The genomic history of Australia

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

^ first joint authors

# these authors contributed equally and are listed alphabetically

§ these authors contributed equally and are listed alphabetically

& these authors contributed equally and are listed alphabetically

* to whom correspondence should be addressed:

Eske Willerslev: ewillerslev@snm.ku.dk
David M Lambert: d.lambert@griffith.edu.au
Laurent Excoffier: laurent.excoffier@iee.unibe.ch
Manjinder S Sandhu: ms23@sanger.ac.uk

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, CH-3012 Bern, Switzerland.
3. Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
4. Environmental Futures Research Institute, Griffith University, Nathan, Australia.
5. CNAG-CRG - Centro Nacional de Análisis Genómico, Parc Científic de Barcelona, Barcelona, 08028, Spain.
6. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

7. Bioinformatics Research Centre, Aarhus University, Denmark.
8. Department of Integrative Biology, University of California, Berkeley, CA, USA.
9. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103, Leipzig, Germany.
10. Interfaculty Bioinformatics Unit University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland.
11. Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Building 208, 2800 Kongens Lyngby, Denmark.
12. Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena.
13. Centre for Bioinformatics, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200, Copenhagen, Denmark.
14. Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Kahlaische Strasse 10, D-07745 Jena, Germany.
15. Dept of Zoology, University of Cambridge, Downing Street, CB2 3EJ, UK.
16. Integrative Systems Biology Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Kingdom of Saudi Arabia.
17. Institute for theoretical physics, ETH Zürich, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland.
18. Jeffrey Cheah School of Medicine& Health Sciences, Monash University, Jalan Lagoon Selatan, Sunway City, 46150 Selangor, Malaysia.
19. Evolutionary Medicine group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse, UMR 5288, Centre National de la Recherche Scientifique, Université de Toulouse 3, Toulouse, France.
20. Novo Nordisk Forundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen N, Denmark.

21. CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal.

22. Explico Foundation, Håvågen 10, 6900 Florø, Norway.

23. UCL Research Department of Genetics, Evolution and Environment, Darwin building, Gower Street, London WC1E 6BT, UK.

24. Papua New Guinea Institute of Medical Research, P.O. Box 60, Goroka, Papua New Guinea.

25. Archaeology, School of Humanities & Social Sciences, University PO Box 320, University of Papua New Guinea & College of Arts, Society & Education, James Cook University, Cairns, Australia.

26. Department of Anatomy, University of Otago, Dunedin, New Zealand.

27. Welcome Trust Centre for Human Genetics, Oxford University, Oxford, OX3 7BN, UK.

28. Estonian Biocentre, Tartu, Estonia.

29. UCL Department of Anthropology, 14 Taviton Street, London WC1H 0BW, UK.

30. Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT.

31. School of Anthropology and Museum Ethnography, Oxford University, Oxford, OX2 6PE, UK.

32. Eureka Archaeological Research and Consulting, University of Western Australia.

33. LCHES, Department of Archaeology and Anthropology, University of Cambridge, Fitzwilliam St, Cambridge CB2 1QH, UK.

34. Department of Linguistics, PO Box 208366 (370 Temple St), New Haven, CT, 06520, USA.

35. Institute for Human Genetics, University of California, San Francisco, CA, USA.

36. Departments of Integrative Biology and Statistics, University of California, Berkeley, CA, USA.

The genomic history of Australia

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

During most of the last 100 ky, 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 (the Wallacean region as defined by biogeographers). Colonisation of Sahul is thought to have required at least 8-10 separate sea crossings between islands[1], potentially constraining the occupation of Australia and New Guinea by earlier hominins[2]. The age of the first occupation of Australia has been disputed, with several archaeological sites dating to 40-45 kya (Figure 1); long argued to represent the age of first occupation[3]. However, recent studies support dates ≥ 50 kya, suggesting that Sahul was first settled by 47.5-55 kya[4–6]. These dates overlap with those for the earliest evidence for modern humans in Sunda[7] (Figure 1). Moreover skeletal remains that share morphological similarities with the ancestors of Aboriginal Australians and Papuans are found in Southeast Asia up until about 3.5 kya[8], suggesting that the ancestors of Aboriginal Australians and Papuans extended from Sahul to Sunda.

Historically, the morphological diversity among Aboriginal Australians was interpreted by some as indicating multiple ancestral migrations[9–11], or descent from Javanese *Homo erectus*, with varying

levels of gene flow from contemporaneous populations[12]. However, statistical analyses indicate that Australian crania show no evidence of *H. erectus* admixture[13]. Irrespective of this, the distinctiveness of the Australian archaeological record has led to the suggestion that the ancestors of Aboriginal Australians and Papuans (hereafter referred to as Australo-Papuans), as well as a small number of other populations, left the African continent earlier than the ancestors of present-day Eurasians[14]. Although some genetic studies support such multiple dispersals from Africa[15,16], others favour only one out of Africa (OoA) event, with one[17] or two[18] independent founding waves into Asia, of which the earlier contributed to Australo-Papuan ancestry[19,20]. Recent genomic results have also shown that both Aboriginal Australian[20] and Papuan[21] ancestors likely admixed with Neanderthal and Denisovan archaic hominins after leaving Africa.

Once in Sahul, contact among groups would have been affected by rising sea-levels that separated the Australian continent from New Guinea and Tasmania 7-14.5 kya through the formation of the Arafura Sea and Bass Strait[22,23](Figure 1). These events still appear to be part of the oral tradition of several Aboriginal Australian communities[24]. Similarly, greater desertification of Australia[25] and more challenging temperature gradients during the last glacial maximum (LGM) 19-26.5 kya, appear to have impacted the number and density of human populations[26,27]. In the same context, unique morphological and physiological adaptations have been identified in Aboriginal Australians living in the desert areas today[28–30], including absence of the increased metabolic rates observed in Europeans when exposed to the freezing night temperatures common in the desert [31,32].

At the time of European contact, Aboriginal Australians spoke over 250 distinct languages[33], two-thirds of which belong to the Pama-Nyungan family. The place of origin of this language family, which covers 90% of the Australian mainland, has been debated[34], as has the effect of its extensive diffusion on its internal phylogenetic structure[33]. The pronounced similarity among Pama-Nyungan languages, together with shared socio-cultural patterns, have been interpreted as resulting from a recent, mid-Holocene, expansion[35]. Other changes in the mid-late Holocene (~4 kya) include the efflorescence of backed blades (microliths[36]) and the introduction of the dingo[37]. The spatial distribution of microliths roughly correlates with that of Pama-Nyungan languages. It has even been suggested that Pama-Nyungan languages, dingoes and backed blades all reflect the same recent migration into Australia[38]. Although an external origin for backed blades has been rejected[36], dingoes were certainly introduced, most likely via island south-east Asia[37]. Rock art traditions also suggest contact between Sulawesi (Indonesia) and Australia[38]. Intriguingly, a recent genetic study found evidence of Indian gene flow into Australia at the approximate time of these Holocene

changes[39]. Finally, substantial contact with Asians and Europeans is well documented in historical times[40–43], suggesting potentially complex admixture among present-day Aboriginal Australians.

After a century of research, the origins and evolutionary history of Aboriginal Australians continue to be debated. To date, only three whole genome sequences have been described - one deriving from a historical tuft of hair from Western Desert Australia[20] and two others from cell lines with limited provenance information[44]. In this study we report the first extensive investigation of Aboriginal Australian genomic diversity by reporting and 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 (S01). We sequenced high-depth genomes (average depth of 60X, range 20X-100X) from 83 Aboriginal Australian individuals representing a wide geographical distribution and a broad range of Pama-Nyungan languages (Figure 1, Extended Data Table 1, S02, S03, S04). Additionally, we sequenced 25 Highland Papuan genomes (38X-53X; S03) from five linguistic groups, and generated genotype data for 45 additional Papuans living or originating in the highlands (Figure 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[44] (ECCAC, Figure 1, S04).

We explored the extent of admixture in the Aboriginal Australian autosomal gene pool by estimating ancestry proportions with an approach based on sparse nonnegative matrix factorization (sNMF)[45]. We found that the genomic diversity of Aboriginal Australian populations is best modelled by a mixture of four main different genetic ancestries that can be assigned to four geographic regions based on their relative frequencies: Europe, East Asia, New Guinea and Australia (Figure 2, Extended Data Figure 1, S05). The degree of admixture varies among groups (S05) with the Ngaanyatjarra speakers from central Australia (WCD) having a significantly higher "Aboriginal Australian component" (median value = 0.95) in their genomes compared to the median value of other Aboriginal Australian groups (median value = 0.64; Mann-Withney rank sum test, one tail p-value = 3.55e-07). The "East Asian" and "Papuan" components are mostly present in northeastern Aboriginal Australian populations (Figure 2b, Extended Data Figure 1, S05),

while the "European component" is widely distributed across groups. 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 component (S06).

Colonisation of Sahul and diversification of Australians and Papuans

The origins of Aboriginal Australians is a source of much debate, as are the nature of the relationships among Aboriginal Australians and between Aboriginal Australians and Papuans. Using f_3 statistics, estimates of genomic ancestry proportions and classical multi-dimensional 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 (Figure 2a, Figure 3a, S05). This is consistent with Aboriginal Australians and Papuans being derived from a common ancestral population, which initially colonised Sahul. Moreover, comparing outgroup f_3 statistics we do not find significant differences between Papuan populations (highland Papuan groups and HGDP-Papuans) in their genetic affinities to Aboriginal Australians (Figure 3b), suggesting that the Papuan groups share common origins after or at the same time as the divergence between Aboriginal Australians and Papuans.

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[46,47](Figure 4a, S07). We compared the HGDP-Papuans to four Aboriginal Australian population samples with low levels of European admixture (Extended Data Figure 1) from both northeastern (CAI and WPA) and southwestern (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 resulted in a better fit to the observed SFS, suggesting that the ancestors of the sampled Aboriginal Australians diverged from a single ancestral population. This scenario is also supported by MDS analyses, even when masking Eurasian tracts, as well as by estimation of ancestry proportion analyses where all Aboriginal Australians form a cluster distinct from the Papuan populations (Figure 2, S05). Additionally, it is supported by f_3 analyses where all Aboriginal Australians share drift to the exclusion of, and are largely equidistant from Papuans, when adjusting for recent admixture (Figure 3c). 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[9–11].

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

Archaic admixture

We characterised the number, timing and intensity of archaic gene flow events using three complementary approaches: SFS-based (Figure 4a, Figure 5c, S07), a goodness-of-fit analysis combining D-statistics (S09), and a method that directly infers putatively derived archaic 'haplotypes' (S11). Aboriginal Australians and Papuan genomes show an excess of putative Denisovan-derived variants (Extended Data Figure 2d, S10), as well as substantially more putative Denisovan-derived haplotypes (PDH) than other non-Africans (Extended Data Figure 3). The number and total length of those putative haplotypes varied considerably across samples. However, the estimated number of PDH correlates almost perfectly ($r^2 = 0.96$) with the estimated proportion of Australo-Papuan ancestry in each individual (Extended Data Figure 3). We also estimated that the values of F_{ST} between autosomal SNPs or PDHs assigned to WCD and Papuans were both around 0.12. Moreover, we found no significant difference in the distribution of the number of PDHs or the average length of PDHs between putatively unadmixed Australians and Papuans (Mann-Whitney U test, $p>0.05$). Taken together, these observations provide strong evidence for a

single Denisovan admixture event that predates the population split between Australians and Papuans (see also[52]) and widespread recent Eurasian admixture in Aboriginal Australians (Figure 2, S05). Furthermore, using the SFS-based approach and constraining Denisovan admixture to have occurred before the Aboriginal Australian-Papuan divergence results in an admixture estimate of ~4% (95% CI 3-5%, Figure 5c, S07), similar to the estimates using D-statistics (~5%, S09). The SFS analyses further suggest that Denisovan/Australo-Papuan admixture took place ~44 kya (95% CI 31-50 kya, S07). We note that the point estimate for the age of the bottleneck overlaps with the confidence interval for the age of admixture, and that a bottleneck could have occurred anywhere along the dispersal route of Australo-Papuan populations from the ancestral source.

The SFS analysis also provides evidence for a primary Neanderthal admixture event (~2%, 95% CI 1-3%, Figure 5c, S07) taking place in the ancestral population of all non-Africans ~60 kya (95% CI 55-84 kya, Figure 5c, S07). Note that, 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 20 putatively unadmixed Australians and 12 putatively unadmixed HGDP-Papuans, the average PNH and PDH length are 33.8 Kb and 37.4 Kb, respectively. These are significantly different from each other ($p = 9.65 * 10^{-6}$ using a conservative sign test), and suggest that the time since Neanderthal admixture was roughly 11% greater than the time since Denisovan admixture, roughly in line with our SFS based estimates for Denisovan pulse (31-50 kya) 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% (95% CI 0.2-2.7%, Figure 5c, S07) pulse of Neanderthal gene flow into the ancestors of Eurasians, and a smaller pulse into the ancestors of Asians (0.2%, 95% CI 0.1-1.0%, Figure 5c, S07). However, there is little evidence for Neanderthal introgression private to Australo-Papuans, potentially limited to ~0.2% (95% CI 0.05-1.3%, Figure 5c, S07). In addition, the fact that the number of Neanderthal-specific introgressed sites increases from Europe to Australia (Extended Data Figure 2d, S10), and then decreases in Amerindians is consistent with recurrent Neanderthal (or Neanderthal-related archaic) gene flow during expansion into Eurasia. Our results are thus indicative of several pulses of Neanderthal gene flow into modern humans, as inferred previously[53–55]. We note however, the apparent high levels in 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 (S10). 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 a model involving discrete Denisovan and Neanderthal admixture pulses[20,21] (S07, S11), supporting the view that the archaic contribution in Australo-Papuans is likely more complex than was previously inferred[20,21] (S07).

Out of Africa

To investigate the relationship of Australo-Papuan ancestors to other world populations, we computed D-statistics[56,57] 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 (S09), suggesting that Africans and Ust'-Ishim - a ~45 kya modern human from Asia[58] - are both closer to 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 (S09). Similarly, the D-statistics for ((H1=Aboriginal Australian, H2=Eurasian), H3=African) becomes 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 (Figure 5a, S09).

To further investigate the timing and number of OoA events giving rise to present-day Australo-Papuan and Eurasians (Sardinians and Han Chinese) we used the observed SFS in a model based composite likelihood framework. When considering only modern human genomes, we find evidence for two waves OoA, with a dispersal of Australo-Papuans ~14 ky before Eurasians (S07). However, when explicitly taking into account archaic Neanderthal and Denisovan introgression into modern humans[44,59], the SFS analysis supports a single origin for the OoA populations marked by a bottleneck ~72 kya (95% CI 60-104 kya, S07). This scenario is reinforced by the observation that the ancestors of Australo-Papuan and Eurasians share a Neanderthal admixture event (95% CI 1.1-3.5%). Our analyses thus indicate that this single OoA ancestral population underwent two expansions at approximately the same time: one involving the ancestors of Australo-Papuan (51-72 kya) and the other, possibly slightly more recent, involving the ancestors of Eurasians (48-68 kya) (Figure 5c). Furthermore, modern humans have both an LD decay rate and a number of predicted

deleterious homozygous mutations (recessive genetic load) that correlates with distance from Africa (S05, S10, and Extended Data Figure 2 a-c), again consistent with a single African origin.

The model estimated from the SFS analysis also suggests an early divergence of Australo-Papuans from the ancestors of all non-Africans, in agreement with two colonisation waves across Asia[20,21,39]. Under our best model, Australo-Papuans began to diverge from Eurasians ~58 kya (95% CI 51-72 kya, Figure5c, S07), whereas Europeans and East Asians diverged from each other ~42 kya (95% CI 29-55 kya, Figure5c, S07) in agreement with previous estimates[19,39,60,61]. 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 Figure 5c) 57-58 kya, the groups remained in close geographical proximity for some time before Australo-Papuan ancestors dispersed eastwards. Furthermore, we infer multiple gene flow events between sub-Saharan Africans and Western Eurasians after ~42 kya. This supports previous findings of extensive contact between African and non-African populations[60–62].

Our 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 a more recent common ancestors (Figure 5b, S08). We show through simulations that these differences cannot be explained by archaic admixture. Moreover, the expected difference in phasing quality is not sufficient to fully explain this pattern (see S08). While a similar separation in cross coalescence rate curves is obtained when comparing Eurasians or Australo-Papuans with Dinka, we find that, when comparing the Australo-Papuans or the Eurasians with the San, the cross coalescence curves are overlapping (S08). We also find that the inferred changes in effective population size through time of Aboriginal Australians, Papuans, and East Asians is very similar until around 50 kya, including a deep bottleneck around 60 kya (Extended Data Figure 7). Taken together, these MSMC results suggest complex population structure in Africa preceding a split of a single non-African ancestral population, combined with gene flow between the ancestors of Yoruba or Dinka (but not San) and the ancestors of Eurasians, which is not shared with Australo-Papuans. These results are qualitatively in line with the SFS-based analyses (see e.g., Figure 5b). While our results do not exclude the possibility of an earlier OoA expansion, they do indicate that any such event left little or no trace in the genomes of modern Australo-Papuans.

Genetic structure of Aboriginal Australians

Uniparental haplogroup diversity in this dataset (Extended Data Table 1, S12) is consistent with previous studies of mitochondrial DNA (mtDNA) and Y chromosome variation in Australia and Oceania, including the presence of typically European, Southeast and East Asian lineages[63–68]. 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[69,70]. 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[71]. Moreover, the inferred European ancestry for the Y chromosome is much greater than that for mtDNA (31.8% vs. 2.4%), reflecting male-biased European gene flow into Aboriginal Australian groups during the colonial era.

On an autosomal level, we find genetic relationships within Australia that mirror geography, with a significant correlation ($r_{GEN,GEO}$ = 0.59, p-value < 0.0005) when comparing the first two dimensions in an MDS analysis (S14**)**. This correlation is higher when genomic regions of putative recent European and East Asian (i.e., Han Chinese) origin are "masked" ($r_{GEN,GEO}$ = 0.77, p-value < 0.0005, Extended Data Figure 5). The main axis of genetic differentiation in the masked Aboriginal Australian genomes was determined using the Bearing correlogram approach. We found that an axis of angle = 65^{o} compared to the equator (i.e., in the southwest to northeast direction) explains most of the genetic differentiation (S14).

Populations from the centre of the continent occupy positions genetically intermediate to this axis (Extended Data Figure 5). A similar result is observed with an F_{ST}-based tree for the masked data (Figure 6a**,** S05) as well as in analyses of genetic affinity based on the f_3 statistic (Figure 3b), suggesting a population division between northeastern and southwestern groups. Such structure is further supported by the SFS analyses showing that populations from southwestern desert and northeastern regions diverged as early as ~31 kya (95% CI 10-32 kya), followed by limited gene flow (estimated $2Nm<0.01$, 95% CI $2<Nm<11.25$). The analysis of the major routes of gene flow within the continent supports the idea that the Australian interior has acted as a barrier to gene flow. Indeed, using a model inspired by principles of electrical engineering where gene flow is represented as a current flowing through the Australian continent and observed F_{ST} values are a proxy for resistance, we infer that gene flow occurred preferentially along the coasts of Australia (Extended Data Figure 6**,** S14). These findings are consistent with a model of expansion followed

by population fragmentation when and the extreme aridity in the interior of Australia[25] formed barriers to population movements during the LGM[22].

We used MSMC based on autosomal data and mtDNA Bayesian Skyline Plots[72](BSP) to estimate changes in effective population sizes 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 kya (Extended Data Figure 7 , S08, S12). This is consistent with archaeological evidence for a population expansion associated with significant changes in socio-economic and subsistence strategies in the Holocene[73,74].

European admixture almost certainly had not occurred before the late 18th century, but earlier East Asian and/or Papuan gene flow into Australia could have taken place. We characterized the mode and tempo of gene flow into Aboriginal Australians using three different approaches (S06, S07, S13). We used approximate Bayesian computation (ABC) to compare the observed mean and variance among Aboriginal Australian individuals in the proportion of European, East Asian and Papuan admixture, to that computed from simulated datasets under various models of gene flow. We estimated European and East Asian admixture to have occurred on the order of ten generations ago (S13), consistent with historical and ethnographic records. Consistent with this, the local ancestry approach based on RFMix suggests that European and East Asian admixture is more recent than Papuan admixture (Extended Data Figure 4a). 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 ($2Nm$=0.4, 95% CI 0.0-20.4, Figure 4a, S07).

To further investigate Papuan gene flow, we conducted follow-up 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, Papuans and Aboriginal Australian (S06). We chose WCD as the representative of Aboriginal Australian ancestry, because it is the least admixed population among our Australian samples (Figure 2, S05). Papuan tract length distribution show a clear geographic pattern, with “younger tracts” (higher median length and variance) in individuals closer to New Guinea and “older” (lower median length and variance) in individuals closer to WCD (Extended Data Figure 4b); there is a strong correlation of Papuan tract

length variance with distance from WCD to other Aboriginal Australian groups (r=0.64, p-value<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, which is consistent with the ABC analysis (S13).We also investigated possible South Asian (Indian related) gene flow into Aboriginal Australian, as reported by a recent study[39]. 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 (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 the discrepancy in the results.

Pama-Nyungan languages and genetic structure

To investigate if linguistic relationships reflect genetic relationships among Aboriginal Australian populations, we built a Bayesian phylogenetic tree for the 28 different Pama-Nyungan languages represented in this sample[75] (Figure 6b, S15). The linguistic and F_{ST}-based genetic trees obtained (Figure 6**)** share several well-supported partitions. For example, both trees indicate that the northeastern (CAI and WPA), and southwestern groups (ENY, NGA, WCD and WON) each form a cluster, while PIL, BDV and RIV are found between them. 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-value on 9,999 permutations = 0.0001). This suggests that differentiation among Pama-Nyungan languages in Australia follows geographic patterns, as observed in other language families elsewhere in the world[15,76]. Furthermore, we find a correlation between linguistics and genetics ($r_{GEN,LAN}$= 0.43, Mantel test p-value < 0.0005) that remains significant when controlling for geography ($r_{GEN,LAN.GEO}$= 0.26, Partial Mantel test p-value < 0.0005). This is consistent with language differentiation after populations lose (genetic) contact with one another[77]. The correlation between the linguistic and genetic trees is all the more striking given the difference in time scales: the Pama-Nyungan family is generally accepted to have diversified within the last 6 ky[78], 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, localised 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 any 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 often used Population-Branch Statistics[79] (PBS, S16). First, we scanned the Aboriginal Australian genomes for loci with an unusually large changes in allele frequency since divergence from Papuans, taking recent admixture with Europeans and Asians into account. Among the top ranked genomic regions (Extended Data Table 2), we identified candidate loci that might be related to cold tolerance and dehydration resistance. One peak of high differentiation (the 7th highest peak) is located near the *NETO1* gene, which harbours alleles that have previously been shown to be associated with thyroid hormone levels. Interestingly, it has been suggested that thyroid hormone levels are associated with Aboriginal Australian specific adaptations to desert cold[80]. We investigated this potential thermoregulatory adaptation further by identifying genomic regions showing high differentiation associated with different ecological regions in Australia (S16). The top candidate gene in this scan is *KCNJ2,* encoding a potassium channel protein harbouring alleles associated with thyrotoxic periodic paralysis[81]. This disease results from complications related to hyperthyroidism, providing additional support for the thyroid hormone system as a target of desert-related natural selection in Aboriginal Australians[80].

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

Discussion

Our findings shed light, but also raise new questions, concerning the population history of Aboriginal Australians. They suggest an early population structure in Sahul likely dating back ~37 kya (25-40 kya), when the ancestors of Highland Papuans and Pama-Nyungan Aboriginal Australians diversified. Intriguingly, despite this, our results also indicate that the population that

370 diverged from Papuans was the ancestor of all the Aboriginal Australian groups sampled in this
371 study; yet, archaeological evidence shows that by 40-45 kya, humans were widespread within
372 Australia (Figure 1). Three non-exclusive demographic scenarios can account for this observation:
373 1) the Aboriginal Australian ancestral population prior to the divergence from Papuans was
374 widespread, maintaining gene flow across the continent; 2) it was deeply structured, and only one
375 group among the early settlers survived to give rise to modern Aboriginal Australians; and 3) other
376 groups survived, but the descendants are not represented in our sample. Additional modern
377 genomes, especially from Tasmania and the Non-Pama-Nyungan regions of the Northern Territory
378 and Kimberley (both regions highly distinct linguistically [83] and not represented in our sample), as
379 well as ancient genomes pre-dating European contact in Australia and other expansions across
380 Southeast Asia[38], should help resolve these questions in the future.

381 To add to this already complex picture, our estimates of ~44 kya (31-50 kya) for the time of
382 admixture between the Australo-Papuan ancestors and an archaic hominin distantly related to
383 Denisovans are recent. In the absence of paleontological evidence that archaic hominins crossed the
384 Wallace Line, combined with evidence of much lower levels of Denisovan ancestry across East
385 Asia and the Americas[52,86], it is possible that the admixture occurred in Southeast Asia or even
386 further to the west, constraining the age when the ancestors of living Australo-Papuan colonised
387 Sahul and/or the actual timing of Denisovan admixture. In this context, it is noteworthy that our
388 SFS based time estimates relies on the use of recently suggested molecular clock (1.25×10^{-8}
389 /generation/site[84]) and generation time for humans (29 years[85]). Should any of these parameters
390 change, our genetic-based time estimates will also need revisions.

391 Interestingly, our results also indicate that southwestern and northeastern Pama-Nyungan
392 populations diverged 10-32kya. Together with the evidence for selection on genes that may have
393 provided an advantage in extreme desert environments, such as those experienced in Western
394 Desert populations during the LGM, these results point to a long-standing genetic structure among
395 Pama-Nyungan Aboriginal Australians that survived post-glacial demographic changes. In other
396 parts of the world, including Southeast Asia, Pleistocene demographic patterns were overlaid by
397 post-glacial and Holocene expansions that left both genetic and linguistic regional signatures[87]. In
398 Australia, the archaeological record also indicates post-glacial expansions[73,74], while the spread of
399 Pama-Nyungan languages across the continent is generally accepted to be mid-to-late Holocene[35].
400 Our genetic findings indicate an early Holocene demographic expansion localized to northeast
401 Aboriginal Australians, as well as gene flow spreading from the northeast across the continent.

These inferences are consistent with a possible origin and spread of the Pama-Nyungan languages from the northeast of Australia to the rest of the continent. Thus, evidence from genetics may add to the linguistic and cultural evidence - such as the spread of large ceremonial gatherings, trade and exchange intensification, broad alliance networks, cross-group male ritual induction, new plant foods, among several others[35] – that the dispersal of Pama-Nyungan languages has been driven by both cultural diffusion and demic expansion.

Data access

The whole genome sequence data and SNP array data generated in this study are available upon request from E.W (ewillerslev@snm.ku.dk) and D.M.L. (d.lambert@griffith.edu.au). The Papuan whole genome sequence data generated in this study are also available under managed access through the EGA database (https://www.ebi.ac.uk/ega) under study accession number EGAS00001001247.

References for the main text

1. Birdsell, J. B. The recalibration of a paradigm for the first peopling of greater Australia. *Sunda Sahul Prehist. Stud. Southeast Asia Melanes. Aust.* 113–167 (1977).
2. Davidson, I. The colonization of Australia and its adjacent islands and the evolution of modern cognition. *Curr. Anthropol.* **51,** S177–S189 (2010).
3. O'Connell, J. F. & Allen, J. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. *J. Archaeol. Sci.* **31,** 835–853 (2004).
4. Summerhayes, G. R. *et al.* Human Adaptation and Plant Use in Highland New Guinea 49,000 to 44,000 Years Ago. *Science* **330,** 78–81 (2010).
5. 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).
6. 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).

7. Barker, G. *et al.* The 'human revolution'in tropical Southeast Asia: the antiquity of anatomically modern humans, and of behavioural modernity, at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52,** 243–261 (2007).
8. Matsumura, H. & Oxenham, M. F. Demographic transitions and migration in prehistoric East/Southeast Asia through the lens of nonmetric dental traits. *Am. J. Phys. Anthropol.* **155,** 45–65 (2014).
9. Topinard, P. *Etude sur les Tasmaniens*. (1869).
10. Birdsell, J. B. Preliminary data on the trihybrid origin of the Australian Aborigines. *Archaeol. Phys. Anthropol. Ocean.* 100–155 (1967).
11. Tbome, A. Morphological contrasts in Pleistocene Australians. *RL Kirk AG Tborne Eds Orig. Aust.* 95–1 (1976).
12. Thorne, A. G. & Wolpoff, M. H. Regional continuity in Australasian Pleistocene hominid evolution. *Am. J. Phys. Anthropol.* **55,** 337–349 (1981).
13. Westaway, M. C. & Groves, C. P. The mark of Ancient Java is on none of them. *Archaeol. Ocean.* **44,** 84–95 (2009).
14. Lahr, M. M. & Foley, R. Multiple dispersals and modern human origins. *Evol. Anthropol. Issues News Rev.* **3,** 48–60 (1994).
15. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes:* (Princeton University Press, 1996).
16. 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.* **111,** 7248–7253 (2014).
17. Consortium, T. H. P.-A. S. Mapping Human Genetic Diversity in Asia. *Science* **326,** 1541–1545 (2009).
18. Liu, H., Prugnolle, F., Manica, A. & Balloux, F. A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. *Am. J. Hum. Genet.* **79,** 230–237 (2006).
19. Wollstein, A. *et al.* Demographic History of Oceania Inferred from Genome-wide Data. *Curr. Biol.* **20,** 1983–1992 (2010).

20. Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334,** 94–98 (2011).
21. 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).
22. Clark, P. U. *et al.* The last glacial maximum. *science* **325,** 710–714 (2009).
23. Lewis, S. E., Sloss, C. R., Murray-Wallace, C. V., Woodroffe, C. D. & Smithers, S. G. Post-glacial sea-level changes around the Australian margin: a review. *Quat. Sci. Rev.* **74,** 115–138 (2013).
24. Nunn, P. D. & Reid, N. J. Aboriginal Memories of Inundation of the Australian Coast Dating from More than 7000 Years Ago. *Aust. Geogr.* 1–37 (2015).
25. 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).
26. Veth, P. Islands in the Interior: A Model for the Colonization of Australia's Arid Zone. *Archaeol. Ocean.* **24,** 81 (1989).
27. Hiscock, P. & Wallis, L. A. in *Desert Peoples* (eds. Veth, P., Smith, M. & Hiscock, P.) 34–57 (Blackwell Publishing Ltd, 2005). at <http://onlinelibrary.wiley.com/doi/10.1002/9780470774632.ch3/summary>
28. Abbie, A. A. & Australian Institute of Aboriginal Studies. *Studies in physical anthropology: volume II*. (Australian Institute of Aboriginal Studies, 1975). at <http://catalog.hathitrust.org/Record/005995683>
29. Kirk, R. L. *Aboriginal Man Adapting: The Human Biology of Australian Aborigines*. (Clarendon Press, 1981).
30. Birdsell, J. B. *Microevolutionary Patterns in Aboriginal Australia: A Gradient Analysis of Clines*. (Oxford University Press, 1993).
31. Scholander, P. F., Hammel, H. T., Hart, J. S., LeMessurier, D. H. & Steen, J. Cold Adaptation in Australian Aborigines. *J. Appl. Physiol.* **13,** 211–218 (1958).

32. Hammel, H. T., Elsner, R. W., Messurier, D. H. L., Andersen, H. T. & Milan, F. A. Thermal and metabolic responses of the Australian aborigine exposed to moderate cold in summer. *J. Appl. Physiol.* **14,** 605–615 (1959).
33. Dixon, R. M. W. *Australian Languages: Their Nature and Development*. (Cambridge University Press, 2002).
34. Williams, A. N. *et al.* A continental narrative: Human settlement patterns and Australian climate change over the last 35,000 years. *Quat. Sci. Rev.* **123,** 91–112 (2015).
35. Evans, N. & McConvell, P. The enigma of Pama-Nyungan expansion in Australia. *Archaeol. Lang. II* 174–191 (1997).
36. Hiscock, P. Review. *Archaeol. Ocean.* **43,** 44–47 (2008).
37. Savolainen, P., Leitner, T., Wilton, A. N., Matisoo-Smith, E. & Lundeberg, J. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12387–12390 (2004).
38. Bellwood, P. *First Migrants: Ancient Migration in Global Perspective*. (Wiley-Blackwell, 2013).
39. 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. U. S. A.* **110,** 1803–1808 (2013).
40. Haddon, A. C. (Alfred C. *et al. Reports of the Cambridge Anthropological Expedition to Torres Straits ..* (Cambridge [Eng.] : The University Press, 1901). at <http://archive.org/details/reportsofcambrid02hadd>
41. Macknight, C. C. Macassans and Aborigines. *Oceania* **42,** 283–321 (1972).
42. Chase, A. 'All Kind of Nation': Aborigines and Asians in Cape York Peninsula. *Aborig. Hist.* 7–19 (1981).
43. Macknight, C. C. Macassans and the Aboriginal Past. *Archaeol. Ocean.* **21,** 69–75 (1986).
44. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).

45. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics* **196,** 973–983 (2014).
46. Nielsen, R. Estimation of Population Parameters and Recombination Rates From Single Nucleotide Polymorphisms. *Genetics* **154,** 931–942 (2000).
47. Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet* **9,** e1003905 (2013).
48. Allen, J. & O'Connell, J. F. Both half right: Updating the evidence for dating first human arrivals in Sahul. *Aust. Archaeol.* 86 (2014).
49. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46,** 919–925 (2014).
50. Moodley, Y. *et al.* The Peopling of the Pacific from a Bacterial Perspective. *Science* **323,** 527–530 (2009).
51. Holt, S. Palaeoenvironments of the Gulf of Carpentaria from the last glacial maximum to the present, as determined by foraminiferal assemblages. (2005).
52. Qin, P. & Stoneking, M. Denisovan Ancestry in East Eurasian and Native American Populations. *Mol. Biol. Evol.* msv141 (2015). doi:10.1093/molbev/msv141
53. Wall, J. D. *et al.* Higher Levels of Neanderthal Ancestry in East Asians than in Europeans. *Genetics* **194,** 199–209 (2013).
54. Vernot, B. & Akey, J. M. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science* **343,** 1017–1021 (2014).
55. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **advance online publication,** (2015).
56. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for Ancient Admixture between Closely Related Populations. *Mol. Biol. Evol.* **28,** 2239–2252 (2011).
57. Patterson, N. J. *et al.* Ancient Admixture in Human History. *Genetics* genetics.112.145037 (2012). doi:10.1534/genetics.112.145037

58. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
59. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338,** 222–226 (2012).
60. 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).
61. Lukić, S. & Hey, J. Demographic Inference Using Spectral Methods on SNP Data, with an Analysis of the Human Out-of-Africa Expansion. *Genetics* **192,** 619–639 (2012).
62. Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci.* **111,** 2632–2637 (2014).
63. gounder Palanichamy, M. *et al.* Phylogeny of Mitochondrial {DNA} Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of South Asia. *Am. J. Hum. Genet.* **75,** 966–978 (2004).
64. Kivisild, T. *et al.* The Role of Selection in the Evolution of Human Mitochondrial Genomes. *Genetics* **172,** 373–387 (2006).
65. Hudjashov, G. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci.* **104,** 8726–8730 (2007).
66. van Holst Pellekaan, S. M., Ingman, M., Roberts-Thomson, J. & Harding, R. M. Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. *Am. J. Phys. Anthropol.* **131,** 282–294 (2006).
67. Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13,** 1600–1606 (2003).
68. Friedlaender, J. *et al.* Expanding Southwest Pacific Mitochondrial Haplogroups P and Q. *Mol. Biol. Evol.* **22,** 1506–1517 (2005).

69. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20,** 278–280 (1998).
70. Lippold, S. *et al.* Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig. Genet.* **5,** 13 (2014).
71. Radcliffe-Brown, A. R. The Social Organization of Australian Tribes. *Oceania* **1,** 34–63 (1930).
72. 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).
73. Haberle, S. G. & David, B. Climates of change: human dimensions of Holocene environmental change in low latitudes of the PEPII transect. *Quat. Int.* **118-119,** 165–179 (2004).
74. Lourandos, H & David, B. in *in Bridging Wallace's Line: the Environmental and Cultural History and Dynamics of the SE Asian-Australasian Region* (ed. A.P. Kershaw, B. David, N. Tapper, D. Penny & J. Brown.) (97-118).
75. Bowern, C. & Atkinson, Q. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* **88,** 817–845 (2012).
76. Excoffier, L., Harding, R. M., Sokal, R. R., Pellegrini, B. & Sanchez-Mazas, A. Spatial differentiation of RH and GM haplotype frequencies in Sub-Saharan Africa and its relation to linguistic affinities. *Hum. Biol.* **63,** 273–307 (1991).
77. Bowern, C. & Evans, B. *The Routledge Handbook of Historical Linguistics*. (Routledge, 2015).
78. Evans, N. & Jones, R. in *Archaeology and linguistics: aboriginal Australia in global perspective* (Oxford University Press Australia, 1997).
79. Yi, X. *et al.* Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* **329,** 75–78 (2010).
80. 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. R. Soc. Lond. B Biol. Sci.* **281,** 20132747 (2014).

81. Cheung, C.-L. *et al.* Genome-wide association study identifies a susceptibility locus for thyrotoxic periodic paralysis at 17q24.3. *Nat. Genet.* **44,** 1026–1029 (2012).
82. 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).
83. Evans, N. *The Non-Pama-Nyungan Languages of Northern Australia: Comparative Studies of the Continent's Most Linguistically Complex Region*. (Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University, 2003).
84. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13,** 745–753 (2012).
85. 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).
86. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl. Acad. Sci.* **108,** 18301–18306 (2011).
87. Bellwood, P. Early Agriculturalist Population Diasporas? Farming, Languages, and Genes. *Annu. Rev. Anthropol.* **30,** 181–207 (2001).

Supplementary Information (see annex)

S01 Ethical approvals in relation to sampling in Australia
S02 Ethnography and linguistics for the Aboriginal Australian individuals
S03 Sample collection, DNA extraction, array genotyping, whole-genome sequencing and processing
S04 Reference panels, relatedness and runs of homozygosity
S05 Linkage disequilibrium (LD) and population structure within Australia
S06 Local ancestry
S07 Demographic inferences
S08 MSMC analysis
S09 D-statistic based tests using sampled reads from sequencing data
S10 Mutation load analysis

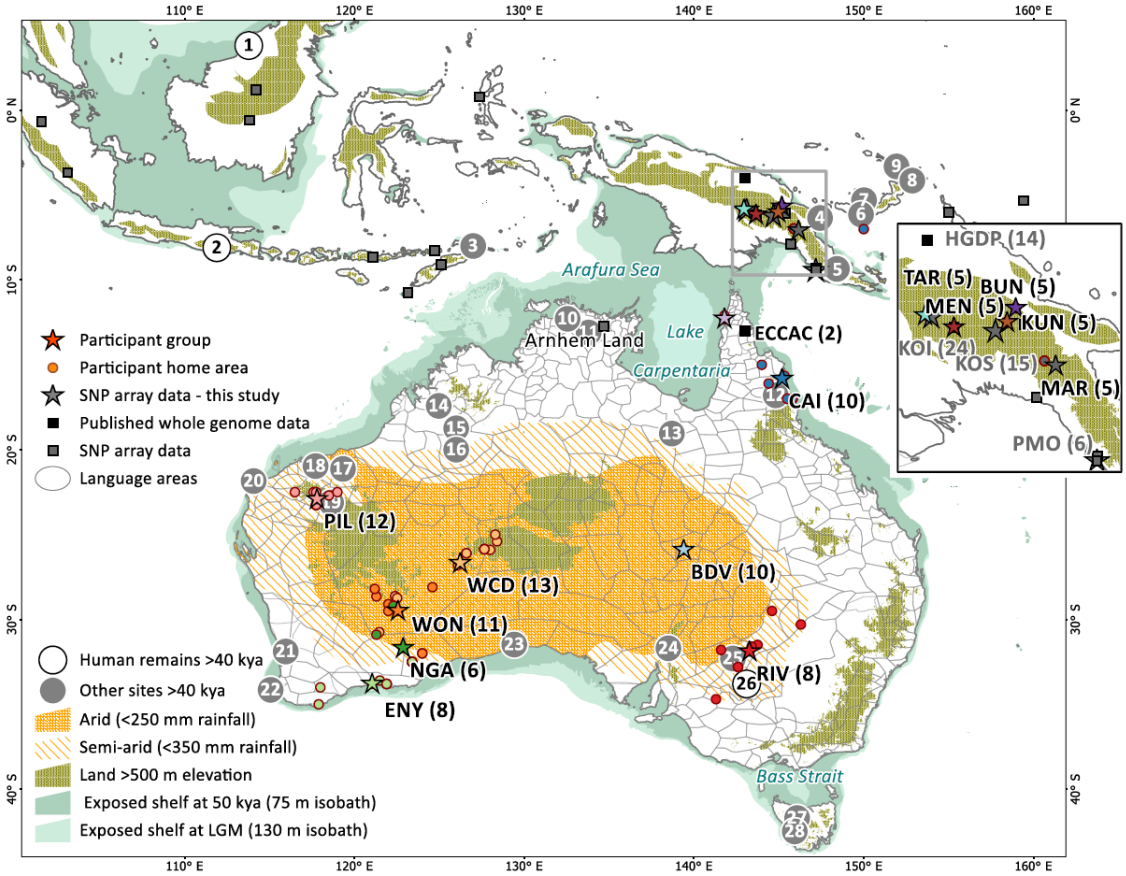


Figure 1 | Archaeological sites and human remains dated to ~40 kya or older in southern Sunda and Sahul and Aboriginal Australians and Papuans samples used in this study. The sites with dated human remains are shown as white circles and the archaeological sites as grey circles. The associated dates can be found in S01. The stars indicate the centroid location for each sampling group. SNP array data are shown in grey, and whole genomes are represented by coloured stars. Also shown on this figure are the coordinates for each participant - when available - computed as the centroid between the parents' birth sites (filled circles). Publicly available genetic data (see S04) used as a reference panel in this study shown as squares. The grey boundaries correspond to territories defined by the language groups provided by the Australian Institute of Aboriginal and Torres Strait Islander studies (Horton 1994). Sampled Aboriginal Australians self-identify primarily as: Yidindji and Gungandji from the Cairns region (CAI, 10, see also S02); Yupangati and Thanakwithi from northwest Cape York (WPA, 6), Wangkangurru and Yarluyandi from the Birdsville region (BDV, 10, 9 sequenced at high depth), Barkindji from southeast (RIV, 8); Pilbara area Yinhawangka and Banjima (PIL, 12), Ngaanyatjarra from central Australia (WCD, 13), Wongatha from WA's northern Goldfields (WON, 11), Ngadju from WA's southern Goldfields (NGA, 6); and Nyungar from southwest Australia (ENY, 8). Papuans include samples from the locations Bundi (BUN, 5), Kundiawa (KUN, 5), Mendi (MEN, 5), Marawaka (MAR, 5) and Tari (TAR, 5) - all whole genome sequenced. Additionally, we generated SNP array data for 45 Papuan samples including 24 Koinambe (KOI) and 15 Kosipe (KOS) - described before (Migliano et al. 2013) - and 6 individuals with highland ancestry sampled in Port Moresby (PMO). Lake Carpentaria, 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.

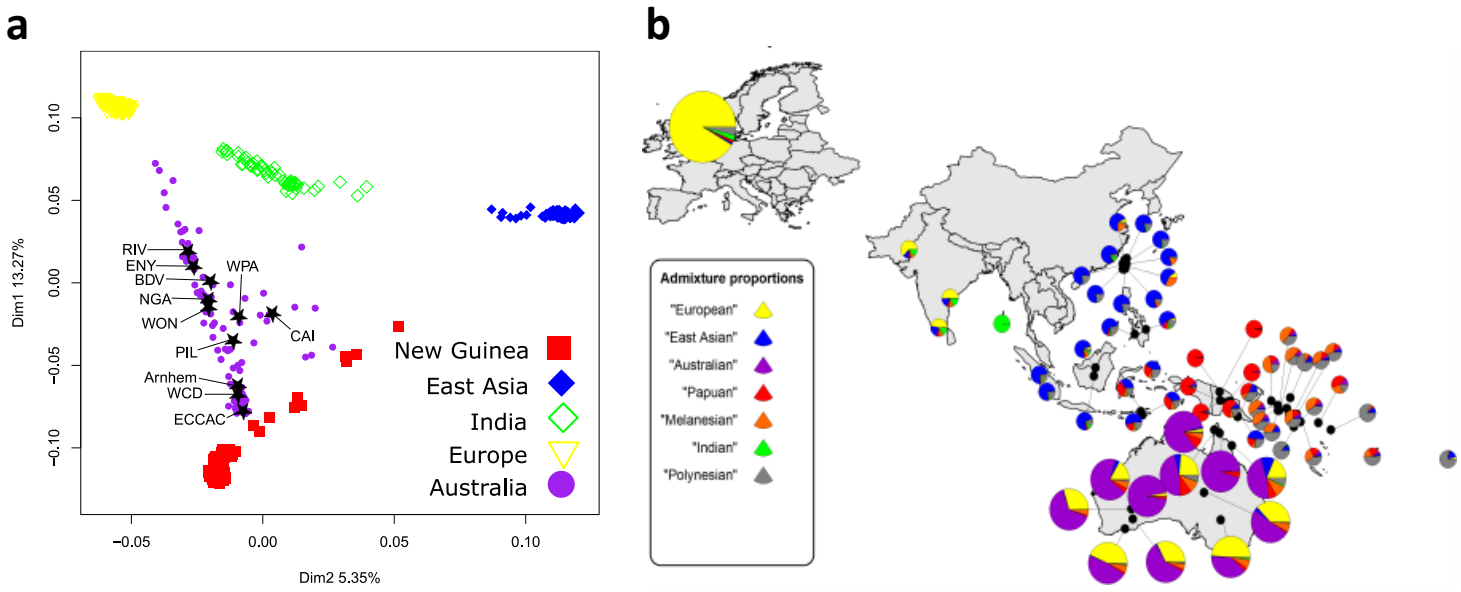


Figure 2 | Genetic ancestry of Aboriginal Australians in a worldwide context. a, MDS plot. Classical Multidimensional scaling (MDS), also known as Principal Coordinate Analysis, plot of first two dimensions based on an identity-by-state (IBS) distance matrix (based on 54,971 SNPs) between individuals from this study and worldwide populations, including publicly available data from (Lazaridis et al. 2014, Qin and Stoneking 2015, Pugach et al. 2013, Reich et al. 2011). The first two dimensions explains 19% of the variance present in the IBS distance matrix. Individuals are colour-coded according to sampling location, grouped into Australia (Arnhem, 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). The stars indicate the centroid for each sampling location of the Aboriginal Australian groups. Aboriginal Australians from this study as well as from previous studies are closest to Papuans, and also show signals of admixture with Eurasians (see S05 for details). **b, Estimation of admixture proportions.** Estimation of genomic ancestry proportions at 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 Figure 1), and publicly available SNP array data (Lazaridis et al. 2014, Qin and Stoneking 2015, Pugach et al. 2013, Reich et al. 2011). Each ancestry component has been labelled according to the geographic region showing the corresponding highest frequency. The genomes of Aboriginal Australian populations are mostly a mixture of European and Aboriginal Australian ancestry components. Northern Aboriginal Australian groups (Arnhem, 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 Papuan populations. A background of 5% “Melanesian” component is observed in all the Aboriginal Australian populations; however, this component is widely spread over all the geographic area shown in this figure, being present as far as Taiwan and India. The “Indian” component is present on average at 1.5% in all the Aboriginal Australian groups while low levels of this “Indian” component is also present in the Europeans; similarly, on average we detected 1.4% “Polynesian” component across the Aboriginal Australian samples. We attribute these residual ancestry components of less than 1.5% to statistical noise as they are present in other southeast Asian populations and are not supported by other analyses (S05).

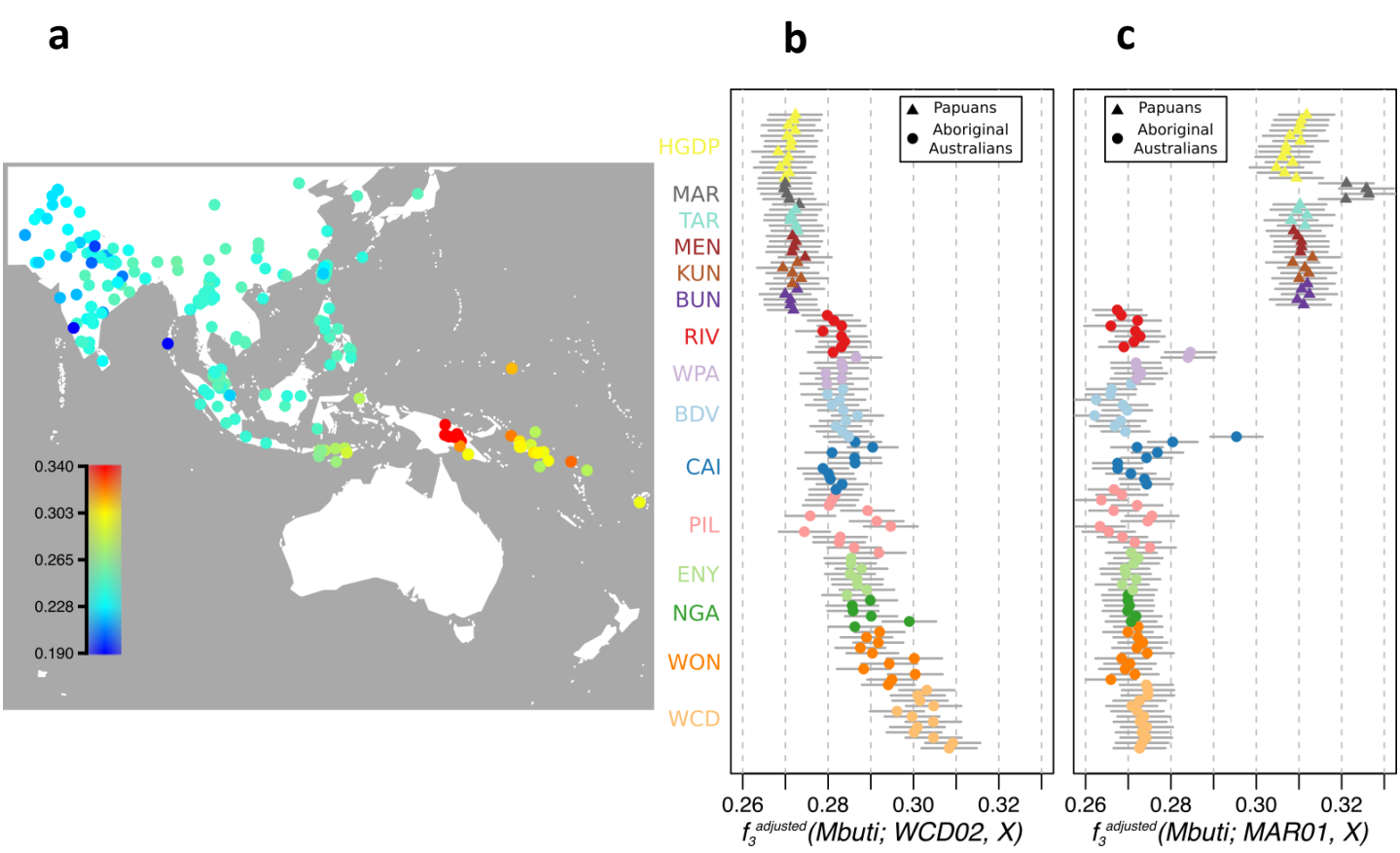


Figure 3 | Genetic affinities of Aboriginal Australian and Highland Papuan genomes. The amount of shared genetic drift between two genomes relative to an outgroup, here the African Mbuti population, was quantified using the outgroup f_3 statistic. Higher values of this statistic implies more shared drift. **a, Australian Aboriginal genetic affinities to populations in southeastern Asia and Oceania.** A heat map displaying f_3 statistics between the individual WCD02, showing no trace of recent admixture and therefore chosen to represent the Australian Aboriginal 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. 760,116 SNPs for which WCD02 had non-missing array genotypes that overlapped with any other datasets were used. **b, Genetic affinities between a Western Central Desert (WCD) genome and Aboriginal Australians and Papuans.** f_3 statistics between the individual WCD02 and all other Australian Aboriginal and Highland Papuan individuals that were whole-genome sequenced for this study, using all genotypes called from the sequencing data. Because the widespread recent admixture in the Australian Aboriginal population has large confounding effects on these 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 (meaning not Australian Aboriginal or Papuan) ancestry in each individual genome. The adjusted f_3 statistics display a genetic gradient that separates western and eastern Australian Aboriginal populations. However, we find no differences between Papuan population samples in their level of Aboriginal Australian affinity (Kruskal-Wallis test, p-value = 0.083). Horizontal lines correspond to ±1 standard error. **c, 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 (**b**). All the Aboriginal Australian populations 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-value = 0.0002, after removing the three outliers), which could be consistent with e.g. low levels of Papuan gene flow into some Aboriginal Australian groups (see S06 and S07), we caution that some of these differences are likely 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.

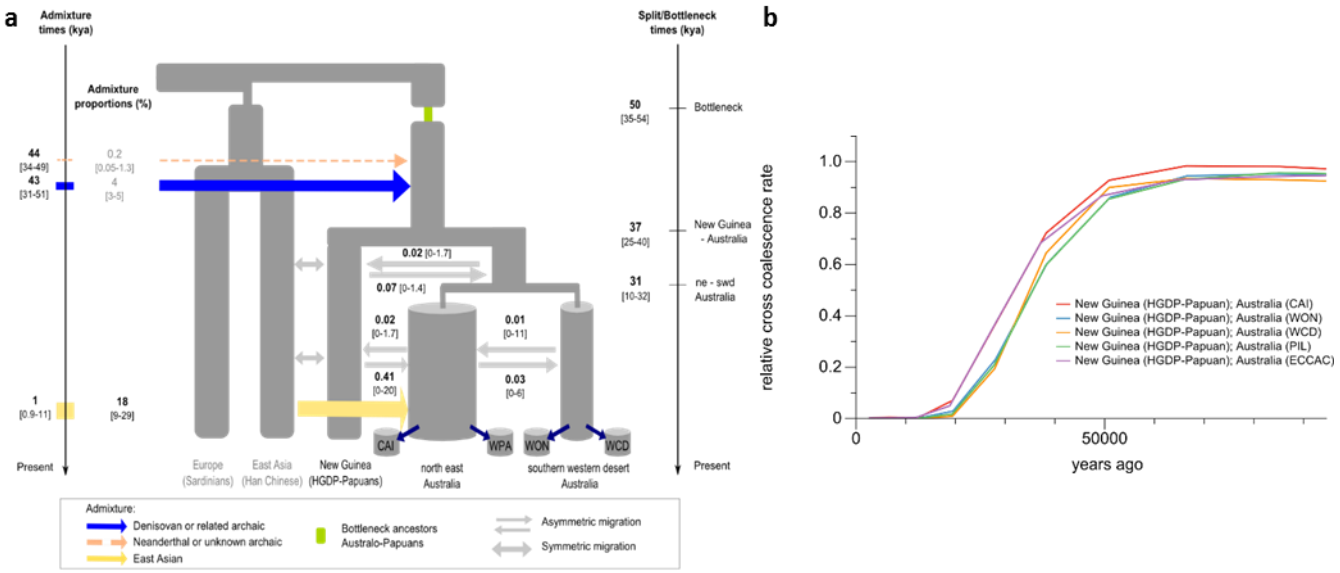


Figure 4 | Settlement of Australia. a, Best supported demographic model of the colonisation of Australia and New Guinea. The demographic history of Aboriginal Australian populations was modelled by considering that sampled individuals are from sub-populations ("islands") that are part of two larger regions ("continents"); these regions geographically match the northeast and the southwestern desert regions. Maximum likelihood parameter estimates (MLEs) were obtained from the joint site frequency spectrum (SFS) of Han Chinese, HGDP-Papuans, CAI, WPA, WON and WCD. Only Aboriginal Australian individuals with low European ancestry were included in this analysis. In this model, we only estimated parameters specific to the settlement of Australia and New Guinea (numerical values shown in black); all the other demographic parameters were set to the point estimates shown in Figure 5 (numerical values shown in grey here). For additional information see S07. **b, MSMC analyses.** Linear interpolation through the midpoints of the time intervals of the relative cross coalescence rate estimates from MSMC (Schiffels and Durbin 2014) using pairs of individuals including one HGDP-Papuan and one other individual as indicated. We used CAI01, PIL06, WCD01, WON03 and an ECCAC for this analysis (see S08 for details). The SFS and MSMC results were scaled using a mutation rate = 1.25×10^{-8} /generation/site as suggested in Scally and Durbin 2012 and a generation time of 29 years corresponding to the average hunter-gatherer generation interval for males and females as reported in Fenner 2005.

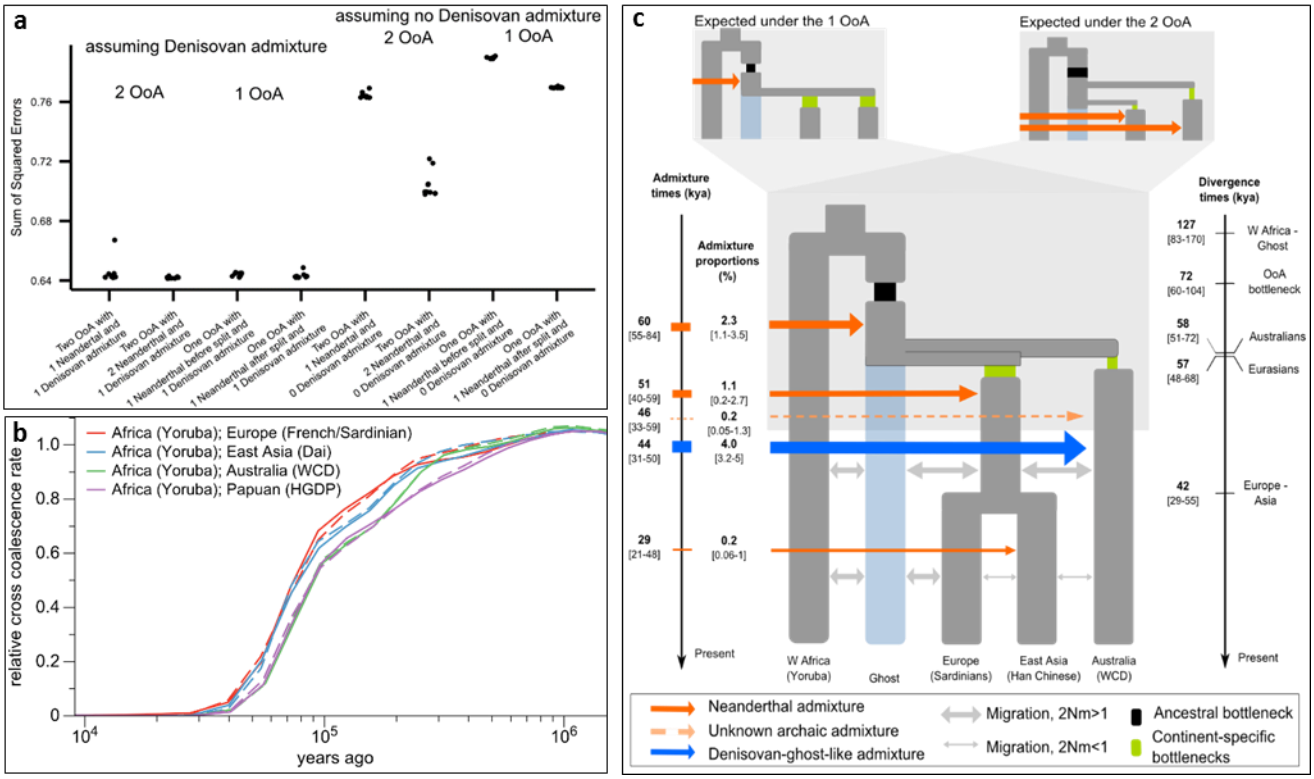


Figure 5 | Out of Africa. **a, Admixture graphs.** Sum of Squared Errors between all observed D-statistics and the expectations for eight admixture graphs. The graphs differ in (i) the number of OoA events, (ii) whether or not Denisovan admixture into modern humans is considered, and (iii) the number of Neanderthal admixture pulses. Each point is the result of the optimization procedure with different starting points. See S09 for details. **b, MSMC analyses.** Linear interpolation through the midpoints of the time intervals of the relative cross coalescence rate estimates from MSMC (Schiffels and Durbin 2014) for pairs of individuals including one Yoruba and one other sample, as indicated in the legend. Solid curves are generated from one individual from each population, and dashed curves are independent replicates using a second individual from each population (see S08 for details). **c, Model comparison and inferred parameters.** We used a likelihood-based model to investigate whether the joint SFS supports the one wave (1 OoA) or the two waves (2 OoA) scenarios. The maximum likelihood estimates (MLEs) 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 that presumably took place outside Africa, and (iii) overlapping divergence times 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 figure 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 a resampling procedure). We assume that the OoA event is associated with the ancestral bottleneck. Note that our model also includes gene flow between neighbouring human populations, and admixture pulses with archaics (Neanderthal and Denisovan). 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 sizes are proportional to the intensity of gene flow and the admixture proportions. See S07 for further details. The SFS and MSMC results were scaled using a mutation rate = 1.25×10^{-8}/generation/site as suggested in Scally and Durbin 2012 and a generation time of 29 years corresponding to the average hunter-gatherer generation interval for males and females as reported in Fenner 2005.

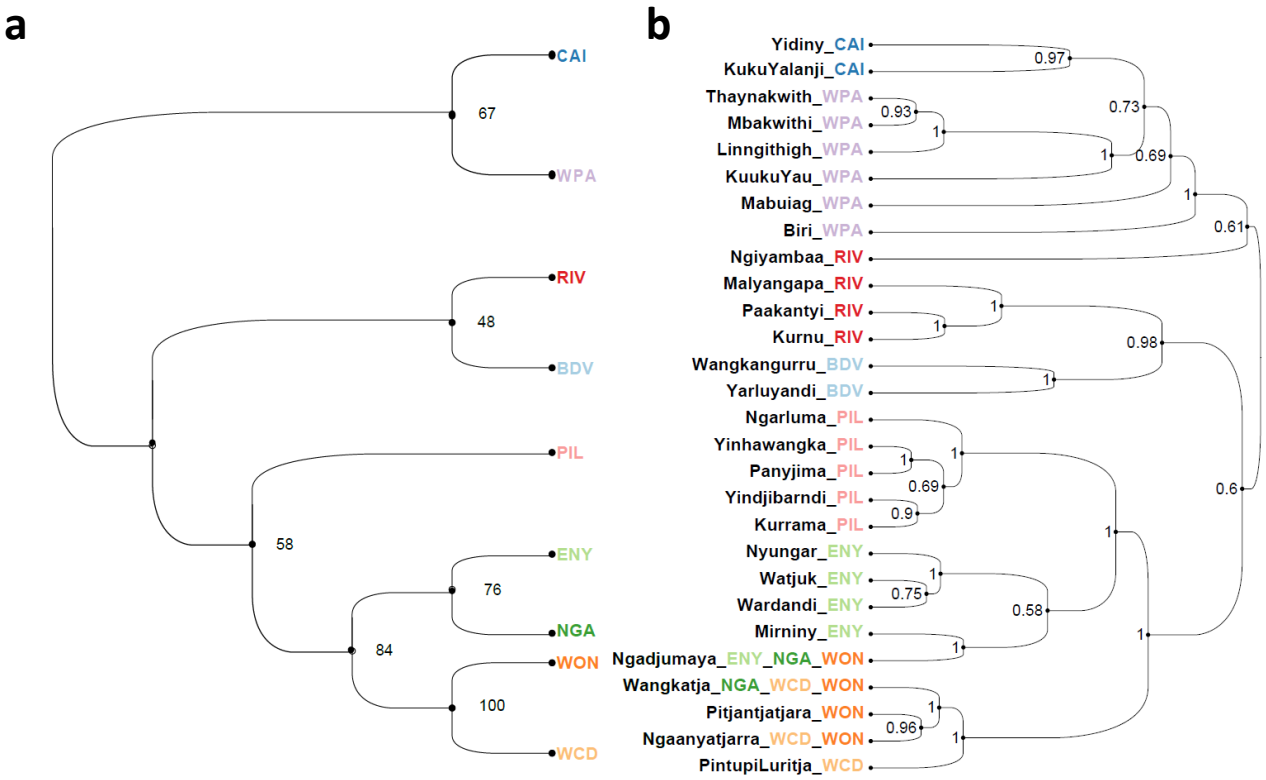
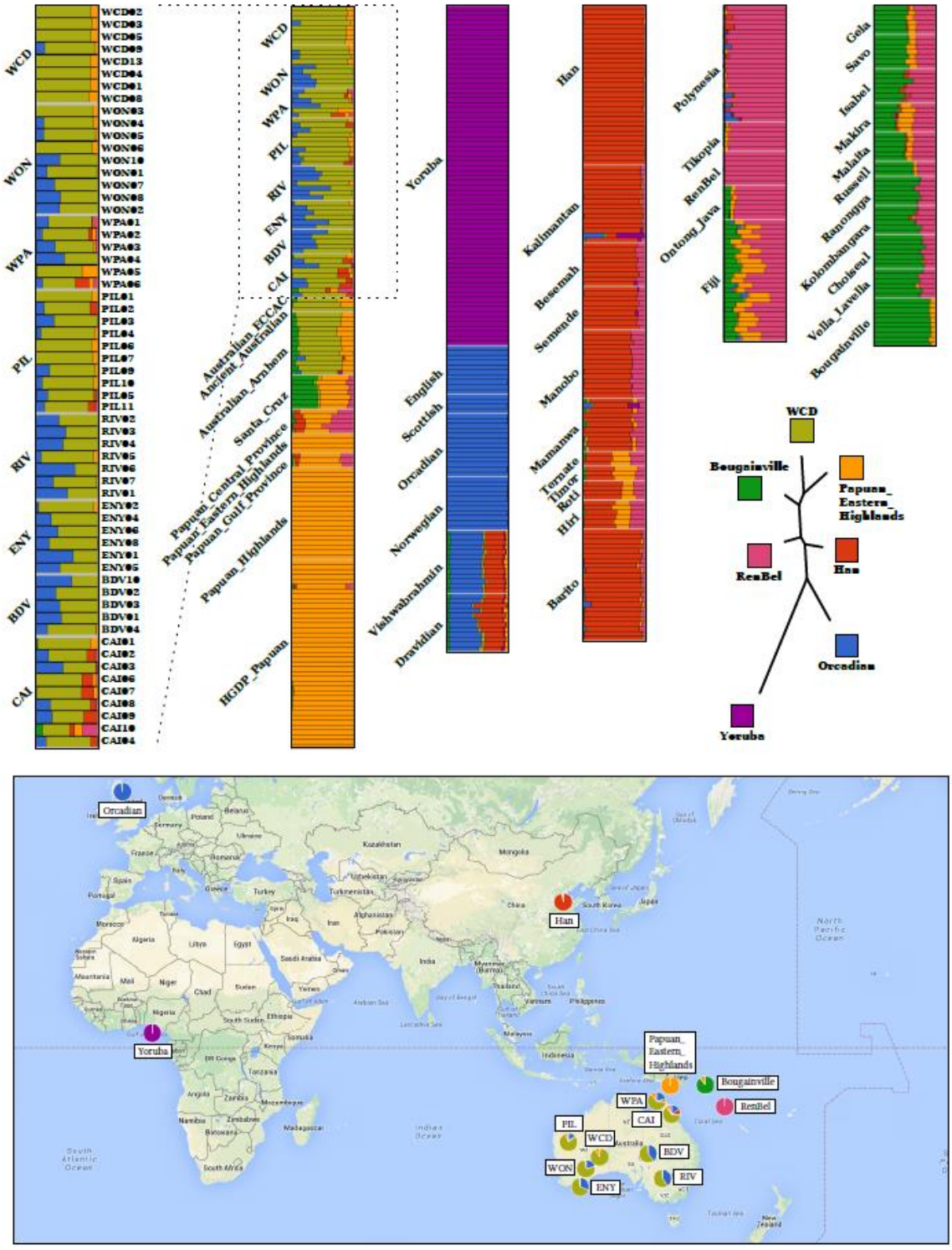
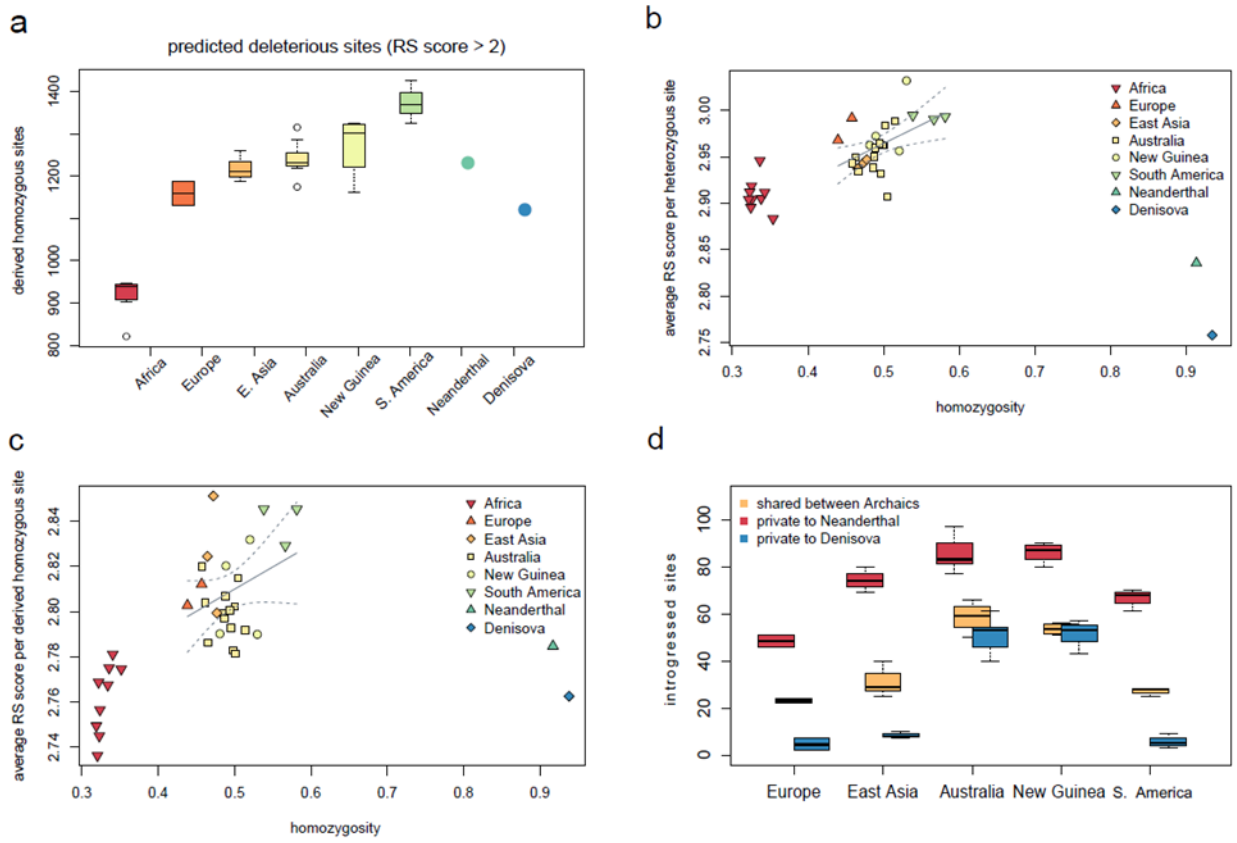


Figure 6 | Correspondence between genetics and linguistics. a, Unrooted Neighbor-Joining (NJ; (Saitou and Nei 1987)) F_{ST}-based genetic tree (cladogram). Weir and Cockerham F_{ST} distance (Weir and Cockerham 1984) was computed between the Aboriginal Australian populations after masking the Eurasian tracts. Statistical robustness of each branch was estimated by means of a bootstrap analysis (1000 replicates). **b,** Bayesian phylogenetic tree for the 28 different Pama-Nyungan languages represented in this sample (from Bowern and Atkinson 2012, see 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 (Drummond and Rambaut 2007).



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 (CV) error is minimized for this value of K for 5-fold CV. The rooted tree supports the shared genetic origin of Aboriginal Australians, Papuans and Bougainvilleans.

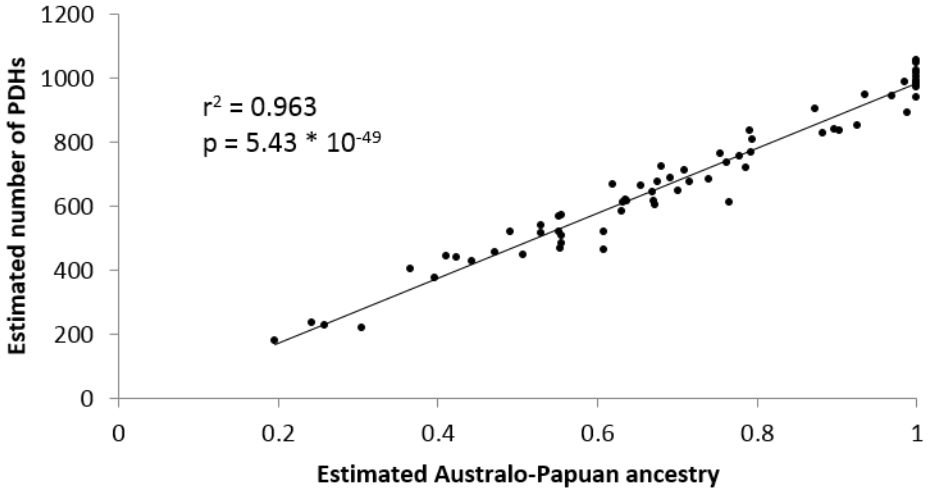


Extended Data Figure 2 | Deleterious and introgressed sites. a, Number of derived homozygous sites that are predicted to be deleterious (average RS score > 2, see S10). Average RS score (see S10) for derived heterozygous sites, **b,** and homozygous sites, **c**. Each coloured symbol corresponds to estimates from a single individual. Homozygosity is calculated as the fraction of derived homozygous sites among all sites at which an individual carries at least one copy of a derived allele. Solid lines show the linear regression of homozygosity against average RS score per site for non-African modern humans. Dashed lines indicate the 95% confidence interval for the linear regression. **d,** Distribution of per individual number of putative introgressed sites from archaic humans.

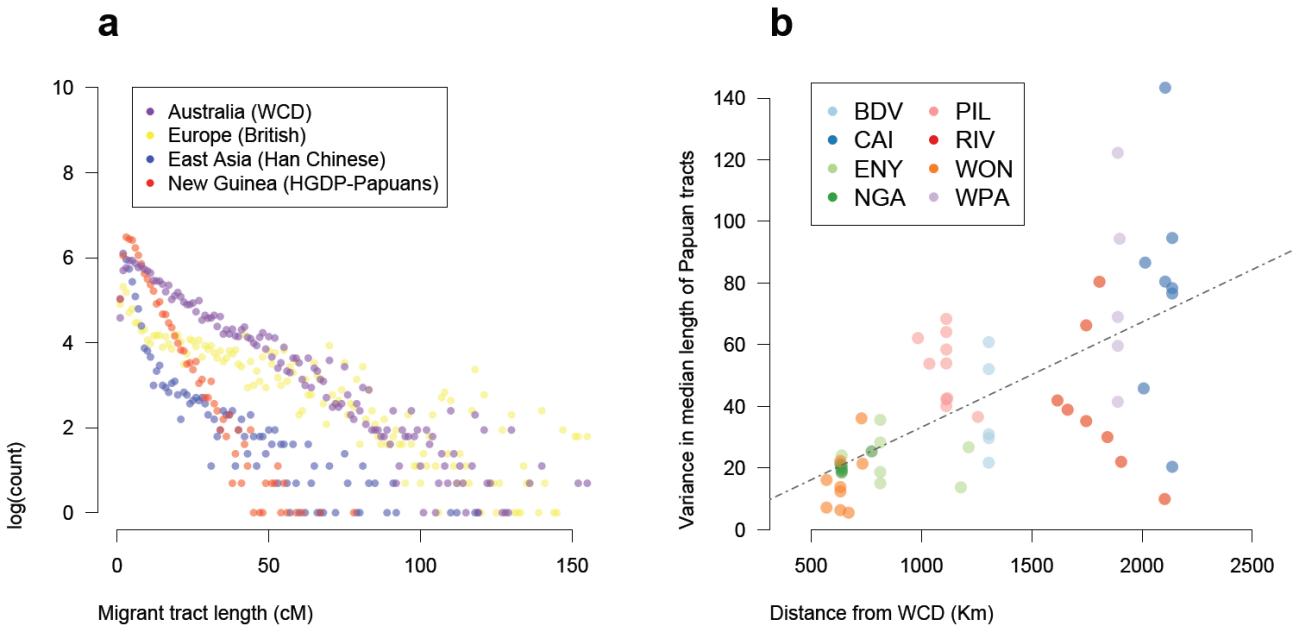
a

Population	Sample size	PDHs	PDH lengths (Mb)	PNHs	PNH lengths (Mb)
European	2	77	1.57	951	38.5
South Asian	6	124	4.47	966	40.2
East Asian	3	128	4.47	1120	44.3
Amerindian	3	115	3.01	1140	45.6
Polynesian	2	241	8.09	1105	43.6
Papuan	16	974	35.8	1183	41.5
BDV	9	602	22.3	1171	42.2
CAI	10	653	23.6	1188	43.6
ENY	8	533	19.7	1149	42.8
NGA	6	682	25.5	1167	42.9
PIL	12	785	28.5	1225	42.6
RIV	8	508	18.2	1093	41.1
WCD	13	991	38.2	1225	39.9
WON	11	717	26.5	1151	40.8
WPA	6	647	24.2	1148	42.1

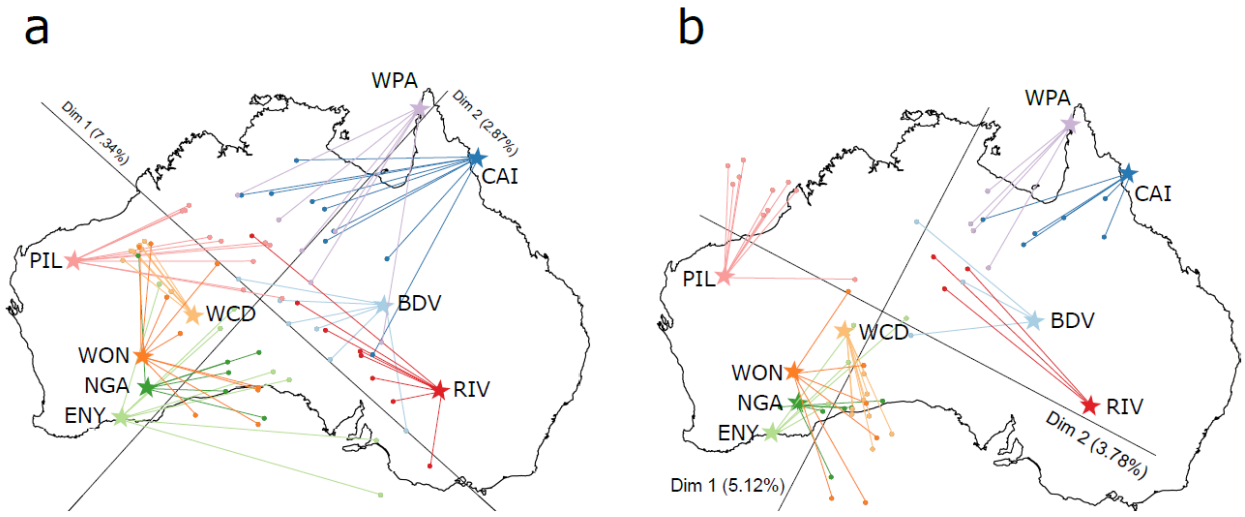
b



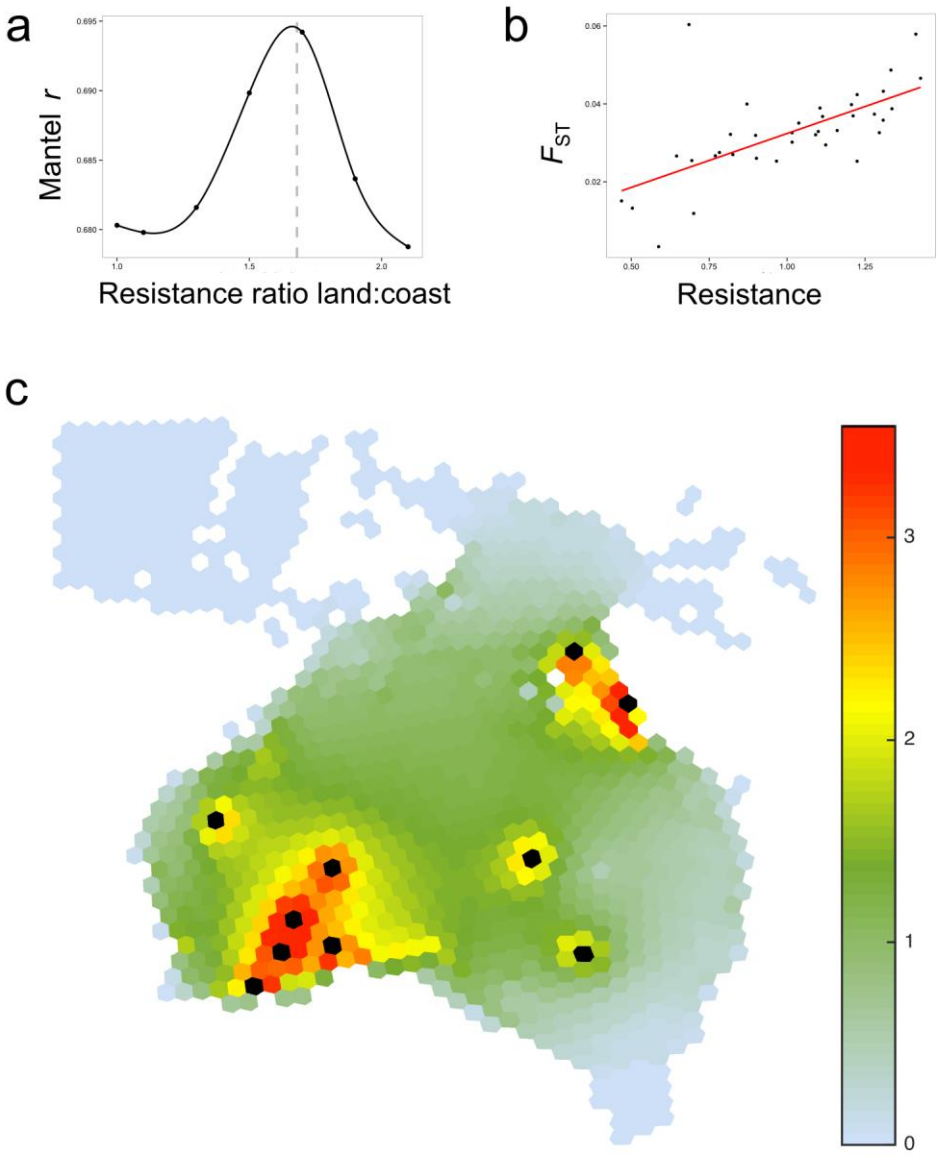
Extended Data Figure 3 | Putative Denisovan (PDH) and Neanderthal haplotypes (PNH). The putative haplotypes correspond to clusters (four or more SNPs spanning at least 4kb) of heterozygous or homozygous genotypes in complete linkage disequilibrium ("diplotypes") that are potentially the result of Neanderthal 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 S11 for details). **a,** Average number of PDHs and PNHs per individual across populations. PDH lengths (and PNH lengths) are the sums of the lengths of PDHs (respectively PNHs) per sample. **b**, Correlation between the estimated amount of Australo-Papuan ancestry (see Figure 2b, S05) and the number of identified PDHs for each Australian sample.



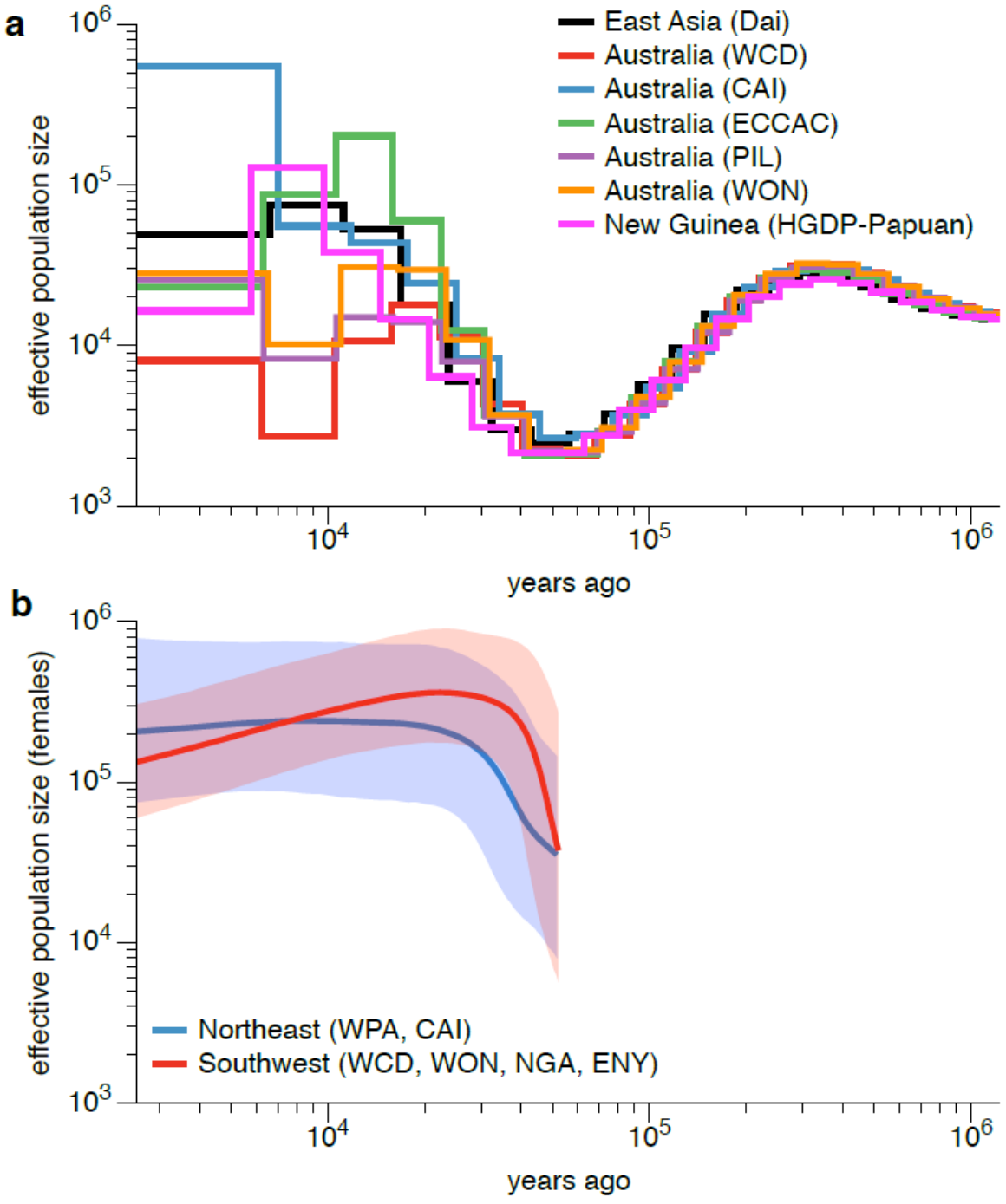
Extended Data Figure 4 | European, East Asian and Papuan genomic tracts in Aboriginal Australians. a, Distribution of the tracts assigned to Aboriginal Australian (WCD), Papuan, East Asian or European ancestry for the 58 unrelated non-WCD Australian samples. Most of the shorter tracts were of Papuan 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 vs. 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-value < 10^{-5}), with "younger tracts" closer to New Guinea (i.e., with a larger variance) and "older tracts" closer to WCD" (i.e., with a smaller variance). This is also consistent with continuous Papuan gene flow spreading from the northeast.



Extended Data Figure 5 | Genetics mirrors geography. First two dimensions of a classical multidimensional scaling (MDS) analysis of the Aboriginal Australian genome sequences. **a,** Considering genetics variants in the whole genomes of Aboriginal Australians. **b,** After masking the tracts assigned to Han Chinese or British (S06). Both MDS plots have been rotated towards the best overlap with geographic sampling locations as defined by Procrustes analysis (Wang et al. 2010). In each plot, the arrows indicate the error of the MDS coordinates towards the assigned population sampling geographic coordinates.



Extended Data Figure 6 | Gene flow across the continent. a, 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. **b,** Best fit of pairwise population genetic differentiation, F_{ST} (computed between the nine Aboriginal Australian groups after masking Eurasian tracts (S06)), versus pairwise connectivity based on the environment (estimated as resistance) when moving inland is 1.7 times harder than moving along coastal nodes. **c,** 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 7 | Effective population size changes over time. a, MSMC analyses. 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, *i.e.*, four haplotypes in each run. MSMC results were scaled using a mutation rate = 1.25×10^{-8}/generation/site as suggested in Scally and Durbin 2012 and a generation time of 29 years corresponding to the average hunter-gatherer generation interval for males and females as reported in Fenner 2005**. b, Bayesian Skyline Plots**. 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 S12).

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

region	pop_id	indiv_id	sex	DoC[1]	mtDNA haplogroup[2]	Ychr haplogroup[3]	Pama-Nyungan language[4]
Australia	BDV	BDV01	F	78	S2	-	Yarluyandi_Wangkangurru
Australia	BDV	BDV02	M	75	S1a	R1b1a2a1a2c1g2a1a2	Yarluyandi_Wangkangurru
Australia	BDV	BDV03	M	-	-	-	Yarluyandi_Wangkangurru
Australia	BDV	BDV04	F	70	O1a	-	Yarluyandi_Wangkangurru
Australia	BDV	BDV05	M	72	S1a	O1a	Yarluyandi_Wangkangurru
Australia	BDV	BDV06	F	70	S1a	-	Yarluyandi_Wangkangurru
Australia	BDV	BDV07	F	70	O1a	-	Yarluyandi_Wangkangurru
Australia	BDV	BDV08	M	70	S1a	R1b1a2a1a2c1g2a1	Yarluyandi_Wangkangurru
Australia	BDV	BDV09	F	74	S1a	-	Yarluyandi_Wangkangurru
Australia	BDV	BDV10	M	72	S1a	I1a2a1d	Yarluyandi_Wangkangurru
Australia	CAI	CAI01	M	84	P	K2b	Yidiny
Australia	CAI	CAI02	M	74	M42	K2b	Yidiny
Australia	CAI	CAI03	F	77	M42a	-	Yidiny
Australia	CAI	CAI04	F	71	P	-	Yidiny_KukuYalanji
Australia	CAI	CAI05	M	80	P	O2a1a	Yidiny
Australia	CAI	CAI06	M	78	P	C1b	Yidiny
Australia	CAI	CAI07	M	71	N13	K2b	KukuYalanji
Australia	CAI	CAI08	M	70	P	K2b	Yidiny
Australia	CAI	CAI09	M	79	P	R1b1a2a1a2b1	Yidiny
Australia	CAI	CAI10	M	73	E1a2	K2b	-
Australia	ENY	ENY01	M	69	H1e1a3	R1b1a2a1a2b1c1	Nyungar
Australia	ENY	ENY02	F	79	R12	-	Ngadjumaya
Australia	ENY	ENY03	F	83	O	-	Mirniny
Australia	ENY	ENY04	F	83	M42	-	Nyungar
Australia	ENY	ENY05	F	78	S2	-	Ngadjumaya
Australia	ENY	ENY06	F	70	M42	-	Wardandi
Australia	ENY	ENY07	M	73	S2	E1b1b1b2a	Watjuk
Australia	ENY	ENY08	M	71	P4b1	C1b	Nyungar_Ngadjumaya
Australia	NGA	NGA01	F	74	O1	-	Ngadjumaya
Australia	NGA	NGA02	F	52	O1a	-	Ngadjumaya
Australia	NGA	NGA03	F	73	O	-	Ngadjumaya
Australia	NGA	NGA04	M	75	O	R1b1a2a1a1b1a1a	Wangkatja
Australia	NGA	NGA05	F	56	R12	-	Ngadjumaya
Australia	NGA	NGA06	F	63	S1a	-	Wangkatja
Australia	PIL	PIL01	M	58	R	C1b	Yinhawangka
Australia	PIL	PIL02	M	61	M42	C1b	Yinhawangka
Australia	PIL	PIL03	F	56	M42	-	Yinhawangka
Australia	PIL	PIL04	F	64	M42	-	Yinhawangka
Australia	PIL	PIL05	M	68	M42	C1b	Yinhawangka
Australia	PIL	PIL06	M	59	O1	K2b	Panyjima
Australia	PIL	PIL07	F	63	O	-	Panyjima
Australia	PIL	PIL08	M	72	M42	C1b	Yindjibarndi_Kurrama
Australia	PIL	PIL09	M	58	S5	R1b1a2a1a2e1	Kurrama
Australia	PIL	PIL10	F	61	R	-	Yinhawangka
Australia	PIL	PIL11	M	57	P3b	C1b	Kurrama
Australia	PIL	PIL12	M	63	P3b	C1b	Yindjibarndi
Australia	RIV	RIV01	F	73	M42a	-	Ngiyambaa
Australia	RIV	RIV02	F	62	P4b1	-	Paakantyi
Australia	RIV	RIV03	F	69	M42a	-	Paakantyi
Australia	RIV	RIV04	M	62	P4b1	I2a1a2a1a	Kurnu
Australia	RIV	RIV05	F	72	P4b1	-	Paakantyi
Australia	RIV	RIV06	M	66	H1bs	J2a1b	Ngiyambaa
Australia	RIV	RIV07	M	70	P4b1	R1b1a2a1a2c1c	Paakantyi
Australia	RIV	RIV08	F	66	P4b1	-	Paakantyi_Malyangapa

Australia	WCD	WCD01	M	62	R12	K2b	Ngaanyatjarra
Australia	WCD	WCD02	M	59	S1a	C1b	Ngaanyatjarra
Australia	WCD	WCD03	M	61	R12	K2b	Wangkatja
Australia	WCD	WCD04	M	52	P3b	K2b	Ngaanyatjarra
Australia	WCD	WCD05	M	60	O1	C1b	Ngaanyatjarra
Australia	WCD	WCD06	M	58	O1a	C1b	Ngaanyatjarra
Australia	WCD	WCD07	F	61	M42	-	Ngaanyatjarra
Australia	WCD	WCD08	F	64	M42	-	Ngaanyatjarra
Australia	WCD	WCD09	M	59	R	J2a1b	Ngaanyatjarra
Australia	WCD	WCD10	F	63	M42	-	Ngaanyatjarra
Australia	WCD	WCD11	M	57	M42	K2b	Ngaanyatjarra
Australia	WCD	WCD12	M	59	M42	C1b	Ngaanyatjarra_PintupiLuritja
Australia	WCD	WCD13	M	67	M14	C1b	Ngaanyatjarra
Australia	WON	WON01	M	71	O	I1a2a1a3a	Wangkatja
Australia	WON	WON02	F	101	O1a	-	Wangkatja
Australia	WON	WON03	F	65	O1a	-	Wangkatja
Australia	WON	WON04	F	58	R	-	Ngaanyatjarra
Australia	WON	WON05	M	56	O1a	I2a2a1a2a2	Wangkatja
Australia	WON	WON06	F	60	R12	-	Wangkatja
Australia	WON	WON07	F	57	O	-	Ngadjumaya
Australia	WON	WON08	F	52	O	-	Wangkatja
Australia	WON	WON09	M	20	O	E1b1b1a1b1a4	Wangkatja
Australia	WON	WON10	M	50	O1	R1b1a2a1a2a	Wangkatja
Australia	WON	WON11	F	58	R12	-	Pitjantjatjara
Australia	WPA	WPA01	F	51	P5	-	Thaynakwith_Linngithigh
Australia	WPA	WPA02	M	50	P	C1b	Mpakwithi_Kaanju
Australia	WPA	WPA03	M	51	M42a	K2b	Thaynakwith_Biri
Australia	WPA	WPA04	F	52	P5	-	Thaynakwith_KukuYau
Australia	WPA	WPA05	F	56	M42	NA	Mabuiag_Thaynakwith
Australia	WPA	WPA06	M	53	P5	O1a	Mpakwithi

[1]The depth of coverage (DoC) is the average number of reads covering every position in the genome (hg19) after duplicate removal (see S03).

[2]The average depth of coverage on the mitochondrial genome (mtDNA) is 3,484 ± 1,515 (mean ± SD) and haplogroups were called with haplogrep (van Oven and Kayser 2009; Kloss-Brandstätter et al. 2011) and haplofind (Vianello et al. 2013), see S12.

[3]The average depth of coverage on the Y chromosome (Ychr) is 28.88 ± 4.5 (mean ± SD). Haplogroup assignment was performed with an in-house script that matched our SNPs with the classification provided in ISOGG (ISOGG) version 10.08, see S12.

[4]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. 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.

Focal Pop	Nearby Gene[a]	Position[b]	rsID	Dist[c]	*PBSn1*[d]	F_{12}[e]	F_{13}	F_{23}	Function of gene product
All	*TMEM86B*	55833076	rs734517	92444	0.7754	0.9309	0.9913	0.0629	Catalyzes the degradation of lysoplasmalogen. Modulates cell membrane proteins.[1]
All	*LRRC52*	165621695	rs4147601	88510	0.7372	0.9615	0.9136	0.0145	Modulates voltage of potassium ion channels. Expressed in testis.[2]
All	*MACROD2*	15209684	rs175279	901	0.7046	0.9195	0.8893	-0.0089	Involved in deacetylase activity.[3,4] Possibly (but not conclusively) causative of Kabuki syndrome.[5]
All	*JRKL*	96747146	rs72959058	507105	0.7390	0.9904	0.8740	0.1466	Homologue to "jerky" gene in mouse.[6]
All	*SPATA20*	48631324	rs73338243	287	0.6969	0.9619	0.8486	0.0946	Spermatid protein.[7]
All	*NAA60*	3537933	rs73503305	970	0.7084	0.9089	0.9060	-0.0169	Histone acetyltransferase required for nucleosome assembly and chromosome segregation during anaphase.[8] Human-specific imprinted gene[9].
All	*CBLN2*	70019066	rs12455116	184848	0.6914	0.9155	0.8669	0.0013	*CBLN2*: cerebellum-specific protein involved in various signaling pathways.[13,14] Possibly associated with pulmonary arterial hypertension.[15]
	NETO1			390482					*NETO1*: brain-specific transmembrane protein involved in the regulation of neuronal circuitry.[10,11] Associated with thyroid function.[12]
All	*SLC2A12*	134391056	rs4896021	17267	0.7574	0.9604	0.9461	-0.0132	Catalyzes sugar absorption[16]. Involved in the pathogenesis of diabetes.[17] Associated with serum urate levels.[18]
All	*LOC101927657*	127358509	rs145200081	16731	0.6522	0.9404	0.7965	0.1302	Unknown (ncRNA).
All	*LOC102724612*	64466486	rs113341339	78446	0.7338	0.9144	0.9538	0.0032	Unknown (ncRNA).
NE	*ZBTB20*	114530679	rs9289004	10658	0.5532	0.6478	0.8172	0.0737	Transcriptional repressor[19] associated with Primrose syndrome.[20]
NE	*ANXA10*	168646016	rs2176513	367671	0.4909	0.6096	0.6125	-0.0136	Calcium-dependent phospholipid-binding annexin.[21]
NE	*TRPC3*	122905041	rs4502701	32132	0.4950	0.5931	0.6408	-0.0129	Non-selective cation channel[22], associated with spinocerebellar ataxia.[23]
NE	*HS3ST1*	11634592	rs7665516	204055	0.4522	0.4526	0.7147	0.0748	Regulates rate of generation of anticoagulant heparan sulfate proteoglycan.[24]
NE	*MIR548C*	65027511	rs2620721	11126	0.5044	0.5506	0.7335	0.0281	Unknown (microRNA).
NE	*STARD13*	33799901	rs7318080	19714	0.4866	0.5430	0.8299	0.2037	Involved in cell proliferation[25] and fibroblast morphology.[26]
NE	*AKAP11*	42931386	rs7319267	33983	0.5314	0.5594	0.8519	0.1317	Directs protein kinase A activity[27] and is involved in cAMP messenger signaling.[28]
NE	*AGMO*	15212231	rs35557899	27711	0.4730	0.5063	0.6770	0.0146	Catalyzes the cleavage of O-alkyl bonds of ether lipids.[29]
NE	*RUNX1T1*	92925296	rs11776341	41898	0.4450	0.5583	0.5415	-0.0026	Involved in transcriptional repression[30]. A translocation involving this gene is associated with acute myeloid leukemia.[31,32]
NE	*FHAD1*	15680451	rs2473358	971	0.4504	0.5966	0.5195	0.0006	Unknown.
SW	*KCNJ2*	68190552	rs35167900	14369	0.5746	0.6093	0.9325	0.2228	Potassium channel[33], associated with familial atrial fibrillation[34] and periodic paralysis.[35-37]
SW	*TACC2*	123754065	rs10159998	5062	0.5035	0.6012	0.6680	-0.0021	Belongs to a family of proteins that interact with the centrosome and microtubules, and that are implicated in cancer.[38]
SW	*LOC101928708*	87228164	rs4843556	17556	0.5788	0.6540	0.8597	0.0713	Unknown (ncRNA).
SW	*C16orf82*	27187689	rs72782349	107202	0.5068	0.6041	0.6925	0.0163	Unknown.
SW	*LOC100507391*	194520805	rs56379930	17908	0.5537	0.6550	0.7515	-0.0072	Unknown (ncRNA).
SW	*HAUS4*	23416252	rs2008951	127	0.4937	0.4984	0.8330	0.1643	A component of a microtubule-binding complex that plays a role in the generation of microtubules in the mitotic spindle.[39,40]
SW	*KNG1*	186438819	rs5029990	815	0.5082	0.5638	0.7172	0.0085	During the inflammatory response, it is involved in vasodilation, coagulation, enhanced capillary permeability and pain induction.[41-43]
SW	*MYDGF*	4657016	rs66891175	540	0.5541	0.6148	0.8808	0.1570	Unknown.
SW	*MSMP*	35757075	rs1951432	2801	0.4822	0.4672	0.8793	0.2652	May be involved in the tumorigenesis of prostate cancer.[44]
SW	*VAV2*	136756316	rs2519771	29762	0.4731	0.5074	0.7267	0.0721	Member of an oncogene family.[45] Involved in T-cell receptor signaling.[46]

a – RefSeq protein coding gene with exon boundary near to windowed-*PBSn1* peak.
b – Genomic position (hg19) of SNP with highest value of *PBSn1* within 200 Mb of the top window.
c– Distance between SNP and the nearest exon boundary of nearest gene.
d– *PBSn1* statistic for top SNP
e – F_{ST}statistics at top SNP for each comparison within the *PBSn1* calculation

References Extended Data Table 2

1 Wu, L. C. *et al.* Purification, identification, and cloning of lysoplasmalogenase, the enzyme that catalyzes hydrolysis of the vinyl ether bond of lysoplasmalogen. *J Biol Chem***286**, 24916-24930, doi:10.1074/jbc.M111.247163 (2011).

2 Yan, J. & Aldrich, R. W. BK potassium channel modulation by leucine-rich repeat-containing proteins. *Proc Natl Acad Sci U S A***109**, 7917-7922, doi:10.1073/pnas.1205435109 (2012).
3 Chen, D. *et al.* Identification of macrodomain proteins as novel O-acetyl-ADP-ribose deacetylases. *J Biol Chem***286**, 13261-13271, doi:10.1074/jbc.M110.206771 (2011).
4 Jankevicius, G. *et al.* A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nat Struct Mol Biol***20**, 508-514, doi:10.1038/nsmb.2523 (2013).
5 Maas, N. M. *et al.* The C20orf133 gene is disrupted in a patient with Kabuki syndrome. *J Med Genet***44**, 562-569, doi:10.1136/jmg.2007.049510 (2007).
6 Zeng, Z. *et al.* Cloning, mapping, and tissue distribution of a human homologue of the mouse jerky gene product. *Biochem Biophys Res Commun***236**, 389-395, doi:10.1006/bbrc.1997.6935 (1997).
7 Shi, H. J. *et al.* Cloning and characterization of rat spermatid protein SSP411: a thioredoxin-like protein. *J Androl***25**, 479-493 (2004).
8 Van Damme, P. *et al.* NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet***7**, e1002169, doi:10.1371/journal.pgen.1002169 (2011).
9 Nakabayashi, K. *et al.* Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. *Hum Mol Genet***20**, 3188-3197, doi:10.1093/hmg/ddr224 (2011).
10 Stöhr, H., Berger, C., Fröhlich, S. & Weber, B. H. A novel gene encoding a putative transmembrane protein with two extracellular CUB domains and a low-density lipoprotein class A module: isolation of alternatively spliced isoforms in retina and brain. *Gene***286**, 223-231 (2002).
11 Michishita, M. *et al.* A novel gene, Btcl1, encoding CUB and LDLa domains is expressed in restricted areas of mouse brain. *Biochem Biophys Res Commun***306**, 680-686 (2003).
12 Porcu, E. *et al.* A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet***9**, e1003266, doi:10.1371/journal.pgen.1003266 (2013).
13 Kavety, B., Jenkins, N. A., Fletcher, C. F., Copeland, N. G. & Morgan, J. I. Genomic structure and mapping of precerebellin and a precerebellin-related gene. *Brain Res Mol Brain Res***27**, 152-156 (1994).
14 Wei, P. *et al.* The Cbln family of proteins interact with multiple signaling pathways. *J Neurochem***121**, 717-729, doi:10.1111/j.1471-4159.2012.07648.x (2012).
15 Germain, M. *et al.* Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. *Nat Genet***45**, 518-521, doi:10.1038/ng.2581 (2013).
16 Rogers, S. *et al.* Identification of a novel glucose transporter-like protein-GLUT-12. *Am J Physiol Endocrinol Metab***282**, E733-738 (2002).
17 Linden, K. C. *et al.* Renal expression and localization of the facilitative glucose transporters GLUT1 and GLUT12 in animal models of hypertension and diabetic nephropathy. *Am J Physiol Renal Physiol***290**, F205-213, doi:10.1152/ajprenal.00237.2004 (2006).
18 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, doi:10.1093/hmg/ddr307 (2011).
19 Sutherland, A. P. *et al.* Zinc finger protein Zbtb20 is essential for postnatal survival and glucose homeostasis. *Mol Cell Biol***29**, 2804-2815, doi:10.1128/MCB.01667-08 (2009).
20 Cordeddu, V. *et al.* Mutations in ZBTB20 cause Primrose syndrome. *Nat Genet***46**, 815-817, doi:10.1038/ng.3035 (2014).
21 Morgan, R. O. *et al.* Novel human and mouse annexin A10 are linked to the genome duplications during early chordate evolution. *Genomics***60**, 40-49, doi:10.1006/geno.1999.5895 (1999).
22 Becker, E. B. *et al.* Candidate screening of the TRPC3 gene in cerebellar ataxia. *Cerebellum***10**, 296-299, doi:10.1007/s12311-011-0253-6 (2011).
23 Fogel, B. L., Hanson, S. M. & Becker, E. B. Do mutations in the murine ataxia gene TRPC3 cause cerebellar ataxia in humans? *Mov Disord***30**, 284-286, doi:10.1002/mds.26096 (2015).
24 Shworak, N. W. *et al.* Molecular cloning and expression of mouse and human cDNAs encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferase. *J Biol Chem***272**, 28008-28019 (1997).
25 Ching, Y. P. *et al.* Deleted in liver cancer (DLC) 2 encodes a RhoGAP protein with growth suppressor function and is underexpressed in hepatocellular carcinoma. *J Biol Chem***278**, 10824-10830, doi:10.1074/jbc.M208310200 (2003).
26 Leung, T. H. *et al.* Deleted in liver cancer 2 (DLC2) suppresses cell transformation by means of inhibition of RhoA activity. *Proc Natl Acad Sci U S A***102**, 15207-15212, doi:10.1073/pnas.0504501102 (2005).
27 Lester, L. B., Coghlan, V. M., Nauert, B. & Scott, J. D. Cloning and characterization of a novel A-kinase anchoring protein. AKAP 220, association with testicular peroxisomes. *J Biol Chem***271**, 9460-9465 (1996).
28 Logue, J. S. *et al.* AKAP220 protein organizes signaling elements that impact cell migration. *J Biol Chem***286**, 39269-39281, doi:10.1074/jbc.M111.277756 (2011).
29 Watschinger, K. *et al.* Identification of the gene encoding alkylglycerol monooxygenase defines a third class of tetrahydrobiopterin-dependent enzymes. *Proc Natl Acad Sci U S A***107**, 13672-13677, doi:10.1073/pnas.1002404107 (2010).
30 Kumar, R. *et al.* CBFA2T3-ZNF652 corepressor complex regulates transcription of the E-box gene HEB. *J Biol Chem***283**, 19026-19038, doi:10.1074/jbc.M709136200 (2008).

31 Erickson, P. *et al.* Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood***80**, 1825-1831 (1992).
32 Miyoshi, H. *et al.* The t(8;21) translocation in acute myeloid leukemia results in production of an AML1-MTG8 fusion transcript. *EMBO J***12**, 2715-2721 (1993).
33 Raab-Graham, K. F., Radeke, C. M. & Vandenberg, C. A. Molecular cloning and expression of a human heart inward rectifier potassium channel. *Neuroreport***5**, 2501-2505 (1994).
34 Xia, M. *et al.* A Kir2.1 gain-of-function mutation underlies familial atrial fibrillation. *Biochem Biophys Res Commun***332**, 1012-1019, doi:10.1016/j.bbrc.2005.05.054 (2005).
35 Tristani-Firouzi, M. *et al.* Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome). *J Clin Invest***110**, 381-388, doi:10.1172/JCI15183 (2002).
36 Plaster, N. M. *et al.* Mutations in Kir2.1 cause the developmental and episodic electrical phenotypes of Andersen's syndrome. *Cell***105**, 511-519 (2001).
37 Cheung, C. L. *et al.* Genome-wide association study identifies a susceptibility locus for thyrotoxic periodic paralysis at 17q24.3. *Nat Genet***44**, 1026-1029, doi:10.1038/ng.2367 (2012).
38 Gangisetty, O., Lauffart, B., Sondarva, G. V., Chelsea, D. M. & Still, I. H. The transforming acidic coiled coil proteins interact with nuclear histone acetyltransferases. *Oncogene***23**, 2559-2563, doi:10.1038/sj.onc.1207424 (2004).
39 Goshima, G., Mayer, M., Zhang, N., Stuurman, N. & Vale, R. D. Augmin: a protein complex required for centrosome-independent microtubule generation within the spindle. *J Cell Biol***181**, 421-429, doi:10.1083/jcb.200711053 (2008).
40 Uehara, R. *et al.* The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells. *Proc Natl Acad Sci U S A***106**, 6998-7003, doi:10.1073/pnas.0901587106 (2009).
41 Cheung, P. P., Kunapuli, S. P., Scott, C. F., Wachtfogel, Y. T. & Colman, R. W. Genetic basis of total kininogen deficiency in Williams' trait. *J Biol Chem***268**, 23361-23365 (1993).
42 Merkulov, S. *et al.* Deletion of murine kininogen gene 1 (mKng1) causes loss of plasma kininogen and delays thrombosis. *Blood***111**, 1274-1281, doi:10.1182/blood-2007-06-092338 (2008).
43 Takagaki, Y., Kitamura, N. & Nakanishi, S. Cloning and sequence analysis of cDNAs for human high molecular weight and low molecular weight prekininogens. Primary structures of two human prekininogens. *J Biol Chem***260**, 8601-8609 (1985).
44 Valtonen-André, C. *et al.* A highly conserved protein secreted by the prostate cancer cell line PC-3 is expressed in benign and malignant prostate tissue. *Biol Chem***388**, 289-295, doi:10.1515/BC.2007.032 (2007).
45 Henske, E. P. *et al.* Identification of VAV2 on 9q34 and its exclusion as the tuberous sclerosis gene TSC1. *Ann Hum Genet***59**, 25-37 (1995).
46 Moores, S. L. *et al.* Vav family proteins couple to diverse cell surface receptors. *Mol Cell Biol***20**, 6364-6373 (2000).

References for Figures and Tables (except Extended Data Table 2)

Allen J, O'Connell JF. 2014. Both half right: Updating the evidence for dating first human arrivals in Sahul. Aust. Archaeol.:86.

Horton D. 1994. The Encyclopedia of Aboriginal Australia. Horton, D. Canberra: Australian Institute of Aboriginal and torres Strait Islander Studies

Bowern C, Atkinson Q. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. Language 88:817–845.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

ISOGG. Y-DNA Haplogroup Tree 2015. Available from: http://www.isogg.org/tree/

Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L, Yang Y, Li S, Tang K, Chen H, et al. 2013. Modeling Recent Human Evolution in Mice by Expression of a Selected EDAR Variant. Cell 152:691–702.

Kayser M, Liu F, Janssens ACJW, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van IJcken WFJ, et al. 2008. Three Genome-wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene. Am. J. Hum. Genet. 82:411–423.

Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. 32:25–32.

Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, et al. 2005. SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans. Science 310:1782–1786.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413.

Matsumoto I, Ohmoto M, Narukawa M, Yoshihara Y, Abe K. 2011. Skn-1a (Pou2f3) specifies taste receptor cell lineage. Nat. Neurosci. 14:685–687.

Ohashi J, Naka I, Tsuchiya N. 2011. The Impact of Natural Selection on an ABCC11 SNP Determining Earwax Type. Mol. Biol. Evol. 28:849–857.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30:E386— – E394.

Pan Y, Luo X, Liu X, Wu L-Y, Zhang Q, Wang L, Wang W, Zuo L, Wang K-S. 2013. Genome-wide association studies of maximum number of drinks. J. Psychiatr. Res. 47:1717–1724.

Pugach I, Delfin F, Gunnarsdóttir E, Kayser M, Stoneking M. 2013. Genome-wide data substantiate Holocene gene flow from India to Australia. Proc. Natl. Acad. Sci. U. S. A. 110:1803–1808.

Qin P, Stoneking M. 2015. Denisovan Ancestry in East Eurasian and Native American Populations. Mol. Biol. Evol.

Ravni A, Qu Y, Goffinet AM, Tissir F. 2009. Planar Cell Polarity Cadherin Celsr1 Regulates Skin Hair Patterning in the Mouse. J. Invest. Dermatol. 129:2507–2509.

Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM-S, Ko Y-C, Jinam TA, Phipps ME, et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am. J. Hum. Genet. 89:516–528.

Russell MR, Penikis A, Oldridge DA, Alvarez-Dominguez JR, McDaniel L, Diamond M, Padovan O, Raman P, Li Y, Wei JS, et al. 2015. CASC15-S Is a Tumor Suppressor lncRNA at the 6p22 Neuroblastoma Susceptibility Locus. Cancer Res. 75:3155–3166.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Sano H, Roach WG, Peck GR, Fukuda M, Lienhard GE. 2008. Rab10 in insulin-stimulated GLUT4 translocation. Biochem. J. 411:89–95.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nat. Genet. 46:919–925.

Soejima M, Tachida H, Ishida T, Sano A, Koda Y. 2006. Evidence for Recent Positive Selection at the Human AIM1 Locus in a European Population. Mol. Biol. Evol. 23:179–188.

Vianello D, Sevini F, Castellani G, Lomartire L, Capri M, Franceschi C. 2013. HAPLOFIND: A New Method for High-Throughput mtDNA Haplogroup Assignment. Hum. Mutat. 34:1189–1194.

Wang C, Szpiech ZA, Degnan JH, Jakobsson M, Pemberton TJ, Hardy JA, Singleton AB, Rosenberg NA. 2010. Comparing spatial maps of human population-genetic variation using Procrustes analysis. Stat. Appl. Genet. Mol. Biol. 9:Article 13.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38:1358–1370.