25] for pairs of individuals including one African sample (Yoruba, Dinka and San) and one other, as indicated in the legend. **d**, Simulation study to assess the effect of archaic admixture on the CCR rates. Relative CCR estimated for data simulated under a simple two-population divergence model where one of the populations admixed at different rates with an archaic population. See Supplementary Information section S08 for details.



Extended Data Figure 5 | Inferred deleterious mutations. a, Box plot of the number of derived homozygous sites per individual for worldwide populations that are predicted to be deleterious. Deleteriousness of SNPs was inferred using genomic evolutionary rate profiling (GERP) rejected substitution scores. Derived alleles with a rejected substitution score larger than 2 were considered to be deleterious, see Supplementary Information section S11. **b**, **c**, Average rejected substitution score per individual calculated across heterozygous sites (**b**), and derived homozygous sites (**c**). Each coloured symbol corresponds to estimates from a single individual. Homozygosity is calculated as the number of derived homozygous sites divided by the number of sites at which an individual carries at least one copy of the derived allele. Solid lines show the linear regression of homozygosity against average rejected substitution score per individual for non-African modern humans. Dashed lines indicate the 95% confidence interval for the linear regression. See Supplementary Information S11 for details.

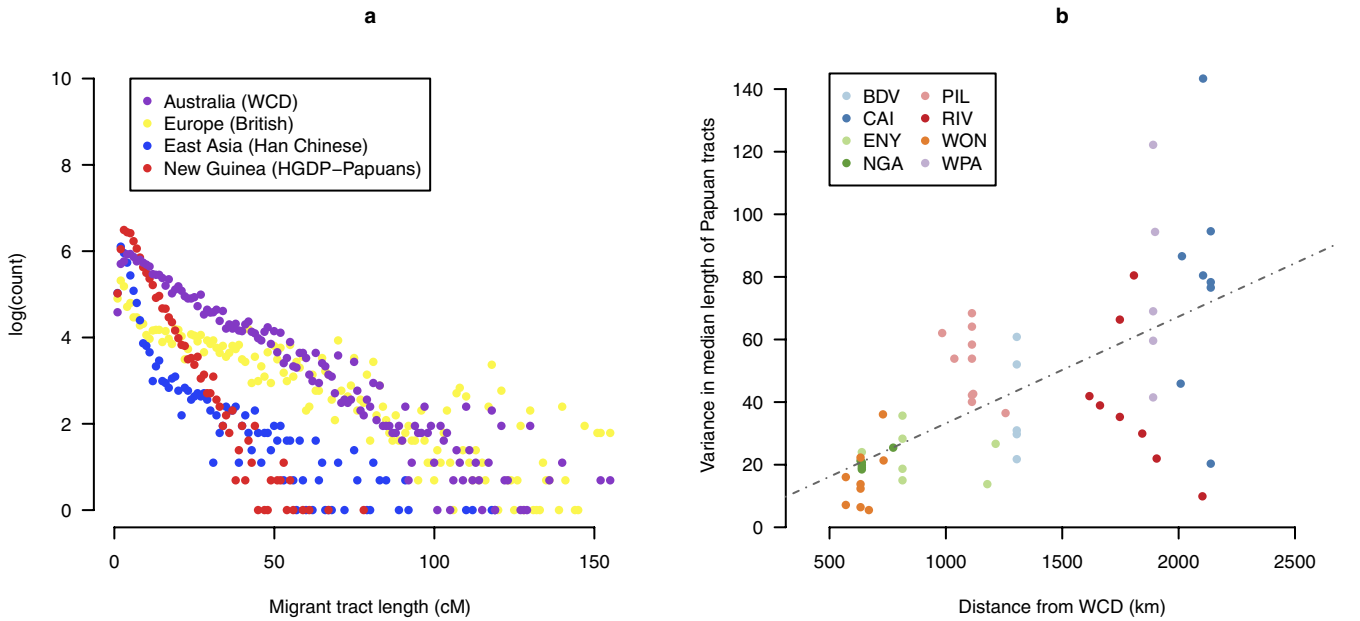


Extended Data Figure 6 | Effective population size changes over time. **a**, Population size estimates from MSMC for pairs of individuals from several populations within and outside of Australia. For each run, we used two individuals from each population, that is, four haplotypes in each run. MSMC results were scaled as in Fig. 3. **b**, Bayesian skyline plots (BSP) calculated from the mtDNA genome sequences, showing the effective population size estimates over time when considering either groups from northeastern Australia (CAI, WPA) or groups from southwestern Australia (ENY, NGA, WCD, WON). Solid lines are the estimates, dashed lines are the corresponding 95% credible intervals (see Supplementary Information section S12).



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Genetics mirrors geography and languages. **a**, **b**, Procrustes analyses of the first two dimensions of a classical multidimensional scaling (MDS) analysis of the Aboriginal Australian genome sequences (autosomes). We considered two cases: an analysis including all variants (**a**), or only the variants remaining after genomic regions of putative recent European and East Asian origin are 'masked' (**b**, Supplementary Information section S06). Both MDS plots have been rotated towards the best overlap with geographic sampling locations as defined by Procrustes analysis[51]. In each plot, the connecting lines indicate the error of the MDS coordinates towards the assigned population-sampling geographic coordinates. We find that the genetic relationships within Australia mirrors geography, with a significant correlation for both cases, that is, $r_{GEN,GEO} = 0.59$, $P < 0.0005$ for all variants and even higher, $r_{GEN,GEO} = 0.77$, $P < 0.0005$, for the masked data. We find using the bearing correlogram approach that the main axis of genetic differentiation in the masked Aboriginal Australian genomes is at an angle of 65° compared to the equator, that is, in the southwest to northeast direction (Supplementary Information section S13). **c**, **d**, Correspondence between genetics and linguistics. **c**, Unrooted neighbour-joining F_{ST}-based genetic tree (cladogram). Weir and Cockerham F_{ST} distance was computed between the Aboriginal Australian populations after masking the Eurasian tracts. Statistical robustness of each branch was estimated by means of a bootstrap analysis (1,000 replicates, Supplementary Information section S05). **d**, Bayesian phylogenetic tree for the 28 different Pama–Nyungan languages represented in this sample (from ref. 13, see Supplementary Information section S15). Posterior probabilities are also indicated. Note that one language group can be shared by different Aboriginal Australian groups. The linguistic tree was built with BEAST[52]. **e**–**g**, Gene flow across the continent. **e**, Mantel non-parametric *r* (estimating the goodness of fit between genetic differentiation and connectivity) versus ratios of resistance of inland to coastal nodes, showing a peak at 1.7. **f**, Best fit of pairwise population genetic differentiation, F_{ST} (computed between the nine Aboriginal Australian groups after masking Eurasian tracts (Supplementary Information section S06)), versus pairwise connectivity based on the environment (estimated as resistance) when moving inland is 1.7 times harder than moving along coastal nodes. **g**, Gene flow across the Australian landscape, quantified as the cumulative current for pairwise connections among Aboriginal Australian groups (black circles), with larger current (warmer colours) representing greater gene flow.



Extended Data Figure 8 | European, East Asian and Papuan genomic tracts in Aboriginal Australians. a, Distribution of the tracts assigned to Aboriginal Australian (WCD), Papuan, East Asian or European ancestry for 58 unrelated non-WCD Aboriginal Australian samples. Most of the shorter tracts were of Papuan origin, suggesting that a large fraction of the Papuan gene flow is much older than that from Europe and East Asia, consistent with a Papuan influence spreading slowly from northeastern to southwestern Australia by ancient migration. **b**, Corresponding scatter plot with fitted line of per-individual variance in Papuan tract length versus geographic distance from WCD, the latter calculated using the great-circle distance formula for pairs of individual GPS coordinates. Papuan tract distribution showed a strong and significant correlation with distance from WCD ($r = 0.64$; $P < 1 \times 10^{-5}$), with 'younger tracts' (that is, with a larger variance) closer to New Guinea and 'older tracts' (that is, with a smaller variance) closer to WCD. This is also consistent with continuous Papuan gene flow spreading from the northeast.

Extended Data Table 1 | Whole genome sequence depth of coverage, haplogroup and language assignments for the Aboriginal Australian samples

indiv.	DoC*	mtDNA haplotype†	Ychr haplotype‡	Pama-Nyungan language§	indiv.	DoC*	mtDNA haplotype†	Ychr haplotype‡	Pama-Nyungan language§
BDV01	78	S2	-	Yarluyandi Wangkangurru	PIL09	58	S5	R1b1a2a1a2e1	Kurrama
BDV02	75	S1a	R1b1a2a1a2c1g2a1a2	Yarluyandi Wangkangurru	PIL10	61	R	-	Yinhawangka
BDV03	-	-	-	Yarluyandi Wangkangurru	PIL11	57	P3b	C1b	Kurrama
BDV04	70	O1a	-	Yarluyandi Wangkangurru	PIL12	63	P3b	C1b	Yindjibarndi
BDV05	72	S1a	O1a	Yarluyandi Wangkangurru	RIV01	73	M42a	-	Ngiyambaa
BDV06	70	S1a	-	Yarluyandi Wangkangurru	RIV02	62	P4b1	-	Paakantyi
BDV07	70	O1a	-	Yarluyandi Wangkangurru	RIV03	69	M42a	-	Paakantyi
BDV08	70	S1a	R1b1a2a1a2c1g2a1	Yarluyandi Wangkangurru	RIV04	62	P4b1	I2a1a2a1a	Kurnu
BDV09	74	S1a	-	Yarluyandi Wangkangurru	RIV05	72	P4b1	-	Paakantyi
BDV10	72	S1a	I1a2a1d	Yarluyandi Wangkangurru	RIV06	66	H1bs	J2a1b	Ngiyambaa
CAI01	84	P	K2b	Yidiny	RIV07	70	P4b1	R1b1a2a1a2c1c	Paakantyi
CAI02	74	M42	K2b	Yidiny	RIV08	66	P4b1	-	Paakantyi Malyangapa
CAI03	77	M42a	-	Yidiny	WCD01	62	R12	K2b	Ngaanyatjarra
CAI04	71	P	-	Yidiny KukuYalanji	WCD02	59	S1a	C1b	Ngaanyatjarra
CAI05	80	P	O2a1a	Yidiny	WCD03	61	R12	K2b	Wangkatja
CAI06	78	P	C1b	Yidiny	WCD04	52	P3b	K2b	Ngaanyatjarra
CAI07	71	N13	K2b	KukuYalanji	WCD05	60	O1	C1b	Ngaanyatjarra
CAI08	70	P	K2b	Yidiny	WCD06	58	O1a	C1b	Ngaanyatjarra
CAI09	79	P	R1b1a2a1a2b1	Yidiny	WCD07	61	M42	-	Ngaanyatjarra
CAI10	73	E1a2	K2b	-	WCD08	64	M42	-	Ngaanyatjarra
ENY01	69	H1e1a3	R1b1a2a1a2b1c1	Nyungar	WCD09	59	R	J2a1b	Ngaanyatjarra
ENY02	79	R12	-	Ngadjumaya	WCD10	63	M42	-	Ngaanyatjarra
ENY03	83	O	-	Mirniny	WCD11	57	M42	K2b	Ngaanyatjarra
ENY04	83	M42	-	Nyungar	WCD12	59	M42	C1b	Ngaanyatjarra PintupiLuritja
ENY05	78	S2	-	Ngadjumaya	WCD13	67	M14	C1b	Ngaanyatjarra
ENY06	70	M42	-	Wardandi	WON01	71	O	I1a2a1a3a	Wangkatja
ENY07	73	S2	E1b1b1b2a	Watjuk	WON02	101	O1a	-	Wangkatja
ENY08	71	P4b1	C1b	Nyungar Ngadjumaya	WON03	65	O1a	-	Wangkatja
NGA01	74	O1	-	Ngadjumaya	WON04	58	R	-	Ngaanyatjarra
NGA02	52	O1a	-	Ngadjumaya	WON05	56	O1a	I2a2a1a2a2	Wangkatja
NGA03	73	O	-	Ngadjumaya	WON06	60	R12	-	Wangkatja
NGA04	75	O	R1b1a2a1a1b1a1a	Wangkatja	WON07	57	O	-	Ngadjumaya
NGA05	56	R12	-	Ngadjumaya	WON08	52	O	-	Wangkatja
NGA06	63	S1a	-	Wangkatja	WON09	20	O	E1b1b1a1b1a4	Wangkatja
PIL01	58	R	C1b	Yinhawangka	WON10	50	O1	R1b1a2a1a2a	Wangkatja
PIL02	61	M42	C1b	Yinhawangka	WON11	58	R12	-	Pitjantjatjara
PIL03	56	M42	-	Yinhawangka	WPA01	51	P5	-	Thaynakwith Linngithigh
PIL04	64	M42	-	Yinhawangka	WPA02	50	P	C1b	Mpakwithi Kaanju
PIL05	68	M42	C1b	Yinhawangka	WPA03	51	M42a	K2b	Thaynakwith Biri
PIL06	59	O1	K2b	Panyjima	WPA04	52	P5	-	Thaynakwith KukuYau
PIL07	63	O	-	Panyjima	WPA05	56	M42	NA	Mabuiag Thaynakwith
PIL08	72	M42	C1b	Yindjibarndi Kurrama	WPA06	53	P5	O1a	Mpakwithi

*The depth of coverage (DoC) is the average number of reads covering every position in the genome (hg19) after duplicate removal (see Supplementary Information section S03).
†The average depth of coverage on the mitochondrial genome (mtDNA) is 3,484 ± 1,515 (mean ± s.d.) and haplogroups were called with haplogrep (http://haplogrep.uibk.ac.at/) and haplofind (https://haplofind.unibo.it/), see Supplementary Information section S12 for details and references.
‡The average depth of coverage on the Y chromosome (Ychr) is 28.9 ± 4.5 (mean ± s.d.). Haplogroup assignment was performed with an in-house script that matched our SNPs with the classification provided in ISOGG version 10.08, see Supplementary Information section S12 for details and references.
§Language group with which the speaker self-identifies, or to which they were assigned. Where more than one language is given, speakers either identified with more than one group, or they could not be assigned to a single group with certainty.

Extended Data Table 2 | Selection scan in Aboriginal Australians

Focal Pop	Nearby Gene*	Position†	rsID	Dist ‡	*PBSn1*§	F_{12} ¶	F_{13}	F_{23}	Function of gene product#
All	*TMEM86B*	55,833,076	rs734517	92,444	0.78	0.93	0.99	0.06	Catalyzes the degradation of lysoplasmalogen. Modulates cell membrane proteins.
All	*LRRC52*	165,621,695	rs4147601	88,510	0.74	0.96	0.91	0.01	Modulates voltage of potassium ion channels. Expressed in testis.
All	*MACROD2*	15,209,684	rs175279	901	0.70	0.92	0.89	-0.01	Involved in deacetylase activity. Possibly (but not conclusively) causative of Kabuki syndrome.
All	*JRKL*	96,747,146	rs72959058	507,105	0.74	0.99	0.87	0.15	Homologue to "jerky" gene in mouse.
All	*SPATA20*	48,631,324	rs73338243	287	0.70	0.96	0.85	0.09	Spermatid protein.
All	*NAA60*	3,537,933	rs73503305	970	0.71	0.91	0.91	-0.02	Histone acetyltransferase required for nucleosome assembly and chromosome segregation during anaphase. Human-specific imprinted gene.
All	*CBLN2*	70,019,066	rs12455116	184,848	0.69	0.92	0.87	0.00	*CBLN2*: cerebellum-specific protein involved in various signaling pathways. Possibly associated with pulmonary arterial hypertension.
	NETO1			390,482					*NETO1*: brain-specific transmembrane protein involved in the regulation of neuronal circuitry. Associated with thyroid function.
All	*SLC2A12*	134,391,056	rs4896021	17,267	0.76	0.96	0.95	-0.01	Catalyzes sugar absorption. Involved in the pathogenesis of diabetes. Associated with serum urate levels.
All	*LOC101927657*	127,358,509	rs145200081	16,731	0.65	0.94	0.80	0.13	Unknown (ncRNA).
All	*LOC102724612*	64,466,486	rs113341339	78,446	0.73	0.91	0.95	0.00	Unknown (ncRNA).
NE	*ZBTB20*	114,530,679	rs9289004	10,658	0.55	0.65	0.82	0.07	Transcriptional repressor associated with Primrose syndrome.
NE	*ANXA10*	168,646,016	rs2176513	367,671	0.49	0.61	0.61	-0.01	Calcium-dependent phospholipid-binding annexin.
NE	*TRPC3*	122,905,041	rs4502701	32,132	0.50	0.59	0.64	-0.01	Non-selective cation channel, associated with spinocerebellar ataxia.
NE	*HS3ST1*	11,634,592	rs7665516	204,055	0.45	0.45	0.71	0.07	Regulates rate of generation of anticoagulant heparan sulfate proteoglycan.
NE	*MIR548C*	65,027,511	rs2620721	11,126	0.50	0.55	0.73	0.03	Unknown (microRNA).
NE	*STARD13*	33,799,901	rs7318080	19,714	0.49	0.54	0.83	0.20	Involved in cell proliferation and fibroblast morphology.
NE	*AKAP11*	42,931,386	rs7319267	33,983	0.53	0.56	0.85	0.13	Directs protein kinase A activity and is involved in cAMP messenger signaling.
NE	*AGMO*	15,212,231	rs35557899	27,711	0.47	0.51	0.68	0.01	Catalyzes the cleavage of O-alkyl bonds of ether lipids.
NE	*RUNX1T1*	92,925,296	rs11776341	41,898	0.45	0.56	0.54	0.00	Involved in transcriptional repression. A translocation involving this gene is associated with acute myeloid leukemia.
NE	*FHAD1*	15,680,451	rs2473358	971	0.45	0.60	0.52	0.00	Unknown.
SW	*KCNJ2*	68,190,552	rs35167900	14,369	0.57	0.61	0.93	0.22	Potassium channel, associated with familial atrial fibrillation and periodic paralysis.
SW	*TACC2*	123,754,065	rs10159998	5,062	0.50	0.60	0.67	0.00	Belongs to a family of proteins that interact with the centrosome and microtubules, and that are implicated in cancer.
SW	*LOC101928708*	87,228,164	rs4843556	17,556	0.58	0.65	0.86	0.07	Unknown (ncRNA).
SW	*C16orf82*	27,187,689	rs72782349	107,202	0.51	0.60	0.69	0.02	Unknown.
SW	*LOC100507391*	194,520,805	rs56379930	17,908	0.55	0.66	0.75	-0.01	Unknown (ncRNA).
SW	*HAUS4*	23,416,252	rs2008951	127	0.49	0.50	0.83	0.16	A component of a microtubule-binding complex that plays a role in the generation of microtubules in the mitotic spindle.
SW	*KNG1*	186,438,819	rs5029990	815	0.51	0.56	0.72	0.01	During the inflammatory response, it is involved in vasodilation, coagulation, enhanced capillary permeability and pain induction.
SW	*MYDGF*	4,657,016	rs66891175	540	0.55	0.61	0.88	0.16	Unknown.
SW	*MSMP*	35,757,075	rs1951432	2,801	0.48	0.47	0.88	0.27	May be involved in the tumorigenesis of prostate cancer.
SW	*VAV2*	136,756,316	rs2519771	29,762	0.47	0.51	0.73	0.07	Member of an oncogene family. Involved in T-cell receptor signaling.

Top 10 peaks of differentiation from genome scans of all Aboriginal Australians combined (All) and two Aboriginal Australians subgroups living in different ecological regions in Australia, the northeast (NE) or southwest (SW).
*RefSeq protein coding gene with exon boundary near to windowed-*PBSn1* peak.
†Genomic position (hg19) of SNP with highest value of *PBSn1* within 200 Mb of the top window.
‡Distance between SNP and the nearest exon boundary of nearest gene.
§*PBSn1* statistic at top SNP.
¶F_{ST} statistics at top SNP for each comparison within the *PBSn1* calculation.
#Please see Supplementary Information section S16 for references.