

# PUBLISHED VERSION

Nano Nagle, Mannis van Oven, Stephen Wilcox, Sheila van Holst Pellekaan, Chris Tyler-Smith, Yali Xue, Kaye N. Ballantyne, Leah Wilcox, Luka Papac, Karen Cooke, Roland A. H. van Oorschot, Peter McAllister, Lesley Williams, Manfred Kayser, R. John Mitchell,

## Aboriginal Australian mitochondrial genome variation - an increased understanding of population antiquity and diversity

Scientific Reports, 2017; 7:1-12

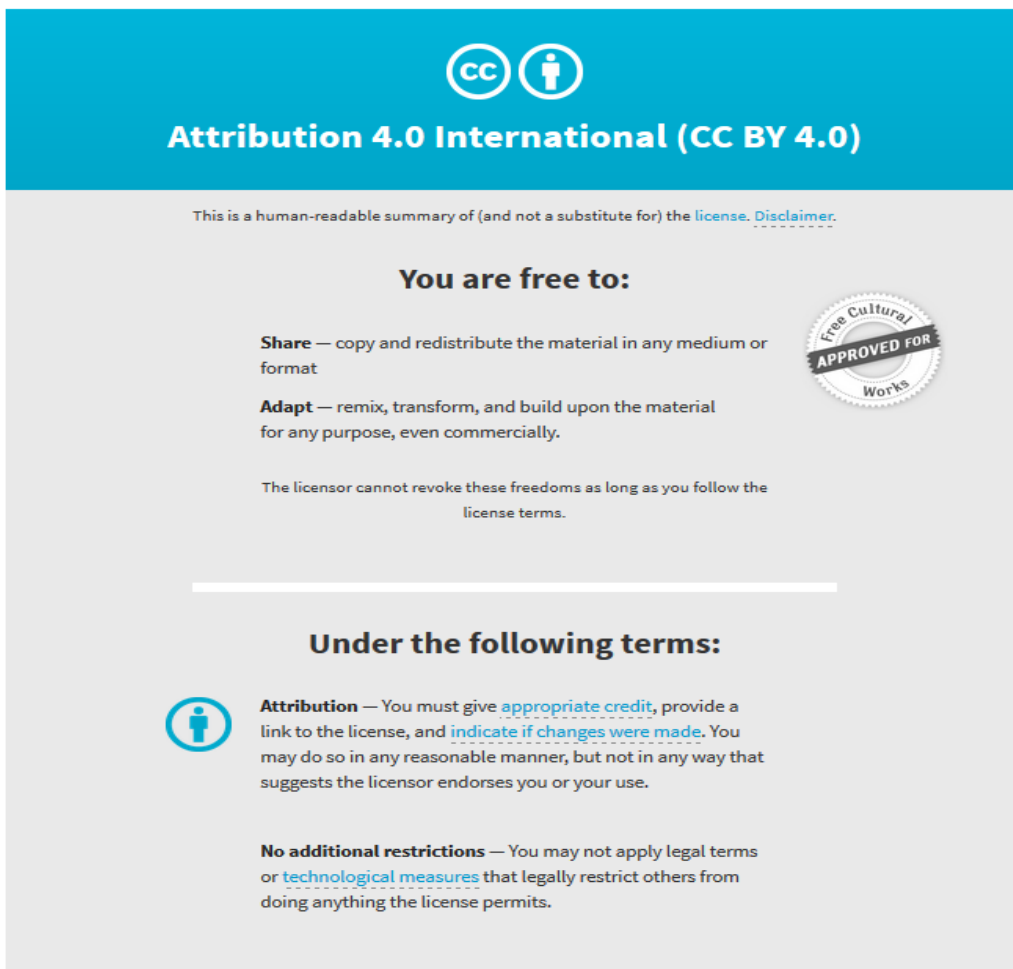
© The Author(s) 2017. This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Originally published at:

<http://doi.org/10.1038/srep43041>

### PERMISSIONS

<http://creativecommons.org/licenses/by/4.0/>



The image is a summary graphic for the Creative Commons Attribution 4.0 International License (CC BY 4.0). It features a blue header with the CC and BY icons and the text "Attribution 4.0 International (CC BY 4.0)". Below the header, it states "This is a human-readable summary of (and not a substitute for) the license. [Disclaimer.](#)". The main content is divided into two sections: "You are free to:" and "Under the following terms:". The "You are free to:" section lists "Share" (copy and redistribute) and "Adapt" (remix, transform, and build upon), with a "Free Cultural Works" seal. The "Under the following terms:" section lists "Attribution" (give appropriate credit) and "No additional restrictions" (do not apply legal terms or technological measures). A horizontal line separates the two sections.

**Attribution 4.0 International (CC BY 4.0)**

This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer.](#)

**You are free to:**

- Share** — copy and redistribute the material in any medium or format
- Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

**Under the following terms:**

- Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

7 November 2017

<http://hdl.handle.net/2440/107808>

# SCIENTIFIC REPORTS



OPEN

## Aboriginal Australian mitochondrial genome variation – an increased understanding of population antiquity and diversity

Received: 11 October 2016

Accepted: 17 January 2017

Published: 13 March 2017

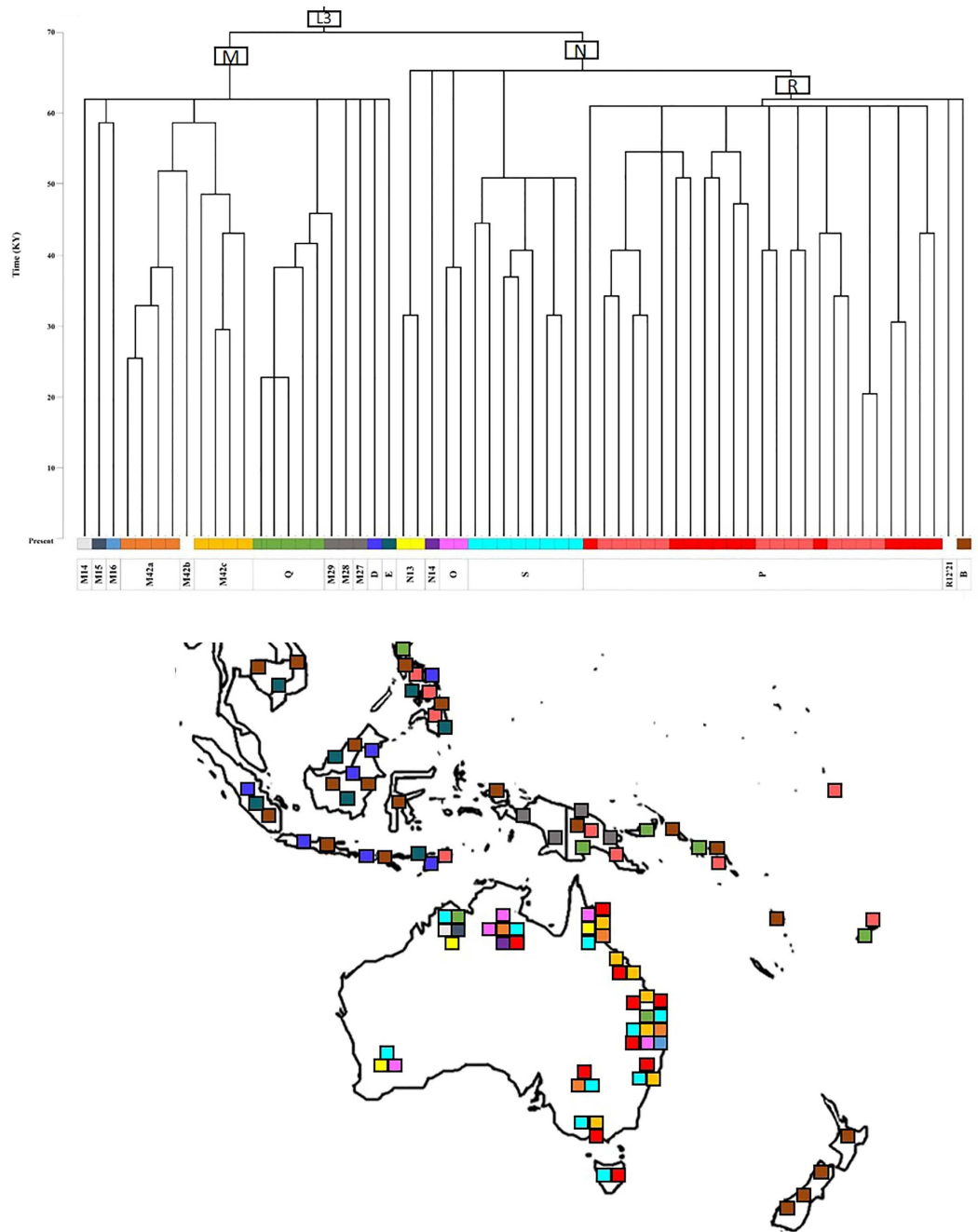
Nano Nagle<sup>1</sup>, Mannis van Oven<sup>2</sup>, Stephen Wilcox<sup>3</sup>, Sheila van Holst Pellekaan<sup>4,5</sup>, Chris Tyler-Smith<sup>6</sup>, Yali Xue<sup>6</sup>, Kaye N. Ballantyne<sup>2,7</sup>, Leah Wilcox<sup>1</sup>, Luka Papac<sup>1</sup>, Karen Cooke<sup>1</sup>, Roland A. H. van Oorschot<sup>7</sup>, Peter McAllister<sup>8</sup>, Lesley Williams<sup>9</sup>, Manfred Kayser<sup>2</sup>, R. John Mitchell<sup>1</sup> & The Genographic Consortium<sup>#</sup>

Aboriginal Australians represent one of the oldest continuous cultures outside Africa, with evidence indicating that their ancestors arrived in the ancient landmass of Sahul (present-day New Guinea and Australia) ~55 thousand years ago. Genetic studies, though limited, have demonstrated both the uniqueness and antiquity of Aboriginal Australian genomes. We have further resolved known Aboriginal Australian mitochondrial haplogroups and discovered novel indigenous lineages by sequencing the mitogenomes of 127 contemporary Aboriginal Australians. In particular, the more common haplogroups observed in our dataset included M42a, M42c, S, P5 and P12, followed by rarer haplogroups M15, M16, N13, O, P3, P6 and P8. We propose some major phylogenetic rearrangements, such as in haplogroup P where we delinked P4a and P4b and redefined them as P4 (New Guinean) and P11 (Australian), respectively. Haplogroup P2b was identified as a novel clade potentially restricted to Torres Strait Islanders. Nearly all Aboriginal Australian mitochondrial haplogroups detected appear to be ancient, with no evidence of later introgression during the Holocene. Our findings greatly increase knowledge about the geographic distribution and phylogenetic structure of mitochondrial lineages that have survived in contemporary descendants of Australia's first settlers.

The human colonisation of Australia occurred relatively soon after the migration of anatomically modern humans out of Northeast Africa some 60 to 80 thousand years ago (KYA)<sup>1–3</sup>. This initial settlement of Australia occurred between 47–55 KYA, based on the dating of archaeological sites dispersed throughout the continent<sup>3–7</sup>, and the analysis of contemporary Aboriginal Australian DNA<sup>8–18</sup>.

Although there is consensus about the time the ancestors of Aboriginal people arrived in the ancient landmass of Sahul (which comprised Australia and New Guinea), there is debate over the route(s) taken by them to reach Sahul<sup>9–11,18–20</sup>. This is mainly because the genetic structure of present-day Aboriginal Australians and New Guineans is different, implying a long separation that started at least 30 KYA<sup>9,15,21</sup>. Did the colonisers enter Sahul via present-day New Guinea and subsequently spread southwards to Australia, or were there different routes into Sahul, such that one or more groups entered Australia via the ancient northwestern coast that is now submerged under the Timor and Arafura Seas<sup>19</sup>? Furthermore, a mix of genetic<sup>22–24</sup>, archeological<sup>25</sup>, anthropological<sup>26,27</sup> and linguistic data<sup>28</sup> have suggested later migration(s) to Australia in the Holocene epoch, particularly from the Asian sub-continent.

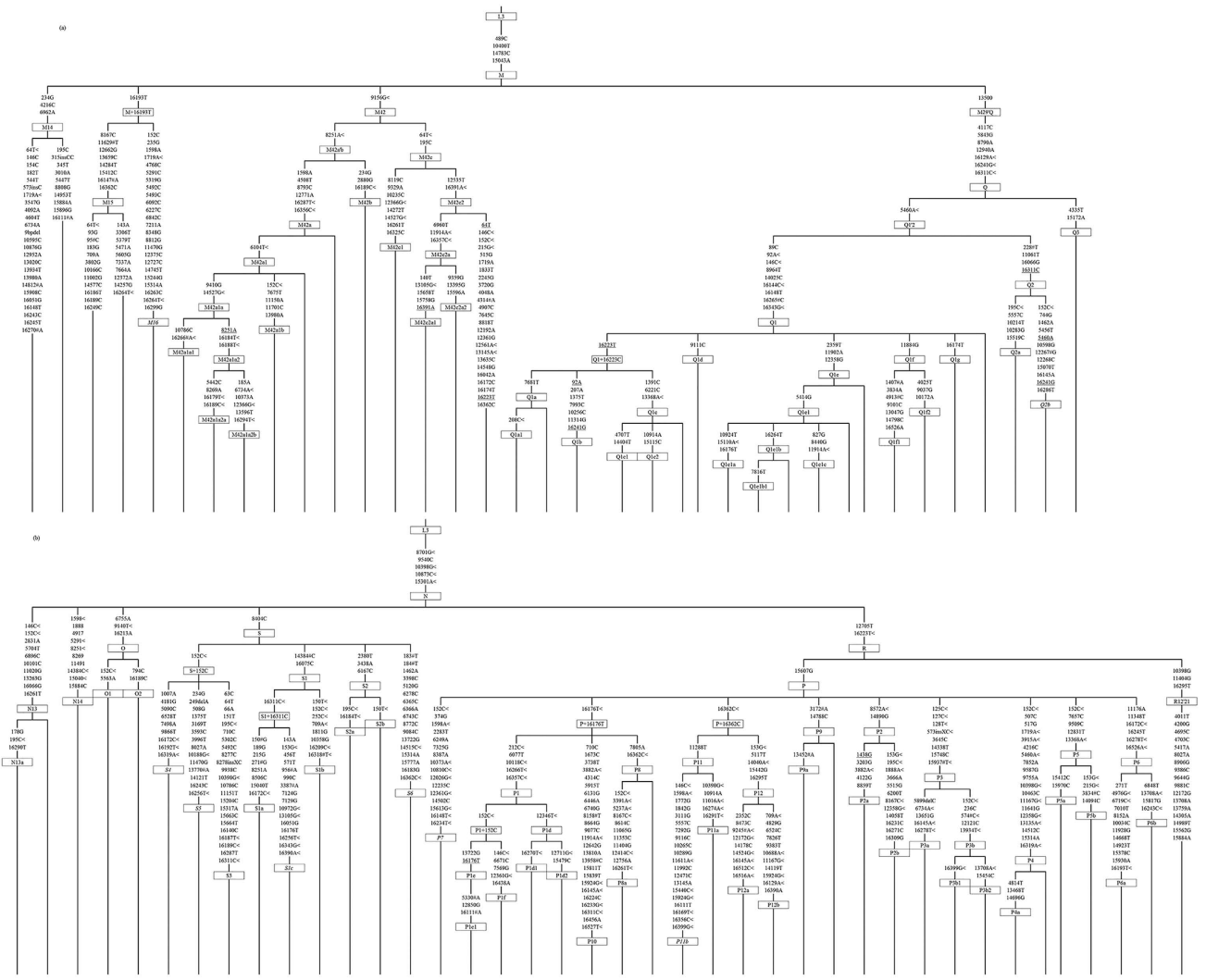
<sup>1</sup>Department of Biochemistry and Genetics, La Trobe Institute for Molecular Sciences, La Trobe University, Melbourne, Victoria, Australia. <sup>2</sup>Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, The Netherlands. <sup>3</sup>Australian Genome Research Facility, Melbourne, Victoria, Australia. <sup>4</sup>Biotechnology and Biomolecular Sciences, University of New South Wales, New South Wales, Australia. <sup>5</sup>School of Biological Sciences, University of Sydney, Sydney, Australia. <sup>6</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom. <sup>7</sup>Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Victoria, Australia. <sup>8</sup>Griffith University, Queensland, Australia. <sup>9</sup>Community Elder and Cultural Advisor, Brisbane, Queensland, Australia. <sup>#</sup>A comprehensive list of authors and affiliations appear at the end of the paper. Correspondence and requests for materials should be addressed to R.J.M. (email: john.mitchell@latrobe.edu.au)



**Figure 1. Phylogeny and geographical distributions of Aboriginal Australian mtDNA lineages and those of the surrounding regions of past and present studies.** Length of branches indicative of time since divergence. Colored squares on map indicate presence of specific haplogroups depicted in phylogeny.

Although mainly based on small sample sizes and few sampling locations, DNA studies have revealed the distinctiveness of Aboriginal Australians. Mitochondrial DNA (mtDNA) lineages of all three major mtDNA clades found outside Africa (macrohaplogroups M, N and R- macrohaplogroup R lineages lie within N; see <http://phylotree.org><sup>29</sup>) are present in Aboriginal Australians. Subsequent to their arrival in Australia, the M founder types diversified into the Australian-specific haplogroups present today, including M42a and M15 (and possibly M14), and the N founder types into N13, N14, O and S as reviewed in ref. 30, whilst the one lineage within macrohaplogroup R, haplogroup P, most probably evolved either in Sunda (the ancient landmass comprising present-day Island Southeast Asia) just prior to the colonisation of Sahul<sup>31–33</sup> or in northern Sahul (New Guinea)<sup>34</sup>. Subsequently, P evolved into sublineages that are unique to New Guinea or to Australia, respectively.

What is apparent from the limited Aboriginal Australian mtDNA data currently available is that it lacks phylogenetic resolution and geographic coverage. Specifically, most studies of mtDNA diversity have focused on populations of northern Australia<sup>9,21,23,35–38</sup>. The exceptions are van Holst Pellekaan, *et al.*<sup>39</sup> and Presser, *et al.*<sup>40</sup>,



**Figure 2.** Schematic tree of (a) macrohaplogroup M and (b) macrohaplogroup N lineages including those specific to Aboriginal Australians. Diagnostic control-region and coding region positions are as indicated.

which analysed samples from New South Wales and Tasmania, respectively. A recent report on the genomic history of Aboriginal Australians<sup>18</sup> included 83 mitogenomes drawn widely from the continent. However, while the mitochondrial haplogroup affiliation of the individuals is given, sequence data were not available.

A relatively large number of Aboriginal mtDNAs have also been reported as haplogroup M, N or P “unclassified” (also denoted as M\*, N\* or P\*). This has occurred because either only the hypervariable control-region segments (HVS-I and HVS-II) were sequenced and/or the mtSNPs genotyped proved insufficient for further subhaplogroup assignment<sup>8,9,17,35,36,41</sup>. The paucity of knowledge is also illustrated by the fact that, to date, there are only 39 Aboriginal Australian mitogenomes in GenBank (0.16% of all human mitogenome sequences as of June, 2016).

The 127 newly described mitogenome sequences of the present study comprise part of the Genographic Project during which samples from Aboriginal Australians were collected for analysis of uniparental DNA variation<sup>14,17</sup>. Participants were drawn especially from those areas of Australia, either previously not sampled (Queensland and Victoria), or very poorly represented (Tasmania). These data were used to further clarify the phylogenies of Aboriginal Australian mitochondrial haplogroups, their distribution within the continent, and potentially make inferences regarding the colonisation and migration routes taken by the maternal ancestors of present-day Aboriginal Australians.

**Results and Discussion**

The samples selected for mitogenome sequencing are a subset within a larger mtDNA sample set described in Nagle, *et al.*<sup>17</sup>. The haplogroup assignments of the present sample based on initial mtSNP genotyping are given in Supplementary Table 1. The frequencies of each haplogroup as well as geographical distributions are very similar to those observed in the much larger Aboriginal sample of mtDNAs reported in Nagle, *et al.*<sup>17</sup> (Supplementary Table 2).

There is uncertainty as to the traditional homeland of many individuals in this study as a result of the great dislocation of Aboriginal society since European colonization started in 1788. This disruption included forced removal from their homelands into settlements, restrictions on marriage and forced removal of children. Accordingly, any attempt at the reconstruction of the historical genetic structure of Aboriginal Australia must be mindful of the past treatment over the last 200 years.

Detailed phylogenetic trees using full mitogenome sequences were constructed for the following haplogroups: a) M14, M15, M16, M42 and Q b) N13, N14, O and S and c) R12 and P (Supplementary Figures 1–3). In addition, a simplified tree showing the overall basic phylogeny and diagnostic mtSNPs of the major Australian sub-branches of macrohaplogroups M and N (including R) are shown in Fig. 2a and b, respectively. Haplogroup nomenclature follows that of PhyloTree Build 17 (van Oven and Kayser, 2009) and logically expands on this nomenclature for newly identified (sub)haplogroups.

**Phylogeny of lineages within macrohaplogroup M.** This tree comprised the 34 haplogroup M mitogenomes of this study and the published M14, M15, M42a, M42b, Q1, Q2 and Q3 mitogenomes<sup>8,9,37,38,42–44</sup> (provisional diagnostic SNPs are presented in Supplementary Figure 1). Notably, construction of the tree revealed a novel subclade within M42 (here labelled M42c) that is, according to available data, unique to Aboriginal Australians. In addition to A9156G, which is a definer of haplogroup M42, M42c is defined by nucleotide positions (np) C64T and T195C, and the absence of G8251A (which is shared by M42b and some M42a mitogenomes). M42c has at least two subtypes, with M42c1 found in Queensland and M42c2 found in Queensland, New South Wales and Victoria (Fig. 1). It is unclear whether the haplogroup labelled 'M42' in Malaspina, *et al.*<sup>18</sup> is identical to M42c because of the lack of sequence data. Distinct subclades were also identified within the previously known haplogroup M42a, namely M42a1 which, in turn, is subdivided into M42a1a and M42a1b (Supplementary Figure 1).

The existence of a novel branch within the M42 clade was first suggested by Ballantyne, *et al.*<sup>35</sup>, in which some samples were labelled M42\* (xM42a) because they carried the transition at np A9156G (the defining mutation of M42) but not G12771A (one of the defining mutations of M42a). Haplogroups M42a and M42c are, by far, the major representatives of haplogroup M in Australia and both appear to be widespread across the continent (Fig. 1).

Haplogroup M42a has previously provoked interest because its apparent sister clade, M42b, was found in tribal group(s) in southern India<sup>42,43</sup> as well as in single individuals from Iran<sup>45</sup>, Saudi-Arabia<sup>46</sup> and Mauritius<sup>46</sup>. Time-to-most-recent-common-ancestor (TMRCA) estimates of M42a and M42b suggest that they diverged some 50 KYA, possibly in India<sup>42</sup>, with the ancestors of Aboriginal Australians continuing southeastward to Sahul<sup>47,48</sup>. However, M42a and M42b are only tenuously linked as they share just two mutations, 9156G and 8251A, with the latter being a well-known mutational hotspot with 26 independent occurrences in PhyloTree Build 17<sup>29</sup>. Indeed, two of the published M42a mitogenomes lack the 8251 transition, suggesting back mutations have occurred at this position<sup>8,42,49</sup>. Hence, the sharing of 8251A between M42b and some of the M42a lineages might be coincidental (identical-by-state) rather than indicative of phylogenetic relatedness (identical-by-descent). In contrast, the A9156G transition appears to be mildly recurrent, with only five occurrences in the entire PhyloTree Build 17. Thus, M42a, and M42c are more likely to be identical-by-descent.

Other indigenous M subhaplogroups in Aboriginal Australians, M15 and possibly M14<sup>9</sup>, appear to be much less frequent than haplogroups M42a and M42c as we did not find haplogroup M14 in our dataset and haplogroup M15 was only observed once. This additional M15 sequence, from Western Australia, together with the only other published haplogroup M15 sequence<sup>9</sup>, helped to refine the diagnostic motif of this haplogroup to nine mutations (Supplementary Figure 1). One individual from Queensland had a novel M mitogenome that shared individual mutations, with existing haplogroups (i.e., M29, M52, M82, M28, M80, M53, M15, M23, M50), but did not show a convincing phylogenetic link to them.

Interestingly, this novel mitogenome shared variant C16193T with haplogroup M15 but there was uncertainty as to whether they shared it by descent and, therefore, this mitogenome was tentatively labelled as haplogroup M16 (Fig. 2a).

Haplogroup Q is of interest because of its estimated age of at least ~37 KY<sup>50</sup> and its relatively high frequency in Australia's nearest neighbours, New Guineans, Timorese and Island Melanesians<sup>21,34,51,52</sup>. This haplogroup has three known subclades; Q1, Q2 and Q3, all of which are found in New Guineans/Island Melanesians, with Q1 and Q3 being additionally present in Timorese<sup>34</sup>. To date, haplogroup Q has been found only in a single Aboriginal Australian person and it was a unique variant within haplogroup Q2, labelled Q2b<sup>9</sup>.

The three haplogroup-Q individuals detected in the current study all belonged to Q1. One of them belonged to a newly proposed subhaplogroup, Q1g, shared with an individual from the Solomon Islands<sup>53</sup> and the other two belonging to haplogroup Q1a (Supplementary Figure 1). While all of them were sampled in Far North Queensland, the three Q individuals were aware they had Torres Strait Island maternal ancestry. Some of the Torres Strait Islands (which lie between Australia and New Guinea) are an integral part of the nation of Australia, with the remainder belonging to Papua New Guinea. Although Aboriginal Australians are culturally and linguistically distinct from Torres Strait Islanders, there are reports of trading and intermarriage<sup>54</sup>. Any such gene flow, however, was not detected in our small sample of mitogenomes.

**Phylogenies of lineages within macrohaplogroup N.** Haplogroups N13 and O show the most geographically restricted distribution of the major observed haplogroups in Australia. Haplogroup N13, however, has a wider distribution than previously thought, with a presence in Queensland as well as Western Australia. One Queensland individual's mitogenome was similar to a sequence from Western Australia<sup>11</sup>, whilst the other individual was more divergent from other N13 mitogenomes. The increased number of haplogroup-N13

mitogenomes has allowed the identification of a novel subclade, N13a, which in turn has a subclade N13a1 (Fig. 2b).

Although haplogroup O has a wide distribution within Australia, being present in Queensland, Western Australia and the Northern Territory, it was not observed in high frequency compared to other widespread Australian haplogroups (Fig. 1). The increased number of haplogroup-O mitogenomes resulted in the identification of a novel subclade within haplogroup O1a (i.e., O1a1) and the newly defined haplogroup O2. Haplogroup O1a was found in all three northern States, while haplogroup O2 was present in the Northern Territory and Queensland (Supplementary Figure 2).

Haplogroup S evolved within Australia and has five recognized subtypes, S1 to S5; although, of these, S4 and S5 are each represented by only one mitogenome sequence<sup>9,37</sup>. However, haplogroup S5 has been reported in Western Australia<sup>18</sup>. Our additional haplogroup-S mitogenomes belonged to S1a, to the newly proposed S1b, to the tentatively defined S1c and to S2 (Fig. 2b). S1 had a wide distribution, being present in Queensland, New South Wales, Western Australia and the Northern Territory. The single S1c subtype was from Queensland, and its provisional diagnostic SNPs are provided in Supplementary Figure 2. S2 also had a wide distribution, being found in all States, including the island of Tasmania (Fig. 3). Novel subclades within haplogroup S2 were identified, and were labelled S2a and S2b (Supplementary Figure 2). Haplogroups S3, S4 and S5 were not observed in our sample. However, we did detect a novel haplogroup S mitogenome (from New South Wales) which we provisionally denoted S6 (Fig. 2b).

**Phylogeny of lineages within macrohaplogroup R.** Haplogroup P is an ancient haplogroup with ten recognized subclades thus far (P1 to P10) (Fig. 2b). The haplogroup is found in Island Southeast Asia (ISEA), in particular the Philippines, and also in New Guinea, Island Melanesia and Australia<sup>9,18,31,32,34,37,38,51,55</sup>.

Although the known two subclades of haplogroup P3 were thought to have separate geographical distributions, with P3a restricted to Australia and the majority of P3b individuals (2 out of 3) to New Guinea<sup>9,37,51</sup>, our results revealed three P3b individuals, who resided in Queensland or New South Wales, but have known Torres Strait Islander maternal ancestry (Supplementary Figure 3).

The subclades of haplogroup P4 have also been thought to show geographical differences, with P4a restricted to New Guinea and P4b to Australia<sup>8,9,37,51</sup>. There was no evidence of P4a in our Aboriginal Australian sample. However, after sequencing five haplogroup P4b individuals it was apparent that P4a and P4b shared too few mutations for both subtypes to remain within the same clade, P4, and that P4b was more parsimoniously joined (through mutation C11288T) with a single haplogroup P lineage from Tasmania, together forming a newly proposed haplogroup, which we designated P11. As a result of this rearrangement, those individuals that were previously assigned to P4b (Aboriginal Australians) have been allocated to the subclade P11 to reflect its uniqueness. Haplogroup P11 has a wide distribution in Australia, with representative mitogenomes from all States where data were available (Fig. 1). The eight mitogenomes of P11 can be further classified into subtypes; P11a and P11b, with P11a found on mainland Australia and P11b represented in a single individual from Tasmania. This new haplogroup assignment also receives support from maximum likelihood analysis (Supplementary Figure 4). As a result of this analysis, the New Guinean-specific haplogroup P4a now becomes P4.

In our sample, there were individuals who carried haplogroup P1e, which was previously found in Timor and New Guinea<sup>34,38</sup>, and New Guinean-specific haplogroup P2<sup>36,37,44,51,55</sup>. Notably, these individuals traced their maternal ancestry to the Torres Strait Islands, and may therefore have (ultimate) maternal ties to the New Guinea mainland. The New Guinean and Torres Strait Islander haplogroup-P1e mitogenomes shared three mutations not present in the Timorese P1e lineage and, therefore, the former has been denoted as a novel clade, P1e1. Interestingly, within haplogroup P2, our new sequences distinguished two subclades, P2a and P2b, with P2a restricted to New Guinea and P2b to the Torres Strait Islanders (Supplementary Figure 3).

This study greatly improved the substructure of haplogroup P5 (previously merely represented by a single sequence from the Northern Territory<sup>51</sup>), with two major subclades which are predominantly found in Queensland: P5a and P5b; the first of which appeared to be more common (Supplementary Figure 3). Four haplogroup-P6 mitogenomes from Victoria and Queensland indicated that this haplogroup is not restricted to the Northern Territory<sup>37</sup>. Our increased sampling further resolved the haplogroup P6 phylogeny, with two subclades (P6a and P6b) evident. There was no evidence of haplogroup P7 in our sample, previously reported in the Northern Territory<sup>37</sup>. The sequencing of an additional haplogroup P8 mitogenome, together with sequence AY289055 previously thought to belong to haplogroup P6, allowed us to propose a basal haplogroup-P8 motif, as well as the recognition of a subtype P8a, and a TMRCA estimate of this haplogroup.

The remaining 13 haplogroup-P mitogenomes belonged to a novel clade, which we denoted P12 (Fig. 2b), with all these individuals residing in Queensland. Two subhaplogroups, P12a and P12b, could also be identified amongst the sequences (Supplementary Figure 3).

In our sample we did not detect haplogroup R12, previously observed in a single individual from the Northern Territory<sup>44</sup> and recently in three individuals in Western Australia<sup>18</sup>.

**TMRCA estimates.** The TMRCA estimates of the Aboriginal Australian haplogroups based on two different mutation rates are shown in Table 1. The mutation rate of Soares, *et al.*<sup>56</sup> reflects an estimated 'evolutionary' mutation rate, whereas that of Fu, *et al.*<sup>57</sup> is calibrated with ancient-DNA samples. The present study's estimates differed from those calculated by Behar, *et al.*<sup>50</sup>, Hudjashov, *et al.*<sup>9</sup> and van Holst Pellekaan<sup>30</sup> due to a difference in methodologies and increased sampling from different regions. Several estimates have large confidence intervals due to increased variability of the mitogenomes as well as low numbers.

The estimates generated by the two methods were similar for most haplogroups, especially for those represented by more than four mitogenomes. Haplogroups represented by less than four mitogenomes showed larger

| Haplogroup | Present Study |                                |              |                                    |             | Other Publications                 |                                       |                                      |                                 |
|------------|---------------|--------------------------------|--------------|------------------------------------|-------------|------------------------------------|---------------------------------------|--------------------------------------|---------------------------------|
|            | n             | Fu <i>et al.</i> <sup>57</sup> |              | Soares <i>et al.</i> <sup>56</sup> |             | Behar <i>et al.</i> <sup>50a</sup> | Hudjashov <i>et al.</i> <sup>9b</sup> | van Holst Pellekaan <sup>30c,d</sup> |                                 |
|            |               | Median (KY)                    | 95% HPD (KY) | Median (KY)                        | 95% CI (KY) | TMRCa ± SD (KY)                    | TMRCa ± SD (KY)                       | TMRCa ± SD (KY) - Coding region      | TMRCa ± SD (KY) - Coding region |
| M15        | 2             | 31                             | 19–43        | 67                                 | 46–89       | —                                  | —                                     | —                                    | —                               |
| M42a       | 15            | 36                             | 27–46        | 39                                 | 31–47       | —                                  | 41 ± 9                                | 34 ± 6                               | 46 ± 9                          |
| M42c       | 19            | 46                             | 35–58        | 43                                 | 29–58       | —                                  | —                                     | —                                    | —                               |
| M42b       | 14            | 46                             | 35–58        | 53                                 | 34–72       | 40 ± 7                             | —                                     | —                                    | —                               |
| M42a'c     | 34            | 53                             | 42–66        | 50                                 | 39–62       | —                                  | —                                     | —                                    | —                               |
| Q          | 34            | 44                             | 34–55        | 53                                 | 37–68       | 37 ± 6                             | 32 ± 6                                | —                                    | —                               |
| Q1         | 18            | 23                             | 17–30        | 24                                 | 17–32       | 18 ± 7                             | 21 ± 6                                | —                                    | —                               |
| Q2         | 11            | 34                             | 24–44        | 27                                 | 20–35       | 29 ± 7                             | 30 ± 9                                | —                                    | —                               |
| Q3         | 5             | 31                             | 22–41        | 35                                 | 27–43       | 31 ± 6                             | 21 ± 5                                | —                                    | —                               |
| N13        | 4             | 32                             | 22–44        | 38                                 | 26–52       | —                                  | —                                     | —                                    | —                               |
| O          | 9             | 37                             | 25–50        | 43                                 | 28–59       | 52 ± 6                             | 17 ± 7                                | —                                    | —                               |
| O1         | 5             | 25                             | 16–36        | 20                                 | 12–29       | 17 ± 8                             | —                                     | —                                    | —                               |
| O1a        | 4             | 12                             | 7–18         | 19                                 | 11–29       | —                                  | —                                     | —                                    | —                               |
| O1a1       | 3             | 7                              | 3–12         | 17                                 | 10–24       | —                                  | —                                     | —                                    | —                               |
| O2         | 4             | 24                             | 14–35        | 22                                 | 13–32       | —                                  | —                                     | —                                    | —                               |
| S          | 48            | 51                             | 40–64        | 49                                 | 39–59       | 53 ± 5                             | 25 ± 5                                | 40 ± 6                               | 54 ± 8                          |
| S1         | 22            | 42                             | 32–53        | 48                                 | 35–62       | —                                  | 22 ± 7                                | —                                    | —                               |
| S1a        | 8             | 22                             | 15–29        | 20                                 | 12–29       | —                                  | —                                     | 32 ± 9                               | 44 ± 12                         |
| S1b        | 13            | 21                             | 15–28        | 29                                 | 22–37       | —                                  | —                                     | —                                    | —                               |
| S1b1       | 5             | 16                             | 10–22        | 20                                 | 11–30       | —                                  | —                                     | —                                    | —                               |
| S1b2       | 5             | 10                             | 5–16         | 9                                  | 2–17        | —                                  | —                                     | —                                    | —                               |
| S1b3       | 3             | 12                             | 7–17         | 12                                 | 4–20        | —                                  | —                                     | —                                    | —                               |
| S2         | 21            | 32                             | 24–42        | 34                                 | 25–44       | 39 ± 9                             | 15 ± 5                                | —                                    | —                               |
| S2a        | 14            | 25                             | 18–34        | 29                                 | 20–38       | —                                  | —                                     | —                                    | —                               |
| S2b        | 7             | 26                             | 18–35        | 30                                 | 19–42       | —                                  | —                                     | —                                    | —                               |
| S3         | 2             | 3                              | 1–7          | 8                                  | 0–17        | 2 ± 2                              | —                                     | —                                    | —                               |
| P          | 112           | 60                             | 50–73        | 62                                 | 54–70       | 55 ± 2                             | 52 ± 6                                | —                                    | —                               |
| P1         | 22            | 38                             | 29–48        | 39                                 | 31–47       | 33 ± 6                             | 30 ± 6                                | —                                    | —                               |
| P1d        | 13            | 32                             | 24–40        | 36                                 | 28–45       | 30 ± 6                             | —                                     | —                                    | —                               |
| P1e        | 8             | 26                             | 16–37        | 30                                 | 17–44       | —                                  | —                                     | —                                    | —                               |
| P1e1       | 3             | 11                             | 5–19         | 14                                 | 6–23        | —                                  | —                                     | —                                    | —                               |
| P1e1a      | 2             | 3                              | 0.3–7        | 3                                  | 0–6         | —                                  | —                                     | —                                    | —                               |
| P2         | 8             | 39                             | 28–50        | 29                                 | 20–38       | —                                  | 13 ± 4                                | —                                    | —                               |
| P2a        | 6             | 23                             | 15–32        | 19                                 | 12–27       | —                                  | —                                     | —                                    | —                               |
| P2b        | 2             | 4                              | 1–8          | 4                                  | 0–8         | —                                  | —                                     | —                                    | —                               |
| P3         | 10            | 43                             | 33–54        | 44                                 | 26–62       | 41 ± 5                             | 39 ± 8                                | —                                    | —                               |
| P3a        | 4             | 26                             | 17–35        | 6                                  | 0–13        | 25 ± 7                             | —                                     | —                                    | —                               |
| P3b        | 6             | 34                             | 26–44        | 35                                 | 24–45       | 35 ± 6                             | —                                     | —                                    | —                               |
| P4         | 5             | 18                             | 12–26        | 22                                 | 14–30       | 53 ± 4                             | 66 ± 13                               | —                                    | —                               |
| P4a        | 3             | 13                             | 8–19         | 20                                 | 10–31       | 19 ± 6                             | 26 ± 7                                | —                                    | —                               |
| P5         | 24            | 31                             | 23–41        | 28                                 | 18–39       | —                                  | —                                     | —                                    | —                               |
| P5a        | 21            | 23                             | 16–30        | 22                                 | 13–30       | —                                  | —                                     | —                                    | —                               |
| P5a1       | 18            | 19                             | 14–25        | 19                                 | 11–27       | —                                  | —                                     | —                                    | —                               |
| P5a2       | 3             | 17                             | 10–24        | 20                                 | 11–30       | —                                  | —                                     | —                                    | —                               |
| P5b        | 3             | 19                             | 12–28        | 21                                 | 10–32       | —                                  | —                                     | —                                    | —                               |
| P6         | 5             | 43                             | 32–56        | 54                                 | 38–70       | 48 ± 7                             | —                                     | —                                    | —                               |
| P6a        | 2             | 12                             | 5–19         | 16                                 | 7–26        | —                                  | —                                     | —                                    | —                               |
| P6b        | 3             | 32                             | 22–43        | 44                                 | 26–63       | —                                  | —                                     | —                                    | —                               |
| P8         | 3             | 49                             | 37–61        | 47                                 | 36–59       | —                                  | —                                     | —                                    | —                               |
| P8a        | 2             | 16                             | 8–25         | 19                                 | 9–30        | —                                  | —                                     | —                                    | —                               |
| P9         | 10            | 40                             | 30–52        | 59                                 | 38–82       | —                                  | —                                     | —                                    | —                               |
| P10        | 4             | 4                              | 1–9          | 7                                  | 2–13        | —                                  | —                                     | —                                    | —                               |
| P11        | 8             | 50                             | 39–62        | 39                                 | 29–48       | —                                  | —                                     | —                                    | —                               |

Continued

| Haplogroup | n  | Present Study                  |              |                                    |             | Other Publications                 |                                       |                                      |                                  |
|------------|----|--------------------------------|--------------|------------------------------------|-------------|------------------------------------|---------------------------------------|--------------------------------------|----------------------------------|
|            |    | Fu <i>et al.</i> <sup>57</sup> |              | Soares <i>et al.</i> <sup>56</sup> |             | Behar <i>et al.</i> <sup>50a</sup> | Hudjashov <i>et al.</i> <sup>9b</sup> | van Holst Pellekaan <sup>30c,d</sup> |                                  |
|            |    | Median (KY)                    | 95% HPD (KY) | Median (KY)                        | 95% CI (KY) | TMRCAs ± SD (KY)                   | TMRCAs ± SD (KY)                      | TMRCAs ± SD (KY) - Coding region     | TMRCAs ± SD (KY) - Coding region |
| P12        | 12 | 46                             | 36–57        | 57                                 | 38–77       | —                                  | —                                     | —                                    | —                                |
| P12a       | 3  | 32                             | 23–42        | 32                                 | 19–45       | —                                  | —                                     | —                                    | —                                |
| P12b       | 9  | 15                             | 9–22         | 15                                 | 7–24        | —                                  | —                                     | —                                    | —                                |

**Table 1. TMRCAs of Aboriginal Australian mitochondrial haplogroups.** <sup>a</sup>One substitution in 3,624 years (Soares *et al.*<sup>56</sup>). <sup>b</sup>One synonymous substitution in 6,764 years (Kivisild *et al.*<sup>44</sup>). <sup>c</sup>One substitution in 3,810 years (Ingman and Gyllensten<sup>37</sup>). <sup>d</sup>One substitution in 5,140 years (Mishmar *et al.*, 2003).

differences. Overall, TMRCAs estimates using Soares, *et al.*<sup>56</sup> mutation rate were older than those generated using that of Fu, *et al.*<sup>57</sup>. The TMRCAs estimates of all the uniquely Aboriginal haplogroups are consistent with the conclusion of the recent review of archaeological evidence that the initial colonisation of Sahul occurred at least 47 KYA<sup>3</sup> with the age estimates using Fu, *et al.*<sup>57</sup> fitting well with this date. The Fu, *et al.*<sup>57</sup> mutation rate may be considered the most accurate since it was calibrated using reliably dated ancient DNA samples taken from modern human remains whose ages span the last 40 KY, and come from different locations throughout the world. Colonisation even earlier than these dates, however, cannot be completely rejected since the large expansion of land formerly comprising the north western part of Australia has been submerged since ~8 KYA<sup>58</sup> and this region may well have been colonised earlier than 47 KYA. In addition, there is an inherent difficulty in finding archaeological evidence of human occupation of Australia before 55 KYA due to the continent's harsh arid environment<sup>3,7,59,60</sup>.

The TMRCAs for haplogroup S is between 49 and 51 KYA and it may have been one of the first Australian-specific haplogroups to diversify after it evolved from its ancestral N type, which is consistent with the fact that the root of S is only one mutation step from N. The recent mitogenome sequencing of the skeletal remains of WLH4 of Lake Mungo, New South Wales<sup>16</sup> has determined that this individual (chronologically dated to the late Holocene by Durband, *et al.*<sup>61</sup>) belonged to haplogroup S2. Our analysis suggests that she belongs to subtype S2a1a, with her closest three maternal living descendants also living in New South Wales, with one individual, W26, residing close to Lake Mungo.

The common ancestor of the M42a and M42c evolved 50–53 KYA (Table 1) and this date may serve as an upper bound to the date of colonization of Australia. The novel haplogroup M42c is 5–10 KY older than M42a.

Haplogroup P is the oldest and most polytomous of the Australian haplogroups, with a TMRCAs estimate of 60 KY (Table 1). In the newly reconstructed P phylogeny comprising 112 mitogenomes, we find that the Australian-specific haplogroups are as ancient as those found in New Guinea and Philippines, but all estimates have large confidence intervals. The 'novel' P subtypes identified in this study P11 (including the former P4b) is 50 KY old and P12 is 46 KY old.

The antiquity of Australian P lineages makes identifying the geographic origin of P more problematic and, perhaps, these findings strengthen the hypothesis it evolved within Sahul. Importantly, given that some of the Australian-specific P lineages (P8, P11 and P12) are at least 45 KY old, these data support a long separation of Australian and New Guinean populations. This deep divergence between New Guineans and Aboriginal Australians is also supported by Y-chromosome data suggesting that the two populations may have separated at least 48 KY ago<sup>14,15</sup>.

Recently, it has been argued, based on genome wide analysis of samples from mainly northern and western Australia, that first settlement of Australia occurred less than 42 KYA<sup>18</sup>. However, the authors acknowledged that this date was not congruent with current archaeological evidence. Our genetic data suggest a much earlier arrival of the ancestors (no later than 45 KYA) as does the recent Y-chromosome study<sup>15</sup>. Further, very recent archaeological evidence supports occupation of the interior of Australia at least 49 KYA<sup>60</sup>.

**Migrations.** While it is presently unknown where P evolved in this region, it is likely that haplogroup P carrying females entered Australia from present-day New Guinea, and this haplogroup rapidly diverged into sub haplogroups. Today, Australia contains a number of unique P subtypes (P5, P6, P7, P8, P11 and P12), while P2 is distantly shared between New Guinea and the Torres Strait Islands and P3 appears in mainland Australia, Torres Strait Islands and New Guinea.

If we assume that haplogroup M42b belongs to a separate clade to that comprising M42a, then the ancestor of the latter most probably evolved in the Near East or South Asia and reached Sahul via the southern route into Sunda<sup>47,48</sup>. Subsequently, the ancestors moved through the string of Indonesian islands, including Nusa Tenggara and Timor. It is possible that this M42a ancestor arrived in Sahul at least 50 KYA, via the coastline of northwest Australia that is now submerged beneath the Timor and Arafura Seas, rather than via New Guinea. Once in the continent, their descendants subsequently diverged into M42a and M42c. Importantly, the female descendants of M42a may never have spread to New Guinea, as the M42 clade has not thus far been detected in present-day New Guineans<sup>23,62,63</sup>.

It is also plausible that the N\* ancestors of equally ancient Aboriginal maternal lineages such as N13, S and O may have followed a similar route of entry to that of the ancestors of M42a and M42c, as these lineages are also not shared, as far as we know, with modern New Guineans<sup>23,37,62</sup>. Whether or not females carrying ancestral N lineages followed a similar route(s) to that taken by the female M ancestors is unknown, but, clearly the descendant haplogroups of both M\* and N\* are distinct from those observed among modern New Guineans.



It has been argued that the M42c and N\* ancestral females may have taken a more northerly route (like that proposed for the Australian haplogroup P ancestors) through Sunda into Sahul via present day New Guinea, and later their descendants extended into Australia, where the M42a, M42c, N13, O, S and Australian-specific P subtypes evolved *in situ*<sup>9,30</sup>. Alternatively, those mitochondrial haplogroups presently found only in Aboriginal Australians may well have been present in ancient New Guineans but have been lost through genetic drift.

There is also considerable debate over whether Australia received migrants from other populations, particularly from South Asia, during the mid-Holocene. The postulated Indian connection arises from some archaeological<sup>25</sup> and linguistic findings<sup>28,64</sup>, as well as DNA analyses<sup>22–24</sup>. In particular, immigrants have been postulated as necessary to explain the introduction of the dingo (Australian dog) ~5 KYA, and the microlith tradition<sup>25,27</sup>. Claims of linguistic similarities between Dravidian languages of southern India and the Pama-Nyungan language of most of Aboriginal Australia have also been used to suggest a connection between the two regions<sup>28</sup>. However, the hypothesized connection to India has been strongly disputed<sup>65</sup>.

Redd and Stoneking<sup>23</sup> had previously suggested mtDNA lineages of Aboriginal Australians were most closely related to those in southern India, and later, Redd, *et al.*<sup>22</sup> claimed that the Y-chromosome haplogroup C-RPS4YT indicated a Holocene connection between the males of Southern India/Sri Lanka and Australia, although both these studies have relatively low data resolution. More recently, using a genome-wide analysis, Pugach, *et al.*<sup>24</sup>, calculated that a considerable component (~11%) of the Aboriginal Australian genome could be attributed to a South Indian migration to Australia some 4 KYA. By contrast, all studies on mtDNA variation in Aboriginal Australians, including the present one, find no evidence of recent gene flow from the Indian sub-continent during the Holocene. In fact, the mitogenomes of the current study show no evidence for sharing of Aboriginal haplogroups with those found in the Asian sub-continent except for haplogroup M42, although the TMRCA estimates for M42a and M42c (Australia) and M42b (South Asia) suggest these clades split long before the Holocene.

Our findings concur with those of other studies, that there is no evidence of back migration of females into New Guinea after novel haplogroups arose in Australia<sup>23,30,37,63,66</sup>. However, what adds to the complexity of any reconstruction of the past is the effects of the Austronesian expansion from (probably) Taiwan through Island Southeast Asia, then via New Guinea into Oceania, and the Pacific some 4 KYA. However, a recent study<sup>67</sup> confirms that this expansion was a rapid coastal migration resulting in little admixture with the indigenous populations of New Guinea.

The traditional Aboriginal marriage pattern of patrilocality and the movement of women along ‘Songline’ routes over large distances<sup>68</sup> for hundreds, if not thousands of generations can explain the wide geographical spread of indigenous haplogroups M42a, M42c, S, N13, O and P subtypes. The explanation for some very low-frequency ancient mitochondrial lineages in Aboriginal Australians remains speculative, but possibly involves factors such as marked isolation and limited resource access of some communities over thousands of years and the consequently marked effects of random genetic drift in small bands. However, it is possible that these low frequencies are due to sample bias and that their frequencies may be higher elsewhere in regions not yet sampled.

As all Aboriginal-specific mitochondrial haplogroups are of great antiquity, show considerable substructure, and are (mostly) very widely dispersed across the Australian continent while not being present outside Australia, it can be inferred that after initial colonisation some 50 KYA there has been a very long period of isolation of humans in Australia.

## Subjects and Methods

**Sample collection.** Saliva samples were collected using Oragene® DNA Collection kits (DNA Genotek Inc. Ontario, Canada) from 127 self-declared Aboriginal Australians who had volunteered to participate in the “Genographic Project”. Informed consent was obtained by all participants of the “Genographic Project”. Further information on sample characteristics and treatment is given in Supplementary Document. Individuals were previously genotyped for selected mtSNPs to assign them to the major mitochondrial haplogroups<sup>17</sup>. Individuals resided in the States of Queensland (n = 103), New South Wales (n = 14), Victoria (n = 6), Tasmania (n = 2) and Western Australia (n = 2), with the majority of participants from the Eastern States of Australia (Supplementary Table 2). Although the majority of samples were collected in Queensland, a number of individuals indicated that their maternal ancestry lay in other regions or States. In particular, participants had maternal ties to Torres Strait Islands, Tasmania and New South Wales. It is important to note that Aboriginal Australian affiliation is culturally based, and not defined by a person’s genetic composition. This study was approved by the Human Ethics Committee of La Trobe University. All methods were carried out in accordance with the relevant guidelines and regulations. All experimental protocols were approved by the Human Ethics Committee of La Trobe University.

**Whole mitogenome sequencing.** An input of 100 ng of genomic DNA extracted from Oragene® collection kits (as per the manufacturer’s recommendations) was prepared and indexed for Illumina sequencing using the TruSeq DNA Sample Prep Kit (Illumina) as per manufacturer’s instruction. The library was quantified using the Agilent TapeStation and the Qubit™ RNA assay kit for Qubit 2.0® Fluorometer (Life Technologies). The indexed libraries were then enriched for mtDNA and prepared for paired-end sequencing on a HiSeq instrument using the v3 100 cycle kit (Illumina) as per manufacturer’s instructions. For more detailed descriptions of the mitogenome data processing, genotyping and filtering, see the Supplemental Experimental Procedures.

**Data Analysis.** The 127 mitogenomes were aligned to the revised Cambridge Reference Sequence (rCRS)<sup>69,70</sup>. Two different mutation rates were utilized to calculate TMRCA estimates. The mutation rate of Soares, *et al.*<sup>56</sup> is based on the analysis of over 2,000 modern human mitogenomes whereas that of Fu, *et al.*<sup>57</sup> is based on the analysis of ten mitogenomes of skeletal remains of ancient Eurasians spanning the last 40 KY. These mutation

rates are based on the recalibration of the molecular clock of modern human and ancient human mitochondrial genomes, respectively.

Phylogenies of the haplogroups were constructed using the Network 5.0 software with the reduced median algorithm<sup>71</sup>. TMRCA estimates of haplogroups were also calculated using the rho ( $\rho$ ) statistic<sup>72</sup> based on phylogenetic rate of one mutation every 3624 years<sup>56</sup>. The mtDNA clock provided by Soares, *et al.*<sup>56</sup> was subsequently used to correct for purification selection.

Bayesian inferences of the TMRCA of haplogroups were calculated using BEAST<sup>73</sup>. The software jModeltest2<sup>74</sup> was used to identify the best-fitting nucleotide substitution model, which was GTR+I+G. For each inference, three independent MCMC runs of 20,000,000 iterations each were performed and combined using LogCombiner v1.8.2 (included in BEAST package). We used one prior distribution for complete mitochondrial genomes incorporating Fu, *et al.*<sup>57</sup> priors,  $2.67 \times 10^{-8}$  ( $2.16-3.16 \times 10^{-8}$ , 95% HPD). TRACER<sup>75</sup> was employed to estimate the TMRCA, to check convergence of the two BEAST runs of 60 million iterations and to compute the effective sample size (ESS) and the 95% confidence intervals for all parameters.

The 127 mitogenome sequences, together with all previously published mitogenome sequences from relevant haplogroups, were analyzed in a phylogenetic context by drawing maximum-parsimony trees using PhyloTree Build 17 (van Oven and Kayser, 2009) as reference. This allowed several phylogenetic improvements to be made and novel (sub)haplogroups identified in the process were given new haplogroup labels. Phylogenetic trees were constructed with the program mtPhyl (<http://eltsov.org>) and PhyloTree (Build 17)<sup>29</sup>. Thirteen of the 127 mitogenomes have also been subject to whole-genome sequencing at the Sanger Institute, using the protocols as described in Bergstrom, *et al.*<sup>15</sup>. These data confirmed the haplogroup assignments of these samples. Aboriginal Australian mitogenome sequences in GenBank that were included in the analysis were from Ingman, *et al.*<sup>38</sup>, Ingman and Gyllensten<sup>37</sup>, Kivisild, *et al.*<sup>44</sup>, van Holst Pellekaan, *et al.*<sup>8</sup>, Friedlaender, *et al.*<sup>51</sup>, Hudjashov, *et al.*<sup>9</sup>, Rasmussen, *et al.*<sup>11</sup> and Heupink, *et al.*<sup>16</sup>. The sequence data for the 83 Aboriginal Australian mitogenomes reported in Malaspinas *et al.*<sup>18</sup> were not available. Mitogenomes from other populations (non-Australians) were taken from Ingman, *et al.*<sup>38</sup>, Redd, *et al.*<sup>22</sup>, Ingman and Gyllensten<sup>37</sup>, Pierson, *et al.*<sup>55</sup>, Friedlaender, *et al.*<sup>51</sup>, Hudjashov, *et al.*<sup>9</sup>, Tabbada, *et al.*<sup>31</sup>, Delfin, *et al.*<sup>32</sup> and Gomes, *et al.*<sup>34</sup>. A Maximum-Likelihood tree was constructed using MEGA v6.0<sup>76</sup> with the GTR+I+G nucleotide model and 1000 bootstraps.

## References

1. Thorne, A. *et al.* Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *J. Hum. Evol.* **36**, 591–612 (1999).
2. Roberts, R. G. *et al.* The Human Colonization of Australia - Optical Dates of 53,000 and 60,000 Years Bracket Human Arrival at Deaf-Adder Gorge, Northern-Territory. *Quat. Sci. Rev.* **13**, 575–583 (1994).
3. O'Connell, J. F. & Allen, J. The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. *J. Archaeol. Sci.* **56**, 73–84 (2015).
4. O'Connor, S. Carpenter's Gap Rockshelter 1:40,000 years of Aboriginal occupation in the Napier Ranges, Kimberley, WA. *Aust. Archaeol.* **40** (1995).
5. Bowler, J. M., Jones, R., Allen, H. & Thorne, A. G. Pleistocene human remains from Australia: a living site and human cremation from Lake Mungo, Western New South Wales. *World archaeology* **2**, 39–60 (1970).
6. Turney, C. S. M. *et al.* Redating the onset of burning at Lynch's Crater (North Queensland): implications for human settlement in Australia. *J. Quaternary Sci.* **16**, 767–771 (2001).
7. Clarkson, C. *et al.* The archaeology, chronology and stratigraphy of Madjedbebe (Malakunanja II): A site in northern Australia with early occupation. *J. Hum. Evol.* **83**, 46–64 (2015).
8. 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).
9. Hudjashov, G. *et al.* Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc. Natl. Acad. Sci. USA* **104**, 8726–8730 (2007).
10. Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
11. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
12. Kayser, M. *et al.* Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* **68**, 173–190 (2001).
13. McEvoy, B. P. *et al.* Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. *Am. J. Hum. Genet.* **87**, 297–305 (2010).
14. Nagle, N. *et al.* Antiquity and diversity of aboriginal Australian Y-chromosomes. *Am. J. Phys. Anthropol.* **159**, 367–381 (2016).
15. Bergstrom, A. *et al.* Deep Roots for Aboriginal Australian Y Chromosomes. *Curr. Biol.* **26**, 809–813 (2016).
16. Heupink, T. H. *et al.* Ancient mtDNA sequences from the First Australians revisited. *Proc. Natl. Acad. Sci. USA* **113**, 6892–6897 (2016).
17. Nagle, N. *et al.* Mitochondrial DNA diversity of present-day Aboriginal Australians and implications for human evolution in Oceania. *Journal of human genetics* (2016).
18. Malaspinas, A. S. *et al.* A genomic history of Aboriginal Australia. *Nature* (2016).
19. Birdsell, J. In *Sunda and Sahul: Prehistoric Studies in Southeast Asia, Melanesia and Australia* (eds J. Allen, J. Golson, & R. Jones) 113–167 (Academic Press, 1977).
20. O'Connell, J. F. & Allen, J. The restaurant at the end of the universe: Modelling the colonisation of Sahul. *Aust. Archaeol.* **74**, 5–17 (2012).
21. Friedlaender, J. *et al.* Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol. Biol. Evol.* **22**, 1506–1517 (2005).
22. Redd, A. J. *et al.* Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. *Curr. Biol.* **12**, 673–677 (2002).
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. 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).
25. Glover, I. C. & Presland, G. In *Recent Advances in Indo-Pacific Prehistory* (eds V. N. Misra & P. Bellwood) 185–195 (Oxford & IBH, 1985).
26. Huxley, T. H. On the Geographical Distribution of the Chief Modifications of Mankind. *Journal of the Ethnological Society of London* **2**, 404–412 (1870).
27. Gollan, K. In *Recent Advances in Indo-Pacific Prehistory* (eds V. N. Misra & P. Bellwood) 439–443 (Oxford & IBH, 1985).

28. Dixon, R. M. W. *The Languages of Australia*. (Cambridge University Press, 1980).
29. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–394 (2009).
30. van Holst Pellekaan, S. Genetic evidence for the colonization of Australia. *Quat. Int.* **285**, 44–56 (2013).
31. Tabbada, K. A. *et al.* Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol. Biol. Evol.* **27**, 21–31 (2010).
32. Delfin, F. *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).
33. 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).
34. Gomes, S. M. *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).
35. Ballantyne, K. N. *et al.* MtDNA SNP multiplexes for efficient inference of matrilineal genetic ancestry within Oceania. *Forensic Sci. Int. Genet.* **6**, 425–436 (2012).
36. 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).
37. Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13**, 1600–1606 (2003).
38. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
39. 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).
40. Presser, J. C., Stoneking, M. & Redd, A. J. Tasmanian aborigines and DNA. *Papers Proc. R. Soc. Tasmania* **136**, 35–38 (2002).
41. Cann, R. L., Stoneking, M. & Wilson, A. C. Mitochondrial DNA and human evolution. *Nature* **325**, 31–36 (1987).
42. Kumar, S. *et al.* Reconstructing Indian-Australian phylogenetic link. *BMC Evol. Biol.* **9**, 173 (2009).
43. Chandrasekar, A. *et al.* Updating phylogeny of mitochondrial DNA macrohaplogroup M in India: dispersal of modern human in South Asian corridor. *PLoS one* **4**, e7447 (2009).
44. Kivisild, T. *et al.* The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373–387 (2006).
45. Derenko, M. *et al.* Complete mitochondrial DNA diversity in Iranians. *PLoS one* **8**, e80673 (2013).
46. Fregel, R. *et al.* Multiple ethnic origins of mitochondrial DNA lineages for the population of Mauritius. *PLoS one* **9**, e93294 (2014).
47. Macaulay, V. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036 (2005).
48. Merriwether, D. A. *et al.* Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc. Natl. Acad. Sci. USA* **102**, 13034–13039 (2005).
49. Bandelt, H. J., Macaulay, V. & Richards, M. B. *Human Mitochondrial DNA and the Evolution of Homo Sapiens*. Vol. 18 (Springer Berlin Heidelberg, 2006).
50. Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
51. Friedlaender, J. S. *et al.* Melanesian mtDNA complexity. *PLoS one* **2**, e248 (2007).
52. Kayser, M. *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).
53. Duggan, A. T. *et al.* Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. *Am. J. Hum. Genet.* **94**, 721–733 (2014).
54. Beckett, J. *Torres Strait Islanders: Custom and Colonialism*. (Cambridge University Press, 1987).
55. Pierson, M. J. *et al.* Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Mol. Biol. Evol.* **23**, 1966–1975 (2006).
56. Soares, P. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
57. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
58. Woodroffe, C. D., Kennedy, D. M., Hopley, D., Rasmussen, C. E. & Smithers, S. G. Holocene reef growth in Torres strait. *Mar. Geol.* **170**, 331–346 (2000).
59. Williams, A. N. A new population curve for prehistoric Australia. *Proceedings. Biological sciences/The Royal Society* **280**, 20130486 (2013).
60. Hamm, G. *et al.* Cultural innovation and megafauna interaction in the early settlement of arid Australia. *Nature* (2016).
61. Durband, A. C., Westaway, M. C. & Rayner, D. R. T. Interproximal grooving of lower second molars in WLH 4. *Aust. Archaeol.* **118**–120 (2012).
62. van Oven, 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).
63. Kayser, M. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**, 281–302 (2003).
64. Reesink, G., Singer, R. & Dunn, M. Explaining the Linguistic Diversity of Sahul Using Population Models. *PLoS biology* **7** (2009).
65. Brown, P. Palaeoanthropology: Of humans, dogs and tiny tools. *Nature* **494**, 316–317 (2013).
66. 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).
67. Soares, P. A. *et al.* Resolving the ancestry of Austronesian-speaking populations. *Human genetics* **135**, 309–326 (2016).
68. Norris, R. P. & Harney, B. Y. Songlines and Navigation in Wardaman and other Australian Aboriginal Cultures. *J. Astron. Hist. Herit.* **17**, 1–15 (2014).
69. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
70. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
71. Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753 (1995).
72. Forster, P., Harding, R., Torroni, A. & Bandelt, H. J. Origin and evolution of Native American mtDNA variation: a reappraisal. *American journal of human genetics* **59**, 935–945 (1996).
73. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
74. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).

75. TRACER v1.6. (<http://beast.bio.ed.ac.uk/Tracer> 2013).

76. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

## Acknowledgements

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. The Genographic Project was supported by National Geographic Society, IBM and the Waitt Family Foundation. Y.L.X. and C.T.-S. were supported by The Wellcome Trust (098051). M.K., M.v.O., and K.N.B. were supported by Erasmus M.C.

## Author Contributions

Project design was carried out by N.N., L. Williams., R.J.M., and The Genographic Consortium; community engagement, ethics, and sampling by L. Williams, P.M. and R.J.M.; Data generation, processing and analysis by N.N., M.v.O., S.W., L. Wilcox, L.P., K.C. and R.J.M.; data interpretation by N.N., M.v.O., S.v.H.P., C.T.-S., Y.X., R.A.H.v.O., M.K. and R.J.M.; manuscript writing and editing by N.N., M.v.O., S.W., S.v.H.P., C.T.-S., Y.X., K.N.B., R.A.H.v.O., M.K. and R.J.M.

## Additional Information

**Accession codes:** GenBank, <http://www.ncbi.nlm.gov/Genbank/> (for sequences [accession numbers KY595546-KY595672])

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Nagle, N. *et al.* Aboriginal Australian mitochondrial genome variation – an increased understanding of population antiquity and diversity. *Sci. Rep.* **7**, 43041; doi: 10.1038/srep43041 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

## Consortia

### The Genographic Consortium

Syama Adhikarla<sup>10</sup>, Christina J. Adler<sup>11</sup>, Elena Balanovska<sup>12</sup>, Oleg Balanovsky<sup>12</sup>, Jaime Bertranpetit<sup>13</sup>, Andrew C. Clarke<sup>14</sup>, David Comas<sup>13</sup>, Alan Cooper<sup>11</sup>, Clio S. I. Der Sarkissian<sup>11</sup>, Matthew C. Dulik<sup>15</sup>, Jill B. Gaieski<sup>15</sup>, ArunKumar GaneshPrasad<sup>10</sup>, Wolfgang Haak<sup>11</sup>, Marc Haber<sup>13,16</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>13</sup>, Elizabeth A. Matisoo-Smith<sup>14</sup>, Marta Melé<sup>13</sup>, Nirav C. Merchant<sup>20</sup>, Amanda C. Owings<sup>15</sup>, Laxmi Parida<sup>18</sup>, Ramasamy Pitchappan<sup>10</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>10</sup>, Fabrício R. Santos<sup>23</sup>, Theodore G. Schurr<sup>15</sup>, Himla Soodyall<sup>17</sup>, David F. Soria Hernanz<sup>24</sup>, Pandikumar Swamikrishnan<sup>25</sup>, Miguel G. Vilar<sup>15</sup>, R. Spencer Wells<sup>24</sup>, Pierre A. Zalloua<sup>16</sup> & Janet S. Ziegler<sup>26</sup>

<sup>10</sup>Madurai Kamaraj University, Madurai, Tamil Nadu, India. <sup>11</sup>University of Adelaide, South Australia, Australia. <sup>12</sup>Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia. <sup>13</sup>Universitat Pompeu Fabra, Barcelona, Spain. <sup>14</sup>University of Otago, Dunedin, New Zealand. <sup>15</sup>Department of Anthropology, University of Pennsylvania, PA, USA. <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. <sup>26</sup>Applied Biosystems, Foster City, CA, USA.