Whole-Genome Genetic Diversity in a Sample of Australians with Deep Aboriginal Ancestry

Brian P. McEvoy,[1,][*] Joanne M. Lind,[2] Eric T. Wang,[3] Robert K. Moyzis,[4] Peter M. Visscher,[1] Sheila M. van Holst Pellekaan,[5,][*] and Alan N. Wilton[5]

Australia was probably settled soon after modern humans left Africa, but details of this ancient migration are not well understood. Debate centers on whether the Pleistocene Sahul continent (composed of New Guinea, Australia, and Tasmania) was first settled by a single wave followed by regional divergence into Aboriginal Australian and New Guinean populations (common origin) or whether different parts of the continent were initially populated independently. Australia has been the subject of relatively few DNA studies even though understanding regional variation in genomic structure and diversity will be important if disease-association mapping methods are to be successfully evaluated and applied across populations. We report on a genome-wide investigation of Australian Aboriginal SNP diversity in a sample of participants from the Riverine region. The phylogenetic relationship of these Aboriginal Australians to a range of other global populations demonstrates a deep common origin with Papuan New Guineans and Melanesians, with little evidence of substantial later migration until the very recent arrival of European colonists. The study provides valuable and robust insights into an early and important phase of human colonization of the globe. A broader survey of Australia, including diverse geographic sample populations, will be required to fully appreciate the continent's unique population history and consequent genetic heritage, as well as the importance of both to the understanding of health issues.

There is strong fossil and genetic evidence that modern humans arose in Africa ~200,000 years ago, with a subset departing the continent much later (~40,000–80,000 years ago) to populate the rest of the world.[1] Mitochondrial DNA (mtDNA) suggests that these migrants exited Africa by the "southern route," across the Red Sea to Arabia, moving relatively rapidly along the coast to reach Southeast Asia and Australia.[2] Indeed, despite its distance from Africa, Australia has some of the earliest reliable evidence of human habitation outside Africa, dating to at least ~46,000 and probably ~60,000 years ago.[3–5] Archeological evidence suggests that New Guinea and Melanesia, the islands immediately north and northeast of Australia, collectively termed Near Oceania, were also settled by ~40,000 years ago.[6] During this late Pleistocene period, sea levels were lower and the first humans entered the region when present day Australia, Tasmania, and New Guinea were part of a single landmass known as the Sahul. However, details of dispersal routes and timing of the settlement remain debated. The common origin hypothesis proposes a single major migration from Eurasia to the Sahul followed by divergence into separate geographic populations. The independent origin model, by contrast, posits a multi-wave early settlement of the Sahul with largely independent migrations to present-day New Guinea and mainland Australia. There is also debate around whether the first settlers were followed by later waves of migrants.

Advances in genotyping technology allow variation along the entire genome to be simultaneously interrogated and have revolutionized the study of human genetic diversity, providing new insights into population history and facilitating gene discovery by genome-wide association studies (GWASs).[7] However, despite its early and unique place in human global colonization, there has been relatively little investigation of Aboriginal Australian genetic diversity. This is partially due to unease felt by some Aboriginal Australians about genetic research because of the legacy of past research experiences.[8] A limited number of previous studies have generally focused on traditional Y chromosome and mtDNA markers and none has surveyed whole genome diversity. GWAS genotyping and analytical approaches are typically geared toward populations of European origin and focused on diseases and conditions that are prevalent in these people. In order to extend gene discovery studies, and their potential longer-term health benefits, to Aboriginal Australians, it will first be necessary to gain a fuller understanding of Australia's population history and the present genetic legacy of that past.[9,10]

According to genealogical information, the Australians involved in this study have assured maternal Aboriginal ancestry with some Aboriginal, European (Scottish), and other non-Aboriginal paternal connections. They come from the Riverine area of western New South Wales and are a subset of those described in previous investigations of mtDNA variation.[11,12] We use the abbreviation AuR for the sample population throughout this report to be consistent with, and for the reasons described in, these publications. Although a single sample population will not necessarily be representative of an entire continent, the Riverine region is especially significant because it

p < 0.05 (22,300 SNPs). The average missing data per individual across the remaining 824,886 SNPs was 0.2%. Most analysis was restricted to 160,337 SNPs (155,166 autosomal and 5,171 X chromosome) that were genotyped in common between the eleven HapMap3 (n = 988) and 51 Human Genome Diversity Panel (HGDP; n = 940) populations[14] (see Table S1 and Table S2, available online, for details).

We first explored the overall relationship of the AuR sample to the HGDP populations. Importantly, the HGDP panel included two samples from Oceania: Papuans (PAP; from New Guinea) and island Melanesians (MEL; from Bougainville), populations thought to be descendents of the region's first inhabitants. The matrix of pairwise interpopulation genetic distances (F_{ST} values) was used to construct a neighbor-joining (NJ) phylogenetic tree that summarizes the relationship of the 52 populations to each other (Figure 1A). The tree divides the populations into five broad groups: African, East Asian, West Eurasian (European, Middle Eastern, and Central and South Asian populations), American, and Oceanic. Although the latter branch contains the MEL, PAP, and AuR groups, AuR show a shorter branch length than the others, placing them closer to the trunk of the tree than the other Oceanic populations. This could be the result of greater genetic drift in the MEL and PAP or admixture of the AuR with populations elsewhere on the tree.

Principal component analysis (PCA) of the same data, which examines individual rather than population-level genetic affinities, favors AuR sample admixture. The most important trends (PC1 and PC2) are sufficient to largely differentiate major continental groups (Figure 2A). Some AuR individuals are close to the Oceanic cluster, composed of MEL and PAP individuals, but most occupy a wide range on PC2 between Europeans and East Asians, generally falling in an area occupied by Central and South Asian

includes the Willandra Lakes and Lake Mungo, where some of the earliest Australian human remains have been found.[3] Personal contact and ongoing negotiation with the participants has been carried out by one of the authors (S.M.v.H.P.) over the past 18 years. Thirty-eight participants (30 female and eight male) gave informed consent to further genetic study. Community agreement has been granted from Maari Ma Health Aboriginal Corporation, and ethical approval has been given by the Aboriginal Health and Medical Research Council Ethics Committee and the University of New South Wales Human Research Ethics Committee. Local and regional communities have been consulted with regard to this report.

Samples were genotyped for ~907,000 SNPs on the Affymetrix Genome-Wide Human SNP Array 6.0 platform with standard protocols (Affymetrix). Genotypes were called from the raw intensity (.cel) files with the Birdsuite software[13] and a confidence threshold score of 0.1. We subsequently conducted additional quality control on the data by excluding any SNP with a missing genotype rate greater than 5% (66,251 SNPs) and those that were out of Hardy Weinberg equilibrium (HWE) applying a stringent cutoff
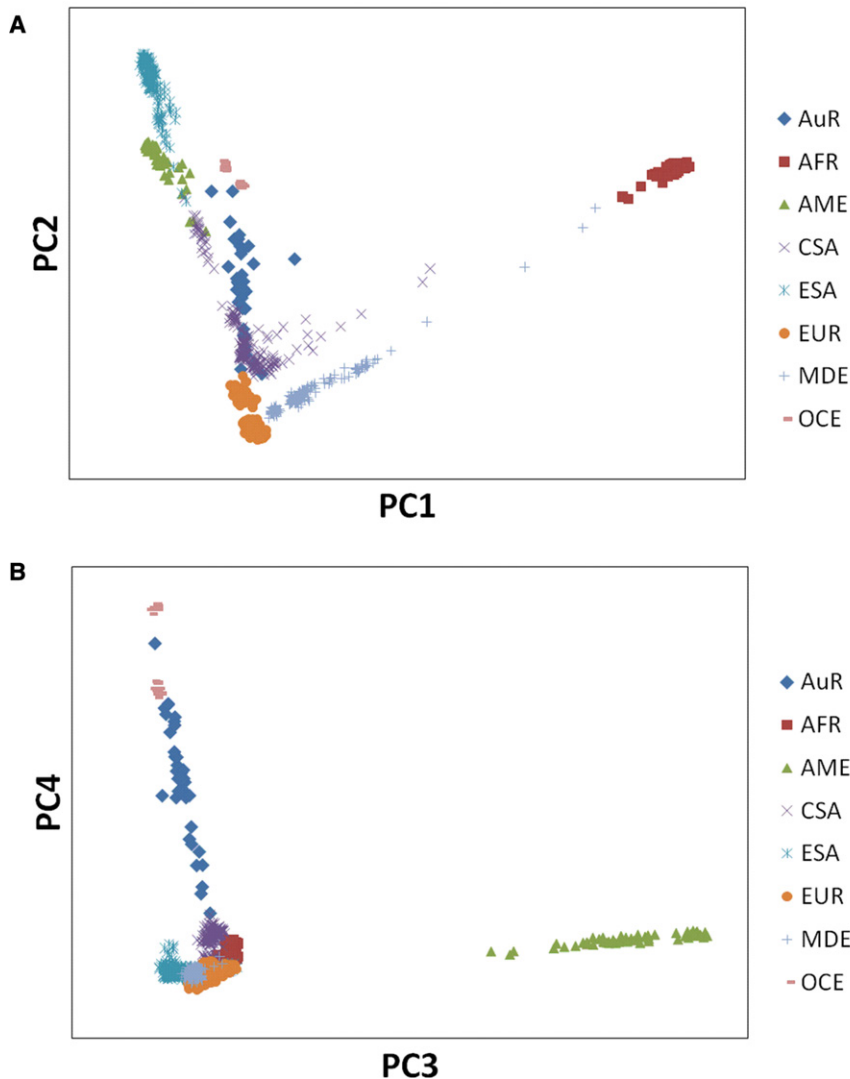
Figure 2. Principal-Component Analysis
(A) PC1 versus PC2 and (B) PC3 versus PC4 derived from the 51 HGDP populations and the AuR sample. HGDP populations have been grouped into seven broader regions: Africa (AFR), America (AME), Central and South Asia (CSA), East Asia (ESA), Europe (EUR), Middle East (MDE), and Oceania (OCE). See Table S2 for further population details. PCA was conducted with EIGENSTRAT[33] (version 2).

to a different subset. Given that results from a series of exploratory runs using the different sets were highly correlated, we focused on just one for our main series of $K = 2$ to $K = 8$ STRUCTURE analyses.

Results from *frappe* and STRUCTURE were highly consistent with each other. At $K = 5$ (Figure 3) there is clear separation of individuals according to the major branches identified in the NJ tree (Figure 1). The Melanesians and Papuans are almost completely assigned to a single Oceanic cluster except for a minority East Asian ancestry component in the Melanesians, probably representing introgression from the Holocene era Austronesian expansion.[18] The AuR sample is a clear mixture of two clusters corresponding to Oceanic ancestry and the majority ancestry component of Western Eurasian populations. Recent admixture with Europeans, who began settling the continent in 1788, is an obvious source for this Western Eurasian component and would be consistent with known genealogical information. However, contact with India earlier in the Holocene has also been proposed on the basis of mtDNA[19] and Y chromosome[20] data. At $K > 5$, further population distinction emerges in the Western Eurasian cluster, with gradual separation of European from Central and South Asian populations (Figure S1). Although the distinction is never complete, the non-Oceanic component in the AuR sample is most consistent with European ancestry. Such a conclusion is anecdotally supported by the presence of evolutionarily recent alleles (for example, the blue-eye-associated allele of the rs12913832 SNP, near the *OCA2* gene[21] [MIM 611409] and the red-hair-associated allele of rs1805007 in the *MC1R* gene[22] [MIM 155555]) that are essentially restricted to (primarily northern) Europeans.

We attempted to quantify individual ancestry more accurately by conducting further STRUCTURE analysis under a model of $K = 2$, assuming that the AuR sample is a mixture of an ancient ancestral Aboriginal population

populations. PC3 separates Amerindians from the other populations. PC4 does the same for the Oceania populations, with MEL and PAP individuals clustered at one pole and all other populations at the opposite extreme (Figure 2B). AuR individuals fall in a broad range between these extremes, supporting substantial admixture from a non-Oceanic source.

To investigate the admixture further, we explored the population structure, without initially considering known geographic origin, by using the *frappe*[15] and STRUCTURE[16,17] methods, which infer population clusters from the genotypes alone and determine the fractional ancestry of each individual derived from these clusters. Both require the prior specification of the number of populations or clusters (K) into which individuals are to be divided. We carried out a series of runs for $K = 2$ to $K = 8$. The *frappe* analysis used the entire set of 155,166 autosomal SNPs. However, because STRUCTURE requires approximate linkage equilibrium between markers and is computationally slow with large numbers of markers, we divided the 155,166 autosomal SNPs into ten equal sets of ~15,500 by assigning every tenth SNP, by order along the genome,
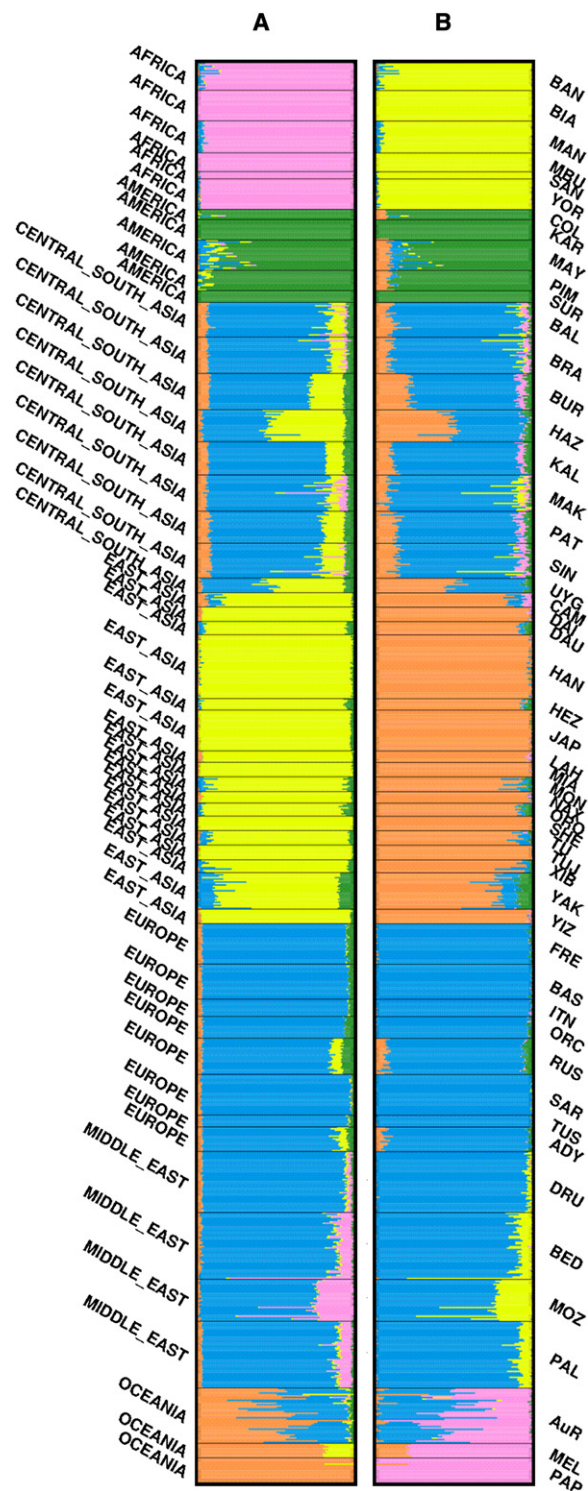
Figure 3. Population Structure Analysis
Individual ancestry proportions in the HGDP and Aboriginal Australian (AuR) samples at $K = 5$, from (A) *frappe* analysis, with all 155,166 autosomal markers, and (B) STRUCTURE analysis, with a one-tenth subset (15,516) of all autosomal SNPs. Each horizontal line represents an individual and is divided into K (number of population clusters) colored segments reflecting the estimated ancestry proportion from each cluster. Different geographic samples are divided by black lines with population and region indicated to the right and left of the plot, respectively. See Table S2 for a full explanation of population codes. *frappe* analysis

and recent European settlers. We therefore only included the AuR sample and European individuals from HapMap3 (TSI and CEU, n = 200, see Table S1) and allowed STRUCTURE to incorporate the origin information of the Europeans in the analysis, effectively rendering them a training set for one parental population. (We used HapMap3 as a discovery data set in this analysis so that we could apply findings downstream in the test HGDP.) STRUCTURE runs were carried out for each of the ten autosomal SNP subsets. Ancestry estimates are highly consistent across all SNP sets, so we averaged across all runs to obtain a single European versus Australian ancestry fraction for each individual. The Australian component ranged from 28% to 100% with an average of 64% (Figure 4). As a result of the strong ascertainment bias of HapMap and Affymetrix SNPs, however, which were largely identified from European ancestry populations, the Aboriginal component might be underestimated.

We also explored the admixture process by examining Y chromosome and mtDNA variation. Previous analysis of AuR mtDNA[11,12] showed that 37 individuals are likely to have ancient Australian maternal origin with deep-rooting mtDNA lineages (haplogroups M42a and b, P4b and S1a), with one mtDNA of probable European origin (haplogroup U5). The 257 Y chromosome SNPs successfully genotyped here were not sufficient to infer in full detail the haplogroup of the eight AuR Y chromosomes.[23] However, five of these Y chromosomes clearly fall into European haplogroups (R1b1 and I),[24] and two are in haplogroup C, one of whose subgroups (C4) is common in Australians.[25,26] The final Y chromosome is most likely haplogroup M (but can only be formally assigned KxPxNO), which is also found in Australians.[27] From these observations (37.5% male and 97.5% female Aboriginal Australian ancestors), the expected autosomal ancestry fraction is ~67%, close to the observed value of 64%. Because each X chromosome spends two-thirds of its history in females, the Australian admixture fraction is expected to be higher when X chromosome markers are used. A value of 70% is observed, close to the ~77% expected from the male/female ancestral bias inferred from Y chromosome and mtDNA markers.

For learning more about the pre-1788 history of the Australian population, a sample with assured Aboriginal paternal and maternal ancestry would obviously be ideal. However, we can use STRUCTURE estimates of parental cluster allele frequencies, which are modeled jointly with ancestry proportions, as an approximation of those in the ancestral population prior to admixture. The STRUCTURE-reconstructed Aboriginal Australian Riverine (AuR*) allele frequencies, along with those observed in the HGDP samples, were therefore used to recalculate pairwise

used 5,000 expectation-maximization (EM) iterations whereas STRUCTURE runs were conducted under the admixture model with a 25,000 replicate burn-in followed by 25,000 Markov chain Monte Carlo (MCMC) iterations.
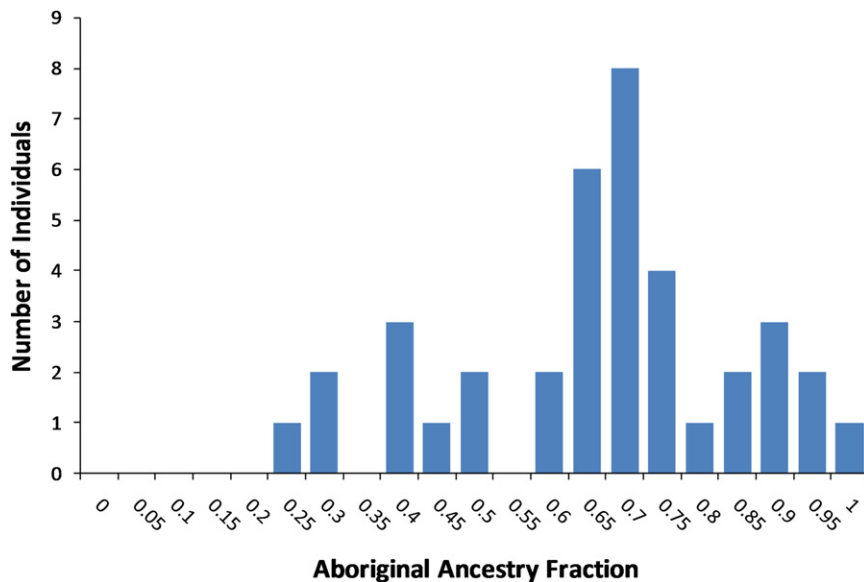
Figure 4. Distribution of Individual Aboriginal Australian Genomic Ancestry Estimates

Values are the fraction of autosomal genomic ancestry assigned to the non-European cluster in STRUCTURE using the 38 AuR individuals, 200 HapMap3 Europeans, and a $K = 2$. STRUCTURE runs consisted of a 25,000 replicate burn-in followed by 25,000 MCMC iterations.

F_{ST} values. We assumed an Australian sample size two-thirds (n = 25) that of the original sample collection in these F_{ST} calculations, in line with admixture proportions. (Note: the HGDP were not used in STRUCTURE runs that estimated the AuR* allele frequencies.) A NJ phylogenetic tree, derived from the interpopulation F_{ST} matrix, again shows five major continental branches (Figure 1B). However, the AuR* population now groups more tightly with the Melanesians and Papuans further toward the terminus of the Oceanic branch. The phylogeny supports the common origin of the earliest indigenous inhabitants of Oceania, followed first by the divergence of Australia from Near Oceania and then by the subsequent split of the latter into Papuan and Melanesian populations.

We also used the reconstructed allele frequencies to explore loci that are particularly differentiated between this sample and the rest of the world by calculating the F_{ST} between AuR* and the entire HapMap3 collection for each autosomal SNP (Figure 5A and 5B). The most highly differentiated SNP (rs12458349) is on chromosome 18q21.33 approximately 30 Kb upstream of the *PHLPP* gene (MIM 609396). Two further SNPs in the same 0.5 Mb region are also in the top 25 most differentiated SNPs. The derived allele of rs12458349 is at its highest frequency in the reconstructed AuR* (61%) and observed AuR (45%) Australian samples (Figure 5C). It occurs at a relatively high frequency in the Papuan sample (30%) but is rare or absent elsewhere in the world. Interestingly, the region has been repeatedly implicated, via linkage and association, in diabetic nephropathy,[28–30] which is one of the principal causes of chronic kidney disease (CKD).[31] Indigenous Australians have ~10-fold-higher rates of CKD than nonindigenous Australians.[32] It is possible that variants in one or more of the genes in this region also contribute to the increased risk of CKD in Aboriginal Australians. The observed differentiation in the 18q21.33 region may simply be the consequence of stochastic sampling error and/or genetic drift. Genetic drift, or random changes in allele frequencies, is expected to be a major force in a population, like Aboriginal Australia, that has been relatively small and/or isolated for a long period of time. Differentiation could also be explained by natural selection. The presence of several highly differentiated SNPs in the region, spanning nearly 0.5 Mb in length, hints at the presence of a long common haplotype that might be indicative of genetic hitch-hiking and recent positive selection. However, it is difficult to distinguish between possible explanations because the sample is small, with extensive admixture hampering phasing and direct investigation of linkage-disequilibrium-based selection signals.

Whereas the admixture present in the AuR sample presents a potential challenge in conducting traditional association methods for disease gene discovery,[33] it opens the possibility of using admixture mapping.[34] Admixture mapping is most suitable for traits, like CKD, that differ in frequency between the two parental populations of an admixture group. The approach essentially looks for genomic regions with an excess of higher-risk population ancestry relative to other regions or controls.[34] A set of markers, spread across the genome, that are highly informative as to ancestry (ancestry informative markers or AIMs) is an essential requirement for admixture mapping. As an exploratory exercise, we identified a set of 100 Oceanic AIMs, from the top 200 by F_{ST} between the AuR* and the full HapMap3 population, such that no AIM was within 2 Mb of another (see Table S3 for a full list). With this marker set and STRUCTURE, under $K = 2$, it was possible to distinguish Oceanic versus non-Oceanic ancestry in the independent HGDP data set with a high degree of accuracy (as judged against earlier *frappe* and STRUCTURE results obtained with more markers) (Figure 6).

Our study of whole-genome diversity in a sample of participants with deep Aboriginal ancestry adds to the genetic, archaeological, anthropological, and linguistic evidence that Australia has had a long, rich, and unique population history. Results from the relatively small number of previous genetic diversity studies, typically using single loci, could not definitively resolve between models of initial Sahul settlement. For example, autosomal
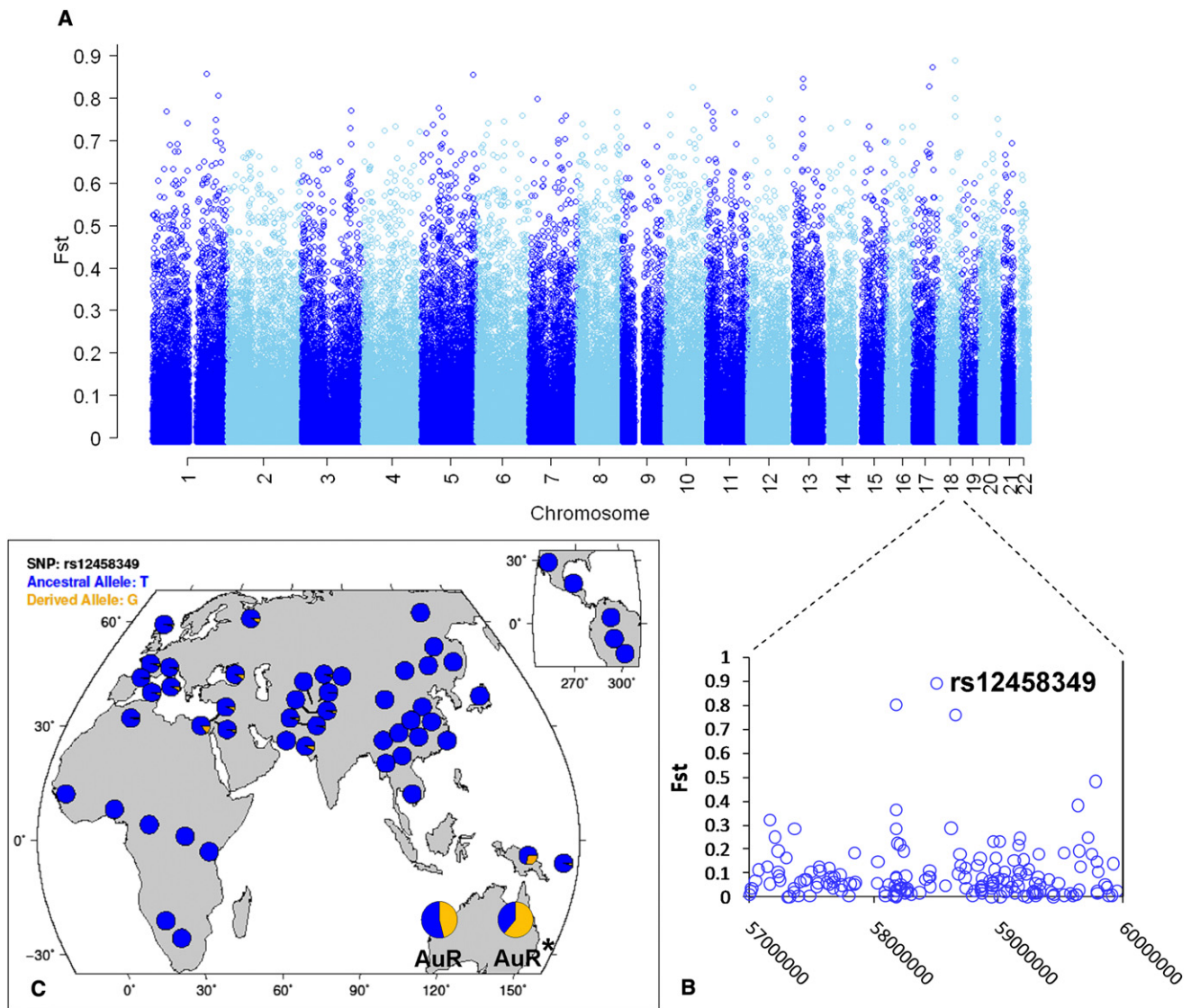
Figure 5. Genomic and Geographic Distribution of Highly Differentiated SNPs
(A) SNP F_{ST} values, calculated between the full HapMap3 sample and the reconstructed Aboriginal Australian (AuR*) allele frequencies, plotted against genomic location.
(B) Detail of the chromosome 18q21.33 region (NCBI-36 coordinates 57,000,000 to 60,000,000) surrounding the highest observed F_{ST} at rs12458349.
(C) Geographic distribution of rs12458349 allele frequencies in the HGDP, observed AuR and reconstructed AuR* population samples. Adapted from a graphic produced by HGDP Selection browser.[49]

α-globin locus haplotypes in Australians and New Guineans produced conflicting evidence for both the independent origin hypothesis[35] and the common origin hypothesis.[36] mtDNA diversity[11,19,25,37–41] indicates very deep and diverse maternal ancestry for Australia and New Guinea. The most recent and comprehensive of these studies[25] found that the defining ancestral node of some subclades (Q and P) were shared between Australia and New Guinea, but there was little or no sharing of more recently evolved derived lineages within these, indicating a single founding Sahul group with subsequent isolation between regional populations. Other ancient lineages are unique to Australia, suggesting the possibility of different entry points and independent origins. On the paternal side, very few studies are available and are limited in language group representation. However, a single Y chromosome lineage (C4a-DYS390.1 del/M347), which Y microsatellite diversity suggests expanded rapidly beginning in the mid-Holocene about 5000 years ago,[20] is apparently unique to the Australian continent.[25,26] There is little evidence of Y chromosome sharing between New Guinea and Australia arguing for at least the relatively recent isolation of the two populations.[26]

Although the Y chromosome and mtDNA are potentially powerful markers of population history, they are each a single locus and potentially prone to high levels of drift that may obscure the window they provide into the very distant past. Our analysis is based on a suite of
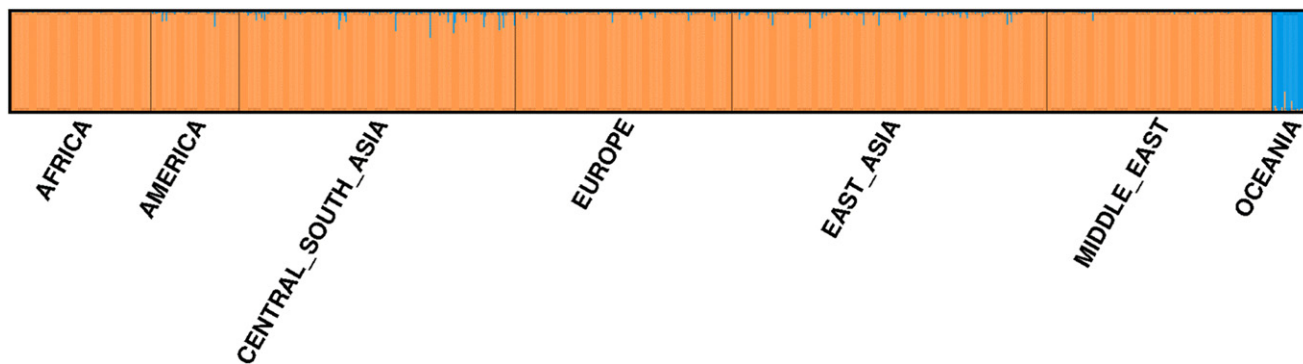
Figure 6. Oceanic Ancestry Informative Markers
Results of a STRUCTURE run ($K = 2$) using 100 AIMs in the HGDP populations. There is clear ability to distinguish the Oceanic (PAPuans and MELanesians) from non-Oceanic populations. See the Figure 3 legend for a fuller description of the plot. STRUCTURE runs consisted of a 25,000 replicate burn-in followed by 25,000 MCMC iterations.

genome-wide SNPs providing a more robust and broad-based insight on Aboriginal genetic affinities. The clear phylogenetic grouping of the Aboriginal Australians with other Near Oceania samples, from New Guinea and Melanesia, favors the common origin hypothesis for the original settlement of the Pleistocene Sahul continent, if not details of routes and possible entry points. The most parsimonious explanation is a single settlement of the Sahul, which archaeological evidence puts at around 50,000–60,000 years ago,[3] followed by differentiation into subregional populations. However, we cannot formally distinguish between this and an initial separation and isolation of the proto-Sahul population in mainland Eurasia followed by multiple ancient migrations to various locations in the Sahul. Our conclusion is consistent with global population genetic affinities gauged from *Alu* insertion polymorphisms[42] and the genetic diversity of the human bacterial parasite, *Helicobacter pylori* (*H. pylori*). The haplotypes of *H. pylori* samples taken from people in New Guinea and Australia form a geographically distinct and phylogenetically ancient group (whose divergence from other global lineages was placed at ~32,000 years on the basis of accumulated diversity) supporting a single-wave initial Sahul settlement.[43]

Analysis of finer population structure in the whole-genome SNP data also indicates that these ancient initial settlers were subsequently isolated and relatively undisturbed by later migrations from the Asian mainland. There is little evidence, for instance, of East Asian ancestry that could be attributed to the Austronesian expansion (beginning ~5,500 years ago), which impacted, to varying degrees, other indigenous Oceanic (Melanesian and Papuan) populations,[18,44,45] nor is there any convincing signal of recent contact between Australia and the Indian subcontinent. The presence of phylogenetically ancient and geographically restricted mtDNA lineages (such as haplogroups S and M42) in Australia together with the absence of Austronesian mtDNA (B4a1a1a) and Y chromosome (O-M110, O-M119, O-M324) lineages supports our whole-genome-based conclusions. The results imply that noted archaeolog-

ical events, such as the mid-Holocene "intensification" (witnessed by increases in both the complexity and density of stone tools at many archaeological sites[46]) or the arrival of the dingo,[47] were not mediated by substantial amounts of migration from mainland Eurasia. It is clear from the data, however, that this period of apparent long-term isolation was ended by the arrival of European settlers beginning in 1788. A significant minority of the biological ancestry of the study participants comes from Europeans and was primarily introduced by males. A marked sex bias or asymmetry has been genetically noted in the history of European miscegenation with indigenous groups—for example, in South Africa.[48]

Although this study massively increases the number of loci used to investigate Aboriginal Australian genetic diversity, the increased power to uncover population history is mitigated by a relatively small sample population from just one part of Australia, albeit an important Pleistocene habitation region. A single sample will probably not reflect the full later history of an entire continent and its people. The long history of Aboriginal Australians and the large physical distances across a climatically shifting, sometimes arid continent, combined with a hunter-gatherer lifestyle, which does not typically support high population densities, may have allowed considerable isolation and genetic drift between Aboriginal groups in different parts of Australia. Some regions, particularly in northern coastal areas, may have experienced migration events that little impacted other areas. A broader survey of Aboriginal Australian groups is justified to capture the full diversity of the continent. As well as providing new insights into the past, a fuller understanding and appreciation of the continent's genetic diversity will be practically important to extend the latest gene discovery methods, and their potential health benefits, to Aboriginal Australians.

Supplemental Data

Supplemental Data include one figure and three tables and can be found with this article online at http://www.cell.com/AJHG/.

Web Resources

The URLs for data presented herein are as follows:

Affymetrix, http://www.affymetrix.com
EIGENSTRAT, http://genepath.med.harvard.edu/~reich/Software.htm
frappe: http://med.stanford.edu/tanglab/software/frappe.html
HGDP Selection Browser: http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
PHYLIP, http://evolution.genetics.washington.edu/phylip.html
STRUCTURE, http://pritch.bsd.uchicago.edu/structure.html

References

1. Cavalli-Sforza, L.L., and Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution. Nat. Genet. *33* (*Suppl*), 266–275.
2. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science *308*, 1034–1036.
3. O'Connell, J.F., and Allen, J. (2004). Dating the colonization of Sahul (Pleistocene Australia–New Guinea): A review of recent research. J. Archaeol. Sci. *31*, 835–853.
4. Balme, J., Davidson, I., McDonald, J., Stern, N., and Veth, P. (2009). Symbolic behaviour and the peopling of the southern arc route to Australia. Quaternary International *202*, 59–68.
5. Oppenheimer, S. (2009). The great arc of dispersal of modern humans: Africa to Australia. Quaternary International *202*, 2–13.
6. Groube, L., Chappell, J., Muke, J., and Price, D. (1986). A 40,000 year-old human occupation site at Huon Peninsula, Papua New Guinea. Nature *324*, 453–455.
7. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nat. Rev. Genet. *9*, 356–369.
8. van Holst Pellekaan, S. (2000). Genetic research: What does this mean for indigenous Australian communities? Australian Aboriginal Studies *1/2*, 65–75.
9. Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I., and Boehnke, M. (2010). Genome-wide association studies in diverse populations. Nat. Rev. Genet. *11*, 356–366.
10. Need, A.C., and Goldstein, D.B. (2009). Next generation disparities in human genomics: Concerns and remedies. Trends Genet. *25*, 489–494.
11. van Holst Pellekaan, S., Frommer, M., Sved, J., and Boettcher, B. (1998). Mitochondrial control-region sequence variation in aboriginal Australians. Am. J. Hum. Genet. *62*, 435–449.
12. van Holst Pellekaan, S.M., Ingman, M., Roberts-Thomson, J., and Harding, R.M. (2006). Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. Am. J. Phys. Anthropol. *131*, 282–294.
13. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K., et al. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. *40*, 1253–1260.
14. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.
15. Tang, H., Peng, J., Wang, P., and Risch, N.J. (2005). Estimation of individual admixture: Analytical and study design considerations. Genet. Epidemiol. *28*, 289–301.
16. Falush, D., Stephens, M., and Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics *164*, 1567–1587.
17. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.
18. Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L.A., Moyse-Faurie, C., Rutledge, R.B., Schiefenhoevel, W., Gil, D., et al. (2006). Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. Mol. Biol. Evol. *23*, 2234–2244.
19. Redd, A.J., and Stoneking, M. (1999). Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am. J. Hum. Genet. *65*, 808–828.
20. Redd, A.J., Roberts-Thomson, J., Karafet, T., Bamshad, M., Jorde, L.B., Naidu, J.M., Walsh, B., and Hammer, M.F. (2002). Gene flow from the Indian subcontinent to Australia: Evidence from the Y chromosome. Curr. Biol. *12*, 673–677.
21. Sturm, R.A., Duffy, D.L., Zhao, Z.Z., Leite, F.P., Stark, M.S., Hayward, N.K., Martin, N.G., and Montgomery, G.W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am. J. Hum. Genet. *82*, 424–431.
22. Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., et al. (2007). Genetic determinants of

hair, eye and skin pigmentation in Europeans. Nat. Genet. *39*, 1443–1452.

23. Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. *18*, 830–838.

24. Jobling, M.A., and Tyler-Smith, C. (2003). The human Y chromosome: An evolutionary marker comes of age. Nat. Rev. Genet. *4*, 598–612.

25. Hudjashov, G., Kivisild, T., Underhill, P.A., Endicott, P., Sanchez, J.J., Lin, A.A., Shen, P., Oefner, P., Renfrew, C., Villems, R., and Forster, P. (2007). Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. Proc. Natl. Acad. Sci. USA *104*, 8726–8730.

26. Kayser, M., Brauer, S., Weiss, G., Schiefenhövel, W., Underhill, P.A., and Stoneking, M. (2001). Independent histories of human Y chromosomes from Melanesia and Australia. Am. J. Hum. Genet. *68*, 173–190.

27. Kayser, M. (2010). The human genetic history of Oceania: Near and remote views of dispersal. Curr. Biol. *20*, R194–R201.

28. McDonough, C.W., Bostrom, M.A., Lu, L., Hicks, P.J., Langefeld, C.D., Divers, J., Mychaleckyj, J.C., Freedman, B.I., and Bowden, D.W. (2009). Genetic analysis of diabetic nephropathy on chromosome 18 in African Americans: Linkage analysis and dense SNP mapping. Hum. Genet. *126*, 805–817.

29. Ewens, K.G., George, R.A., Sharma, K., Ziyadeh, F.N., and Spielman, R.S. (2005). Assessment of 115 candidate genes for diabetic nephropathy by transmission/disequilibrium test. Diabetes *54*, 3305–3318.

30. Iyengar, S.K., Abboud, H.E., Goddard, K.A., Saad, M.F., Adler, S.G., Arar, N.H., Bowden, D.W., Duggirala, R., Elston, R.C., Hanson, R.L., et al; Family Investigation of Nephropathy and Diabetes Research Group. (2007). Genome-wide scans for diabetic nephropathy and albuminuria in multiethnic populations: The family investigation of nephropathy and diabetes (FIND). Diabetes *56*, 1577–1585.

31. Meguid El Nahas, A., and Bello, A.K. (2005). Chronic kidney disease: The global challenge. Lancet *365*, 331–340.

32. Thomas, N., MacRae, A., Burns, J., Catto, M., Debuyst, O., Krom, I., Midford, R., Potter, C., Ride, K., Stumpers, S., and Urquhart, B. (2008). Overview of Australian Indigenous health status, April 2010. http://www.healthinfonet.ecu.edu.au/health-facts/overviews.

33. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

34. Seldin, M.F. (2007). Admixture mapping as a tool in gene discovery. Curr. Opin. Genet. Dev. *17*, 177–181.

35. Tsintsof, A.S., Hertzberg, M.S., Prior, J.F., Mickleson, K.N., and Trent, R.J. (1990). Alpha-globin gene markers identify genetic differences between Australian aborigines and Melanesians. Am. J. Hum. Genet. *46*, 138–143.

36. Roberts-Thomson, J.M., Martinson, J.J., Norwich, J.T., Harding, R.M., Clegg, J.B., and Boettcher, B. (1996). An ancient common origin of aboriginal Australians and New Guinea highlanders is supported by alpha-globin haplotype analysis. Am. J. Hum. Genet. *58*, 1017–1024.

37. Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G., Babb, P., Cerchio, S., Kaestle, F., Schanfield, M., et al. (2005). Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol. Biol. Evol. *22*, 1506–1517.

38. Huoponen, K., Schurr, T.G., Chen, Y., and Wallace, D.C. (2001). Mitochondrial DNA variation in an aboriginal Australian population: Evidence for genetic isolation and regional differentiation. Hum. Immunol. *62*, 954–969.

39. Ingman, M., and Gyllensten, U. (2003). Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res. *13*, 1600–1606.

40. Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., et al. (2006). The role of selection in the evolution of human mitochondrial genomes. Genetics *172*, 373–387.

41. van Holst Pellekaan, S.M., and Harding, R.M. (2006). Excavating the mitochondrial genome identifies major haplogroups in Aboriginal Australians. Before Farming *2006*, article 3.

42. Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., Saha, N., Jenkins, T., Tahir, M.A., Deininger, P.L., and Batzer, M.A. (1997). Alu insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. Genome Res. *7*, 1061–1071.

43. Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.Y., Maady, A., Bernhoft, S., Thiberge, J.M., Phuanukoonnon, S., et al. (2009). The peopling of the Pacific from a bacterial perspective. Science *323*, 527–530.

44. Kimura, R., Ohashi, J., Matsumura, Y., Nakazawa, M., Inaoka, T., Ohtsuka, R., Osawa, M., and Tokunaga, K. (2008). Gene flow and natural selection in oceanic human populations inferred from genome-wide SNP typing. Mol. Biol. Evol. *25*, 1750–1761.

45. Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nürnberg, P., Trent, R.J., and Stoneking, M. (2008). Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. Am. J. Hum. Genet. *82*, 194–198.

46. Mulvaney, D.J., and Kamminga, J. (1999). Prehistory of Australia (St. Leonards, NSW: Allen & Unwin).

47. Savolainen, P., Leitner, T., Wilton, A.N., Matisoo-Smith, E., and Lundeberg, J. (2004). A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. Proc. Natl. Acad. Sci. USA *101*, 12387–12390.

48. Quintana-Murci, L., Harmant, C., Quach, H., Balanovsky, O., Zaporozhchenko, V., Bormans, C., van Helden, P.D., Hoal, E.G., and Behar, D.M (2010). Strong maternal Khoisan contribution to the South African coloured population: A case of gender-biased admixture. Am. J. Hum. Genet. *86*, 611–620.

49. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., and Pritchard, J.K. (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Res. *19*, 826–837.