

## ORIGINAL ARTICLE

# Mitochondrial DNA diversity of present-day Aboriginal Australians and implications for human evolution in Oceania

Nano Nagle<sup>1</sup>, Kaye N Ballantyne<sup>2,3</sup>, Mannis van Oven<sup>3</sup>, Chris Tyler-Smith<sup>4</sup>, Yali Xue<sup>4</sup>, Stephen Wilcox<sup>5</sup>, Leah Wilcox<sup>1</sup>, Rust Turkalov<sup>5</sup>, Roland AH van Oorschot<sup>2</sup>, Sheila van Holst Pellekaan<sup>6,7</sup>, Theodore G Schurr<sup>8</sup>, Peter McAllister<sup>9</sup>, Lesley Williams<sup>10</sup>, Manfred Kayser<sup>3</sup>, R John Mitchell<sup>1</sup> and The Genographic Consortium<sup>27</sup>

Aboriginal Australians are one of the more poorly studied populations from the standpoint of human evolution and genetic diversity. Thus, to investigate their genetic diversity, the possible date of their ancestors' arrival and their relationships with neighboring populations, we analyzed mitochondrial DNA (mtDNA) diversity in a large sample of Aboriginal Australians. Selected mtDNA single-nucleotide polymorphisms and the hypervariable segment haplotypes were analyzed in 594 Aboriginal Australians drawn from locations across the continent, chiefly from regions not previously sampled. Most (~78%) samples could be assigned to mtDNA haplogroups indigenous to Australia. The indigenous haplogroups were all ancient (with estimated ages > 40 000 years) and geographically widespread across the continent. The most common haplogroup was P (44%) followed by S (23%) and M42a (9%). There was some geographic structure at the haplotype level. The estimated ages of the indigenous haplogroups range from 39 000 to 55 000 years, dates that fit well with the estimated date of colonization of Australia based on archeological evidence (~47 000 years ago). The distribution of mtDNA haplogroups in Australia and New Guinea supports the hypothesis that the ancestors of Aboriginal Australians entered Sahul through at least two entry points. The mtDNA data give no support to the hypothesis of secondary gene flow into Australia during the Holocene, but instead suggest long-term isolation of the continent.

*Journal of Human Genetics* advance online publication, 1 December 2016; doi:10.1038/jhg.2016.147

## INTRODUCTION

Aboriginal Australians represent one of the oldest continuous cultures in the world, as their ancestors were among the first groups to leave Africa and travel through the Middle East Asia, and then Southeast Asia (i.e., the ancient landmass called Sunda).<sup>1–3</sup> A recent review of the archeological record of the ancient landmass of Sahul (which comprised present-day New Guinea and Australia) concluded that anatomically modern humans colonized Sahul at least 47 thousand years ago (KYA).<sup>4</sup> The timing of this migration, the route(s) taken and the entry point(s) into Sahul have been subject to ongoing debate by archeologists, anthropologists and geneticists for many years, with evidence both for and against a single colonization event.<sup>4–13</sup> Unfortunately, the adverse impacts of European colonization on Aboriginal society since the eighteenth century seriously confound attempts at reconstructing the past genetic structure of Aboriginal society.<sup>14</sup>

Analyses of the uniparentally inherited Y chromosome and mitochondrial DNA (mtDNA) and the biparentally inherited autosomes of Aboriginal Australians has been crucial in identifying the possible migration routes and the timing and number(s) of potential colonization events their ancestors undertook to reach the continent. Whole-genome studies using the DNAs of both a 100-year-old Aboriginal hair sample<sup>15</sup> and a number of present-day Aboriginal Australians<sup>13,16,17</sup> support the great antiquity of the settlement of Australia by the ancestors of the Aboriginal Australians, as well as their deep connection with indigenous populations of New Guinea. However, one genome-wide single-nucleotide polymorphism (SNP) study claimed that, in addition to the initial settlement, substantial gene flow from India into Australia occurred during the Holocene.<sup>8</sup>

Y-chromosome and mtDNA analyses also support an ancient settlement of Australia. However, inferences have generally been limited by the low sample numbers of 10 to 100 individuals and/or

<sup>1</sup>Department of Biochemistry and Genetics, La Trobe University, Melbourne, VIC, Australia; <sup>2</sup>Office of the Chief Forensic Scientist, Victorian Police Forensic Services Department, Melbourne, VIC, Australia; <sup>3</sup>Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands; <sup>4</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK; <sup>5</sup>Australian Genome Research Facility, Melbourne, VIC, Australia; <sup>6</sup>School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia; <sup>7</sup>School of Biological Sciences, University of Sydney, Camperdown, NSW, Australia; <sup>8</sup>Department of Anthropology, University of Pennsylvania, Philadelphia, PA, USA; <sup>9</sup>Griffith University, Nathan, QLD, Australia and <sup>10</sup>Community Elder and Cultural Advisor, Brisbane, QLD, Australia

<sup>27</sup>Genographic Consortium members are listed above references.

Correspondence: Dr RJ Mitchell, Department of Biochemistry and Genetics, La Trobe University, Plenty Road, Melbourne, VIC 3086, Australia.

E-mail: john.mitchell@latrobe.edu.au

Received 19 August 2016; revised 21 October 2016; accepted 27 October 2016

the few geographic regions within Australia that have been surveyed.<sup>6,7,13,18–38</sup> Despite this, a recent comprehensive study of Y-chromosome variation in Aboriginal males suggested a settlement of Australia at least 45 KYA,<sup>39</sup> with this date being subsequently supported by the deep sequencing of Aboriginal Y chromosomes.<sup>40</sup>

Most previous mtDNA data, by far, have been obtained from groups in the Northern Territory, northern Western Australia and western New South Wales.<sup>6,21–23,33,41</sup> There are limited data from South Australia and Queensland (except for the samples reported by Malaspinas *et al.*<sup>13</sup> and Ballantyne *et al.*<sup>41</sup> which the latter we reanalyze here), and none are available from Victoria, although these States contain large populations of Aboriginal people. For example, one-fifth of all Aboriginal people reside in Queensland.<sup>42</sup> Also, as not all of the studies typed the same mtDNA SNPs or sequenced hypervariable segments I and II (HVS-I and HVS-II), it is difficult to compare some of these data sets, although van Holst Pellekaan<sup>14</sup> attempted to do so.

mtDNA haplogroups acknowledged as uniquely Aboriginal Australian fall within macrohaplogroups M and N (including R). Haplogroups presently considered autochthonous to Australia are N13, O, M42a, M14, M15, S and numerous P subtypes (P3a, P4b, P5, P6, P7, P8). Current time to most recent common ancestor (TMRCA) estimates indicate great antiquity for most of them.<sup>6,14,21,22,33,43,44</sup>

The absence of mtDNA data for most of Australia limits inferences about the distribution of these haplogroups throughout the continent, Aboriginal Australians' relationships with neighboring populations and the maternal genetic history of Australia as a whole. A common misconception about Aboriginal Australians is that they are culturally and linguistically a homogenous population, as discussed in Hiscock.<sup>45</sup> Before European settlement in the eighteenth century, there was a well-defined Aboriginal language structure, as described by Horton.<sup>46</sup> Consequently, a genetic survey of one region would not necessarily encapsulate their overall diversity. van Holst Pellekaan *et al.*<sup>21</sup> for example, described significant differences in the HVS-I haplotypes between the Aboriginal people of the Darling River in New South Wales and the Yuendumu of Central Australia (the Northern Territory).

Given the lack of whole mtDNA genome sequences of Aboriginal Australians (there are 38 in GenBank as of June 2016), and, therefore, the paucity of detailed phylogenetic knowledge regarding Aboriginal haplogroups, assignment to haplogroups based on limited SNP information is problematic for many samples. As only one representative mitogenome sequence is available for certain haplogroups; for example, M15, S5 and P5, HVS-I motifs are inadequate for Australian-specific haplogroup assignment. Also, given the antiquity of these haplogroups, the HVS, due to its elevated average evolutionary rate compared with the mtDNA coding region, has most likely accumulated recurrent mutations that may obscure phylogenetic signatures. Furthermore, given the antiquity of Australian-specific haplogroups, their HVS haplotypes are characteristically more divergent than those of younger haplogroups. These factors render haplogroup assignment made from HVS data alone challenging or impossible. However, as emphasized by van Holst Pellekaan,<sup>14</sup> studies that have limited coverage of mtDNA sequences nonetheless provide important information about family genealogies as well act as a useful sorting method for subsequent whole mtDNA genome sequencing.<sup>6,22,47</sup>

To expand knowledge of mtDNA diversity among this poorly studied population, we assembled a large sample of self-reported Aboriginal Australian individuals for the analysis of both mtDNA SNPs and HVS variation. This sample set was drawn from widespread locations across Australia, especially from either areas not previously sampled, such as Victoria, or those previous poorly sampled, such as

Queensland and Tasmania. Additionally, the sample included individuals whose ancestry lay in one of the Torres Strait Islands, which lie between Cape York in Far North Queensland and Papua New Guinea, and represent the remnants of the former land bridge between Australia and New Guinea.

This study aims to investigate the distribution of indigenous mitochondrial haplogroups and associated HVS variation across the continent, to infer ages of these lineages relevant for studying human history and to explore the maternal genetic relationship of the Aboriginal Australian people with neighboring populations.

## MATERIALS AND METHODS

### Sample collection

A total of 502 buccal swab, mouthwash or Oragene samples were collected from self-declared Aboriginal Australians residing in the States of Queensland, New South Wales, Victoria, Western Australia and Tasmania, as well as the Northern Territory. It is important to note that Aboriginal Australian is a culturally based affiliation and not one defined by a person's genetic composition. The current residence of participants may not represent the homeland or even birthplace of the individual, but, in the absence of genealogical information from the majority of the participants, assignment of samples by the State of residence was the best possible treatment of the data. This study was conducted with the approval of the La Trobe University Human Ethics Committee.

In addition, the 92 Aboriginal DNA samples from the Northern Territory ( $n=76$ ), Queensland ( $n=10$ ) and South Australia ( $n=6$ ) previously reported in Ballantyne *et al.*<sup>41</sup> were available for HVS sequencing and included in these analyses, giving an overall total of 594 samples.

### DNA extraction

DNA was extracted from buccal swabs and mouthwash samples following the method of de Vries *et al.*<sup>48</sup> and from the Oragene saliva collection tubes following the manufacturer's protocol.

### MtDNA SNP typing and HVS sequencing

Selected coding-region mtDNA SNPs were genotyped using custom TaqMan assays (Applied Biosystems, Foster City, CA, USA). The SNP hierarchical genotyping protocol was as follows (and depicted in Figure 1). All samples were typed for basal haplogroup L (nucleotide positions 3594, 13 650 and 7256), M (nucleotide position 14 783), N (nucleotide positions 10 873, 8701 and 9540) and R (nucleotide position 12 705) macrohaplogroup markers. Individuals who belonged to macrohaplogroup N were subsequently typed for the 8404 SNP, which denotes haplogroup S, and SNP 9140, which denotes haplogroup O. Those individuals who belonged to haplogroup R were subsequently typed for nucleotide position 15 607, which characterizes both haplogroups P and T. Those individuals who belonged to macrohaplogroup M were typed for M42a (nucleotide position 12 771) and M29/Q (nucleotide position 13 500).

mtDNA haplogroups considered to be indigenous to Australia that are supported by observations are M14, M15, M42a, N13, O and subtypes, S and subtypes and specific P subtypes: P3a, P4b, P5, P6, P7 and P8.<sup>6,14,21,22,33,43</sup> It is unclear whether haplogroup Q is indigenous to Australia as discussed in Hudjashov *et al.*<sup>6</sup> All other mtDNA haplogroups typed are considered non-indigenous as they could be explained by historical Eurasian admixture. The 92 samples of Ballantyne *et al.*<sup>41</sup> had previously been SNP genotyped using a specifically designed SNaPshot assay that comprised the major mtDNA haplogroups of Oceania.

In all samples, the first HVS-I from nucleotide position 16 000 to 16 400 and HVS-II from nucleotide position 40 to 310 were sequenced at the Australian Genome Research Facility, Melbourne, Australia, with the ABI Big Dye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed on an ABI 3730 DNA Analyzer (Applied Biosystems). Primers are described in Supplementary Table 1. The 25  $\mu$ l PCR reactions included 1  $\mu$ l DNA, 1  $\mu$ l 10  $\mu$ M forward primer, 1  $\mu$ l 10  $\mu$ M reverse primer, 1  $\mu$ l 5 U  $\mu$ l<sup>-1</sup> Taq polymerase, 5  $\mu$ l PCR master mix and 16  $\mu$ l H<sub>2</sub>O. The PCR protocol was: 37 cycles of denaturation at 95 °C for 30 s, primer annealing at 63 °C for 35 s and extension

at 72 °C for 1 min and 30 s. Mutations in the HVS region were defined by aligning and comparing the sequences to the revised Cambridge Reference Sequence<sup>49,50</sup> using MEGA v.6.<sup>51</sup>

**Haplogroup assignment**

All HVS haplotypes and associated mtDNA SNPs were also analyzed with the haplogroup prediction program, HaploGrep,<sup>52</sup> based on Build 16 of the reference human mtDNA phylogeny <http://phyloftree.org>,<sup>53</sup> and confirmed via manual interpretation using the same PhyloTree Build. Haplogroups were initially assigned using HaploGrep, although, upon inspection, some dubious haplogroup assignments emerged. Each sample was rechecked manually to improve and verify the predictions made by HaploGrep (e.g., haplogroup D for some of those samples within macrohaplogroup M). In cases where mtDNA sequences did not possess the necessary diagnostic SNPs and HVS motifs for adequate haplogroup assignment, samples were denoted by their basal haplogroup status; that is, S\* and P\*.

**Data analysis**

Median-joining networks were constructed from HVS haplotype and mtDNA SNP data using Network version 4.6.1.2 (<http://www.fluxus-engineering.com/sharenet.htm>; Bandelt *et al.*<sup>54</sup>). HVS sequences were combined with published Aboriginal Australian HVS data.<sup>6,21,33,38</sup>

Population pairwise genetic distances ( $F_{ST}$  values) were calculated from HVS-I haplotypes using the Arlequin v.3.5 package<sup>55</sup> and their significance was assessed by 1000 bootstrap simulations. Haplotype diversity and sampling variance were calculated using the formula of Nei.<sup>56</sup> Multidimensional scaling (MDS) analysis of indigenous mtDNA haplogroup frequencies of the States of Australia was performed using SPSS v.21.0 software (SPSS, Inc., Chicago, IL, USA).<sup>57</sup> An MDS analysis of haplogroup frequencies of the entire Aboriginal Australian sample compared with those from neighboring Island Southeast Asian (ISEA) and Oceanian populations,<sup>47,58,59</sup> African populations<sup>60</sup> and European populations<sup>61</sup> was also performed.

Estimation of the TMRCA of haplogroups was performed using the  $\rho$ -statistic, as calculated in the software Network.<sup>62</sup> The  $\rho$ -statistic calculates TMRCA estimates based on an average mutation rate and an inferred ancestral haplotype, and do not incorporate a specific demographic model. However, the resultant TMRCA estimates can be underestimates since the method sometimes fails to distinguish between haplotypes that are identical by State rather than by descent. The mutation rate used was that of Soares *et al.*<sup>63</sup> where 1 transition occurs every 16 677 years for the region 16 051 to 16 400 (excluding nucleotide positions 16 182, 16 183 and 16 194).

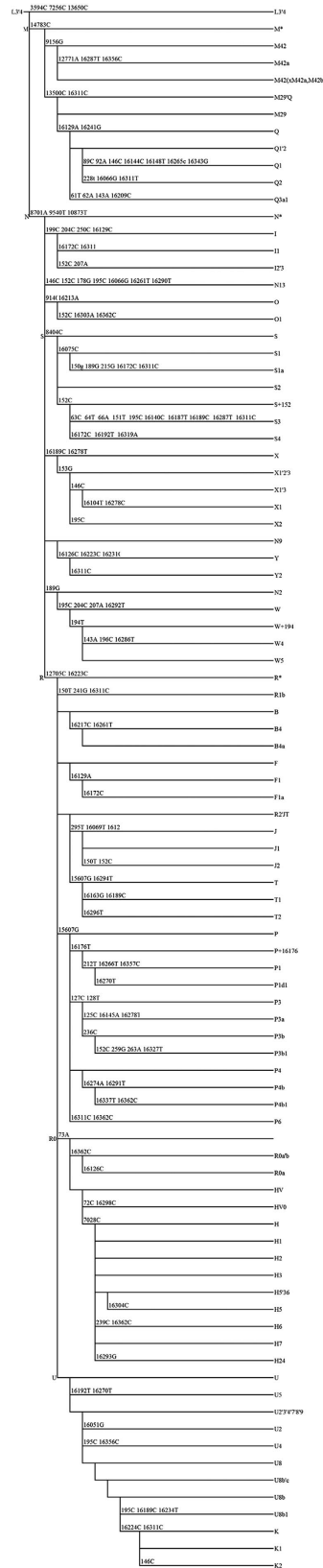
**RESULTS**

**Non-indigenous mtDNA haplogroups and admixture in Aboriginal Australians**

Sequencing the HVS and genotyping coding region SNPs in 594 Aboriginal Australian mtDNAs revealed that 133 (22.4%) individuals could be assigned to non-indigenous haplogroups. Of these, the vast majority (98%) was of European origin (Table 1a and Supplementary Table 2), which is explained by historical admixture. The frequency of non-indigenous haplogroups varied across Australia, with New South Wales having the highest (of States with >10 sampled individuals) and the Northern Territory the lowest (Table 1a). An MDS plot based on  $F_{ST}$  distances of mtDNA haplogroup frequencies in ISEA, Oceanian, European and African populations confirmed the close affinities of Australian individuals with these non-indigenous haplogroups to Europeans (Supplementary Figure 1). The most frequent non-indigenous haplogroup was H (9.3% of the total sample).

**Indigenous mtDNA haplogroups**

The remaining 461 samples (77.6%) were classified as belonging to indigenous Aboriginal Australian mtDNA haplogroups (Table 1b), and are the subject of the subsequent analyses.



**Figure 1** mtDNA classification tree used for haplogroup assignment in the present study.

**Macrohaplogroup M.** Macrohaplogroup M lineages accounted for 122 (26.4%) of the 461 samples classified as indigenous. Forty-three individuals (9.3% of indigenous sequences) were assigned to haplogroup M42a and 12 individuals (2.6%) to haplogroup Q. Haplogroup M42a was present in all eastern mainland States and the Northern Territory (Figure 2). The 12 haplogroup Q individuals from New South Wales and Queensland were subclassified as either Q1 or Q3. HVS sequence data for 67 (54.9%) M individuals, however, did not allow subhaplogroup assignment with confidence. These individuals were labeled M\*(xM42a,Q), but considered indigenous given that the HVS haplotypes could not be assigned to Melanesian M subhaplogroups M27, M28 or M29.<sup>6,64,65</sup> All States, except South Australia and Tasmania, had individuals belonging to this unresolved paralog.

In the network, HVS-I haplotypes belonging to haplogroups M42a and Q formed well-defined clusters (Figure 3). Distinct clades could be identified within M42a, suggesting the existence of distinct subhaplogroups within it. Q haplotypes also clustered and were separated by a long branch from the rest of the haplotypes in this network. The M\*(xM42a,Q) haplotypes were widely dispersed, despite some limited clustering being evident. Ten individuals from Queensland shared HVS-I (and HVS-II haplotypes) with two of the M42\*(xM42a,Q) individuals from the Northern Territory reported by Ballantyne *et al.*<sup>41</sup> Also, two M\*(xM42a,Q) individuals, one from the Northern Territory and the other from Western Australia, clustered with the HVS-I and HVS-II haplotype of the published haplogroup M15 individual.<sup>6</sup> Overall, geographical clustering of haplotypes within M (including Q) at the State/Territory level was substantial (Figure 3).

**Macrohaplogroup N.** Haplogroups N\*, N13, O, S and P within macrohaplogroup N together comprised the majority (339 or 73.5%) of the 461 individuals with indigenous haplogroups. Eleven individuals (3%; from the Northern Territory and Queensland) could not be assigned to a recognized N subtype, and were denoted N\*(xN13, O,S,R). Haplogroup N13 was detected in three individuals from Queensland; one shared its haplotype with a previously reported individual from Western Australia,<sup>6</sup> while the other two shared a novel haplotype. Haplogroup O had a wide distribution, being present in New South Wales, the Northern Territory, Queensland and South Australia (Figure 2), with some geographic structure evident (Figure 4). The majority of haplogroup O haplotypes (88.8%) could not be assigned to the O1 subtype (so far, the only defined subhaplogroup of O), suggesting that a novel subtype(s) may exist within this clade.

Haplogroup S comprised 104 individuals (22.6%) of indigenous samples. The most frequent subtype was S1 (52.8% of S individuals). However, a large proportion (33.6%) could not be assigned to any of the currently known S subtypes (S1–S5) based on HVS data, suggesting that either these known subtypes have diverged further than currently recognized and/or novel S subtype(s) exist. The two distinct S1 clusters in the haplotype network (Figure 5) suggest the existence of at least two subhaplogroups within S1 (so far, one subclade, S1a, is defined in PhyloTree). Haplogroup S5 was absent from the present study, but S3 and S4 were found at low frequencies. There was some clustering of the haplotypes within S sublineages, in particular for S1a. Geographically, haplogroup S is a very widely dispersed haplogroup, being observed in all States and the Northern Territory and with minimal sharing of haplotypes across the States (Figure 5).

The most common indigenous haplogroup was P, which comprised 40.9% of the individuals with indigenous haplogroups after the

**Table 1a Distribution of mtDNA haplogroups in Aboriginal Australian populations after SNP typing**

Population sample	n	Indigenous mitochondrial haplogroups										Non-indigenous mitochondrial haplogroups													
		M*(xM42a,Q)		M42a		Q		N* and N13		O		S		P		Total		H		L		R and others <sup>a</sup>		Total	
		No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
New South Wales	90	9	10.0	3	3.3	1	1.1	0	0.0	4	4.4	13	14.4	19	21.1	49	54.4	15	16.7	0	0.0	26	28.9	41	45.6
Northern Territory	77	5	6.5	1	1.3	0	0.0	3	3.9	4	5.2	25	32.5	37	48.1	75	97.4	1	1.3	0	0.0	1	1.3	2	2.6
Queensland	391	45	11.5	37	9.5	11	2.8	11	2.8	8	2.0	58	14.8	140	35.8	310	79.3	36	9.2	1	0.3	44	11.3	81	20.7
Tasmania	7	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	14.3	2	28.6	3	42.9	2	28.6	0	0.0	2	28.6	4	57.1
Victoria	18	5	27.8	2	11.1	0	0.0	0	0.0	0	0.0	4	22.2	4	22.2	15	83.3	0	0.0	0	0.0	3	16.7	3	16.7
Western Australia	5	3	60.0	0	0.0	0	0.0	0	0.0	0	0.0	1	20.0	4	80.0	4	80.0	1	20.0	0	0.0	0	0.0	1	20.0
South Australia	6	0	0.0	0	0.0	0	0.0	0	0.0	2	33.3	2	33.3	1	16.7	5	83.3	0	0.0	0	0.0	1	16.7	1	16.7
Total	594	67	11.3	43	7.2	12	2.0	14	2.4	18	3.0	104	17.5	203	34.2	461	77.6	55	9.3	1	0.2	77	13.0	133	22.4

Abbreviations: mtDNA, mitochondrial DNA; SNP, single-nucleotide polymorphism.

<sup>a</sup>Others: haplogroups W, X, Y, B, F, J, I, T, K and U.

**Table 1b Distribution of indigenous mtDNA haplogroups in Aboriginal Australian populations after SNP typing**

Population sample	n	Indigenous mitochondrial haplogroups													
		M*(xM42a,Q)		M42a		Q		N* and N13		O		S		P	
		No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
New South Wales	49	9	18.4	3	6.1	1	2.0	0	0.0	4	8.2	13	26.5	19	38.8
Northern Territory	75	5	6.7	1	1.3	0	0.0	3	4.0	4	5.3	25	33.3	37	49.3
Queensland	310	45	14.5	37	11.9	11	3.5	11	3.5	8	2.6	58	18.7	140	45.2
Tasmania	3	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	33.3	2	66.7
Victoria	15	5	33.3	2	13.3	0	0.0	0	0.0	0	0.0	4	26.7	4	26.7
Western Australia	4	3	75.0	0	0.0	0	0.0	0	0.0	0	0.0	1	25.0	0	0.0
South Australia	5	0	0.0	0	0.0	0	0.0	0	0.0	2	40.0	2	40.0	1	20.0
Total	461	67	14.5	43	9.3	12	2.6	14	3.0	18	3.9	104	22.6	203	44.0

Abbreviations: mtDNA, mitochondrial DNA; SNP, single-nucleotide polymorphism.

removal of individuals with haplogroups P1 and P3b of assumed New Guinean origin. Haplogroup P was found in all States except Western Australia (its absence there can be attributed to the small sample size for that State since Hudjashov *et al.*<sup>6</sup> reported its presence there). The contribution of New Guinean sublineages P1 and P3b was minor (6.3% of haplogroup P individuals). Notably, all those individuals carrying P1 or P3b mtDNAs were individuals with maternal ancestry tracing back to the Torres Strait Islands, which lie between New Guinea and the Australian mainland. The most common P subtype was P3a (28.3% of haplogroup P individuals). Most importantly, a significant proportion of Australian haplogroup P individuals (88 or 43%) could not be assigned to a known P subhaplogroup based on HVS data (Figure 6), and were accordingly denoted as P\*. There was a lack of clustering of haplotypes within particular P sublineages (excluding haplogroups P1, P3a and P3b) due to the high level of diversity of the haplotypes within the subtypes. Overall, many P haplotypes were shared among individuals of the same region, with > 80% being geographically localized across Australia.

**Diversity indices**

When comparing the regions within Australia (i.e., those with sample sizes > 10), the Northern Territory had the highest level of mtDNA diversity and Victoria the lowest. Among the 461 HVS-I sequences belonging to indigenous haplogroups, there were 235 haplotypes (50.9% discrimination), and of these, 119 were singletons (50.6%) and another 57 (24.3%) were shared only within a State or the Northern Territory. Combined, these 176 (74.9%) haplotypes were restricted to a particular State. The remaining 59 haplotypes were observed in multiple States and/or the Northern Territory, with the majority shared between Queensland and the Northern Territory. The most diverse (in terms of HVS variation) haplogroups were S and P, which had values > 0.940, respectively (Supplementary Table 3a).

Analysis of molecular variance based on haplogroup frequencies demonstrated that variation among the States was small but statistically significant ( $F_{ST} = 0.013$ ,  $P < 0.05$ ) (Tasmania and Western Australia were excluded because of low sample sizes). Analysis of molecular variance based on HVS sequence data gave a similarly small value that was statistically highly significant ( $F_{ST} = 0.0113$ ,  $P < 0.0001$ ). An MDS analysis of the  $F_{ST}$  values based on haplogroup frequencies between States indicated that on dimension 1, the largest distance was between the Victoria and Northern Territory populations, with those of Queensland and New South Wales approximately in the middle, while in dimension 2 the Queensland population was separated from the rest (Supplementary Figure 2).

**TMRCAs estimates**

In all instances, the ages of indigenous Australian haplogroups were very old, but depending on the mutation rate used there were marked differences among some of them (Table 2). For example, the TMRCAs estimates of haplogroups S and O using the rate of Soares *et al.*<sup>63</sup> the median ages were 49 KY (range 39–60 KY) for S and 39 KY (range 26–53 KY) for O. The Australian-specific P lineages, including the P\* paragrroup, were estimated to be greater than 55 KY (range 48–69 KY) in age. The TMRCAs of the haplogroups were compared with published estimates, where these were available (Table 2).

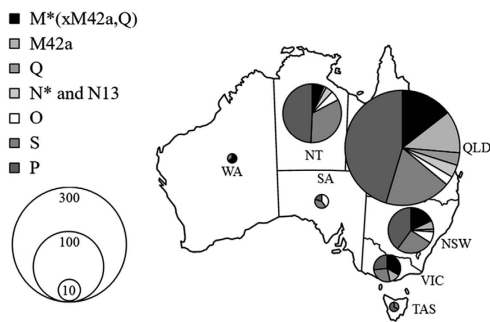
**DISCUSSION**

This study represents the largest mtDNA database of Aboriginal Australians yet assembled with respect to both sample size and geographical coverage. It includes samples from the hitherto very poorly represented or unrepresented States of Queensland (including

the Torres Strait Islands), Victoria and Tasmania, and with additional samples from the Northern Territory.

The level of introgression of European mtDNA lineages in this Aboriginal sample is considerably smaller (~20%) than the level reported in Y-chromosome studies of ~80%.<sup>37,39</sup> There may have been a loss of Aboriginal Australian mtDNA diversity as a result of European settlement since the late eighteenth century, although the full extent remains unknown. The introduction of exotic disease, the forced removal of Aboriginal Australians from their ancestral homelands into government or religious settlements, restrictions on marriage and the progressive exodus from rural areas into cities for employment have all influenced the findings of this study.

As a result of these processes, much of the language/clan structure illustrated by Horton<sup>46</sup> has dissolved, especially in the eastern States of Australia. For example, Aboriginal Australians of Queensland were removed from their native homelands and placed into large government settlements. Field research by Catherine Tennant-Kelly<sup>66</sup> recorded that, in a population of 900 Aboriginal Australians, 28 different tribal groups were represented in the government Cherbourg settlement (located in southeast Queensland), a finding that illustrates the disturbance of traditional population structure. Any attempts at reconstructing the past through interpretations of genetic patterns in



**Figure 2** Geographic distribution and frequencies of indigenous Australian mtDNA haplogroups. NSW, New South Wales; NT, Northern Territory; QLD, Queensland; SA, South Australia; TAS, Tasmania; VIC, Victoria; WA, Western Australia. The size of the pie charts are proportional to the sample size. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Aboriginal diversity today must be cognizant of the treatment of Aboriginal people by European settlers.<sup>14</sup>

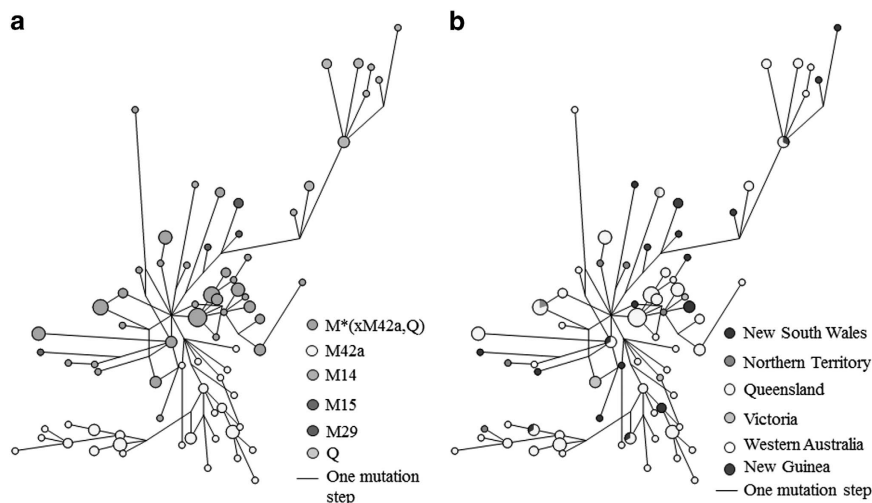
Of some concern was the possible inclusion of close maternal relatives in the sample. Although there was some sharing of haplotypes within haplogroups, the impact of relatives in our sample was probably minimal. In cases where haplotype sharing within haplogroups was observed, it was not possible to clarify the degree of relatedness of the individuals, which could be close or distant given the relatively low mtDNA mutation rate. The widespread sharing of mtDNA haplogroups across regions can be most readily explained by the traditional practice of patrilocality, in which the bride move to the place of her husband and his clan upon marriage.<sup>67</sup> Although individuals were characterized broadly by State or Territory, rather than by language or social groups as defined by Horton,<sup>46</sup> there was evidence for some geographic clustering of haplotypes within the States, with some signals of female movement between them.

The Aboriginal population showed a high level of HVS-I sequence diversity, which is most probably a function of the merging of groups from all over the continent. However, only one diversity measure was used in this study and different measures will give different values. Aboriginal Australians exhibited one of the highest levels of HVS-I diversity in ISEA and Oceania, with only Bali and Banjarmasin (Borneo) having diversity values >0.987 (Supplementary Table 3b).

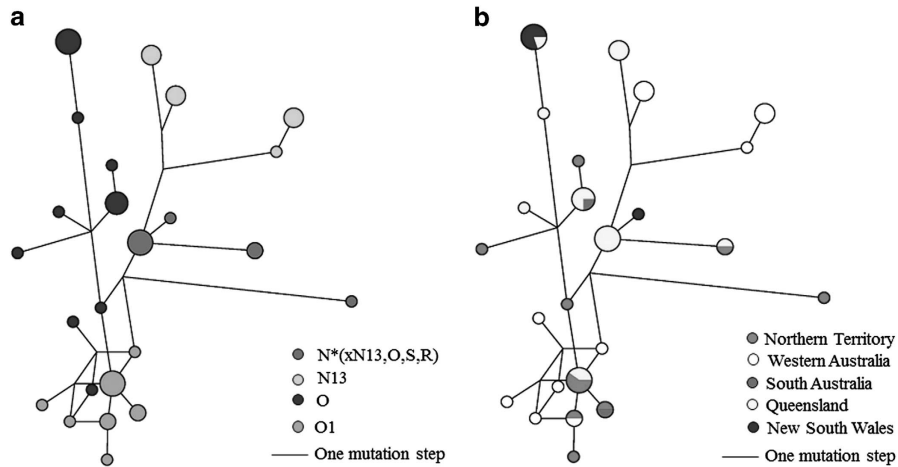
#### TMRCA estimates

The TMRCA of haplogroups covered by the present study suggest that the mtDNA haplogroups in Aboriginal Australians are older than previous calculations, but closest to those reported most recently by van Holst Pellekaan,<sup>14</sup> which in turn were themselves much older than prior estimates (Table 2). The estimates of Behar *et al.*<sup>44</sup> van Holst Pellekaan<sup>14</sup> and Hudjashov *et al.*<sup>6</sup> were based on coding-region sequences only (~15 kb) and not the HVS-I, and used a variety of mutation rates.<sup>31,35,68</sup> The estimates of van Holst Pellekaan<sup>14</sup> and Hudjashov *et al.*<sup>6</sup> were calculated using  $\rho$ -statistic, whereas Behar *et al.*<sup>44</sup> implemented PAML 4.4.<sup>69</sup>

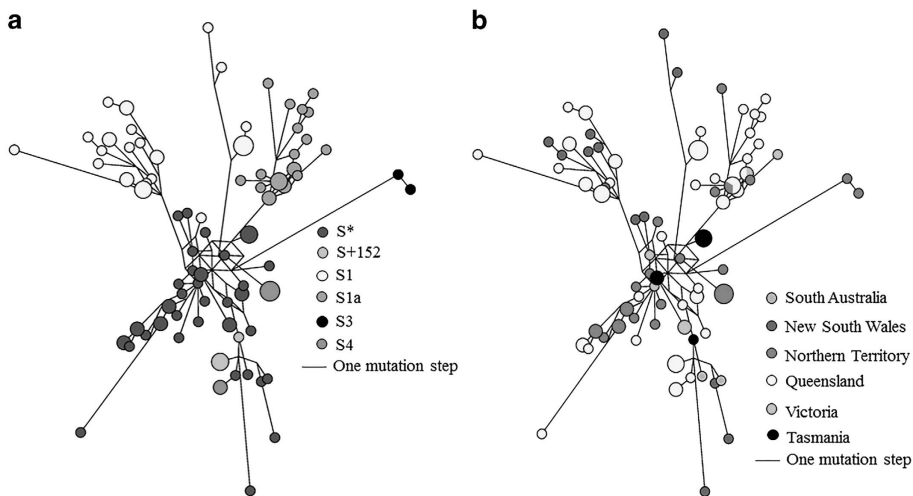
The coalescence of the haplogroups based on the rate of Soares *et al.*<sup>63</sup> fit reasonably well with the currently accepted latest date of the colonization of Sahul based on archeological evidence, 47 KYA.<sup>4</sup> However, haplogroups S and M42a are older than 47 KY



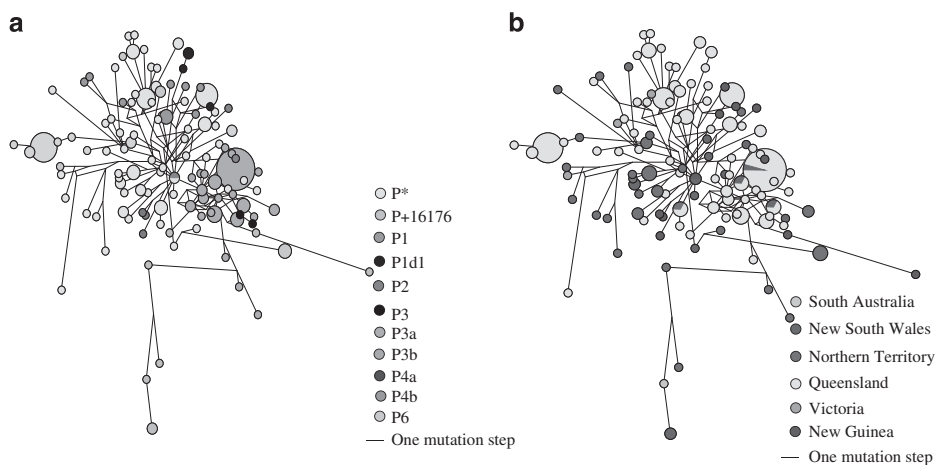
**Figure 3** Median-joining networks based on 68 haplogroup M and Q HVS-I haplotypes from the present study and published M14, M15 and M42a haplotypes. Nodes represent haplotypes, with areas proportional to their frequency. Lines represent mutational steps between haplotypes. Nodes are colored according to (a) (sub)haplogroup affiliation and (b) geographic origin. A full color version of this figure is available at the *Journal of Human Genetics* journal online.



**Figure 4** Median-joining networks based on 24 haplogroup N and O HVS-I haplotypes from the present study and published haplotypes. Nodes represent haplotypes, with areas proportional to their frequency. Lines represent mutational steps between haplotypes. Nodes are colored according to (a) (sub) haplogroup affiliation and (b) geographic origin. A full color version of this figure is available at the *Journal of Human Genetics* journal online.



**Figure 5** Median-joining networks based on 77 haplogroup S HVS-I and HVS-II haplotypes from the present study and published haplotypes. Nodes represent haplotypes, with areas proportional to their frequency. Lines represent mutational steps between haplotypes. Networks are colored according to (a) (sub) haplogroup affiliation and (b) geographic origin. A full color version of this figure is available at the *Journal of Human Genetics* journal online.



**Figure 6** Median-joining networks based on 109 haplogroup P HVS-I and HVS-II haplotypes from the present study and published haplotypes. Nodes represent haplotypes, with areas proportional to their frequency. Lines represent mutational steps between haplotypes. Networks are colored according to (a) (sub)haplogroup affiliation and (b) geographic origin. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

**Table 2 TMRCA estimates for indigenous haplogroups analyzed in this and other published studies**

Haplogroup	n	$\rho \pm \sigma^a$	van Holst Pellekaan <sup>1,4,b</sup>	van Holst Pellekaan <sup>1,4,c</sup>	Behar <i>et al.</i> <sup>44,d</sup>	Hudjashov <i>et al.</i> <sup>6,e</sup>
S*	103	49 500 ± 10 200	40 000 ± 6000	54 000 ± 8000	53 500 ± 5500	25 000 ± 500
S1	33	34 600 ± 12 100	—	—	—	22 000 ± 8000
S1a	22	20 800 ± 11 000	32 000 ± 9000	44 000 ± 12 000	—	—
S+152	5	13 000 ± 8100	—	—	—	—
M*(xM42a)	66	45 700 ± 9900	—	—	—	—
M42a	43	55 200 ± 16 700	34 000 ± 6000	46 000 ± 9000	—	40 600 ± 9000
Q	12	53 100 ± 12 900	—	—	37 500 ± 5600	—
O	18	39 500 ± 13 400	42 800 ± 9000	57 800 ± 13 000	52 000 ± 6400	16 900 ± 7000
P*	201	57 400 ± 11 700	—	—	54 800 ± 2500	—
P+16 176	21	33 300 ± 16 700	—	—	—	—
P1	6	38 900 ± 13 000	—	—	32 900 ± 6200	—
P3	66	15 400 ± 6700	—	—	41 300 ± 5300	—
P3a	57	12 700 ± 6700	—	—	25 000 ± 7100	—
P4b	9	48 400 ± 15 500	52 100 ± 10 000	70 300 ± 14 000	40 100 ± 7100	47 300 ± 12 000

Abbreviation: TMRCA, time to most recent common ancestor.  
<sup>a</sup>One mutation every 16 677 years (Soares *et al.*<sup>63</sup>).  
<sup>b</sup>One substitution in 3810 years (Ingman and Gyllenstein<sup>31</sup>).  
<sup>c</sup>One substitution in 5140 years (Mishmar *et al.*<sup>68</sup>).  
<sup>d</sup>One substitution in 3624 years (Soares *et al.*<sup>63</sup>).  
<sup>e</sup>One synonymous substitution in 6764 years (Kivisild *et al.*<sup>35</sup>).

(Table 2). The age of haplogroup P is consistent with it being one of the oldest mtDNA lineages in the region, evolving within Sunda just before human populations entering Sahul via New Guinea.

The TMRCA of indigenous haplogroups that are much older than 47 KY can be explained by at least two scenarios. In the first, haplogroups S, O and M42a haplogroups evolved in group(s) living in Sunda shortly before migration to Australia and subsequently became extinct, or are yet to be found in modern populations in ISEA. Alternatively, Australia was settled considerably earlier than 47 KYA and evidence of this settlement may lie submerged under the Timor and Arafura Seas off the northern West Australian coast, because of a subsequent rise in sea level. We consider the second scenario to be much more likely. All Australian-specific mtDNA haplogroups analyzed in the present study are considered ancient and geographically widespread across the continent. Resolving the competing hypotheses on settlement dates requires improvement in the accuracy of TMRCA estimates, and this may be done through the resequencing of more mtDNA genomes and the use of other mutation rates, such as that of Fu *et al.*<sup>70</sup> which is based on an ancient DNA calibration.

The maternal ancestors of the Aboriginal Australians carried M\* and N\* (including R\*) lineages through Sunda into Sahul most probably between 40 and 60 KYA (according to median values calculated here), as supported by both archeological and paleontological evidence of human activity in Sunda before 40 KYA.<sup>71,72</sup> M haplogroups observed in Aboriginal Australians today appear to be only very distantly related to M lineages outside Australia, including those observed in New Guinea<sup>6,64,73</sup> and India.<sup>74,75</sup> Haplogroup M42a is widely accepted as indigenous to Australia, with M42a distantly related to M42b, which is found in the South Indian tribal groups, but with a divergence time of at least 55 KYA.<sup>74,75</sup>

However, the findings of the present study indicated that the status of haplogroup Q as indigenous to Australia, at least to the Australian mainland, remains contentious.<sup>6</sup> All of the haplogroup Q individuals in this study identified themselves as possessing Torres Strait Islander maternal ancestry, mainly from the islands of Mer (Murray), Darnley and St Pauls. Torres Strait Islanders have a distinct culture, combining elements from New Guinea, Aboriginal Australia, the Solomon Islands and Melanesia.<sup>76</sup> Although haplogroups Q1 and Q3 were present in Torres Strait Islanders, we did not find the West Australian Aboriginal Q2b subtype observed in a single individual by Hudjashov *et al.*<sup>6</sup>

The presence of Q1 and Q3 in Australia thus appears to be limited to groups with ancestry in the Torres Strait and possibly communities of Far North Queensland that have regular contact with the former.

Those haplotypes that could not be assigned to a particular M subhaplogroup and were denoted M\*(xM42a,Q) are considered indigenous to Australia, as they could not be assigned to M lineages found in any of the neighboring populations (such as M29 found in New Guinea).<sup>6,34,64</sup> M lineages are also virtually absent from the European populations that have settled Australia during the last three centuries. These M\*(xM42a,Q) individuals most likely carry completely novel M subhaplogroups restricted to Australia, or possess variants of known Australian M subhaplogroups, such as M14 or M15. Some, if not all, individuals denoted as M\*(xM42a,Q) may belong to the M42\*(xM42a) haplogroup identified by Ballantyne *et al.*<sup>41</sup> as supported by the matching of haplotypes of the two respective haplogroups. However, the present study did not screen samples for the nucleotide position 9156 SNP that characterizes haplogroup M42, and, therefore, it is uncertain whether these individuals belong to this lineage. Accordingly, these samples require full mitogenome sequencing to determine their detailed haplogroup affiliation.

Haplogroups within macrohaplogroup N account for the majority of Aboriginal lineages. Haplogroups N13, S and O have so far not been found outside Australia, although some studies focusing on populations in proximity to Australia have neglected to type SNPs diagnostic of these haplogroups because of their assumed restricted distribution.<sup>58,77</sup> Given the high frequency and wide distribution of haplogroup S within Australia, it is thought that this haplogroup evolved on the background of an ancestral haplogroup N\* mitogenome within Australia, and has spread throughout the continent, to as far south as Tasmania,<sup>38</sup> which was connected to the Australian mainland by a land bridge until ~ 12 KYA.<sup>78</sup> Tasmanian haplogroup S HVS-I and HVS-II haplotypes observed in the present study are identical to that of 'Fanny Cochrane'.<sup>38</sup> Interestingly, this Tasmanian haplogroup S haplotype was not found in mainland Australia, suggesting the potential isolation of Tasmanian Aboriginals from communities of mainland Australia.<sup>78</sup>

Haplogroup P evolved in Sunda just before its carriers moved into Sahul, with the P lineages (P9 and P10) being found in the Aeta or Agta people of the Philippines.<sup>77,79,80</sup> It is, therefore, most likely that P was brought into Sahul via present-day New Guinea some 55–60 KYA



and, eventually, some of the descendants spread into Australia, which today contains a number of indigenous P subtypes (P5, P6, P7 and P8), as well as some distantly shared with New Guinea (P3 and P4).<sup>6,22,31,81</sup> The genotyping protocol used in this study was such that P5, P6, P7 and P8 were not discernable from mtDNA SNP and HVS sequence data. The two Tasmanian haplogroup P HVS-I and HVS-II sequences were identical to that of the 'Wyerlooberer' lineage (Pipers River tribe) of Tasmania.<sup>38</sup>

Haplogroup P3 and P4 subtypes have been found in Australia as well as in New Guinea, with P3a and P4b being restricted to Australia and P3b and P4a restricted to New Guinea, suggesting that these respective subtypes evolved *in situ* within Australia and New Guinea separately.<sup>6,14</sup> The divergence of New Guinean and Aboriginal Australian P mtDNAs is estimated to have occurred at least 30 KYA,<sup>22,31,36,81</sup> despite that the two groups inhabiting the single landmass of Sahul until around 6 to 8 KYA, when they became separated.<sup>82,83</sup> The resequencing of Aboriginal Australian Y chromosomes supported the ancient connection between New Guineans and Aboriginal Australians implied by mtDNA evidence, indicating that they had diverged at least 48 KYA.<sup>40</sup> Additionally, whole-genome analysis of present-day Aboriginal Australian and New Guinean populations suggests that the two populations diverged at least 37 KYA.<sup>13</sup> The persistence of haplogroups P and Q and their subtypes in the indigenous populations of Sunda and Sahul,<sup>58,79,81</sup> therefore, suggests a shared but deep ancestry between populations of the two former landmasses.

There are two main hypotheses concerning the routes taken by the founding populations to colonize Sahul. The postulated Northern Route took seafarers from Sunda via passage across Wallacea to Sulawesi from which the colonists crossed a series of smaller islands before arriving at either the Bird's Head Peninsula (New Guinea) or the exposed Sahul shelf near the islands of Aru, to the west of New Guinea.<sup>11</sup> Under the Southern Route hypothesis, seafarers from the southern tip of Sunda moved along the island chain of Nusa Tenggara to the island of Timor. From Timor, the ancestors crossed the ~ 80 km wide Timor Trough to reach the then exposed northwestern tip of Australia (now submerged under the Timor and Arafura Seas),<sup>11</sup> hence bypassing New Guinea.<sup>4,84</sup>

Hudjashov *et al.*<sup>6</sup> postulated a single route of migration with entry to New Guinea and subsequent spread into Australia based on both Y-chromosome and mtDNA data. The findings of the present study give support to settlement via both the Northern and Southern routes, on the basis that some mtDNA lineages are restricted to Australia, while others are shared with New Guinea. A similar picture is seen for Y-chromosome variation in Aboriginal Australians, with both Australian-specific Y lineages (C-M347 and S-P308) and lineages shared with New Guinea (K-M526 and M-M186)<sup>20,39,40</sup> being observed.

Using ethnographic, anthropological, archeological and genetic data, Williams<sup>5</sup> modeled the number of initial colonizers of Australia to be at least 1000 to 2000 individuals who arrived ~ 50 KYA. Under this scenario, autochthonous lineages such as M42a, S and O and potentially M14, M15 and N13 evolved within Australia some time after the arrival of the ancestors carrying M\* and N\* mitogenomes into Northern Australia. However, the entry point to Australia of the ancestors carrying P\*, P3\* and P4\* is much more likely to have been via present-day New Guinea. Most importantly, under either model of migration, there is no evidence of back migration of females carrying Australian-specific mtDNA lineages into New Guinea after they arose within Australia.<sup>14,23,25,31,84</sup>

What adds to the complexity of any postulated scenario is the possible effects of the expansion of Austronesian-speakers throughout ISEA and then via coastal New Guinea into Oceania and the Pacific some 8 to 12 KYA.<sup>58</sup> This movement of people was so extensive that it might have replaced many of the ancient mtDNA genomes in those regions, further confounding attempts to retrace migration routes of the ancestors of Aboriginal Australians through these regions. The data from this study suggest a marked isolation between New Guinea and Australia despite their physical contiguity. When comparing the mtDNA haplogroup composition of populations of ISEA,<sup>58</sup> East Timor,<sup>47</sup> New Guinea<sup>59</sup> and Australia, the MDS analysis suggested that Aboriginal Australians are distant from ISEAs and Timorese, with their nearest neighbors being New Guineans (Supplementary Figure 1).

Controversy remains as to the extent that Aboriginal Australians remained isolated until historical times. It has been proposed that, during the mid-Holocene, the appearance and extensive use of microliths, the spread of the Pama-Nyungan language throughout the Australian continent (except for the far north of the Northern Territory) and the arrival of the dingo could be attributed to a South Asian invasion event.<sup>85-87</sup> A recent genetic analysis also suggested that ~ 11% of Aboriginal Australian genomes could be attributed to an Indian invasion event that dated to ~ 4 KYA.<sup>8</sup> However, this Indian connection has been challenged by both anthropological and archeological evidence, as discussed in Brown.<sup>88</sup> Archeological evidence suggests that microlith use by Aboriginal Australians occurred well before the mid-Holocene and was not introduced from outside the region.<sup>45</sup> In addition, the origin and spread of the Pama-Nyungan language family ~ 5 KYA remains uncertain, although its origin is thought to be indigenous as it shares distant ties with languages in New Guinea, rather than those of India.<sup>89</sup>

Furthermore, there has been increased interest in deciphering the origins of the Australian dog, the dingo.<sup>90-95</sup> Skeletal evidence from sites in New South Wales, Victoria and Western Australia suggested that the dingo was introduced into Australia during the Holocene ~ 3.5 KYA,<sup>96-98</sup> possibly by the Toolean or other hunter-gathers from South Sulawesi.<sup>92</sup> Yet, it is possible that those responsible for the introduction of the dingo left human genetic footprints as well. The findings from mtDNA analyses (including the present study) and Y-chromosome studies<sup>7,39,40</sup> clearly exclude introgression of these loci from groups living in Sulawesi, southern India or elsewhere during the mid-Holocene into the Aboriginal Australians sampled.

Evidence from mtDNA and Y-chromosomal DNA further indicates that the settlement of Australia is both ancient and complex. In short, it cannot be fully explained by postulating a single ancestral population entering Sahul some ~ 50 KYA, most likely via present-day New Guinea, and subsequently dispersing into the rest of the landmass.<sup>6,7</sup> Instead, it is more likely that the founding populations that moved into Sunda fissioned, and the resulting sub-populations went in different directions, resulting in a variety of entry points into Sahul. Consequently, these ancient events are reflected today in the observation that different regions of Sahul have unique as well as shared mtDNA lineages.

## CONCLUSION

In summary, this report presents the largest and most comprehensive study of mtDNA lineages of Aboriginal Australians, one that has revealed high levels of mtDNA diversity, with structure found among the States and the Northern Territory. We found no evidence for subsequent introgression events by people from outside Australia (except New Guinea), before the start of European settlement in the eighteenth century, either by Austronesians and/or Southern Indians as postulated previously. Full mitogenome sequencing is required

to determine whether many unresolved mtDNA haplogroup M\*(xM42a,Q), N\*, S\* and P\* individuals found in our study comprise other (sub) novel haplogroups. This will provide a fuller understanding of the mtDNA history of Aboriginal Australians.

### Web Resources

Accession numbers and URLs for data presented herein are as follows: Fluxus-engineering.com, <http://www.fluxus-engineering.com/sharenet.htm> (for Network 4.6 software package). GenBank, <http://www.ncbi.nlm.nih.gov/genbank/> (for sequences (accession numbers KY192710–KY193685)).

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

This work was supported by National Geographic Society, IBM, Waitt Family Foundation. We gratefully acknowledge the participation of Aboriginal Australians from Victoria, Queensland, the Northern Territory, South Australia, Western Australia and Tasmania whose collaboration made this study possible. We owe Tammy Williams and Jason Tatipata many thanks for their support throughout this study. We also thank the jurisdictional forensic departments for access to their samples. YLX and CTS were supported by The Wellcome Trust (098051), and MK, MvO and KNB were supported by Erasmus MC.

### THE GEOGRAPHIC CONSORTIUM

Syama Adhikarla<sup>11</sup>, Christina J Adler<sup>12</sup>, Elena Balanovska<sup>13</sup>, Oleg Balanovsky<sup>13</sup>, Jaime Bertranpetit<sup>14</sup>, Andrew C Clarke<sup>15</sup>, David Comas<sup>14</sup>, Alan Cooper<sup>12</sup>, Clio SI Der Sarkissian<sup>12</sup>, Matthew C Dulik, Jill B Gaieski, ArunKumar GaneshPrasad<sup>11</sup>, Wolfgang Haak<sup>12</sup>, Marc Haber<sup>14,16</sup>, Angela Hobbs<sup>17</sup>, Asif Javed<sup>18</sup>, Li Jin<sup>19</sup>, Matthew E Kaplan<sup>20</sup>, Shilin Li<sup>19</sup>, Begoña Martínez-Cruz<sup>14</sup>, Elizabeth A Matisoo-Smith<sup>15</sup>, Marta Melé<sup>14</sup>, Nirav C Merchant<sup>20</sup>, Amanda C Owings, Laxmi Parida<sup>18</sup>, Ramasamy Pitchappan<sup>11</sup>, Daniel E Platt<sup>18</sup>, Lluís Quintana-Murci<sup>21</sup>, Colin Renfrew<sup>22</sup>, Ajay K Royyuru<sup>18</sup>, Arun Varatharajan Santhakumari<sup>11</sup>, Fabrício R Santos<sup>23</sup>, Himla Soodyall<sup>17</sup>, David F Soria Hernanz<sup>24</sup>, Pandikumar Swamikrishnan<sup>25</sup>, Miguel G Vilar<sup>24</sup>, R Spencer Wells<sup>16</sup>, Pierre A Zalloua<sup>26</sup> and Janet S Ziegler<sup>26</sup>  
<sup>11</sup>Madurai Kamaraj University, Madurai, Tamil Nadu, India; <sup>12</sup>University of Adelaide, South Australia, Australia; <sup>13</sup>Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia; <sup>14</sup>Universitat Pompeu Fabra, Barcelona, Spain; <sup>15</sup>University of Otago, Dunedin, New Zealand; <sup>16</sup>Lebanese American University, Chouran, Beirut, Lebanon; <sup>17</sup>National Health Laboratory Service, Johannesburg, South Africa; <sup>18</sup>IBM, Yorktown Heights, NY, USA; <sup>19</sup>Fudan University, Shanghai, China; <sup>20</sup>University of Arizona, Tucson, AZ, USA; <sup>21</sup>Institut Pasteur, Paris, France; <sup>22</sup>University of Cambridge, Cambridge, United Kingdom; <sup>23</sup>Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; <sup>24</sup>National Geographic Society, Washington, DC, USA; <sup>25</sup>IBM, Somers, NY, USA and <sup>26</sup>Applied Biosystems, Foster City, CA, USA.

- 7 Kayser, M., Brauer, S., Weiss, G., Schiefenhover, W., Underhill, P. A. & Stoneking, M. Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* **68**, 173–190 (2001).
- 8 Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M. & Stoneking, M. Genome-wide data substantiate Holocene gene flow from India to Australia. *Proc. Natl Acad. Sci. USA* **110**, 1803–1808 (2013).
- 9 Huxley, T. H. On the geographical distribution of the chief modifications of mankind. *J. Ethnol. Soc. Lond.* **2**, 404–412 (1870).
- 10 Bellwood, P. *Man's Conquest of the Pacific: The Prehistory of Southeast Asia and Oceania* (Collins, Auckland, New Zealand, 1978).
- 11 Birdsell, J. in *Sunda and Sahul: Prehistoric Studies in Southeast Asia, Melanesia and Australia* (eds Allen, J., Golson, J. & Jones, R.) 113–167 (Academic Press, London, UK, 1977).
- 12 O'Connell, J. F. & Allen, J. The restaurant at the end of the universe: mthe colonisation of Sahul. *Aust. Archaeol.* **74**, 5–17 (2012).
- 13 Malaspinas, A. S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I. *et al.* A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016).
- 14 van Holst Pellekaan, S. Genetic evidence for the colonization of Australia. *Quat. Int.* **285**, 44–56 (2013).
- 15 Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
- 16 Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- 17 McEvoy, B. P., Lind, J. M., Wang, E. T., Moyzis, R. K., Visscher, P. M., van Holst Pellekaan, S. M. *et al.* Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. *Am. J. Hum. Genet.* **87**, 297–305 (2010).
- 18 Vandenberg, N., van Oorschot, R. A., Tyler-Smith, C. & Mitchell, R. J. Y-chromosome-specific microsatellite variation in Australian aboriginals. *Hum. Biol.* **71**, 915–931 (1999).
- 19 Redd, A. J., Roberts-Thomson, J., Karafet, T., Bamshad, M., Jorde, L. B., Naidu, J. M. *et al.* Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. *Curr. Biol.* **12**, 673–677 (2002).
- 20 Karafet, T. M., Mendez, F. L., Sudoyo, H., Lansing, J. S. & Hammer, M. F. Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. *Eur. J. Hum. Genet.* **23**, 369–373 (2015).
- 21 van Holst Pellekaan, S., Frommer, M., Sved, J. & Boettcher, B. Mitochondrial control-region sequence variation in aboriginal Australians. *Am. J. Hum. Genet.* **62**, 435–449 (1998).
- 22 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).
- 23 Redd, A. J. & Stoneking, M. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. *Am. J. Hum. Genet.* **65**, 808–828 (1999).
- 24 Cann, R. L., Stoneking, M. & Wilson, A. C. Mitochondrial DNA and human evolution. *Nature* **325**, 31–36 (1987).
- 25 Stoneking, M., Jorde, L. B., Bhatia, K. & Wilson, A. C. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**, 717–733 (1990).
- 26 Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507 (1991).
- 27 Lum, J. K., Rickards, O., Ching, C. & Cann, R. L. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum. Biol.* **66**, 567–590 (1994).
- 28 Betty, D. J., Chin-Atkins, A. N., Croft, L., Sraml, M. & Easta, S. Multiple independent origins of the COII/tRNA(Lys) intergenic 9-bp mtDNA deletion in aboriginal Australians. *Am. J. Hum. Genet.* **58**, 428–433 (1996).
- 29 Lum, J. K. & Cann, R. L. mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *Am. J. Phys. Anthropol.* **113**, 151–168 (2000).
- 30 Tommaseo-Ponetta, M., Attimonelli, M., De Robertis, M., Tanzariello, F. & Saccone, C. Mitochondrial DNA variability of West New Guinea populations. *Am. J. Phys. Anthropol.* **117**, 49–67 (2002).
- 31 Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13**, 1600–1606 (2003).
- 32 Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
- 33 Huoponen, K., Schurr, T. G., Chen, Y. & Wallace, D. C. Mitochondrial DNA variation in an aboriginal Australian population: evidence for genetic isolation and regional differentiation. *Hum. Immunol.* **62**, 954–969 (2001).
- 34 Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62 (2001).
- 35 Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K. *et al.* The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373–387 (2006).
- 36 Friedlaender, J. S., Friedlaender, F. R., Hodgson, J. A., Stoltz, M., Koki, G., Horvat, G. *et al.* Melanesian mtDNA complexity. *PLoS ONE* **2**, e248 (2007).
- 37 Taylor, D., Nagle, N., Ballantyne, K. N., van Oorschot, R. A., Wilcox, S., Henry, J. *et al.* An investigation of admixture in an Australian Aboriginal Y-chromosome STR database. *Forensic Sci. Int. Genet.* **6**, 532–538 (2012).
- 38 Presser, J. C., Stoneking, M. & Redd, A. J. Tasmanian aborigines and DNA. *Pap. Proc. R. Soc. Tasmania* **136**, 35–38 (2002).

- 1 Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036 (2005).
- 2 Oppenheimer, S. A single southern exit of modern humans from Africa: before or after Toba? *Quat. Int.* **258**, 88–99 (2012).
- 3 Stringer, C. Palaeoanthropology. Coasting out of Africa. *Nature* **405**, 24–25 27 (2000).
- 4 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).
- 5 Williams, A. N. A new population curve for prehistoric Australia. *Proc. Biol. Sci.* **280**, 20130486 (2013).
- 6 Hudjashov, G., Kivisild, T., Underhill, P. A., Endicott, P., Sanchez, J. J., Lin, A. A. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl Acad. Sci. USA* **104**, 8726–8730 (2007).

- 39 Nagle, N., Ballantyne, K. N., van Oven, M., Tyler-Smith, C., Xue, Y., Taylor, D. *et al.* Antiquity and diversity of aboriginal Australian Y-chromosomes. *Am. J. Phys. Anthropol.* **159**, 367–381 (2016).
- 40 Bergstrom, A., Nagle, N., Chen, Y., McCarthy, S., Pollard, M. O., Ayub, Q. *et al.* Deep roots for Aboriginal Australian Y chromosomes. *Curr. Biol.* **26**, 809–813 (2016).
- 41 Ballantyne, K. N., van Oven, M., Ralf, A., Stoneking, M., Mitchell, R. J., van Oorschot, R. A. *et al.* mtDNA SNP multiplexes for efficient inference of matrilineal genetic ancestry within Oceania. *Forensic Sci. Int. Genet.* **6**, 425–436 (2012).
- 42 ABS. *ABoS 2011 Estimates of Aboriginal and Torres Strait Islander Australians* (Australian Bureau of Statistics (ABS), Canberra, 2011).
- 43 van Holst Pellekaan, S. M. *Origins of the Australian and New Guinean Aborigines. Encyclopedia of Life Sciences* (John Wiley and Sons, Chichester, UK, 2008).
- 44 Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogvali, E. L., Silva, N. M. *et al.* A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
- 45 Hiscock, P. *The Archaeology of Ancient Australia* (Routledge, Sydney, NSW, Australia, 2008).
- 46 Horton, D. R. *Cartographer Aboriginal Australia Wall Map* (Aboriginal Studies Press, Canberra, ACT, Australia, 1996).
- 47 Gomes, S. M., Bodner, M., Souto, L., Zimmermann, B., Huber, G., Strobl, C. *et al.* Human settlement history between Sunda and Sahul: a focus on East Timor (Timor-Leste) and the Pleistocene mtDNA diversity. *BMC Genomics.* **16**, 70 (2015).
- 48 de Vries, H. G., Collee, J. M., van Veldhuizen, M. H., Achterhof, L., Smit Sibinga, C. T., Scheffer, H. *et al.* Validation of the determination of deltaF508 mutations of the cystic fibrosis gene in over 11 000 mouthwashes. *Hum. Genet.* **97**, 334–336 (1996).
- 49 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
- 50 Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
- 51 Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
- 52 Kloss-Brandstatter, A., Pacher, D., Schonherr, S., Weissensteiner, H., Binna, R., Specht, G. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).
- 53 van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E338–E394 (2009).
- 54 Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
- 55 Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
- 56 Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, NY, USA, 1987).
- 57 IBM. *SPSS v.21.0 for Windows* (IBM, Chicago IL, 2012).
- 58 Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P. B. *et al.* A mitochondrial stratigraphy for island southeast Asia. *Am. J. Hum. Genet.* **80**, 29–43 (2007).
- 59 van Oven, M., Brauer, S., Choi, Y., Ensing, J., Schiefelhovel, W., Stoneking, M. *et al.* Human genetics of the Kula Ring: Y-chromosome and mitochondrial DNA variation in the Massim of Papua New Guinea. *Eur. J. Hum. Genet.* **22**, 1393–1403 (2014).
- 60 Chen, Y.-S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am. J. Hum. Genet.* **66**, 1362–1383 (2000).
- 61 Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefansson, K., Ward, R. *et al.* mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet.* **68**, 723–737 (2001).
- 62 Forster, P., Harding, R., Torroni, A. & Bandelt, H. J. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* **59**, 935–945 (1996).
- 63 Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
- 64 Merriwether, D. A., Hodgson, J. A., Friedlaender, F. R., Allaby, R., Cerchio, S., Koki, G. *et al.* Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc. Natl Acad. Sci. USA* **102**, 13034–13039 (2005).
- 65 Pierson, M. J., Martinez-Arias, R., Holland, B. R., Gemmell, N. J., Hurles, M. E. & Penny, D. Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Mol. Biol. Evol.* **23**, 1966–1975 (2006).
- 66 Tennant-Kelly, C. Tribes on Cherbourg settlement, Queensland. *Oceania* **5**, 461–473 (1935).
- 67 Radcliffe-Brown, A. R. Former numbers and distribution of Australian Aborigines. *Off. Yearbook* **23**, 687–696 (1930).
- 68 Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hossaini, S. *et al.* Natural selection shaped regional mtDNA variation in humans. *Proc. Natl Acad. Sci. USA* **100**, 171–176 (2003).
- 69 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- 70 Fu, Q., Mitnick, A., Johnson, P. L., Bos, K., Lari, M., Bollongino, R. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- 71 Barker, G., Barton, H., Bird, M., Daly, P., Datan, I., Dykes, A. *et al.* The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52**, 243–261 (2007).
- 72 Demeter, F., Shackelford, L. L., Bacon, A. M., Durringer, P., Westaway, K., Sayavongkhamdy, T. *et al.* Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc. Natl Acad. Sci. USA* **109**, 14375–14380 (2012).
- 73 Kayser, M., Choi, Y., van Oven, M., Mona, S., Brauer, S., Trent, R. J. *et al.* The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol. Biol. Evol.* **25**, 1362–1374 (2008).
- 74 Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B. N., Urade, B. P., Mallick, S. *et al.* Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS ONE* **4**, e7447 (2009).
- 75 Kumar, S., Ravuri, R. R., Koneru, P., Urade, B. P., Sarkar, B. N., Chandrasekar, A. *et al.* Reconstructing Indian-Australian phylogenetic link. *BMC Evol. Biol.* **9**, 173 (2009).
- 76 Greer, S., Henry, R. & McIntyre-Tamwoy, S. Mainland magic: interpreting cultural influences across Cape York–Torres Strait. *Quat. Int.* **385**, 69–78 (2015).
- 77 Tabbada, K. A., Trejaut, J., Loo, J. H., Chen, Y. M., Lin, M., Mirazon-Lahr, M. *et al.* Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol. Biol. Evol.* **27**, 21–31 (2010).
- 78 Jones, R. Tasmanian archeology: establishing the sequences. *Annu. Rev. Anthropol.* **24**, 423–446 (1995).
- 79 Delfin, F., Min-Shan Ko, A., Li, M., Gunnarsdottir, E. D., Tabbada, K. A., Salvador, J. M. *et al.* Complete mtDNA genomes of Filipino ethnolinguistic groups: a melting pot of recent and ancient lineages in the Asia-Pacific region. *Eur. J. Hum. Genet.* **22**, 228–237 (2014).
- 80 Heyer, E., Georges, M., Pachner, M. & Endicott, P. Genetic diversity of four Filipino Negrito populations from Luzon: comparison of male and female effective population sizes and differential integration of immigrants into Aeta and Agta communities. *Hum. Biol.* **85**, 189–208 (2013).
- 81 Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G. *et al.* Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol. Biol. Evol.* **22**, 1506–1517 (2005).
- 82 Chappell, J. Sea level changes forced ice breakouts in the Last Glacial cycle: new results from coral terraces. *Q. Sci. Rev.* **21**, 1229–1240 (2002).
- 83 Lambeck, K. & Chappell, J. Sea level change through the last glacial cycle. *Science* **292**, 679–686 (2001).
- 84 Kayser, M., Brauer, S., Weiss, G., Schiefelhovel, W., Underhill, P., Shen, P. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Genet.* **72**, 281–302 (2003).
- 85 Glover, I. C. & Presland, G. in *Recent Advances in Indo-Pacific Prehistory* (eds Misra, V. N. & Bellwood, P.) 185–195 (Oxford and IBH, New Delhi, India, 1985).
- 86 Gollan, K. in *Recent Advances in Indo-Pacific Prehistory* (eds Misra, V. N. & Bellwood, P.) 439–443 (Oxford and IBH, New Delhi, India, 1985).
- 87 Dixon, R. M. W. *The Languages of Australia* (Cambridge University Press, New York, NY, 1980).
- 88 Brown, P. Palaeoanthropology: of humans, dogs and tiny tools. *Nature* **494**, 316–317 (2013).
- 89 Reesink, G., Singer, R. & Dunn, M. Explaining the linguistic diversity of Sahul using population models. *PLoS Biol.* **7**, e1000241 (2009).
- 90 Ardalan, A., Oskarsson, M., Natanaelsson, C., Wilton, A. N., Ahmadian, A. & Savolainen, P. Narrow genetic basis for the Australian dingo confirmed through analysis of paternal ancestry. *Genetica* **140**, 65–73 (2012).
- 91 Crowther, M. S., Fillios, M., Colman, N. & Letnic, M. An updated description of the Australian dingo (*Canis dingo* Meyer, 1793). *J. Zool.* **293**, 192–203 (2014).
- 92 Fillios, M. A. & Taçon, P. S. C. Who let the dogs in? A review of the recent genetic evidence for the introduction of the dingo to Australia and implications for the movement of people. *J. Archaeol. Sci.* **7**, 782–792 (2016).
- 93 Oskarsson, M. C., Klutsch, C. F., Boonyaparakob, U., Wilton, A., Tanabe, Y. & Savolainen, P. Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proc. Biol. Sci.* **279**, 967–974 (2012).
- 94 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. USA* **101**, 12387–12390 (2004).
- 95 Smith, B. P. & Litchfield, C. A. A review of the relationship between Indigenous Australians, Dingoes (*Canis dingo*) and domestic dogs (*Canis familiaris*). *Anthrozoos* **22**, 111–128 (2009).
- 96 Milham, P. & Thompson, P. Relative antiquity of human occupation and extinct fauna at Madura Cave, southeastern Western Australia. *Mankind* **10**, 175–180 (1976).
- 97 McBryde, I. *Coast and Estuary: Archaeological Investigations on the North Coast of New South Wales at Wombah and Schnapper Point* (Australian Institute of Aboriginal Studies, Canberra, 1982).
- 98 Mulvaney, D., Lawton, G., Twidaie, C., Macintosh, N., Mahoney, J. & Wakefield, N. Archaeological excavation of rock shelter no. 6 Fromm's Landing, South Australia. *Proc. R. Soc. Vic.* **77**, 479–516 (1964).