Antiquity and Diversity of Aboriginal Australian Y-Chromosomes

Nano Nagle,[1] Kaye N. Ballantyne,[2,3] Mannis van Oven,[3] Chris Tyler-Smith,[4] Yali Xue,[4] Duncan Taylor,[5,6] Stephen Wilcox,[7] Leah Wilcox,[1] Rust Turkalov,[7] Roland A.H. van Oorschot,[2] Peter McAllister,[8] Lesley Williams,[9] Manfred Kayser,[3] Robert J. Mitchell,[1]* and The Genographic Consortium

[1]*Department of Biochemistry and Genetics, La Trobe Institute of Molecular Sciences, La Trobe University, Melbourne, VIC, Australia*
[2]*Victorian Police Forensic Services Department, Office of the Chief Forensic Scientist, Melbourne, VIC, Australia*
[3]*Department of Forensic Molecular Biology, Erasmus MC University Medical Center, Rotterdam, The Netherlands*
[4]*The Wellcome Trust Sanger Institute, Welcome Trust Genome Campus, Hinxton, Cambridgeshire, UK*
[5]*Forensic Science South Australia, 21 Divett Place, Adelaide, SA, 5000, Australia*
[6]*School of Biological Sciences, Flinders University, Adelaide, SA, 5001, Australia*
[7]*Australian Genome Research Facility, Melbourne, VIC, Australia*
[8]*Griffith University, Brisbane, QLD, Australia*
[9]*Department of Communities, Child Safety and Disability Services, Queensland Government, Brisbane, QLD, Australia*

KEY WORDS Y-chromosome; haplogroup; haplotype; Australia; Sahul; Aboriginals

ABSTRACT

Objective: Understanding the origins of Aboriginal Australians is crucial in reconstructing the evolution and spread of *Homo sapiens* as evidence suggests they represent the descendants of the earliest group to leave Africa. This study analyzed a large sample of Y-chromosomes to answer questions relating to the migration routes of their ancestors, the age of Y-haplogroups, date of colonization, as well as the extent of male-specific variation.

Methods: Knowledge of Y-chromosome variation among Aboriginal Australians is extremely limited. This study examined Y-SNP and Y-STR variation among 657 self-declared Aboriginal males from locations across the continent. 17 Y-STR loci and 47 Y-SNPs spanning the Y-chromosome phylogeny were typed in total.

Results: The proportion of non-indigenous Y-chromosomes of assumed Eurasian origin was high, at 56%. Y lineages of indigenous Sahul origin belonged to haplogroups C-M130*(xM8,M38,M217,M347) (1%), C-M347 (19%), K-M526*(xM147,P308,P79,P261,P256,M231,M175,M45,P202) (12%), S-P308 (12%), and M-M186 (0.9%). Haplogroups C-M347, K-M526*, and S-P308 are Aboriginal Australian-specific. Dating of C-M347, K-M526*, and S-P308 indicates that all are at least 40,000 years old, confirming their long-term presence in Australia. Haplogroup C-M347 comprised at least three sub-haplogroups: C-DYS390.1del, C-M210, and the unresolved paragroup C-M347*(xDYS390.1del,M210).

Conclusions: There was some geographic structure to the Y-haplogroup variation, but most haplogroups were present throughout Australia. The age of the Australian-specific Y-haplogroups suggests New Guineans and Aboriginal Australians have been isolated for over 30,000 years, supporting findings based on mitochondrial DNA data. Our data support the hypothesis of more than one route (via New Guinea) for males entering Sahul some 50,000 years ago and give no support for colonization events during the Holocene, from either India or elsewhere. Am J Phys Anthropol 159:367–381, 2016. © 2015 Wiley Periodicals, Inc.

INTRODUCTION

Under most scenarios of human prehistory Australia is viewed as a continent that was settled very soon after the expansion "out of Africa," some 60–100 thousand years ago (kya) (Roberts and Jones, 1994; Thorne et al., 1999), by anatomically and behaviorally modern humans. Genetic and archaeological evidence currently suggest that the ancestors of present-day Aboriginal Australians arrived at least 45 kya (O'Connell and Allen, 2004; Summerhayes et al., 2010; Williams, 2013; O'Connell and Allen, 2015). This early settlement is

Additional Supporting Information may be found in the online version of this article.

The authors declare no conflict of interest.

Genographic Consortium members are listed in the Acknowledgments section.

Grant sponsor: National Geographic Society, IBM, Waitt Family Foundation.

*Correspondence to: Robert J. Mitchell; Department of Biochemistry and Genetics, La Trobe Institute of Molecular Sciences, La Trobe University, Melbourne, Victoria, Australia.
E-mail: john.mitchell@latrobe.edu.au

Received 1 April 2015; revised 1 October 2015; accepted 8 October 2015

DOI: 10.1002/ajpa.22886
Published online 30 October 2015 in Wiley Online Library (wileyonlinelibrary.com).

© 2015 WILEY PERIODICALS, INC.

supported by archaeological sites widely spread across the continent, such as Devil's Lair in Western Australia (Turney et al., 2001a), Carpenter's Gap in northern Western Australia (O'Connor, 1995), Lake Mungo in New South Wales (Bowler et al., 1970) and Lynch's Crater in Queensland (Turney et al., 2001b), as well as by DNA evidence (van Holst Pellekaan et al., 2006; Hudjashov et al., 2007; Rasmussen et al., 2011; Reich et al., 2011). Their long-term in situ presence makes Aboriginal Australians central to the study of modern human origins and early dispersals.

While there may exist some consensus about the date of the first arrival of modern humans in Sahul (the ancient continent comprising the then connected islands of New Guinea and Australia in the Pleistocene epoch), there is still controversy over the route(s) taken and their points of entry into Sahul (Birdsell, 1977; Dixon, 1980; Redd et al., 2002; Hudjashov et al., 2007; Reich et al., 2011; Pugach et al., 2013). There is also conflicting evidence regarding whether or not the ancestors of present-day New Guineans and Aboriginal Australians entered as a single founder population into Sahul with subsequent dispersal within it, or if there were separate founder populations for these two groups (Tsintsof et al., 1990; Roberts-Thomson et al., 1996; Evans, 2005; Hudjashov et al., 2007). Further, there is still ongoing debate surrounding the number and source of any subsequent incursions into Australia that may have occurred prior to the well-documented European invasions of historical times (Kayser et al., 2001; Redd et al., 2002; Hudjashov et al., 2007; McEvoy et al., 2010; Rasmussen et al., 2011; Pugach et al., 2013).

One factor contributing to the long-running controversy over Aboriginal origins and the extent of diversity is that they are among the most poorly studied groups of humans with respect to their genetic structure. This dearth of information reflects the experiences of Aboriginal people participating in earlier anthropological/genetic studies (Dodson, 2000; van Holst Pellekaan, 2000a,b,2008; Kowal, 2012).

Controversy over Aboriginal origins has a long history, extending back to Tindale and Birdsell's trihybrid theory in which three distinct migrant groups were identified, each colonizing at different times, but all within the last 10 ky (Birdsell, 1977). No genetic or anthropometric evidence has substantiated this trihybrid hypothesis (Kirk et al., 1963, 1972; Gao and Serjeantson, 1991; White, 1997). It was not until the discovery of evolutionary significant DNA markers that the uniqueness and the long isolation of Aboriginal Australians from other groups became evident. In particular, mitochondrial DNA (mtDNA) variation revealed a distinctive array of haplogroups or lineages, most of which are unique to the Aboriginal population (Cann et al., 1987; van Holst Pellekaan et al., 1998; Redd and Stoneking, 1999; Huoponen et al., 2001; Ingman and Gyllensten, 2003; Friedlaender et al., 2005; van Holst Pellekaan et al., 2006; Hudjashov et al., 2007).

Most mtDNA data demonstrate strong (although ancient) affinities between Aboriginal Australians and New Guineans (van Holst Pellekaan et al., 1998; Hudjashov et al., 2007), but other mtDNA data have been used to suggest that the closest affinities lie with populations of India (Redd and Stoneking, 1999; Kumar et al., 2009).

Analysis of the whole genome of a single non-admixed Aboriginal Australian strongly supported the hypothesis that present-day Aboriginal Australians represent one of the oldest continuous populations found outside Africa, with little evidence of genetic introgression from other populations until historical times (Rasmussen et al., 2011; Reich et al., 2011).

There are few studies that focus on Y-chromosome diversity in Aboriginal Australians (Vandenberg et al., 1999; Kayser et al., 2001; Underhill et al., 2001b; Redd et al., 2002; Hudjashov et al., 2007; Taylor et al., 2012). Apart from the Taylor et al. (2012) analysis of Aboriginal males from South Australia, all Y-chromosome studies have been restricted to the northern areas of Australia and, therefore, provided an inadequate coverage of the continent. Sample sizes were usually very small and/or sample locations were not reported. Some of these studies also had a risk of sampling relatives or resampling individuals who participated in preceding genetic studies (as discussed in White, 1997; van Holst Pellekaan, 2013).

Earlier Y-chromosome studies were also greatly limited by the low numbers of markers used. Some studies reported Y-SNPs only (Underhill et al., 2001b; Hudjashov et al., 2007; Karafet et al., 2014), others typed only Y-STRs (Forster et al., 1998; Vandenberg et al., 1999), whilst others typed differing combinations of Y-SNPs and Y-STRs (Kayser et al., 2001; Redd et al., 2002). Consequently, there is great difficulty in comparing datasets. In addition, European admixture in Aboriginal Australians is considerable, especially in the paternal line (Taylor et al., 2012), and has been so since first contact with Europeans (Radcliffe-Brown, 1930; Williams, 2013), but its extent is variable between regions (Walsh et al., 2007). Accordingly, small sample sizes, limited geographical coverage, the difficulty of comparing existing Y-chromosome datasets, plus the great, but unquantified, reduction in the frequencies of indigenous Y-chromosomes through two centuries of European admixture, make attempts at reconstructing past patterns of male mediated diversity at a tribal or group level, as well as possible routes of migration, extremely difficult (van Holst Pellekaan, 2008).

To further our knowledge of Y-chromosome diversity among this poorly studied population, we assembled a large sample of Y-chromosomes from individuals who self-reported as Aboriginal Australian for analysis of both Y-SNP and Y-STR variation. Seventeen Y-STR loci and 47 Y-SNPs were typed in total. This dataset was drawn widely from across the continent, especially from areas previously unsampled, such as Queensland and Victoria, and also the forensic samples included in this study are less likely to contain resampled individuals as autosomal STR profiles were available for clarification that the same individual was not included more than once. This study aims to investigate the distribution of indigenous Y-chromosome haplogroups and associated Y-STR variation of 657 individuals, across the continent in order to decipher possible origins, ages of paternal lineages, and relationships with neighboring populations.

MATERIALS AND METHODS

Subjects

Four distinct sample groups make up the total dataset analyzed in the present study. Firstly, the Genographic Project database comprised 144 self-declared Aboriginal males drawn from Australia: Queensland ($n = 119$), New South Wales ($n = 19$), Victoria ($n = 4$), Western Australia ($n = 1$), and Tasmania ($n = 1$). Information on the paternal history of these participants was noted, if known,

and many were aware that they had non-indigenous fathers/grandfathers but had Aboriginal ancestry in other lines. These samples were collected between 2006 and 2013. The second and third sample sets were drawn from the Forensic Science South Australia (FSSA) and Victorian Police Forensic Services Department (VPFSD) Aboriginal databases. The allocation of ethnic affiliation for these samples was chiefly by self-declaration or, occasionally, by information from investigating officers, but individuals were not asked specifically about their paternal ancestry. There were a total of 757 males in the South Australian sample and 182 males in the Victorian collection. Finally, the Northern Territory samples were drawn from the Northern Territory Forensic Center's Aboriginal database and selected for Aboriginality with higher stringency than the other forensic samples (Vandenberg et al., 1999). No locality information was available, but of the 92 males, 77 came from the Northern Territory, 5 from South Australia, and 10 from Queensland. Combined, the total of these databases is 1175 males (further described in Supporting Information Table S1). It is important to note that Aboriginal Australian is a culturally based affiliation, and not one defined by a person's genetic composition.

The FSSA and VPFSD samples came with a place name that was either the location of the offense or the residence of the donor. Reference to Horton's map of Aboriginal languages and traditional regions (Horton, 1996) allowed contemporary locations to be converted to a location within traditional Aboriginal regions. This knowledge allowed for some regional analysis of these samples. The South Australian samples were divided into four regions; Desert, Spencer, Riverine, and Urban. Further information is provided in Taylor et al. (2012). The Victorian samples were divided into two regions; South East Victoria and Riverine. Where no placeholder or location information was available, samples were assigned to the State or Territory of the individual's residence. This, of course, may not correspond with the homeland or birthplace of the individual. Assignment of samples by State or Territory, however, was considered the best treatment for the data given the lack of more precise information. Combined, these four datasets provide good geographical coverage of Aboriginal Australia, especially compared with previous Y-chromosome studies.

This study was approved by La Trobe University Human Ethics Committee and the Government institutions involved.

METHODS

DNA processing

Most Genographic Project samples were collected using Oragene® saliva collection kits (DNA Genotek Inc. Ontario, Canada) and DNA was extracted following the manufacturer's protocol. Some samples were collected using mouth swabs and DNA was extracted using the method described in de Vries et al. (1996). DNA extraction from the samples of South Australia and the Northern Territory regions is described in Taylor et al. (2012) and Vandenberg et al. (1999), respectively. The Victorian samples, consisting of blood deposits on cloth or filter paper, were extracted with the Qiagen DNA Blood Mini Kit (Qiagen, Victoria, Australia) according to the manufacturer's instructions.

Y-STR typing

All 1175 DNA samples in this study were amplified using the Applied Biosystems Amp*Fl*STR® Yfiler™ PCR amplification kit as per the manufacturer's specifications. The Yfiler™ 17 loci comprised DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and YGATA H4. The FSSA South Australian Y-STR profiles are available on the Y-STR haplotype reference database (YHRD; https://yhrd.org). The YHRD accession number for the Aboriginal database is YA003697.

Y-SNP typing

Two different methods were used for Y-SNP typing of the samples; TaqMan® (Applied Biosystems®) and SNaPshot® (Applied Biosystems®). This study has followed the Y-SNP nomenclature of the Y phylogenetic tree devised by van Oven et al. (2014b).

South Australian and The Genographic Project. The Genographic Project and South Australian samples were SNP typed using TaqMan analysis (Applied Biosystems®).

The Yfiler profiles of the South Australian sample of 757 were submitted to a Y-haplogroup predictor program (www.hprg.com/hapest5) (Athey, 2006). These Yfiler profiles were compared with all three population databases associated with the Y-haplogroup predictor program; northwest Europe, Eastern Europe, and the Mediterranean. Those profiles that could not be assigned to a Eurasian haplogroup with a high degree of probability (fitness score >70) were subjected to hierarchical Y-SNP typing as described in Taylor et al. (2012). As a result of the Y-haplogroup predictor program interrogation, 265 South Australian males were SNP typed and the remaining 492 South Australian Y-STR profiles were removed from the investigation as they were considered to be Eurasian and were subsequently not Y-SNP typed. All Genographic Project samples were SNP typed using the same hierarchical protocol of Taylor et al. (2012). Y-haplogroups considered to be indigenous to Australia that are supported by observations are C-M130*, C-M347, K-M526, and S-P308 (Kayser et al., 2001; Redd et al., 2002; Hudjashov et al., 2007; Karafet et al., 2014). It is unclear as to whether haplogroup M-M186 is indigenous to Australia, given that the Aboriginal Australians of Queensland, the closest State to New Guinea (in which M-M186 is found), have not been investigated. In this study M-M186 was considered a potential indigenous haplogroup. All other Y-haplogroups typed are considered non-indigenous as they could be explained by historical Eurasian male introgression.

Those samples that were derived for the Y-SNP marker KLT-M9 but did not belong to a Eurasian subhaplogroup were subsequently typed for markers K-M526 and S-P308. Only the Genographic Project males were typed for the additional markers K-P261 and S-P79 which define other subtypes of K-M526 and S-P405 (Scheinfeldt et al., 2006; Karafet et al., 2008, 2014; van Oven et al., 2014b).

Northern Territory and Victorian samples. All the Northern Territory and Victorian samples were Y-SNP typed by the single-nucleotide primer extension

(*SNaPshot*) technology utilizing the global multiplexes and methodologies described in van Oven et al. (2011). Samples that were found to belong to haplogroup C-M216 were additionally typed for markers C-M38, C-M217, C-M347, and C-M210 using an updated version of the haplogroup-C multiplex described in van Oven et al. (2014a). A Y-phylogeny adapted from van Oven et al. (2014b) comprising all the Y-SNPs typed within the entire dataset is given in Supporting Information Figure 1.

West New Guinea, Papua New Guinea and Nusa Tenggara. Median-joining network analysis, comprising a total of 300 7-locus Y-STR haplotypes from West New Guinea, Papua New Guinea and Nusa Tenggara (Kayser et al., 2003, 2006, 2008; Mona et al., 2009; Delfin et al., 2012; van Oven et al., 2014a) and the equivalent 7-locus haplotypes of the Aboriginal Australians, was undertaken to determine whether or not there were shared or close paternal haplotypes. Twenty six males from Papua New Guinea and Nusa Tenggarra (assigned to haplogroups KLT-M9* and C-M130*) were selected for additional Y-SNP typing of K-M526, K-M147, S-P308, K-P261, M-P256 and LT-P326 (haplogroup K*), and C-M347 (haplogroup C*) based on either the shared and/or close proximity of their haplotypes to known K-M526* and S-P308 Aboriginal Australian males.

Data analysis

Population pairwise genetic distances (R_{ST} values) were calculated from Y-STR haplotypes using the Arlequin v3.5 package (Excoffier and Lischer, 2010) and their significance was assessed from 1000 bootstrap simulations. Haplotype diversity was calculated using $H = \frac{N}{N-1}\left(1-\sum x_i^2\right)$ where N is the population size and x_i is the frequency of the ith haplotype (Nei, 1973). Multidimensional scaling (MDS) analysis of population pairwise distances was performed using SPSS v21.0 software (IBM, 2012) and the values of stress were verified according to the findings reported by Sturrock and Rocha (2000). Haplogroup distributions across regions were compared using the χ^2 test of independence. Median-joining haplotype networks were created using Network version 4.6.1.2 (http://www.fluxus-engineering.com/). Y-STR loci were weighted in accordance with the inverse of the mutation rate of each locus reported by Goedbloed et al. (2009). The duplicated locus DYS385 was excluded from these analyses since the constituent loci are not distinguished in YFiler. DYS389 was treated as two separate loci, DYS389I and DYS389b (whereby DYS389b = DYS389II–DYS389I). Since locus DYS390 in Aboriginal Australians violates the stepwise mutation model (Forster et al., 1998) implemented in Bayesian analysis of trees with internal node generation program (otherwise known as BATWING), it was omitted in BATWING, Network and average squared distance (ASD) analyses (Goldstein et al., 1995; Forster et al., 1996; Saillard et al., 2000; Behar et al., 2003; Wilson et al., 2003), resulting in a haplotype comprising 14 Y-STR loci in all analyses.

Estimation of time to the most recent common ancestor (TMRCA) of Aboriginal Australian Y-haplogroups was performed using BATWING, Rho statistic as calculated within Network and ASD as calculated in the Ytime program (Behar et al., 2003). Rho and ASD statistics calculate TMRCA estimates based on an average Y-STR mutation rate and an inferred ancestral haplotype and do not incorporate a specific demographic model into the estimation. TMRCA estimates via Rho statistic can sometimes be underestimates because the method fails to distinguish between individuals who are identical by state rather than by descent. The ancestral haplotype for the Rho and ASD calculations was the most central node of the haplogroup. Bayesian analysis via BATWING enables the user to place prior conditions on the data and set individual Y-STR mutation rates. However, all methods appear to underestimate long time periods of divergence, given the inherent nature of the STR loci (Wei et al., 2013).

Both the evolutionary mutation rates (EMR) of 6.9×10^{-4} per locus per generation (Zhivotovsky et al., 2004) and the observed mutation rate (OMR) of 2.1×10^{-3} per locus per generation (Goedbloed et al., 2009) of Y-STR loci were used to estimate the TMRCAs of haplogroups in Network, BATWING and Ytime. Average generation time was set at 25 years for EMR and at 30 years for OMR TMRCA estimates (Dupuy et al., 2004; Zhivotovsky et al., 2004; Fenner, 2005). The difference in average generation times for the OMR and EMR is due to the parameters by which the average mutation rates were devised, with the OMR based on counts of mutations between fathers and sons (or more distant male relatives) and the EMR based events dated by history (Dupuy et al., 2004; Zhivotovsky et al., 2004; Fenner, 2005).

Inferences about Y-chromosome lineage histories and TMRCA estimates were made by including unique-event polymorphisms (i.e., SNPs/indels) with STR data in BATWING (Wilson et al., 2003). Y-STR mutation rates (Zhivotovsky et al., 2004; Gusmao et al., 2005; Sanchez-Diz et al., 2008) and population/coalescent priors described in Xue et al. (2006) and Kayser et al. (2001) were used. TRACER (Rambaut and Drummond, 2013) was used to check that convergence of the three BATWING runs of 500 million iterations (with a 10% burn in) and that the effective sample size (ESS) were within acceptable ranges. Trace and density plots of the MCMC chains indicated good mixing of the chains.

RESULTS

Non-indigenous Y-chromosomes and admixture in Aboriginal Australians

Y-SNP typing of a total of 657 self-reported Aboriginal males revealed that 365 or 55.6% of the Y-chromosomes could be assigned to a non-indigenous haplogroup (defined as not C-M130*, C-M347, K-M526*, S-P308, or M-M186), and, of these, the vast majority was of Eurasian origin (Table 1). The frequency of non-indigenous haplogroups varied across Australia, with Victoria having the highest frequency and the Northern Territory, the lowest (Table 1). The much lower frequency in the Northern Territory reflects the difference in sample ascertainment (as mentioned above). A large proportion belonged to Eurasian haplogroup R-M207 (32.4%), which was the most common non-indigenous Y-haplogroup in all States. The second most frequent was the Eurasian haplogroup I-M170 (8.2%), although it was absent in the Northern Territory. There is a highly significant difference in the frequency distribution of non-indigenous haplogroups among the three States ($\chi^2 = 69.2$, d.f. = 8, $P < 0.0001$). The Northern Territory was excluded in this analysis because of the difference in sample selection.

TABLE 1. Distribution of Y-haplogroups in Aboriginal Australian populations after SNP typing

Population Sample	*n*	Indigenous Y-haplogroups												Non-indigenous Y-haplogroups									
		C-M130*		C-M347[a]		C-DYS390.1del[b]		K-M526*		S-P308		M-M186		Total		R-M207		I-M170		J-M304		O-M175		Others		Total	
		No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
South Australia	265	1	0.4	71	26.8	32	12.1	38	14.3	57	21.5	2	0.8	169	63.8	25	9.4	22	8.3	11	4.2	19	7.2	19	7.2	96	36.2
Northern Territory	77	3	3.9	44	57.1	35	45.5	5	6.5	15	19.5	1	1.3	68	88.3	6	7.8	0	0.0	1	1.3	1	1.3	1	1.3	9	11.7
Victoria	186	0	0.0	2	1.1	0	0.0	16	8.6	4	2.2	0	0.0	22	11.8	119	64.0	20	10.8	8	4.3	4	2.2	13	7.0	164	88.2
Queensland	129	1	0.8	5	3.9	2	1.6	20	15.5	4	3.1	3	2.3	33	25.6	63	48.8	12	9.3	1	0.8	14	10.9	6	4.7	96	74.4
Total	657	5	0.8	122	18.6	69	10.5	79	12.0	80	12.2	6	0.9	292	44.4	213	32.4	54	8.2	21	3.2	38	5.8	39	5.9	365	55.6

[a] C-M347 column inclusive of C-DYS390.1del samples.
[b] C-DYS390.1del derived from combined Y-SNP and DYS390 Y-STR analysis.

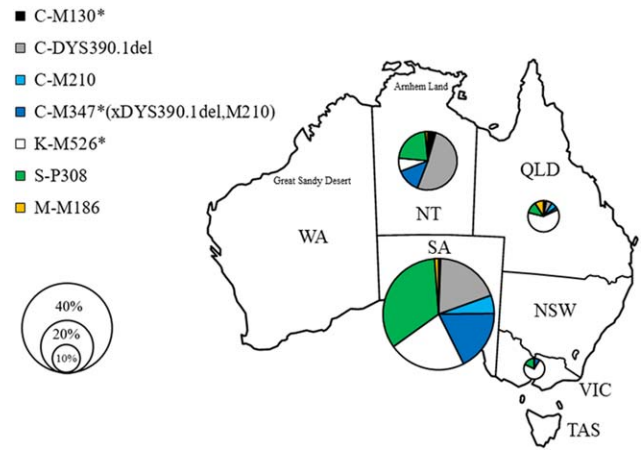

Fig. 1. Geographic distribution and relative frequencies of indigenous Australian Y-haplogroups. WA, Western Australia; NT, Northern Territory; SA, South Australia; QLD, Queensland; NSW, New South Wales; VIC, Victoria; and TAS, Tasmania. Size of pie charts proportional to haplogroup frequencies.

Haplogroup O-M175 (of East/Southeast Asian origin) was found in less than 6% of the total dataset, whilst haplogroups C-M38 (Oceanic) and C-M217 (predominantly Asian) were found in 0.6% and 0.2% of the sample, respectively. Haplogroups C-M38 and C-M217 were absent in the Northern Territory but present in other States (data not shown). Collectively, haplogroups A, B, D, G, H, N, P, and Q comprised 6% of the 657 Y-chromosomes. Although the diagnostic SNP M356 which is restricted to South Asia, was only typed in a single Australian C-M130/M216*(xM8,M38,M347,M217) male from the Genographic Project Queensland sample (who was found to be C-M356 ancestral), there were no 17-locus haplotype matches for all five Australian C-M130* haplotypes after interrogation of the Y-STR YHRD (Willuweit et al., 1998) and after comparison with the Indian C-M130* and C-M356 haplotypes of Arunkumar et al. (2012). In fact, the Australian C-M130* Y-STR haplotypes were markedly distinct (data not shown).

The single haplogroup H-M69 male in the sample knew that his great-grandfather was an immigrant from Europe and his Yfiler Y-STR profile indicated matches for his "minimum" haplotype of the YHRD with males from Europe, West Asia, and South Asia. As our study design only tested samples for haplogroup R-M207 in general (not its subhaplogroups), we cannot be certain that R-M479 (also known as R2), which is typical for South Asia, is absent from our sample (Table 1). However, all 125 haplogroup R 17-locus Y-STR haplotypes that were not allocated to R-M17 or R-M412 after SNP typing were interrogated via the YHRD and could be assigned to either R-M420 or R-M343 with strong likelihood of European ancestry.

Indigenous Y-chromosomes of Australia

Two hundred and ninety two or 44.4% of all 657 males SNP typed carried a Y-chromosome belonging to haplogroups C-M347, K-M526*(xM147,P308,P79,P261, P256,M231,M175,M45,P202), S-P308, or M-M186 (Table 1 and Fig. 1)—all considered indigenous in Aboriginal Australians. In addition, five of the sampled Aboriginal

TABLE 2. Distributions of Indigenous Y-haplogroups in Australia

		Haplogroup																	
		C-M130*		C-M347 (Total)		C-M210		C-DYS390. 1del		C-M347* (xDYS390. 1del,M210)		K-M526 (Total)		K-M526*		S-P308		M-M186	
State	*N*	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Victoria	**22**			**2**	**9.1**					**2**	**9.1**	**20**	**90.9**	**16**	**72.7**	**4**	**18.2**		
Riverine	5			1	*20.0*					1	*20.0*	4	*80.0*	4	*80.0*				
South East	17			1	*5.9*					1	*5.9*	16	*94.1*	12	*70.6*	4	23.5		
Queensland	**33**	**1**	**3**	**5**	**15.2**	**2**	**6.1**	**2**	**6.1**	**1**	**3**	**24**	**72.7**	**20**	**60.6**	**4**	**12.1**	**3**	**9.1**
Queensland (other)	15	1	*6.7*	2	*13.3*	1	*6.7*	1	*6.7*			10	*66.7*	9	60.0	1		2	13.3
Far North Queensland	18			3	*16.7*	1	*5.6*	1	*5.6*	1	5.6	14	*77.8*	11	61.1	3	16.7	1	5.6
South Australia	**169**	**1**	**0.6**	**71**	**42**	**9**	**5.3**	**32**	**18.9**	**30**	**17.8**	**97**	**57.4**	**38**	**22.5**	**57**	**33.7**	**2**	**1.2**
Desert	69			30	*43.5*	4	*5.8*	11	*15.9*	15	*21.7*	39	*56.5*	6	*8.7*	33	*47.8*		
Spencer	57	1	*1.8*	26	*45.6*	3	*5.3*	14	*24.6*	9	*15.8*	30	*52.6*	15	*26.3*	15	*26.3*		
Riverine	3											3	*100.0*	3	*100.0*				
Urban	37			13	*35.1*	2	*5.4*	6	*16.2*	5	*13.5*	24	*64.9*	14	*37.8*	8	*21.6*	2	*5.4*
South Australia (Other)	3			2	*66.7*			1	*33.3*	1	*33.3*	1	*33.3*			1	*33.3*		
Northern Territory	**68**	**3**	**4.4**	**44**	**64.7**			**35**	**51.5**	**9**	**13.2**	**20**	**29.4**	**5**	**7.4**	**15**	**22.1**	**1**	**1.5**
Total	**292**	**5**	*1.7*	**122**	*41.8*	**11**	*3.8*	**69**	*23.6*	**42**	*14.4*	**159**	*54.5*	**79**	*27.1*	**80**	*27.4*	**6**	*2.1*

The text that is in bold in the table represents the States of Australia and the aggregate totals of the regions within them (non-bold text below States). The text in italics is the percentage of individuals who possess that particular haplogroup in each region.

men carried a Y-chromosomes that was derived for marker C-M130 but ancestral for the SNPs C-M8, C-M38, C-M217, and C-M347—all defining major C subhaplogroups. These were considered indigenous and labeled C* (van Oven et al., 2014b). These five males were geographically widely dispersed: 1 from Queensland (also ancestral for C-M526), 1 from South Australia and 3 from the Northern Territory. Haplogroup M-M186 was found in Queensland ($n = 3$), South Australia ($n = 2$), and the Northern Territory ($n = 1$).

Haplogroups K-M526*, S-P308, and M-M186. The most common Aboriginal haplogroups were K-M526* and S-P308, accounting for 159 (54%) of indigenous male lineages (Table 2). The highest frequency of S-P308 was in South Australia (34% of the indigenous South Australian sample) and the lowest in Queensland (12% of the Queensland sample). The highest frequency of K-M526* was in Victoria (73% of the indigenous Victorian sample) and the lowest in the Northern Territory (7% of the indigenous sample). The four haplogroup M-M186 (equivalent of New Guinean specific haplogroup M-M4 ; Kayser et al., 2003) males in Figure 2 were represented by two nodes. The larger node represented haplotypes from South Australia, Northern Territory and Queensland and the other represented a single South Australian male. Both nodes were characterized by long connecting branches.

The branch lengths connecting the nodes of the median-joining haplotype network of K-M526* and S-P308 show that both haplogroups are highly diverse (Figs. 2a, b). All Victorian S-P308 haplotypes (Fig. 2b) laid outside the main cluster of S-P308 haplotypes. The S-P308 main cluster shows a more compact shape and structure than that seen for K-M526*, although there is no clear central node or any haplotype of high frequency. However, haplotypes of K-M526* are less diverse than S-P308 (Supporting Information Table 1.). The K-M526* network is even more visually complex and appears to lack structure, with no central node and no haplotype of high frequency, indicating considerable antiquity. The distribution of haplotypes within this network also strongly suggests that additional K-M526 sublineages are yet to be identified. The low amount of sharing of haplotypes within these haplogroups suggested that they are of great age (Supporting Information Table S2).

Haplogroup C-M347. Haplogroup C-M347 accounted for 122 (41.8%) of the indigenous Australian Y-chromosomes, with the greatest proportion found in the Northern Territory (65% of the Northern Territory indigenous sample) and the lowest in Victoria (9% of the indigenous Victorian sample) (Table 2). Three sub-groups of C-M347 were identified: haplogroup C-DYS390.1del (defined by a unique deletion at the DYS390 Y-STR locus; Forster et al., 1998; Kayser et al., 2001), haplogroup C-M210 and the paragroup C-M347*(xDYS390.1del,M210). Haplogroup C-M210 was found only in Queensland and South Australia and at low frequencies, 1.6% and 7.4%, respectively, of the total C-M347 (Table 2). C-DYS390.1del occurred at its highest frequencies in South Australia and Northern Territory (26% and 28%, respectively, of the total C-M347), was present in Queensland at a low level (2% of the total C-M347), and was absent from Victoria. The most widely distributed C-M347 subhaplogroup (although not the most frequent) was C-M347*(xDYS390.1del,M210). This haplogroup was

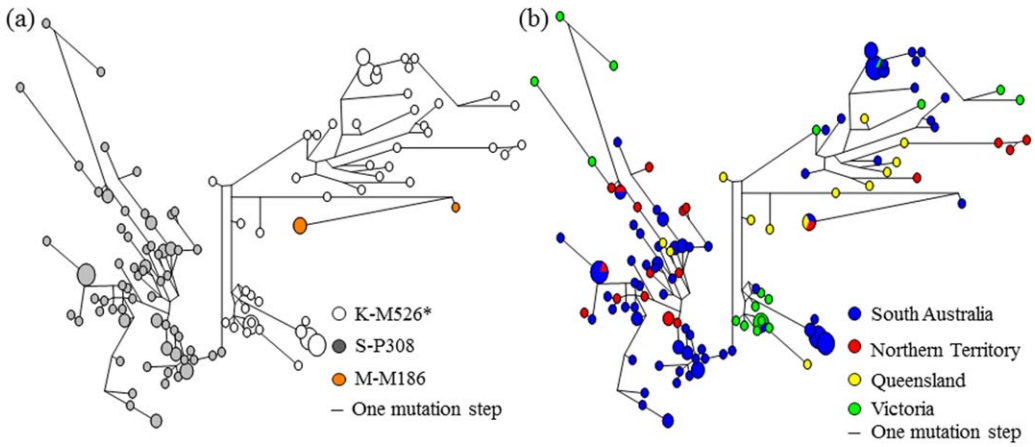

Fig. 2. Networks of K-M526*, S-P308, and M-M186 haplotypes of Aboriginal Australians. Networks based on 14 loci of Yfiler. *Circles* represent haplotypes, with area proportional to frequency. Networks colored according to (**a**) K sub-lineages and (**b**) geography by State or Territory.

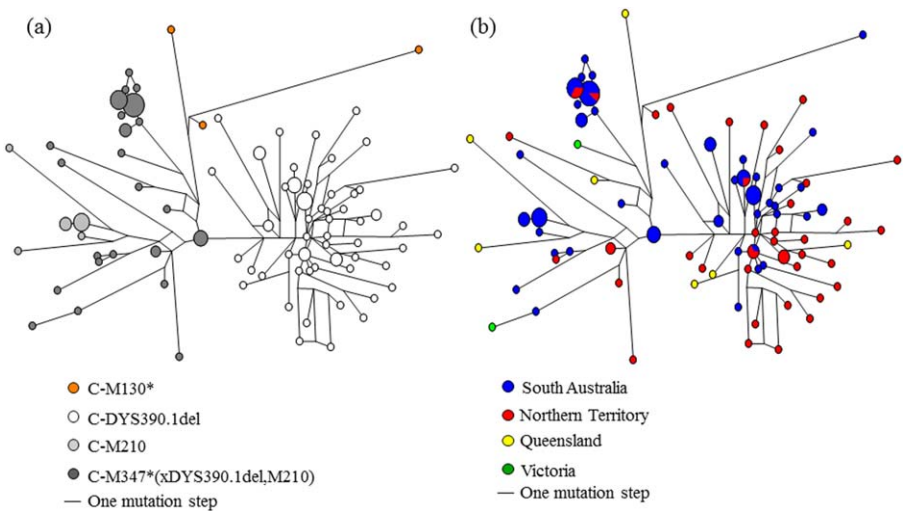

Fig. 3. Networks of C haplotypes of Aboriginal Australians. Networks based on 14 loci of Yfiler. *Circles* represent haplotypes, with area proportional to frequency. Networks colored according to (**a**) C sub-lineages and (**b**) geography by State or Territory.

found in all States, but predominated in South Australia (24% of the total C-M347) and the Northern Territory (7% of the total C-M347).

Networks of haplogroup-C associated Y-STR haplotypes strongly supported the antiquity of this haplogroup, with long branches between haplotypes (Fig. 3). The three C-M130*(xM8,M387,M347,M217) haplotypes from three different States in Figure 3 are interspersed throughout the network with no evidence of clustering, and suggest deep coalescence time given the lengths of their relative connecting branches. The large expansive cluster of C-DYS390.1del haplotypes in the network lacked a distinctive central node. The C-M210 cluster does have a central node with some geographical clustering. C-M347*(xDYS390.1del,M210) haplotypes comprised most of the rest of the network, connected by long branches with some clustering and sharing of haplotypes between regions evident. The C-M210 and C-DYS390.1del clusters arose on the background of the M347 mutation, and were anchored to C-M347*(xDYS390.1del,M210) haplotypes and not to each other, indicating that C-M347*(xDYS390.1del,M210) chromosomes are a paragroup to C-DYS390.1 del and C-M210 chromosomes.

Diversity of indigenous Y-haplogroups

There was a highly significant difference in indigenous haplogroup frequencies across Australia ($\chi^2 = 93.5$, d.f. = 12, $P < 0.0001$). Pairwise analysis revealed that differences between Victoria and Queensland were non-significant, but all other comparisons were significant (haplogroups C*, C-M210, and M-M186 were excluded from this analysis because of their small numbers). Analysis of Molecular Variance (AMOVA) of the indigenous Y-haplogroups demonstrated that variation among the States was significant ($F_{ST} = 0.14$, $P < 0.0001$). AMOVA of the 14-locus Y-STR haplotypes among the States was also significant ($R_{ST} = 0.12$, $P < 0.0001$).

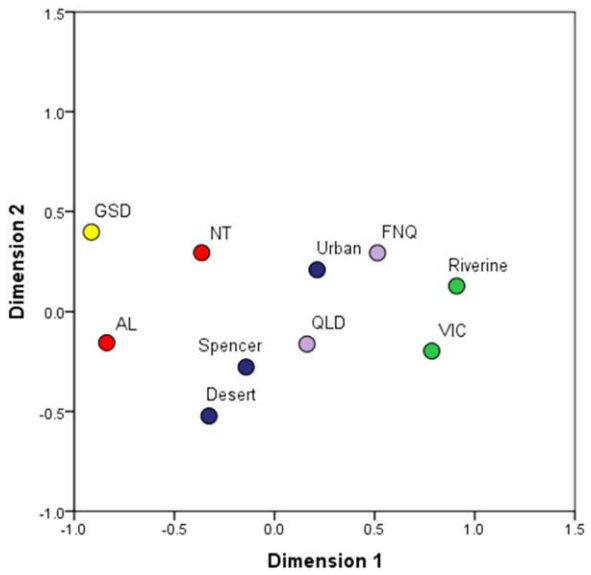

Fig. 4. MDS plot based on R_{ST} values derived from Aboriginal Australian 6-locus Y-STR haplotypes (DYS393, DYS19, DYS391, DYS389I, DYS392, and DYS389II); including data of Kayser et al. (2001). *Red*: Northern Territory, *Yellow*: Western Australia (Great Sandy Desert), *Blue*: South Australia, *Purple*: Queensland, *Green*: Victoria. Normalized stress: 0.02055.

Haplotype diversity

Haplotype diversity within most haplogroups was very high with moderate sharing of haplotypes within haplogroups, suggesting that some haplogroups, such as C-DYS390.1del and S-P308, are of great age (Supporting Information Table S2.). Among the 278 complete 14-locus Y-STR haplotype profiles available for the 292 indigenous Y-chromosomes, 210 unique haplotypes were identified (75.5% discrimination). Of these, 164 haplotypes were singletons (78.1%) and 25 haplotypes (11.9%) were shared only within a State or the Northern Territory ("multiple unique" haplotypes). When combined, these 189 haplotypes represented the total that was specific to a particular State/region (90.0%). The remaining 21 haplotypes (10.0%) were observed in multiple States or the Northern Territory. The majority (70.0%) of the shared haplotypes occurred between South Australia and the Northern Territory.

South Australia has the highest proportion of multiple unique haplotypes ($n = 21$, 18.9%) and the proportion of total unique haplotypes was high (above 85%) for all States. Haplotypes within haplogroups C-M347*(xDYS390.1del,M210) ($n = 2$), C-DYS390.1del ($n = 2$), and S-P308 ($n = 2$) were shared between South Australia and the Northern Territory. Two haplotypes within K-M526* were shared between South Australia and Victoria.

Regional analysis

Interestingly, there was no sharing of any K-M526* or S-P308 haplotypes between the Queensland regions and elsewhere in Australia (Fig. 2b). Sharing of K-M526* or S-P308 haplotypes is seen between South Australia, Victoria, and the Northern Territory. Evidence of microevolution of both K-M526* and S-P308 haplogroups was indicated by the small clusters of haplotypes dispersed throughout the network that belonged to males from the same region. Similarly, there appeared to be some microevolution of C-M347 haplotypes within regions, in particular its subhaplogroup C-M210 in South Australian regions and branches of its other subhaplogroup, C-DYS390.1del, in the Northern Territory (Fig. 3). Haplotype sharing was observed between South Australian regions and the Northern Territory, for both the C-M347*(xDYS390.1del,M210) and C-DYS390.1del haplogroups. Although some sharing of haplotypes is observed there is minimal geographical clustering of haplotypes, at least at the regional level.

To achieve greater geographic coverage of the Australian continent, haplotypes defined by 6 Y-STR loci (DYS19, DYS389I, DYS389II, DYS391, DYS392, and DYS393) in the present indigenous sample were amalgamated with comparable haplotypes reported by Kayser et al. (2001) from Arnhem Land in the Northern Territory and the Great Sandy Desert in Western Australia. These are the only Aboriginal Y-STR datasets with known haplogroup affiliation. Multidimensional scaling of the R_{ST} distances between regions indicated considerable differentiation in the first dimension, with the smallest distance between those of the Northern Territory, South Australia, and Queensland. The Victorian regions and the Great Sandy Desert lay on the fringe of the plot in Figure 4. There were significant differences between the Northern Territory and both Arnhem Land and the Great Sandy Desert. Also, there were significant differences between the Northern Territory and Victoria (associated R_{ST} and P-values given in Supporting Information Table S3).

Additional Y-SNP typing of 26 K* or C* Y-chromosomes reported in previous studies (Kayser et al., 2006; Mona et al., 2009; van Oven et al., 2014a), performed as part of the present study, found no evidence of C-M347 or S-P308 in either New Guinea or in Nusa Tenggara. However, the equivalent K-M526* Y-STR haplotypes in the present sample were compared with the 7-locus KLT-M9* haplotypes of New Guinea typed by Kayser et al. (2014). Six KLT-M9* males of the Massim region in south eastern Papua New Guinea were found to share 7-locus Y-STR profiles with six Aboriginal Australians from Queensland, Victoria, and South Australia. The reduction from 15 to 7 Y-STR loci however considerably decreased the discrimination capacity within the Aboriginal sample from 75% to 48% and influenced the likelihood of finding shared haplotypes between populations.

The equivalent 285 Y-STR haplotypes in the indigenous Aboriginal sample were compared with both the 9-locus haplotypes of 11 Indonesian populations investigated by Tumonggor et al. (2013) (Supporting Information Fig. 2) and the 7-locus haplotypes of 15 Filipino groups investigated by Delfin et al. (2011) (Supporting Information Fig. 3). The Aboriginal haplotypes form a distinct cluster in the MDS plots, being separated from all other groups in dimension 1, strongly suggesting no close paternal ties between the populations. The genetic distances of the Aboriginal Australians from both Indonesian and Filipino populations were statistically significant (Supporting Information Tables S4 and S5). Furthermore, the addition of Arnhem Land and Great Sandy Desert (Kayser et al., 2001) results in a more distinct separation of the northern Australian States and Territory from the rest of the continent in dimension 2, reinforcing the clinical distribution not only at the Y-

TABLE 3. TMRCAs of Aboriginal Australian Y-haplogroups

			Observed mutation rate (OMR) (30 years per generation)					Evolutionary mutation rate (EMR) (25 years per generation)				
				ASD		Bayesian Analysis			ASD		Bayesian Analysis	
Haplogroup	No.	Prior	Network $\rho \pm \sigma$	Median	95% CI	Median	95% CI	Network $\rho \pm \sigma$	Median	95% CI	Median	95% CI
C-M130*	3		6,800 ± 1,416	11,338	(6,137–21,460)			24,840 ± 5,175	41,400	(22,131–80,469)		
		a				22,514	(14,133–36,163)				76,783	(48,500–125,453)
		b				19,565	(8,501–43,587)				75,509	(31,324–158,835)
C-M347	123		8,175 ± 1,358	12,000	(6,697–20,929)			29,865 ± 4,964	43,830	(25,314–76,299)		
		a				14,568	(9,730–22,471)				49,879	(33,471–77,692)
		b				13,638	(5,750–31,541)				54,252	(21,395–111,120)
C-M347*(xDYS390.1del,M210)	42		6,577 ± 1,481	8,160	(4,241–14,714)			24,026 ± 5,412	29,800	(16,068–53,628)		
		a				8,280	(5,462–12,300)				27,966	(18,405–41,934)
		b				8,667	(3,393–21,059)				34,297	(13,197–76,813)
C-DYS390.1del	70		4,062 ± 576	9,633	(5,284–17,286)			14,840 ± 2,105	35,184	(19,539–62,055)		
		a				11,882	(7,656–18,788)				40,462	(26,466–64,656)
		b				11,212	(4,640–28,196)				44,608	(16,521–93,961)
C-M210	11		2,240 ± 551	4,251	(1,852–8,503)			8,185 ± 2,015	15,526	(6,717–31,083)		
		a				5,951	(3,286–10,014)				20,244	(11,604–33,660)
		b				6,134	(2,215–15,643)				23,901	(8,178–58,809)
K-M526*	79		18,103 ± 2,450	19,501	(11,560–33,623)			66,131 ± 8,950	71,247	(41,433–121,683)		
		a				16,270	(11,035–25,832)				57,128	(37,405–92,206)
		b				16,026	(6,466–36,423)				63,223	(24,613–133,519)
S-P308	78		10,275 ± 1,737	10,956	(6,090–19,546)			37,535 ± 6,348	40,014	(22,566–71,085)		
		a				10,748	(7,070–17,082)				37,057	(24,078–58,902)
		b				11,007	(4,463–25,988)				43,644	(16,864–95,470)

[a]Kayser et al. (2001).
[b]Xue et al. (2006).

haplogroup, but also at the Y-haplotype level (Supporting Information Fig. 3).

TMRCA estimates of indigenous Y-haplogroups

The time to the most recent common ancestor (TMRCAs) of chosen sets of Y-chromosomes were estimated using three methods: the rho statistic, ASD, and via the BATWING algorithm. The results from applying each were somewhat similar, despite their differing methodologies. Given the threefold difference between the observed mutation rate (OMR) and the evolutionary mutation rate (EMR), there was a corresponding difference between the estimates using the two rates (Table 3). Application of the OMR indicated that the TMRCA of haplogroup C* was approximately 22 kya, using Kayser et al. (2001) priors, which are specific to Australia and New Guinea. However, rho estimates for this haplogroup are approximately a third of the OMR TMRCA estimate using BATWING. The Bayesian TMRCA of haplogroup C-M347 was approximately 14 kya and that of the paragroup C-M347*(xDYS390.1del,M210) was 8 kya. Haplogroup C-DYS390.1del was approximately 11 kya, while C-M210 was the youngest C-M347 sub-lineage with a TMRCA of approximately 5 kya (Table 3). The TMRCA estimate of K-M526* at 16 kya is greater than that of C-M347 lineage. It can be inferred that S-P308 is at least 10 ky old. Collectively, Rho estimates using the OMR were younger than both ASD and Bayesian analyses.

Employing the EMR, the coalescence estimate of haplogroup C* chromosomes was approximately 76 kya using Kayser et al. (2001)'s priors, however Rho and ASD estimates suggest that the haplogroup is considerably younger (Table 3.) The TMRCA estimates of haplogroup C-M347 suggest that the haplogroup is at least 49 kya based also on Kayser et al. (2001)'s priors. Paragroup C-M347*(xDYS390.1del,M210) and haplogroup C-DYS390.1del were 27 and 40 kya, respectively. C-M210 emerged at least 20 kya. TMRCA estimates for K-M526* chromosomes were at least 57 kya and S-P308 evolved at least 37 kya. The different priors of Kayser et al. (2001) and Xue et al. (2006) gave similar TMRCA estimates for most of the haplogroups but it must be emphasized, however, that some estimates have very broad 95% confidence intervals (Table 3). The TMRCA estimates calculated using ASD suggest that Rho underestimated the age of these haplogroups, as most ASD estimates are closer to the estimates using the Bayesian method, rather than Rho.

DISCUSSION

Here we present Y-chromosome SNP and STR data for contemporary Aboriginal Australians from previously unsampled populations of Queensland, Victoria, and South Australia, and we have analyzed in greater detail samples from the Northern Territory. The resulting database of Aboriginal Australians is not only the largest of its kind but also has greater coverage of the continent than previous studies (Forster et al., 1998; Vandenberg et al., 1999; Kayser et al., 2001; Underhill et al., 2001b; Redd et al., 2002; Hudjashov et al., 2007; Taylor et al., 2012).

In discussing the findings about Aboriginal Australian population structure, it is imperative to recognize the significance of introgression of Eurasian male lineages. The proportion of Eurasian Y-chromosomes in the largest self-declared Aboriginal forensic sample, that of

South Australia, is approximately 80% (Taylor et al 2012), and our data from the Victorian forensic database show similar proportions. High levels of Eurasian Y-chromosome introgression is not uncommon in other populations of Oceania, as Hurles et al. (1998) reported that 27% of Polynesian males from the Cook Islands carried Eurasian Y-chromosomes. Also, Underhill et al. (2001a) found that approximately 48% of self-declared Māori males of New Zealand had European Y-chromosomes. Clearly, there has been a considerable loss of indigenous Australian Y-chromosome diversity since the late 18th century, and inferences from our dataset are mindful of this loss. It has been estimated that at least 50% of the total indigenous population died shortly after European settlement due to the introduction of foreign diseases (e.g., smallpox) (Kirk and Thorne, 1976). The first decades of European colonization dramatically reduced the Aboriginal Australian population. The Aboriginal Protection Board was formed and empowered in 1869 to manage reserves for Aboriginal people and to remove "neglected" or "unprotected" children to these reserves or institutions for their better care (Robson, 1983; Broome, 2005). Later Acts of Parliament, known as the "half caste acts," explicitly prohibited unmarried or half caste Aboriginal people from living on reserves, from visiting "full-blood" family members and from gaining access to Aboriginal welfare (Broome, 2005). More often than not Aboriginal males and females were placed in separate reserves. The repercussions of these policies are still felt today by the children of Aboriginal people who do not know their grandparents or may have only limited knowledge of their Aboriginal heritage and culture (Broome, 2005; van Holst Pellekaan, 2013). A series of bottleneck events, not just one large event, and the introgression of Eurasian males over a period of two centuries, resulted in a significant loss of indigenous Y haplotypes if not haplogroups.

Of some concern was the possibility of including a high number of relatives in the sample especially those with indigenous Y-chromosomes, as may have happened in earlier studies, especially when sample collection was focused on a single location (Redd et al., 2002; Hudjashov et al., 2007). However, the haplotype diversity of the indigenous Y-chromosomes was high, and there was limited sharing of 14-locus Y-STR haplotypes within Y-haplogroups (excluding C-M210 and C4-M347* (xDYS390.1del,M210) indicating that the impact of close relatives is minimal in our sample. In cases where haplotype sharing within haplogroups occurred, it was not possible to clarify the degree of relatedness of the individuals bearing these Y-chromosomes.

Distribution of K-M526*, S-P308, and C-M347 haplogroups

Y-chromosomes falling within the unresolved paragroup K* have been reported within Australia and also in the surrounding regions of New Guinea, Melanesia and Micronesia (Kayser et al., 2006; Scheinfeldt et al., 2006; Karafet et al., 2008a,b; Mona et al., 2009; Delfin et al., 2012; Karafet et al., 2014; van Oven et al., 2014a). Recently, several phylogenetically intermediate K-M526 submarkers were reported altering the internal structure of the haplogroup and resolving its polytomous structure into a dichotomous one (Karafet et al., 2014) (http://www.phylotree.org/Y; van Oven et al., 2014b). The results of genotyping more than 4,000 worldwide K-M526 samples (including 21 Warlpiri males from Central Australia; Redd et al., 2002) for a variety of Y-SNPs, including S-P60, which is phylogenetically equivalent to S-P308, have been recently reported (Karafet et al., 2008a,b, 2014). Of the 21 K-M526 males from the Northern Territory, Karafet et al. (2014) found 18 (85%) belonged to S-P308, whereas in the present Aboriginal sample its frequency was lower (ranging from 8% of Queensland to 71% of the Northern Territory K-M526* chromosomes). The present study confirms that that S-P308 is common among Aboriginal Australians, has a wide geographic distribution throughout Australia, and has accumulated considerable Y-STR diversity, as depicted in Figure 2.

The exceptionally high frequency of S-P308 in the Warlpiri (Karafet et al., 2014) could indicate the presence of (close) paternal relatives in the sample or may reflect a strong paternal bottleneck with subsequent isolation of the Warlpiri population. Interestingly though, our Desert sample (from South Australia), who are in close proximity to the Warlpiri displayed a similar frequency of S-P308. The configuration of the K-M526* network (Fig. 2) also suggests that there may be more substructure characterized by yet unknown Y-SNPs within this haplogroup, which may be followed-up in future resequencing studies.

Combined, K-M526* and S-P308 are the most common Y-haplogroups in our Aboriginal Australian database and this is probably a reflection of the wider geographic coverage of our sample compared with that of earlier studies. These previous studies focused on locations in northern Australia where haplogroup C-M347 has its highest frequencies, and is more common than haplogroup KLT-M9* (Kayser et al., 2001; Hudjashov et al., 2007). Given the wide distribution and antiquity of K-M526* and S-P308 in Australia (and the restriction of the latter to Aboriginal Australians), it is difficult to determine whether the arrival of males carrying K-M526 chromosomes in Sahul represents a single colonization event, with K-M526 chromosomes diverging further within Sahul, or multiple colonization events at differing times in prehistory, bringing a diverse set of K-M526* Y-chromosomes with them.

The existence of a class of Y-chromosomes that is a paragroup to both C-DYS390.1 del and C-M210 haplogroups, that is restricted to Australia, is significant. Haplogroup C-M347*(xDYS390.1del,M210) is widely dispersed throughout the Australian continent, with no single haplotype of great frequency and some evidence of differentiation in South Australia and Northern Territory. The accumulation of mutations along C-M347* (xDYS390.1del,M210) lineages, reflected in the long branches in the Y-STR haplotype networks, indicates that this is a haplogroup of great age.

The C-DYS390.1 del haplogroup is the most discussed Y-haplogroup in Aboriginal Australians, and it has been previously suggested that this lineage (or its ancestor) arrived in Australia with Holocene immigrants from South Asia, with India being the most likely origin (Redd et al., 2002). The association of this haplogroup with an immigration event (Redd et al., 2002) was partly based on its estimated age (based on a small number of Y-STRs) of between 4 and 7 kya (Kayser et al., 2001; Redd et al., 2002). It was postulated that these immigrants introduced the native dog, the dingo, to Australia from India (Gollan, 1985) and are associated with the first presence of a new tool tradition, microliths, in the

archaeological record (Glover and Presland, 1985). Subsequently, based on a genome-wide SNP study, Pugach et al. (2013) suggested that there was a substantial signal (~11%) of gene flow between India and northern Australia during the Holocene, approximately 4,230 years ago. Given this high autosomal percentage of Indian genes one would expect some, perhaps high, level of Indian male introgression into Aboriginal Australians. However, the absence of South Asian-specific C-M356 chromosomes in our large Aboriginal sample, as well as other South Asian haplogroups L-M20 and R-M207, indicated there is no evidence for an Indian male-mediated gene flow into Australia in the Holocene. From the limited Aboriginal Australian mtDNA data available, there is also no evidence of an Indian connection during the Holocene. The uniquely Australian mitochondrial haplogroup M42a has been phylogenetically associated to the Indian M42b haplogroup; however, these diverged greater than 50 kya, probably in South Asia (Kumar et al., 2009). Provided that the signal of Indian ancestry detected by Pugach et al. (2013) reflects a true element of Australian prehistory, the absence of Indian-specific Y and mitochondrial lineages needs to be explained by genetic drift acting more strongly upon the uniparental than the bi-parental parts of the genome.

In addition, recent analyses of canine mitochondrial and Y-chromosome diversity have suggested that the dingo originated in South China rather than in India, along with its closest relative, the New Guinean singing dog (Savolainen et al., 2004; Ardalan et al., 2012; Oskarsson et al., 2012). Thus, there are serious doubts concerning the association between the proposed Indian origin of human Y-haplogroup C-DYS390.1del and the introduction of the dingo as suggested by Redd et al. (2002). Kayser et al. (2001) linked the expansion of Y-chromosome diversity with the period of archaeological "intensification" within Australia beginning about 4 kya including the arrival of the dingo. More recent analysis of archaeological data suggests, however, that there was more than one such period of intensification in Australia's prehistory (Williams, 2013). There is also evidence that microliths were already in use in Australia already before the arrival of the dingo, and developed within existing Aboriginal tool-making traditions (Mulvaney and Kamminga, 1999; Hiscock, 2008; Brown, 2013). Rather, the more parsimonious explanation is that the C-DYS390.1del evolved in situ in a long-settled Australian Aboriginal ancestor on the background of C-M347 Y-chromosome as suggested by Kayser et al. (2001). Clearly, whole-genome analysis of additional Aboriginal Australian samples would help resolve the issue of any Indian connection during the Holocene.

Given its clinal distribution (decreasing inversely with latitude) and relatively high diversity, the C-DYS390.1del haplogroup may have arisen in the Northern Territory, possibly in Arnhem Land. Of the three C-M347 subtypes, the youngest, C-M210, has the most restricted geographical distribution, being found only in South Australia and Queensland. We cannot rule out, however, that C-M210 may be present at a higher frequency at a location(s) not covered in this study. Unfortunately, only two small studies have investigated the SNP M210. One reported its absence (0 of 6 Aboriginal Australian males) (Hudjashov et al., 2007) and the other found the haplogroup at high frequency but in an equally small sample (3 out of 7 Aboriginal Australian males) (Underhill et al., 2001b). However, Underhill et al. (2001b) did not genotype the Y-STRs of the samples, thereby precluding TMRCA estimates for the haplogroup.

What factor(s) can explain the high frequency and continent-wide distribution of the C-DYS390.1del haplogroup versus the much lower frequency and geographically more restricted haplogroup C-M210? Aboriginal people traditionally lived in small groups comprising family members descended through male lines and females moved to the husband's clan at marriage (virilocality). Warfare between groups and/or the success of particular clans could result in a proliferation or steep decline of a particular Y-haplogroup. Also, polygamy was practiced in traditional Aboriginal society and more frequently in the clans that had access to greater and more reliable resources than those living in other less resource-rich parts of the continent (Townsend and Brown, 1978; White, 1997; Scelza and Bird, 2008).

Dating of indigenous Y-chromosomes

Previous estimates of the TMRCA of C-DYS390.1del chromosomes from Northern Territory Aboriginal males (Kayser et al., 2001; Redd et al., 2002) were based on six and seven Y-STRs respectively. Both used average Y-STR mutation rates of 2.8×10^{-3} mutations per locus per generation, values similar to that of the OMR (Gusmao et al., 2005; Sanchez-Diz et al., 2008) rather than that of the EMR (Zhivotovsky et al., 2004). Wei et al. (2013) investigated phylogenies and time estimates for Y-chromosomal lineages based on the analysis of Y-SNP and Y-STR data and compared them with Y-chromosome resequencing data. The dating of ancient Y-haplogroups relying on Y-STR data is prone to error due to the inherent nature of the loci. The mutation rate of Y-STRs is relatively high and recurrence may confound any age estimate of Y-haplogroups in Aboriginal Australians, given their antiquity (Wei et al., 2013). Also, caution must be exercised when interpreting peopling processes via the estimation of TMRCAs of haplogroups, as TMRCAs can typically be older than the age of the population split.

Wei et al. (2013) discovered that estimating TMRCAs using the EMR with BATWING gave age estimates that were more similar (but still an underestimate) to age estimates calculated from resequencing data of the Y-chromosome, than to those using the OMR. Nevertheless, in the current study age estimates were calculated using both the EMR and OMR, but given the findings of Wei et al. (2013) and our own estimates, we are confident that the previous age estimations of C-DYS390.1del of Kayser et al. (2001) and Redd et al. (2002) are greatly underestimated and do not reflect the true antiquity of the haplogroup in Australia.

Origins and migration routes of C-M347 and K-M526

The observation that haplogroup C-M347 and its subtypes are unique to Aboriginal Australians has implications for reconstructing the routes that their ancestors took into Sahul some 50 kya. If these C-M130 chromosomes carried by some of the founders reached Sahul via New Guinea, it is necessary to explain the absence of C-M347 in this region today. There was no major land or sea barrier between Australia and New Guinea until the Holocene, however, there may have been cultural and/or

linguistic barriers to gene flow between them (Evans, 2005).

It is possible that the founders, who carried C-M130 among other possible haplogroups, followed a route along the present Indonesian chain of islands, including Nusa Tenggara, split into two or more groups before entering Sahul. One group entered the region approximating to the coastline of northern Australia some 50 kya, whilst another group entered present day New Guinea, in which subsequently C-M38 and other C subtypes emerged (Scheinfeldt et al., 2006; Zhong et al., 2010). Within Australia, C-M347 arose, in conjunction with other Australian-specific mitochondrial and Y-chromosome haplogroups.

Alternatively, a single group of males carrying C-M130 entered the northern part of Sahul (i.e., New Guinea), while other C-M130 descendants migrated further south into the regions that after the separation became Australia, in which C-M347 arose approximately 44 kya (Table 3). This initial colonization was followed by a long period of isolation between New Guinea and Australia until the Austronesian expansion into New Guinea and the European colonization of Australia. This scenario is supported by autosomal evidence of McEvoy et al. (2010). Full sequencing of the Aboriginal Australian Y-haplogroup C-M130* chromosomes is required to identify their evolutionary relationships to the other Aboriginal C chromosomes (i.e., those within C-M347) and determine their age. The limited sampling of New Guinea for Y-chromosome variation compounds the difficulties in delineating the evolution of these haplogroups.

Resolving the routes and point of entry of the colonizers carrying K-M526* Y-chromosomes without sub haplogroup markers is equally complex. K-M526* chromosomes are present in New Guinea, Melanesia, and Australia today, even though none of the subtypes are shared (Karafet et al., 2014). Thus, it is very possible that similar to C-M130 carriers, the founders carrying K-M526 could have initially settled Sahul via New Guinea and subsequent dispersal throughout the rest of Sahul, with those reaching Australia becoming differentiated from those in New Guinea with time and distance. If so, then the uniquely Australian S-P308 lineage diverged relatively early and has been isolated from New Guinea for over 30 ky. This hypothesis of an early separation of Aboriginal Australians from New Guineans receives support from mtDNA data in the two populations, as haplogroup P mtDNAs shared between New Guineans and Aboriginal Australians (P3 and P4) have been separated for over 30 ky (Ingman and Gyllensten, 2003; Friedlaender et al., 2005; van Holst Pellekaan et al., 2006; Friedlaender et al., 2007).

Alternatively, the founders carrying K-M526 could have followed a similar route and entry point as those founders carrying C-M130. Whereas mtDNA haplogroup P is shared distantly between Australians and New Guineans, other ancient mitochondrial haplogroups are unique to Australians including M42a, M14, M15, S, N13 and O (Ingman et al., 2000; Huoponen et al., 2001; Ingman and Gyllensten, 2003; Kivisild et al., 2006; van Holst Pellekaan et al., 2006; Hudjashov et al., 2007) and their ancestors may well have entered Sahul without passing through New Guinea. Given these ambiguities, it is clearly imperative that more extensive Y-chromosome analyses of New Guinea, Melanesia, and Indonesia are conducted, as many different routes and points of entry by the ancestors of Aboriginal Australians into Sahul are possible.

Haplogroup M-M186 (more commonly known as haplogroup M-M4) today is predominantly found in New Guinea and Melanesia, with modest frequencies in East Indonesia and Micronesia (Kayser et al., 2003). It is thought to have arisen in Melanesia approximately 8.2 kya (Kayser et al., 2003) and its widespread occurrence across New Guinea including throughout the highlands suggests that this haplogroup represents an ancient New Guinea Y lineage. It is unclear whether the Australian M-M186 chromosomes are indigenous, reflecting a shared Sahul Y-chromosome heritage with New Guinea, or their presence reflects (recent) migration from New Guinea. Notably, we found M-M186 in Aboriginal Australians in the Northern Territory, South Australia, and Queensland, albeit in very low frequencies, contrasting with its common occurrence across New Guinea. Further investigation into the history of M-M186 is needed in order to make further conclusions about its presence in Australia.

Limitations

Indigenous samples were not available for the States of New South Wales, Western Australia, and Tasmania and, therefore, the study lacks a comprehensive coverage of the whole continent. The use of de-identified forensic samples meant there was a lack of information about paternal ancestry and language affiliation which, in turn, did not allow an in-depth analysis of structure at a tribal level. Re-sequencing of (a portion of) some of the indigenous Y-chromosomes will reveal more details about their phylogenetic structure, and will either confirm or alter the TMRCA estimates presented here.

We acknowledged that the use of States as regional divisions is not ideal as these are European constructs and do not reflect boundaries defined by Aboriginal languages or clans (Horton, 1996). States and Territory boundaries were employed because most samples have come from State forensic databases, with no genealogical data attached to them. Except in the case of the Northern Territory, there was some location information for South Australia, Queensland, and Victorian samples. Most participants in the Genographic Project sample were unsure of where their father or grandfather' birthplace or tribal affiliation, due to the breakdown of traditional society after European settlement, which resulted in this part of their history being lost. It is therefore difficult to reconstruct Aboriginal history at the regional level from present-day samples for much of Australia, especially the south and eastern part of the continent (as discussed by van Holst Pellekaan, 2013).

CONCLUSIONS

This study is the largest Y-chromosome study of Aboriginal Australians to date, covering four geographic regions within Australia (i.e., three States and one Territory). The detection of C-M347*(xDYS390.1del,M210) chromosomes is of great significance as we can now confirm that haplogroup C-M347 arose within Australia over 40 kya and was not part of a putative South Asian invasion event during the Holocene. The C-DYS390.1del chromosomes should also no longer be associated with the proposed arrival of the dingo and/or presence of microliths in Australia. We can further confirm that the Australian-specific S-P308 arose before 35 kya in a long-

settled population. In addition, we found no evidence of characteristic South Asian Y-chromosomes in Aboriginal Australians. Most importantly, the Y-chromosome findings support those from analyses of mtDNA diversity in identifying the great antiquity of the settlement of the continent by the ancestors of the present-day Aboriginal Australians, with very little or no genetic introgression from outside populations until historic times.

ACKNOWLEDGMENTS

We gratefully acknowledge the participation of Aboriginal Australians from Victoria, Queensland, Northern Territory, South Australia, Western Australia and Tasmania whose collaboration made this study possible. We also thank the jurisdictional forensic departments for access to their samples. "We owe Tammy Williams many thanks for her support throughout this study". In addition, we thank Sheila van Holst Pellekaan and Neville White for their constructive comments on the manuscript.

The participants of the Genographic Consortium are arranged by surname alphabetically in the following list: Syama Adhikarla[a], Christina J. Adler[b], Elena Balanovska[c], Oleg Balanovsky[c], Jaume Bertranpetit[d], Andrew C. Clarke[e], David Comas[d], Alan Cooper[b], Clio S. I. Der Sarkissian[b], Matthew C. Dulik[f], Jill B. Gaieski[f], ArunKumar GaneshPrasad[a], Wolfgang Haak[b], Marc Haber[d,g,] Angela Hobbs[h], Asif Javed[i], Li Jin[j], Matthew E. Kaplan[k], Shilin Li[j], Begoña Martínez-Cruz[d], Elizabeth A. Matisoo-Smith[e], Marta Melé[d], Nirav C. Merchant[k], Amanda C. Owings[f], Laxmi Parida[i], Ramasamy Pitchappan[a], Daniel E. Platt[i], Lluis Quintana-Murci[l], Colin Renfrew[m], Ajay K. Royyuru[i], Arun Varatharajan Santhakumari[a], Fabrício R. Santos[n], Theodore G. Schurr[f], Himla Soodyall[h], David F. Soria Hernanz[o], Pandikumar Swamikrishnan[p], Miguel G. Vilar[f], R. Spencer Wells[o], Pierre A. Zalloua[g] and Janet S. Ziegle[q].

[a]Madurai Kamaraj University, Madurai, Tamil Nadu, India;[b]University of Adelaide, South Australia, Australia; [c]Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia; [d]Universitat Pompeu Fabra, Barcelona, Spain; [e]University of Otago, Dunedin, New Zealand; [f]Department of Anthropology, University of Pennsylvania, PA, USA; [g]Lebanese American University, Chouran, Beirut, Lebanon; [h]National Health Laboratory Service, Johannesburg, South Africa; [i]IBM, Yorktown Heights, NY, USA; [j]Fudan University, Shanghai, China; [k]University of Arizona, Tucson, AZ, USA; [l]Institut Pasteur, Paris, France; [m]University of Cambridge, Cambridge, United Kingdom; [n]Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [o]National Geographic Society, Washington, DC, USA; [p]IBM, Somers, NY, USA and [q]Applied Biosystems, Foster City, CA, USA

LITERATURE CITED

Ardalan A, Oskarsson M, Natanaelsson C, Wilton AN, Ahmadian A, Savolainen P. 2012. Narrow genetic basis for the Australian dingo confirmed through analysis of paternal ancestry. Genetica 140(1-3):65–73.

Arunkumar G, Soria-Hernanz DF, Kavitha VJ, Arun VS, Syama A, Ashokan KS, Gandhirajan KT, Vijayakumar K, Narayanan M, Jayalakshmi M, Ziegle JS, Royyuru AK, Parida L, Wells RS, Renfrew C, Schurr TG, Smith CT, Platt DE, Pitchappan R, Genographic C. 2012. Population differentiation of southern Indian male lineages correlates with agricultural expansions predating the caste system. PLoS One 7(11):e50269.

Athey TW. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal: 34–39.

Behar DM, Thomas MG, Skorecki K, Hammer MF, Bulygina E, Rosengarten D, Jones AL, Held K, Moses V, Goldstein D, Bradman N, Weale ME. 2003. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. American Journal of Human Genetics 73(4):768–779.

Birdsell J. 1977. The recalibration of a paradigm for the first peopling of Greater Australia.

Bowler JM, Jones R, Allen H, Thorne AG. 1970. Pleistocene human remains from Australia: a living site and human cremation from Lake Mungo, Western New South Wales. World Archaeol 2(1):39–60.

Broome R. 2005. Aboriginal Victorians: A History Since 1800. Sydney, Australia: Allen & Unwin.

Brown P. 2013. Of humans, dogs and tiny tools. Nature 494:6–7.

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and Human Evolution. Nature 325:31–61.

de Vries HG, Collèe JM, van Veldhuizen MH, Achterhof L, Smit Sibinga CT, Scheffer H, Buys CH, ten Kate LP. 1996. Validation of the determination of deltaF508 mutations of the cystic fibrosis gene in over 11 000 mouthwashes. Human genetics 97(3).

Delfin F, Myles S, Choi Y, Hughes D, Illek R, van Oven M, Pakendorf B, Kayser M, Stoneking M. 2012. Bridging near and remote Oceania: mtDNA and NRY variation in the Solomon Islands. Mol Biol Evol 29(2):545–64.

Delfin F, Salvador JM, Calacal GC, Perdigon HB, Tabbada KA, Villamor LP, Halos SC, Gunnarsdottir E, Myles S, Hughes DA, Xu S, Jin L, Lao O, Kayser M, Hurles ME, Stoneking M, De Ungria MC. 2011. The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups. Eur J Hum Genet 19(2):224–30.

Dixon RMW. 1980. The Languages of Australia. New York: Cambridge University Press.

Dodson M. 2000. Human genetics: control of research and sharing of benefits. Australian Aboriginal Studies 1/2:56–64.

Dupuy BM, Stenersen M, Egeland T, Olaisen B. 2004. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum Mutat 23(2):117–24.

Evans N. 2005. Australian languages reconsidered: a review of Dixon (2002). Oceanic Linguistics 44(1):243–286.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564–7.

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am J Phys Anthropol 128(2):415–23.

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. American journal of human genetics 59(4):935–45.

Forster P, Kayser M, Meyer E, Roewer L, Pfeiffer H, Benkmann H, Brinkmann B. 1998. Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. Mol Biol Evol 15(9):1108–1114.

Friedlaender J, Schurr T, Gentz F, Koki G, Friedlaender F, Horvat G, Babb P, Cerchio S, Kaestle F, Schanfield M, Deka R, Yanagihara R, Merriwether DA. 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol Biol Evol 22(6):1506–17.

Friedlaender JS, Friedlaender FR, Hodgson JA, Stoltz M, Koki G, Horvat G, Zhadanov S, Schurr TG, Merriwether DA. 2007. Melanesian mtDNA complexity. PLoS One 2(2):e248.

Gao X, Serjeantson SW. 1991. Diversity in HLA-DR4-related DR, DQ haplotypes in Australia, Oceania, and China. Human Immunology 32(4):269–276.

Glover IC, Presland G. 1985. Microliths in Indonesian flaked stone industries. In: Misra VN, Bellwood P, editors. Recent Advances in Indo-Pacific Prehistory. New Delhi, India: Oxford & IBH. p. 185–195.

Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Kruger C, Roewer L, Lessig R, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M. 2009. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. Int J Legal Med 123(6):471–82.

Goldstein DB, Linares aR, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics 139(1):465–471.

Gollan K. 1985. Prehistoric dogs in Australia: an Indian origin? In: Misra VN, Bellwoood P, editors. Recent Advances in Indo-Pacific Prehistory. New Delhi, India: Oxford & IBH. p. 439–443.

Gusmao L, Sanchez-Diz P, Calafell F, Martin P, Alonso CA, Alvarez-Fernandez F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Lonardo AM, Espinheira R, Fagundes de Carvalho E, Farfan MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza Goes AC, Santapa OA, Sumita DR, Vallejo G, Vidal Rioja L, Vide MC, Vieira da Silva CI, Whittle MR, Zabala W, Zarrabeitia MT, Alonso A, Carracedo A, Amorim A. 2005. Mutation rates at Y chromosome specific microsatellites. Hum Mutat 26(6):520–8.

Hiscock P. 2008. The Archaeology of Ancient Australia. Routledge.

Horton DR. 1996. Aboriginal Australia Wall Map. Aboriginal Studies Press.

Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R, Forster P. 2007. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. Proc Natl Acad Sci U S A 104(21):8726–30.

Huoponen K, Schurr TG, Chen Y, Wallace DC. 2001. Mitochondrial DNA variation in an Aboriginal Australian population: evidence for genetic isolation and regional differentiation. Human Immunology 62(9):954–969.

Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling Ma, Sykes BC. 1998. European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. American Journal of Human Genetics 63(6):1793–1806.

IBM. 2012. SPSS v21.0 for Windows. Chicago IL.

Ingman M, Gyllensten U. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res 13(7):1600–6.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408(6813):708–713.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18(5):830–8.

Karafet TM, Mendez FL, Sudoyo H, Lansing JS, Hammer MF. 2014. Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia. European Journal of Human Genetics.

Kayser M, Brauer S, Cordaux R, Casto A, Lao O, Zhivotovsky LA, Moyse-Faurie C, Rutledge RB, Schiefenhoevel W, Gil D, Lin AA, Underhill PA, Oefner PJ, Trent RJ, Stoneking M. 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. Mol Biol Evol 23(11):2234–44.

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. American journal of human genetics 72(2):281–302.

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill PA, Stoneking M. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. Am J Hum Genet 68(1):173–190.

Kayser M, Choi Y, van Oven M, Mona S, Brauer S, Trent RJ, Suarkia D, Schiefenhovel W, Stoneking M. 2008. The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. Mol Biol Evol 25(7):1362–74.

Kirk R, Thorne A. 1976. The Origin of Australians. Atlantic Highlands, NJ: Humanities Press.

Kirk RL, Cleve H, Bearn AG. 1963. The Distribution of Gc-Types in Sera From Australian Aborigines. American Journal of Physical Anthropology 21:215–223.

Kirk RL, Sanghvi LD, Balakrishnan V. 1972. A further study of genetic distance among Australian aborigines: nine tribes in the Northern Territory. Humangenetik 14(2):95–102.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ. 2006. The role of selection in the evolution of human mitochondrial genomes. Genetics 172(1):373–87.

Kowal EE. 2012. Genetic research in Indigenous health: significant progress, substantial challenges. Med J Aust 197(1):19–20.

Kumar S, Ravuri RR, Koneru P, Urade BP, Sarkar BN, Chandrasekar A, Rao VR. 2009. Reconstructing Indian-Australian phylogenetic link. BMC Evol Biol 9:173.

McEvoy BP, Lind JM, Wang ET, Moyzis RK, Visscher PM, van Holst Pellekaan SM, Wilton AN. 2010. Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. Am J Hum Genet 87(2):297–305.

Mona S, Grunz KE, Brauer S, Pakendorf B, Castri L, Sudoyo H, Marzuki S, Barnes RH, Schmidtke J, Stoneking M, Kayser M. 2009. Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. Mol Biol Evol 26(8):1865–77.

Mulvaney DJ, Kamminga J. 1999. Prehistory of Australia. Washington, D.C: Smithsonian Institution Press.

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70(12):3321–3.

O'Connell JF, Allen J. 2015. The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. Journal of Archaeological Science 56:73–84.

O'Connell JF, Allen JC. 2004. Dating the colonization of Sahul (Pleistocene Australia–New Guinea): a review of recent research. Journal of Archaeological Science 31(6):835–853.

O'Connor S. 1995. Carpenter's Gap Rockshelter 1: 40,000 years of Aboriginal occupation in the Napier Ranges, Kimberley, WA. Australian Archaeology 40(58-59).

Oskarsson MC, Klutsch CF, Boonyaprakob U, Wilton A, Tanabe Y, Savolainen P. 2012. Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. Proc Biol Sci 279(1730):967–74.

Pugach I, Delfin F, Gunnarsdottir E, Kayser M, Stoneking M. 2013. Genome-wide data substantiate Holocene gene flow from India to Australia. Proc Natl Acad Sci U S A 110(5):1803–8.

Radcliffe-Brown AR. 1930. Former numbers and distribution of Australian Aborigines. Official Year Book 23:687–696.

Rambaut A, Drummond AJ. 2013. TRACER v1.6. http://beast.bio.ed.ac.uk/Tracer.

Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T, Kivisild T, Zhai W, Eriksson A, Manica A, Orlando L, De La Vega FM, Tridico S, Metspalu E, Nielsen K, Avila-Arcos MC, Moreno-Mayar JV, Muller C, Dortch J, Gilbert MT, Lund O, Wesolowska A, Karmin M, Weinert LA, Wang B, Li J, Tai S, Xiao F, Hanihara T, van Driem G, Jha AR, Ricaut FX, de Knijff P, Migliano AB, Gallego Romero I, Kristiansen K, Lambert DM, Brunak S, Forster P, Brinkmann B, Nehlich O, Bunce M, Richards M, Gupta R, Bustamante CD, Krogh A, Foley RA, Lahr MM, Balloux F, Sicheritz-Ponten T, Villems R, Nielsen R, Wang J, Willerslev E. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334(6052):94–8.

Redd AJ, Roberts-Thomson J, Karafet T, Bamshad M, Jorde LB, Naidu JM, Walsh B, Hammer MF. 2002. Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. Current Biology 12(8):673–677.

Redd AJ, Stoneking M. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet 65(3):808–28.

Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM, Ko YC, Jinam TA, Phipps ME, Saitou N, Wollstein A, Kayser M, Paabo S, Stoneking M. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet 89(4):516–28.

Roberts-Thomson JM, Martinson JJ, Norwich JT, Harding RM, Clegg JB, Boettcher B. 1996. An ancient common origin of aboriginal Australians and New Guinea highlanders is supported by alpha-globin haplotype analysis. American Journal of Human Genetics 58(5):1017–1024.

Roberts RG, Jones R. 1994. Luminescence dating of sediments: new light on the human colonisation of Australia. Australian Aboriginal Studies 1994(2):2–17.

Robson LL. 1983. A History of Tasmania. Volume 1 Van Diemen's Land from the earliest times to 1855. Melbourne: OUP.

Saillard J, Forster P, Lynnerup N, Bandelt HJ, Nørby S. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. American journal of human genetics 67(3):718–26.

Sanchez-Diz P, Alves C, Carvalho E, Carvalho M, Espinheira R, Garcia O, Pinheiro MF, Pontes L, Porto MJ, Santapa O, Silva C, Sumita D, Valente S, Whittle M, Yurrebaso I, Carracedo A, Amorim A, Gusmao L, Gep I. 2008. Population and segregation data on 17 Y-STRs: results of a GEP-ISFG collaborative study. Int J Legal Med 122(6):529–33.

Savolainen P, Leitner T, Wilton AN, Matisoo-Smith E, Lundeberg J. 2004. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. Proc Natl Acad Sci U S A 101(33):12387–90.

Scelza B, Bird RB. 2008. Group Structure and Female Cooperative Networks in Australia's Western Desert. Human Nature 19(3):231–248.

Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet T, Hammer M, Lorenz J. 2006. Unexpected NRY chromosome variation in Northern Island Melanesia. Mol Biol Evol 23(8):1628–41.

Sturrock K, Rocha J. 2000. A multidimensional scaling stress evaluation table. Field Methods 12(1):49–60.

Summerhayes GR, Leavesley M, Fairbairn A, Mandui H, Field J, Ford A, Fullagar R. 2010. Human adaptation and plant use in highland New Guinea 49,000 to 44,000 years ago. Science 330(6000):78–81.

Taylor D, Nagle N, Ballantyne KN, van Oorschot RA, Wilcox S, Henry J, Turakulov R, Mitchell RJ. 2012. An investigation of admixture in an Australian Aboriginal Y-chromosome STR database. Forensic Sci Int Genet 6(5):532–8.

Thorne A, Grün R, Mortimer G, Spooner Na, Simpson JJ, McCulloch M, Taylor L, Curnoe D. 1999. Australia's oldest human remains: age of the Lake Mungo 3 skeleton. Journal of Human Evolution 36(1976):591–612.

Townsend GC, Brown T. 1978. Inheritance of tooth size in Australian aboriginals. American journal of physical anthropology 48(3):305–14.

Tsintsof AS, Hertzberg MS, Prior JF, Mickleson KN, Trent RJ. 1990. Alpha-globin gene markers identify genetic differences between Australian aborigines and Melanesians. American Journal of Human Genetics 46(1):138–143.

Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, Cox MP. 2013. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. J Hum Genet 58(3):165–73.

Turney CSM, Bird MI, Fifield LK, Roberts RG, Smith M, Dortch CE, Grün R, Lawson E, Ayliffe LK, Miller GH, Dortch J, Cresswell RG. 2001a. Early Human Occupation at Devil's Lair, Southwestern Australia 50,000 Years Ago. Quaternary Research 55(1):3–13.

Turney CSM, Kershaw AP, Moss P, Bird MI, Fifield LK, Cresswell RG, Santos GM, Di Tada ML, Hausladen PA, Zhou Y. 2001b. Redating the onset of burning at Lynch's Crater (North Queensland): implications for human settlement in Australia. Journal of Quaternary Science 16(8):767–771.

Underhill Pa, Passarino G, Lin aa, Marzuki S, Oefner PJ, Cavalli-Sforza LL, Chambers GK. 2001a. Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. Human Mutation 17(4):271–280.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001b. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Annals of Human Genetics 65(Pt 1):43–62.

van Holst Pellekaan S. 2013. Genetic evidence for the colonization of Australia. Quaternary International 285:44–56.

van Holst Pellekaan SM. 2000a. Genetic Research: What does this mean for Indigenous Australian communities? Australian Aboriginal Studies 1 & 2:65–75.

van Holst Pellekaan SM. 2000b. Human Genome Diversity: Ethics and Practice in Australia. Human Evolution 19(2):131–144.

van Holst Pellekaan SM. 2008. Origins of the Australian and New Guinean Aborigines. Encyclopedia of Life Sciences. Chichester: John Wiley and Sons.

van Holst Pellekaan SM, Frommer M, Sved J, Boettcher B. 1998. Mitochondrial control-region sequence variation in aboriginal Australians. American Journal of Human Genetics 62(2):435–449.

van Holst Pellekaan SM, Ingman M, Roberts-Thomson J, Harding RM. 2006. Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. Am J Phys Anthropol 131(2):282–94.

van Oven M, Brauer S, Choi Y, Ensing J, Schiefenhovel W, Stoneking M, Kayser M. 2014a. Human genetics of the Kula Ring: Y-chromosome and mitochondrial DNA variation in the Massim of Papua New Guinea. Eur J Hum Genet.

van Oven M, Ralf A, Kayser M. 2011. An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. Int J Legal Med 125(6):879–85.

van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH. 2014b. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum Mutat 35(2):187–91.

Vandenberg N, van Oorschot RAH, Tyler-Smith C, Mitchell RJ. 1999. Y-chromosome-specific microsatellite variation in Australian Aboriginals. Human Biology 71:915–931.

Walsh SJ, Mitchell RJ, Watson N, Buckleton JS. 2007. A comprehensive analysis of microsatellite diversity in Aboriginal Australians. J Hum Genet 52(9):712–28.

Wei W, Ayub Q, Xue Y, Tyler-Smith C. 2013. A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. Forensic Sci Int Genet 7(6):568–72.

White NG. 1997. Genes, languages and landscapes in Australia. In: McConvell P, Evans N, editors. Archaeology and Linguistics: Aboriginal Australia in Global Perspective. p. 45–83.

Williams AN. 2013. A new population curve for prehistoric Australia. Proc Biol Sci 280(1761):20130486.

Willuweit S, Roewer L, International Forensic YCUG. 2007. Y chromosome haplotype reference database (YHRD): update. Forensic Sci Int Genet 1(2):83–7.

Wilson IJ, Weale ME, Balding DJ, Wilson J. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. Journal of the Royal Statistical Society. Series A (Statistics in Society) 166(2):155–201.

Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, Yang H, Tyler-Smith C. 2006. Male demography in East Asia: a north-south contrast in human population expansion times. Genetics 172(4):2431–9.

Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. American journal of human genetics 74(1):50–61.

Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, Su B. 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. J Hum Genet 55(7):428–35.