

# A comprehensive analysis of microsatellite diversity in Aboriginal Australians

Simon J. Walsh · R. John Mitchell ·  
Natalie Watson · John S. Buckleton

Received: 19 March 2007 / Accepted: 18 June 2007 / Published online: 13 July 2007  
© The Japan Society of Human Genetics and Springer 2007

**Abstract** Indigenous Australians have a unique evolutionary history that has resulted in a complex system of inter and intra-tribal relationships. While a number of studies have examined the population genetics of indigenous Australians, most have used a single sample to illuminate details of the global dispersal of modern humans and few studies have focussed on the population genetic features of the widely dispersed communities of the indigenous population. In this study we examine the largest Aboriginal Australian sample yet analysed ( $N = 8,868$ ) at fifteen hypervariable autosomal microsatellite loci. A comprehensive analysis of differentiation indicates different levels of heterogeneity among indigenous peoples from traditional regions of Aboriginal Australia. The most genetically differentiated populations inhabit the North of the country, in particular the Tiwi of Melville and Bathurst islands, Arnhem Land (itself divided into West and East Arnhem), and Fitzmaurice regions. These tribal groups are

most differentiated from other Aboriginal Australian tribes, especially those of the Central Desert regions, and also show marked heterogeneity from one another. These genetic findings are supportive of observations of body measurements, skin colour, and dermatoglyphic features which also vary substantially between tribes of the North (e.g. Arnhem Land) and Central Australian regions and, more specifically, between the Tiwi and West and East Arnhem tribes. This study provides the most comprehensive survey of the population genetics of Aboriginal Australia.

**Keywords** DNA profiling · Autosomal STRs/ microsatellites · Australian aborigines ·  $F_{ST}$  · Phylogeny · Sub-population diversity · Population genetics

## Introduction

The archaeological record in the Australian continent provides evidence for a long period of human habitation and, together with linguistic evidence, of marked isolation from its near neighbours. Dispersal of a branch of modern humans from Africa is believed to have progressed along a southern coastal route into south Asia and down into present day Papua New Guinea (PNG) and Australia (Underhill et al. 2001). The entry of modern humans into Sahul (the single Pleistocene landmass encompassing Australia, New Guinea, and Tasmania) is estimated to have occurred at least 50 kya (Cavalli-Sforza et al. 1994). A similar period of occupation has been given for present day Australia itself (White and O'Connell 1982; Flood 1983; Roberts et al. 1990, 1994; Thorne et al. 1999; Turney et al. 2001). The isolation of the Australian population was maintained for millennia and was only disturbed after

---

S. J. Walsh (✉)  
Forensic and Technical Services, Australian Federal Police,  
GPO Box 401, Canberra, ACT 2601, Australia  
e-mail: simonjoseph.walsh@afp.gov.au

S. J. Walsh  
Centre for Forensic Science, UTS, PO Box 123,  
Broadway, NSW 2007, Australia

R. J. Mitchell  
Department of Human Genetics, LaTrobe University,  
Melbourne, Australia

N. Watson  
Forensic Services Group, NSW Police, Sydney, Australia

J. S. Buckleton  
ESR Ltd, Private Bag 92-021, Auckland, New Zealand  
e-mail: john.buckleton@esr.cri.nz

European colonisation and permanent settlement in 1788 and the subsequent massive immigration from Europe and, more recently, Asia.

At the time of European colonisation, Australia was at a stage of economic development, that of hunter–gatherer, which had been replaced thousands of years earlier in most parts of the world. It is thought there were between 400 and 600 tribal groups each containing an average of 500–1,000 individuals (Birdsell 1993). Tribes tended to be defined on linguistic and/or territorial criteria, although there is greater support for linguistic delineation. Horton (Horton (creator) 1996) allocated the tribes, on the basis of linguistics, to eighteen Aboriginal Australian regions (Fig. 1).

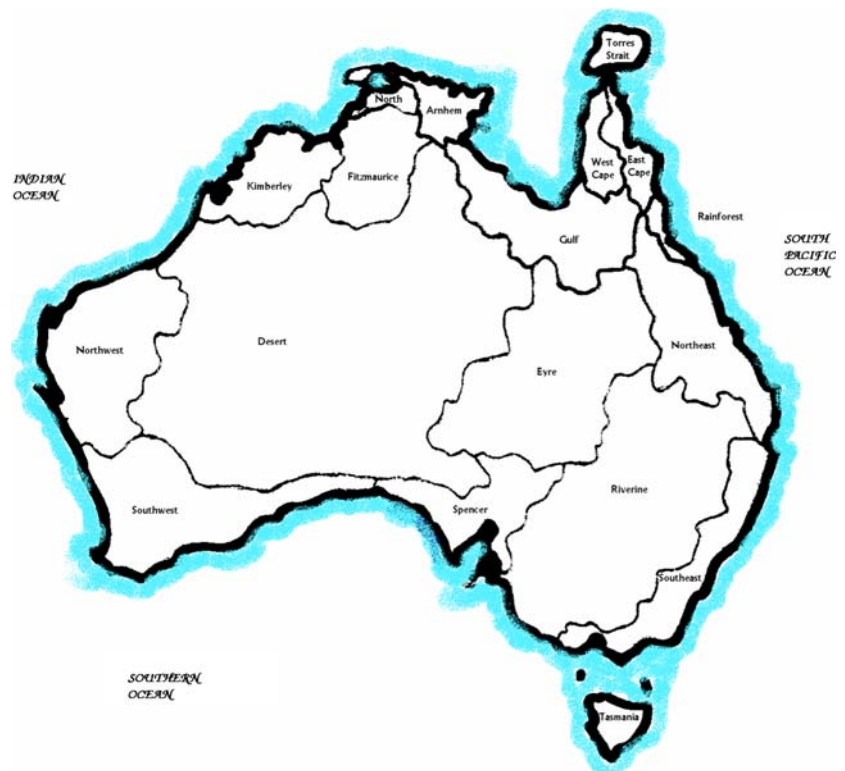
European colonisation disrupted the traditional lives of Aboriginal Australians in a massive way. In the first 130 years after 1788 Aboriginal population numbers fell drastically, from the original population of approximately 250,000 (Jones 1970) to 61,000 recorded at the 1921 census (Kirk 1983). Since that time the recorded Aboriginal population has increased and now approaches 500,000 which is approximately 2% of the current Australian population of 19.6 million. At present, 30% of the indigenous population live in major cities, approximately 44% in regional areas and 26% in remote areas (ABS 1998, 2004).

Studies of DNA polymorphisms in Aboriginal populations are meagre compared with those for most other major groups of the world. Those that have been conducted have

usually been limited to analysis of a sample of a single tribe and the purpose has been to illuminate the dispersal of modern humans and/or attempt to date the founding event(s) or points of entry to the continent and/or detect Aboriginal affinities with neighbouring groups. Only a limited investigation, using DNA markers, has been conducted of diversity among Australians. Studies, based on “classical markers” (blood group antigens, HLA, serum proteins, and red cell enzymes) investigated genetic diversity among some Aboriginal groups and found that linguistic differences best explained the differences among populations (Balakrishnan et al. 1975; Kirk 1989) and tended to support other evidence suggesting indigenous Australians and inhabitants of highland PNG were more closely related to each other than either was to other populations of Asia and Polynesia.

Phylogenetic studies of Australasian and Pacific populations using mitochondrial DNA (mtDNA) have variously included Aboriginal samples but often without any provenance details (Cann et al. 1987; Stoneking 1994; Merriwether et al. 1999; Redd and Stoneking 1999; Friedlaender et al. 2002). Commonly some affinity between Australian and PNG Highland samples has been observed (van Holst Pellekaan et al. 1997, 1998, 2006; Lum et al. 1998; Huoponen et al. 2001), but there are exceptions to this view (Stoneking et al. 1990; Redd and Stoneking 1999). Compared with other global mtDNA haplogroup data those

**Fig. 1** The eighteen linguistically defined regions of Aboriginal Australia as defined by Horton (Horton (creator) 1996)



from indigenous Australians tend to form distinct clusters. Huopenen et al. (2001) found significant mtDNA sequence differences between a Central Desert group (the Walbiri) and a sample from Southeast Australia, and in both samples found evidence of uniqueness of the Australians compared with their near neighbours indicative of long isolation. In the most recent study of the mtDNA genome in Aborigines (van Holst Pellekaan et al. 2006) the five Australian-specific haplogroups were indicative of genetic isolation for many millennia. Further, results from coalescence analysis of some of these sequences supported evidence for a continuity of presence of descendants of a founding population from at least 40,000 years ago. Whereas two haplogroups were distributed widely throughout the continent the other three were more restricted with two found only in the Central Desert region.

There have been even fewer studies of Y-chromosome-specific polymorphisms. The results generally support an Asian origin of the founder male population but there are many differences between present day Australians and their near neighbours, again indicating substantial isolation (Forster et al. 1998; Vandenberg et al. 1999; Kayser et al. 2001, 2003; Underhill et al. 2001).

The Aboriginal hunter-gatherer lifestyle and isolation for approximately 50,000 years led geneticists to expect much microdifferentiation among the contemporary indigenous inhabitants. The carrying capacity of the majority of the land meant the density per square km would be relatively low; this condition is highly conducive to low effective population size,  $N_e$ , and hence the opportunity for random drift of alleles. Since European settlement it has been known that natural disasters could severely reduce the population size of a band or a tribe. The Kaiadilt of Bentinck Island, for example, suffered a population crash, principally because of drought and this left an  $N_e$  of approximately 20 individuals, which is certainly small enough for random drift to operate significantly (Simmons et al. 1962). These natural calamities would have had greatest repercussions on groups inhabiting the marginal lands of the interior. Those inhabiting the relatively resource-rich riverine and coastal areas would, to some extent, have been protected from the harshest conditions because they had a more diverse range of resources to rely upon. In particular, coastal foods would be less affected by terrestrial climatic stress.

In Central Australia (CA) tribes traditionally occupied large territories in arid to semi-desert areas with low population densities. For example, the Pintubi numbered less than 1 per 200 km<sup>2</sup> (Kimber 1990) and suffered considerable mortality from drought (Kimber 1990). In such circumstances hunter-gatherer bands stressed inter-tribal marriage alliances which allowed access rights in the territories (and hence resources) of other groups. Arnhem

Land (AL), in the North of Australia, by contrast, is a region characterised by a monsoonal and markedly seasonal climate with greater habitat diversity and a wider range, as well as predictability, of food resources, particularly in coastal regions. Population densities range from 2 per km<sup>2</sup> on the coast to 1 per 15 km<sup>2</sup> in inland AL, with tribes having fewer members and occupying much smaller territories than in CA. The ecological and demographic characteristics of AL and the Cape York regions (in the North Eastern corner of the continent) favour smaller, more stable population units, shorter marriage distances and marriage systems characterised by polygyny. These would be expected to generate higher levels of diversity among intramarrying groups, as a result of drift. This is in contrast with the larger and more mobile populations of CA (White 1997). These large ecological and anthropological differences between CA and AL have implications for the demographic genetics of the tribes.

Complicating any attempt to comprehend genetic differentiation in Australian Aborigines is the issue of admixture with exotic populations. The coastal people in the north of the continent have been exposed to admixture with sea-faring peoples from the north. In particular, there was seasonal movement in historical times (only ceasing in the early twentieth century) of Macassan seafarers to the northern coast to collect the sea cucumber (trepang) (MacKnight 1972). Evidence of this contact is copious—both in the oral stories of the Aborigines themselves and the incorporation of Macassan words into the Yolngu language (one of AL's largest tribes) (Walker and Zorc 1981).

Although there have been extremely few studies of genetic variation among Aboriginal groups at the DNA level, this is not so for classical markers (Balakrishnan et al. 1975; Cavalli-Sforza et al. 1994). Balakrishnan et al. (1975), reviewed classical marker (blood group, red cell enzyme, and serum protein polymorphisms) data for a variety of communities in three States of Australia. The main conclusion was that linguistic diversity most closely, but by no means perfectly, matches genetic diversity, and that diversity is greatest along the northern coast of Australia, from the northwest, through AL and into Cape York. A comprehensive review of Aboriginal population diversity on the basis of classical markers (Cavalli-Sforza et al. 1994) suggests a clinal gradient of differentiation that is most pronounced in the North-Eastern tip of the continent (Cape York Peninsula).

In summary, the extent of genetic differentiation among Australia's indigenous population remains poorly understood and what knowledge there is on the topic exists, at the DNA level of analysis especially, as a collection of fairly disparate data. Further, these data have commonly been gathered to address questions surrounding population differentiation on a regional and/or global scale, rather than

within the continent itself. These data also comprise small samples (usually <100 individuals) and are often not well provenanced by geography, tribe, or language; they were also tested for a small number of genetic markers only.

Given our current understanding of the experience of the indigenous population, both before and after European colonisation, one might expect considerable genetic differentiation in the contemporary population. It is also very likely that the continent's remarkably varied ecological zones played a prominent role in shaping any such diversity, by acting on any initial genetic diversity brought in by the founding group(s). At present, however, we do not have an adequate measure of genetic structure in the Aboriginal population and, therefore, have limited means of assessing differing environmental, customary, cultural, and/or post-colonisation effects.

In an attempt to address some of these uncertainties this study has assembled the largest Aboriginal Australian dataset to date. The data have been analysed, wherever possible, using traditional regional and tribal affiliations in order to detect substructure within the population. Although the study is limited to fifteen autosomal microsatellites it is, nonetheless, the most comprehensive analysis of genetic diversity within Australia's indigenous population.

## Materials and methods

### Population data

Anonymous DNA autosomal microsatellite information from Aboriginal Australians collected during forensic casework was contributed by six Australian forensic agencies representing Western Australia (WA,  $N = 336$ ), South Australia (SA,  $N = 335$ ), Queensland (QLD,  $N = 545$ ), New South Wales (NSW,  $N = 2,558$ ), Victoria (VIC,  $N = 363$ ), and the Northern Territory (NT) which maintains both a "Pure" ( $N = 586$ ) and a "Declared" ( $N = 5,378$ ) dataset. In the State jurisdictions of WA, QLD, NSW, SA, and VIC all samples were collected during forensic casework and ethnicity was assigned by self-declaration. As mentioned, NT maintains two Aboriginal Australian datasets described as "Pure" and "Declared". The Pure dataset includes individuals who met a number of criteria: they live in a remote district, have a skin name, or were assigned as pure blood on the basis of information from the investigating officers. The Declared dataset comprises those who self-declared themselves as Aboriginal. Each sample came with the location of the offence with which the donor was associated. Not all tribes in the NT are represented by a separate Pure dataset. The present sample of indigenous

Australians, although very large, is heavily biased to NSW and the NT, with these two jurisdictions supplying more than 80% of all the samples.

The fifteen autosomal microsatellites used in this analysis were: D3S1358, HumvWA/vWF31A, HumFIBRA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, D2S1338, D19S253, TH01, TPOX, and CSF1PO. All samples were scored using one, or a combination, of the nine-locus AMPF/STR Profiler Plus (Applied Biosystems, Foster City, CA, USA) the three-locus GenePrint CTT Multiplex (Promega, Madison, WI, USA); and/or the fifteen-locus AMPF/STR Identifiler (Applied Biosystems, Foster City, CA, USA) multiplex PCR systems.

### Partitioning the data

All samples were provided with an accompanying geographic placeholder. This placeholder information most commonly indicated either the place of residence or the location of the forensic matter with which the individual was associated and referred to contemporary geographical locations, for example town or a remote community centre. Reference to Horton's map of Aboriginal Australia (Horton (creator) 1996) allowed a given placeholder to be converted to a location within a traditional region and also a tribal territory (Table 1). It is unlikely that all samples come from the regional or tribal populations assigned to them, but we consider it a reasonable treatment of the data. It would be wrong to suppose, for example, that the location of the forensic matter corresponds exactly with the birthplace or homeland of the donor.

After allocation of all individuals to a location, any grouping of fifteen or more individuals was used in the next phase of the analysis, which was allocation to the fourteen regional populations identified by Horton (Fig. 1, Table 1). These are labelled: Arnhem, North, Kimberley, Northwest, Desert, Spencer, Riverine, Southeast, Northeast, Rainforest, Torres Strait, West Cape, Gulf, and Fitzmaurice. Eight groups were also drawn from Aboriginal individuals resident in one of the urban centres. The urban populations were Darwin Urban (pure,  $N = 153$ ), Darwin Urban (declared,  $N = 1,249$ ), Brisbane Urban ( $N = 120$ ), Adelaide Urban ( $N = 182$ ), Perth Urban ( $N = 168$ ), Sydney Urban (Central,  $N = 406$ ), Sydney Urban (North,  $N = 49$ ), and Sydney Urban (West,  $N = 229$ ). It was expected these urban populations would contain individuals with a mixture of regional and tribal backgrounds. Samples were then assigned to one of 65 traditional tribal territories identified by Horton (Horton (creator) 1996). These tribal territories are referred to as "remote", to distinguish them from urban samples. These 65 tribal populations (Table 2) formed the most informative sample set.

**Table 1** Summary of Aboriginal data used in the analysis of genetic differentiation and the segregation of these data into 14 regional and 65 tribal populations

Region				Tribe		
No.	Name	Abbr.	$N_R$	No.	Name	$N_T$
1	Arnhem	Arn	419	1	Gunwinggu (Pure)	19
				2	Gunwinggu (Decl)	34
				3	Nakara	67
				4	Ngalakan	49
				5	Warnindilyakwa (Pure)	16
				6	Warnindilyakwa (Decl)	117
				7	Yolngu (Pure)	31
				8	Yolngu (Decl)	76
2	Gulf	Gulf	97	9	Binbinga	38
				10	Ganggalida	21
				11	Lardil	27
3	Desert	Des	2,516	12	Arrernte (Pure)	172
				13	Arrernte (Decl)	1,903
				14	Jinjili	23
				15	Luritja	19
				16	Pitjantjatjara	45
				17	Tjupany	23
				18	Wangkathaa	22
				19	Warumungu (Pure)	57
				20	Warumungu (Decl)	224
				4	Fitzmaurice	Fz
22	Jawoyn (Pure)	43				
23	Jawoyn (Decl)	803				
24	Malak Malak	18				
25	Murrinh-patha (Pure)	21				
26	Murrinh-patha (Decl)	125				
5	Southeast	SE	817	27	Awakabal	36
				28	Biripi	64
				29	Boonwurrung	44
				30	Bundjalung	68
				31	Dainggatti	60
				32	Gunbainggir	81
				33	Kurnai	38
				34	Nganyaywana	45
				35	Ngurraillam	21
				36	Tharawal	128
				37	Wathaurong	34
				38	Woiworung	126
				39	Wonnarua	24
				40	Yuin	37

**Table 1** continued

Region				Tribe		
No.	Name	Abbr.	$N_R$	No.	Name	$N_T$
6	Kimberley	Kim	62	41	Jukun	31
				42	Nyikina	31
7	North	Nth	161	43	Kundjey'mi	91
				44	Tiwi	57
				45	Bindal	20
8	Northeast	NE	74	46	Darumbal	33
				47	Waka Waka	21
				48	Kariyarra	28
9	Northwest	NW	49	49	Yinggarda	21
				50	Nyawaygi	24
10	Rainforest	Rf	48	51	Yidindji	24
				52	Barkindji	54
11	Riverine	Riv	1,084	53	Kamilaroi	212
				54	Kureinji	25
				55	Latje Latje	24
				56	Ngarrindjeri	20
				57	Queensland <sup>a</sup>	28
				58	Wailwan	138
				59	Waveroo	21
				60	Wiradjuri	408
12	Spencer	Sp	59	61	Wongaibon	58
				62	Narangga	20
13	Torres Strait	TS	21	63	Nukunu	39
				64	Combined <sup>b</sup>	21
14	West Cape	WC	33	65	Combined <sup>b</sup>	33

The number of samples in each region ( $N_R$ ) does not always equal the sum of sample numbers of the tribes shown ( $N_T$ ), because the regional datasets contain some tribes with  $N < 15$  and these tribes have not been listed here. The Riverine–Queensland tribe (marked with <sup>a</sup>) is a conglomerate of neighbouring tribes of the northern Riverine region. Because of the limited numbers of samples from the Torres Strait and West Cape regions (marked with <sup>b</sup>) these datasets were included as a collated regional set and analysed at both the regional and tribal level

### Independence testing

Fisher's exact test for allelic association (Guo and Thompson 1992) was used to determine within-locus dependence (indicating departures from Hardy–Weinberg equilibrium) and between-locus dependence (indicating evidence of linkage) with 10,000 permutations. This number is expected to give a 95% confidence interval of  $\pm 0.0043$  for a  $p$  value of approximately 0.05. Independence testing using the genetic data analysis (GDA) software program (Lewis and Zaykin 2001) was undertaken on datasets from each of the fourteen Aboriginal regions and on the combined Aboriginal dataset.

**Table 2** Maximum inter-tribe pairwise distances at the overall level

	Arnhem Gunwinggu Decl	Arnhem Wamindilyakwa Pure	Arnhem Wamindilyakwa Decl	Arnhem Yolngu Pure	Arnhem Yolngu Decl	Fitzmaurice MalakMalak Decl	Fitzmaurice Murrinh patha Pure	Fitzmaurice Murrinh patha Decl	Desert Luritja Decl	Desert Jingili Decl	Desert Tjupany	North Kundjei mi Decl	North Tiwi Decl	Gulf Ganggalida	Gulf Lardi	Riverine Ngarrindjeri	Riverine Barkindji	Riverine Kureinji	Riverine Latje-Latje	Riverine Queensland	Riverine Waveroo	Riverine Wiradjuri	Southeast Boonwurrung	Southeast Bundjalung	Southeast Dainggati	Southeast Gumbainggir	Southeast Nganyaywana	Southeast Ngurrailam	Southeast Tharawal	Southeast Wathaurong	Southeast Woioorong	Southeast Wonnarua	Southeast Yuim	Torres Strait	Spencer Narangga		
Arnhem Guwinggu Pure	6.2	4.4	4.4	3.7																																	
Arnhem Wamindilyakwa Pure						3.6				3.5		3.1		3.2		3.0																				3.1	
Arnhem Wamindilyakwa Decl						3.9	3.2	3.2		3.3		3.9		3.6		3.7	3.2	3.0				3.6	3.1	3.2						3.2		3.4			4.1	3.0	
Arnhem Yolngu Pure	3.8					4.8	4.2	4.3			3.0	3.1	5.6			3.1	3.0				3.8	3.1	3.0	3.4	3.3	3.1	3.1	3.5	3.4	3.1	3.0	3.0	3.6	3.0	3.4		
Arnhem Yolngu Decl						4.2	3.1	3.1					4.6									3.0															
Arnhem Ngalakan Decl						3.3							3.0																								
Fitzmaurice MalakMalak Decl														3.5	3.3		3.0																			3.8	
North Tiwi Decl									4.1	3.1				3.5	4.1		4.3																			3.7	4.3

The 68 instances of pairwise distance values where  $\theta \geq 0.03$  (or 3%) are shown

*Genetic distances and phylogenetic trees*

A series of neighbour-joining (NJ) trees was developed by using the GDA software package based on  $F_{ST}$  (or  $\theta$ ) estimates generated by the method of Weir and Cockerham (1984). A tree was initially constructed from  $\theta$  estimates for the fourteen regional sets (minus urban groups) ( $n_{pops} = 14$ ,  $N = 6,513$ ). Trees were then constructed for the dataset partitioned at the tribal level. The first tree contained only the remote tribal populations, ( $n_{pops} = 65$ ,  $N = 6,312$ ). The second tree contained the full set of tribal populations plus the Aboriginal urban populations ( $n_{pops} = 73$ ,  $N = 8,868$ ). Principal-components analysis (PCA) plots were generated from the full matrices of genetic distances from the phylogenetic trees using Minitab statistical software. To avoid repetition only the PCA plot of the regional differentiation and the tree of full set of tribal populations plus the Aboriginal urban populations are presented here.

To evaluate the extent of genetic structuring among different population clusters, hierarchical analysis of variance (AMOVA) was undertaken using the package Arlequin ver. 2.000 (Schneider et al. 2000). This approach estimates the percentage of the genetic variation that is explained:

- 1 among groups of populations defined a priori;
- 2 between populations in the same group; and
- 3 within the population.

In this study population clustering was imposed on the basis of inter-regional genetic distances and assessed using AMOVA.

*Analysis of heterozygosity*

The expected heterozygosity ( $H_e$ ) calculated by GDA is the unbiased estimator obtained by multiplying the sample expected heterozygosity ( $1 - \sum_u p_u^2$ ) by the factor  $(2n)/(2n - 1)$  (Nei's variance calculation). By estimating this parameter the relative effect of isolation and genetic drift as causal factors for observed genetic differentiation in outlier tribal groups are examined.

**Results**

The use of microsatellites for inferring deep population histories and structure has been questioned because of their relatively faster rate of evolution than most other types of DNA, polymorphisms such as SNPs, and indels. Several studies have shown, however, that when a reasonably large number of microsatellites are scored the inferences from those samples, even at the worldwide level, are sound (Bowcock et al. 1994; Jorde et al. 2000; Rosenberg et al. 2002; Ayub et al. 2003; Zhivotovsky et al. 2003; Shepard et al. 2005; Zabala Fernandez et al. 2005). For examining structure and variation within a single population or ethnic group they may be the marker of choice (Brinkmann et al. 1998; Destro-Bisol et al. 2000; Perez-Miranda et al. 2005; Li et al. 2006). This study of structure and variation among Australian Aboriginals using fifteen autosomal microsatellites fulfils these requirements.

## Independence testing

The results of independence testing demonstrated that many of the  $p$  values were significant at  $<0.05$ , indicating departure from Hardy–Weinberg and linkage equilibrium for some samples. Graphical representation of the distributions of these independence data enables visual comparison across regions (Buckleton 2005). If the hypotheses of Hardy–Weinberg and linkage equilibrium were true the  $p$  values should be distributed uniformly between 0 and 1;  $p \approx U[0,1]$ . As an example, the  $x = y$  trend-line in the  $p$ – $p$  plots (Fig. 2) represents equilibrium and the 95% confidence interval generated by simulation samples from a uniform distribution is also displayed as the region within the two curved lines.

The  $p$ – $p$  plots for the combined Aboriginal dataset (Fig. 2a) show strong evidence of departure from independence because values plot well away from the linear trendline and mostly outside the 95% confidence limit envelope. The  $p$ – $p$  plots for the regional Aboriginal datasets (Figs. 2b–2o) show evidence of departure from independence in many instances, especially in Arnhem, Desert, Fitzmaurice, and Riverine (all  $N > 400$ ). The limited power of Fisher's Exact test to find departures from independence in smaller datasets (Curran et al. 2003) means that some of these regions are uninformative (Gulf, Kimberley, Northeast, Northwest, Rainforest, Spencer, West Cape, and Torres Strait Islands; all  $N < 100$ ), and the lack of evidence of departure should not be taken to imply that none exists. For the relatively large Southeast ( $N = 817$ ) sample values are close to equilibrium. The evidence of disequilibria at the regional level implies that substructure exists at the level of the tribe or community.

## Genetic distance

The dataset was initially subdivided into the fourteen regional populations (Table 2). The NJ tree (not shown) for these regions shows a grouping that reflects geography and overall population density (or degree of urbanisation). This pattern is more clearly apparent in a PC plot of the full distance matrix (Fig. 3). The data generate at least three major clusters which we labelled “Combined”, “South-Eastern”, and “Northern”. Overall, geographically neighbouring regions tend to cluster in the same space of the plot. The Combined cluster contains nine regions: Desert, Fitzmaurice, Rainforest, Northeast, Northwest, Gulf, Spencer, Kimberley and West Cape. It is possible to also impose a cluster comprising the Gulf and West Cape regions, because they lie on the fringe of the main cluster. This would maintain the trend of geographically close regions clustering (Fig. 1). South-Eastern contains those regions (South-east and Riverine) comprising the most populous areas of

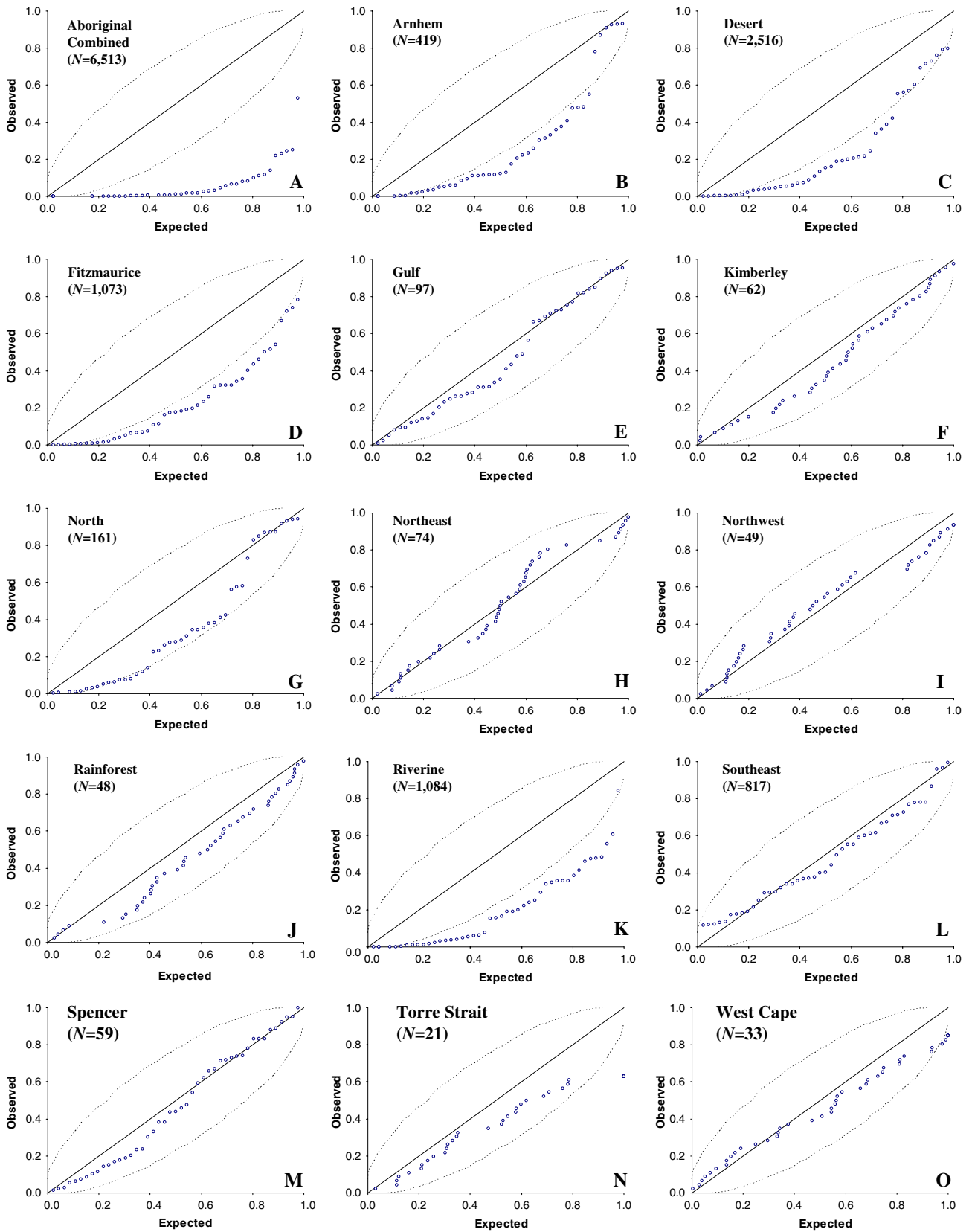
Australia. Northern comprises those regions having the most “remote” tribes in the NT (Arnhem and North).

Figure 4 shows the NJ tree for the full set of 73 Aboriginal samples including the 65 tribal populations and the eight urban populations. The tree follows the pattern of clustering at the regional level (Fig. 3). Tribes of the same region tend to have small observed genetic distances from each other, and this is also true of tribes from different, but geographically close, regions. An important observation is that where Declared and Pure datasets were available for a tribe from the NT these samples plot very close together. This finding supports the suggestion that any subpopulation structure revealed by this analysis is not being compromised by any differences among samples possibly induced by these classifications. It is clear from Fig. 4 there is little evidence many tribes comprise distinct sub-populations and cluster close to the midpoint of the tree. This is particularly evident for tribes in the South-East, whether remote or urban in origin; it is also true for tribes located in the Riverine region, (except for the Kureinji and Riverine Queensland, for both of which differentiation is slightly greater than for the other tribes) and also for the two tribes in the Spencer region. All Urban groups except Darwin show minimal evidence of sub-populations, plotting near the midpoint of the tree.

The tribes of the Desert region all lie in a distinct grouping at the edge of the tree, with most (Lurtija, Pitjantjatjara, Tjupany, Arrernte, and Wangkatha) lying on one branch and the Waramungu and Jingli on a neighbouring branch. Both these branches are longer than those seen for Riverine, Southeast, Spencer, and Urban populations, indicating that the Desert peoples are more differentiated.

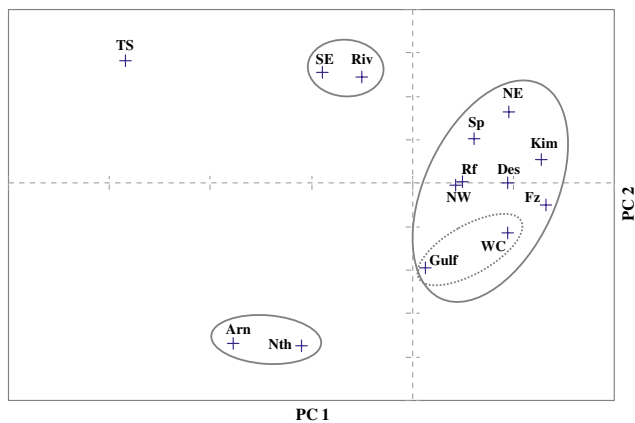
Population differentiation is greatest for Northern Australia. By far, the most differentiated tribes are those of Arnhem, Fitzmaurice, and also the Tiwi. The longest branches of all are for the Warnindilyakwa and Yolngu, both of East Arnhem Land; these two tribes lie on a different branch to the Guwinggu and Nakara of West Arnhem Land, however (Fig. 4). The tribes of the Fitzmaurice region are located on a number of branches but the most differentiated, the Malak-Malak, is closest to the Tiwi. The tribes of the Gulf region lie on the same branch, but vary substantially from other tribes and are closest to the North and (Western) Arnhem Land tribes.

To examine inter-tribal differentiation at the Australia-wide level the pairwise distances for all 2,080 pairs of the 65 tribal populations were evaluated. The largest distances ( $\theta F_{ST}$  values  $>0.03$ ) are displayed in Table 2. Interestingly, all 68 instances of  $F_{ST} > 0.03$  involve only six tribal populations (Fig. 5); either one of four Arnhem tribes (Guwinggu, Yolngu, Warnindilyakwa, Ngalakan), or North Tiwi, or Fitzmaurice Malak Malak. The most differentiated tribes are the Yolngu, the Warnindilyakwa, and the Tiwi.



**Fig. 2** *p-p* plots for the combined and fourteen regional Aboriginal datasets





**Fig. 3** PC plot of the full genetic distance matrix for the fourteen Aboriginal regional samples. The cluster of nine regions in the right hand quadrants we have referred to as the “combined” cluster. The cluster in the lower left quadrant containing the Arnhem (*Arn*) and North (*Nth*) regions we have referred to as the “Northern” cluster. The cluster in the upper left quadrant containing the Southeast (*SE*) and Riverine (*Riv*) regions we have referred to as the “South Eastern” cluster

The largest pairwise difference ( $F_{ST} = 0.062$ ) occurs between two tribes of the Arnhem region, the Guwinggu (from West Arnhem Land) and the Warnindilyakwa of Groote Eylandt, an island off the East coast of Arnhem Land.

In Fig. 6 the distance matrix of all 2,080 pairwise combinations of the 65 populations is represented in a series of PC plots. Within the South-East and Riverine regions the tribal populations cluster and are minimally different. For the Arnhem, North, Fitzmaurice, and Gulf regions, however, there are large differences among the tribes within each region. For the other regions the pattern of tribal variation is between these extremes.

A genetic classification is supported when AMOVA analysis reveals maximum genetic variance among groups ( $F_{CT}$ ) and minimum variation within groups ( $F_{SC}$ ). The AMOVA results are presented in Table 3. The overall  $F_{ST}$  value of 0.01295 when data are analysed at the regional level approximates values reported elsewhere from studies on similar microsatellite data (Buckleton et al. 2005). Most variation exists at the individual level with the remainder evenly spread among groups and among populations within groups ( $F_{ST} = 0.01295$ ,  $F_{SC} = 0.00659$ ,  $F_{CT} = 0.00640$ ). The among-group variation increases ( $F_{CT} = 0.00755$ ) when the fourteen regions are grouped into the three clusters described in Fig. 3. The within-group variation also increased ( $F_{SC} = 0.00746$ ), however, indicating there is more heterogeneity within one or more of the imposed clusters. Although the North and Arnhem regions cluster in Fig. 3 analyses at the tribal level (Figs. 4 and 6; Table 2) reveal substantial heterogeneity within these regions. When the AMOVA performed again with four clusters (South Eastern, Combined, Arnhem, and North) the among-group

variation increased ( $F_{CT} = 0.01002$ ) and the among-population variation decreased ( $F_{SC} = 0.00601$ ).

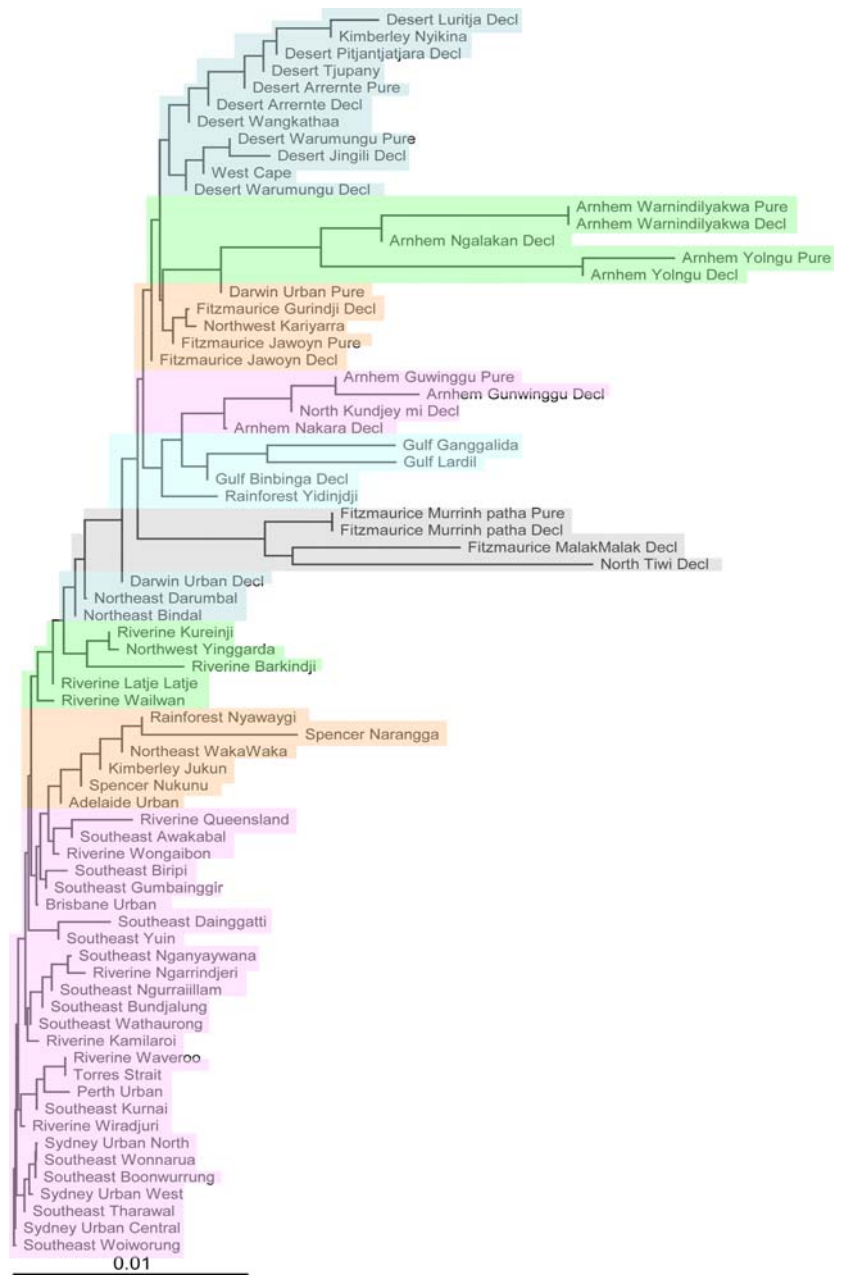
When each of the three clusters is compared against the rest of the data a portion of the variation understandably shifted to among populations within groups (overall increase in  $F_{SC}$ ). On the whole, however, these  $F$  values are uninformative, because among-population variation is consistently high ( $F_{SC} \approx 0.01$ ) for the combined clusters (labelled as “rest”). Again, however, they imply that the Arnhem and North regions are most differentiated from the remaining data (maximised  $F_{CT}$  values, 0.00891 and 0.01168, respectively). By contrast, the pairwise values are more informative. Generally among populations within group variation is reduced, suggesting the clusters imposed are reasonable. In each pairwise measure involving the Arnhem cluster the  $F_{SC}$  values increase marginally, however. This again shows the effect of tribal differentiation within the Arnhem region. Minimum overall diversity is seen in the pairwise comparison of the South Eastern and Combined clusters ( $F_{ST} = 0.01266$ ,  $F_{SC} = 0.00425$ ,  $F_{CT} = 0.00844$ ) whereas the most differentiated clusters are South Eastern and Arnhem ( $F_{CT} = 0.02037$ ) and South Eastern and North ( $F_{CT} = 0.01565$ ). Although there is evidence of substantial differentiation between the North and Arnhem regions ( $F_{CT} = 0.01495$ ), more noticeable is the increase in among-populations within-group variation ( $F_{SC} = 0.02526$ ). Once again this result reflects the extent of differentiation among the tribes of these two regions.

### Heterozygosity

Heterozygosity analysis was performed to assess the relative effect of isolation and genetic drift as causal factors for any observed genetic diversity (Table 4). Because  $H_o = (1 - f)H_e$  is true for the parametric values the sample value of  $(H_e - H_o)/H_e$  is expected to estimate  $f$ . A density plot of the distribution of  $f$  values shows they seem to spread symmetrically about zero (Fig. 7). The actual mean is 0.00061. There seems to be little evidence from the overall data of excess homozygosity in the indigenous tribal populations.

Although the distribution of observed heterozygosity (and the parameter  $f$ ) does not reveal any excess homozygosity within the overall indigenous population it is necessary to assess whether there is evidence of excess in those tribal populations significantly differentiated from the remainder. Interestingly, when  $H_o$  and  $H_e$  values are considered at the tribal level there are four regions, all in northern Australia, where greater than 50% of the tribes from the region have an excess of homozygotes. These are Arnhem (seven of eight tribes, 87.5%), North (two of two tribes, 100%), Gulf (three of three tribes, 100%) and Northwest (two of two tribes, 100%).

**Fig. 4** Neighbour-joining tree based on pairwise distances for the full set of Aboriginal samples. Major clusters are coloured to assist interpretation

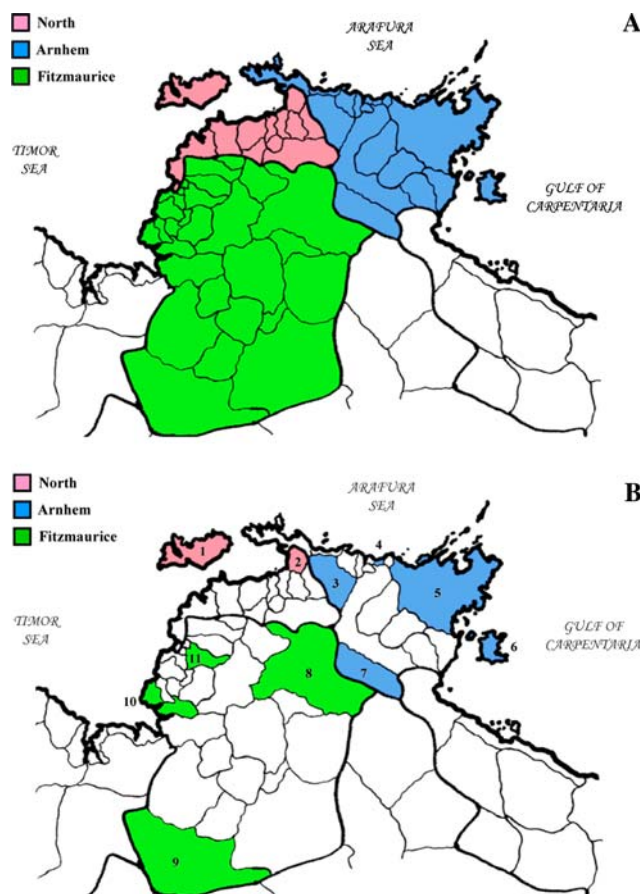


**Discussion**

This study investigated the extent of structuring among Australian indigenous populations using available autosomal microsatellite datasets held by Forensic Laboratories in Australia. Although the sample size in this study is extremely large ( $N = 8,868$ ) it lacks adequate representation from the States of WA and QLD, each of which has a large number of Aboriginal communities. In particular this study is deficient in groups from Cape York Peninsula in QLD and the Kimberleys of WA. This deficiency makes any findings we present less comprehensive for the total population than desired. Further, the suitability of microsatellites as markers

for measuring genetic structure at a continent level has been questioned, although there is now much evidence these markers are appropriate for use in the current context.

The key finding of this study is the high level of genetic structuring evident among indigenous Australians, and that the most structured groups are those that inhabit northern Australia (Fig. 5). The reason(s) for these differences is/are not easy to identify and, further, are likely to vary (and have varied) in intensity from region to region. The causes could include, but are not restricted to, extremes of isolation and drift, admixture with exotic populations in the past, or recent admixture with Caucasians or other immigrant groups.



**Fig. 5** Expanded map of Aboriginal regions of the Central North of Australia, after Horton (Horton (creator) 1996). **A.** The three regions Arnhem, North, and Fitzmaurice. The *heavy borders* are the boundaries of the regions and the *light borders* are the boundaries of tribal territories within these regions. Genetic distance estimates have shown that the tribes of these three regions are most differentiated from other Aboriginal tribes and from each other. **B.** The tribal populations from these regions that were available for this study. They are 1. North Tiwi, 2. North Kundjey'mi, 3. Arnhem Guwinggu, 4. Arnhem Nakara, 5. Arnhem Yolngu, 6. Arnhem Warnindilyakwa, 7. Arnhem Ngalakan, 8. Fitzmaurice Jawoyn, 9. Fitzmaurice Gurindji, 10. Fitzmaurice Murrinh'patha, and 11. Fitzmaurice Malak Malak

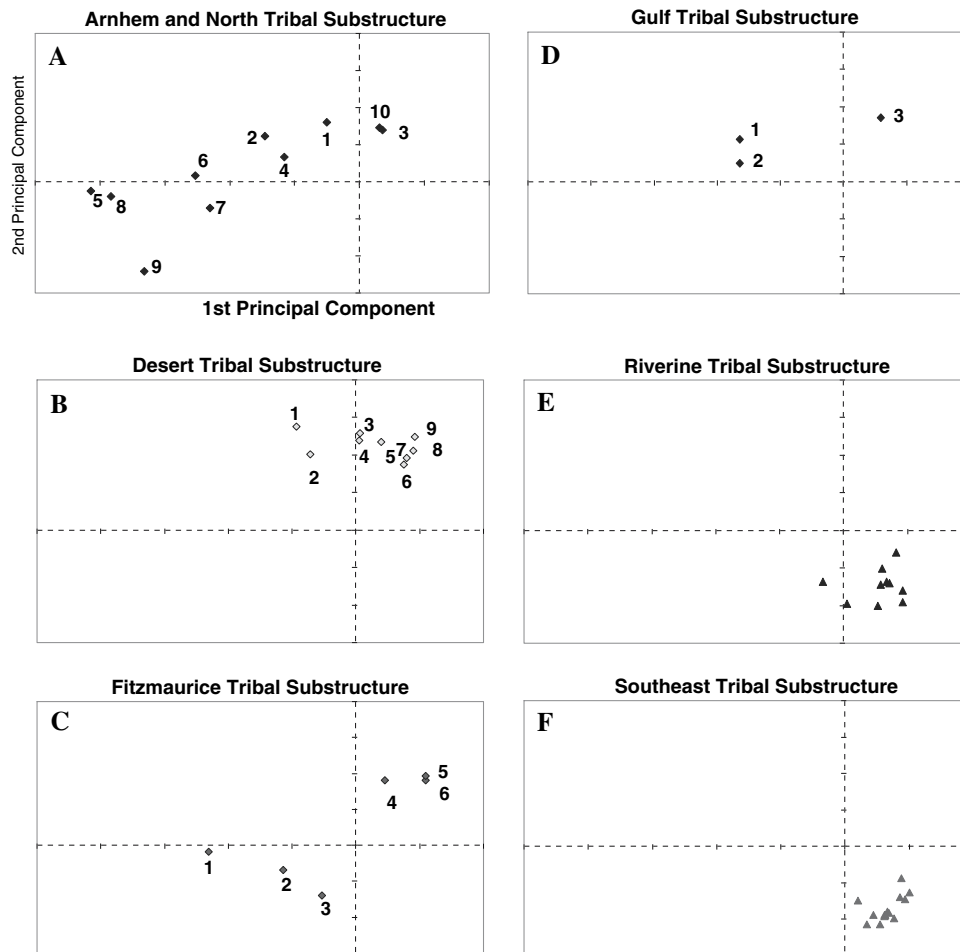
The minimum structuring seen in urban groups (except Darwin) is, perhaps, best explained by the fact that these samples most likely comprise individuals from a wide variety of tribal backgrounds and also are those with the highest levels of admixture. Pooling of groups and/or admixture with exotics reduces the variation detected by our method of sampling. In this study large datasets were available for the Southeast and Riverine regions yet minimum differentiation was observed between tribes of these regions. This limited variability is also likely to be the result of mixing of individuals from different tribes, i.e. different gene pools, and/or admixture with Europeans, than an indication of homogeneity.

Observations on body measurements (Macho and Freedman 1987) and skin colour and dermatoglyphic features (Parsons and White 1976) reveal substantial variation among Australian Aborigines, especially among those of the North (e.g. Arnhem Land) and those of Central Australia (White 1997). Such a clear North–South cline in variation suggests that either there was a period of separation between peoples of Arnhem Land and those of Central Australia long enough for genetic differences to occur, presumably as a result of genetic adaptation to the different ecological zones, or that these two groups have different genetic origins (ancestors). Aboriginal people were known to be in Central Australia at least 20,000 years ago (Jones 1987; Smith 1989) and their presence in Arnhem Land extends back to at least 50,000 years ago (Roberts et al. 1990). White (1997) found that tribes (language units) are also highly differentiated in dermatoglyphics. This is especially true for coastal groups and is consistent with their high endogamy rates. Also, it may have been the same resource scarcity that made inland groups actively seek marriage partners from other groups to maximise their survival and economic opportunities. In addition, resource abundance could lead to increased territorialism and hence increased drift.

#### Discussion of Arnhem land variation

Approximately one-eighth of the Australian continent, in a band from the SW corner of the Gulf of Carpentaria through Arnhem Land to the Kimberley region (in the Northern part of Western Australia), is the most linguistically diverse area in Australia. Within this belt, anthropological and linguistic data clearly show that Arnhem Land can be divided into two main regions West Arnhem Land (or WAL) and East Arnhem Land (or EAL). The languages of WAL are all of the so-called prefixing type. The Warnindilyakwa of Groote Eylandt (and island in the Gulf of Carpentaria) in EAL speak a prefixing language but one that is substantially different from the mainland prefixing languages. The prefixing languages are strikingly different typologically from the suffixing languages found over nearly seven-eighths of the continent, including EAL, and which belong to a single phyletic family, Pama-Nyungan. Differences between WAL and EAL are also evident in social organisation, because the moiety system in WAL is basically matrilineal whereas in EAL it is patrilineal (Berndt and Berndt 1968). In addition, circumcision, probably the most important of the male initiation rites (Berndt and Berndt 1968) is not practised over most of WAL but it is performed in EAL.

Total ridge count (TRC) analysis has been undertaken for 3,260 full descent Aboriginal Australians from 36 tribes in the Northern Territory (mostly from AL) (White 1997).



**Fig. 6** PCA plots of inter-tribal genetic distance at the regional level. The plots show varied levels of heterogeneity amongst tribal populations within a region. **A.** Tribes of the Arnhem and North regions: 1. Arnhem Guwinggu (pure), 2. Arnhem Guwinggu (decl), 3. Arnhem Nakara (decl), 4. Arnhem Ngalakan (decl), 5. Arnhem Yolngu (pure), 6. Arnhem Yolngu (decl), 7. Arnhem Warnindilyakwa (pure), 8. Arnhem Warnindilyakwa (decl), 9. North Tiwi (decl), 10. North Kudjey'mi (decl). **B.** Tribes of the desert region: 1. Luritja (decl), 2. Jingili (decl), 3. Tjupany (decl), 4. Pintjatjantjara (decl), 5. Arrernte (pure), 6. Waramungu (decl), 7. Arrernte (decl), 8.

Waramungu (pure), 9. Wangkathaa. **C.** Tribes of the Fitzmaurice region: 1. Malak Malak (decl), 2. Murrinh'patha (decl), 3. Murrinh'patha (pure), 4. Gurindji (decl), 5. Jawoyn (decl), 6. Jawoyn (pure). **D.** Tribes of the gulf region: 1. Ganggailida, 2. Lardil, 3. Binbinga (decl). **E, F** Tribes of the Riverine and Southeast regions, respectively. Clearly the tribes in plots **A** to **D** have varying levels of heterogeneity. The tribes of the Riverine and Southeast regions (plots **E** and **F**) are not numbered because they plotted in such close proximity to each other

The tribes White surveyed came from two markedly different ecological zones, Central Australia (CA) and Arnhem Land (AL). The dermatoglyphic results showed that tribes (language units) are highly differentiated in mean TRC. Fingerprint pattern intensity index (PII) analysis has also showed that the WAL tribes (Tiwi and Guwinggu) cluster together and are quite separate from EAL groups (Yolngu and Warnindilyakwa), which also cluster. Combining PII and TRC showed that the AL tribes form two distinct clusters—WAL and EAL—and that differentiation is most pronounced among coastal groups especially, consistent with their high endogamy.

In this study the Tiwi sample is the most differentiated of all tribes in the NJ tree (Fig. 4); this is consistent with findings from investigations of body shape (Macho and Freedman 1987), dermatoglyphics (Parsons and White 1976), and other non-DNA genetic data (Balakrishnan et al. 1975). The Tiwi are one of the most isolated groups, living on Melville and Bathurst Islands (in the Arafura Sea), and this is reflected in their linguistic isolation (White 1997). Another island group, the Warnindilyakawa of Groote Eylandt, is also highly differentiated from other Aboriginal groups. White (1997) uses evidence of particular alleles having either been lost (RH allele RH\*cDe in

**Table 3** AMOVA results revealing the structure of microsatellite variation among Australian Aboriginal regions

Grouping system	Within populations		Among populations within groups		Among groups	
	%	$F_{ST}$	%	$F_{SC}$	%	$F_{CT}$
Regional level	98.705	0.01295	0.654	0.00659	0.640	0.00640
Three regions (combined, South Eastern, and Northern)	98.504	0.01495	0.740	0.00746	0.754	0.00755
Four regions (combined, South Eastern, North, and Arnhem)	98.404	0.01596	0.595	0.00601	1.002	0.01002
Combined versus rest	98.577	0.01423	0.981	0.00986	0.442	0.00422
South Eastern versus rest	98.387	0.01613	0.848	0.00854	0.765	0.00765
North versus rest	97.970	0.02030	1.139	0.01149	0.891	0.00891 <sup>a</sup>
Arnhem versus rest	97.759	0.02241	1.073	0.01086	1.168	0.01168
South Eastern versus Combined	98.734	0.01266	0.422	0.00425	0.844	0.00844
South Eastern versus Arnhem	97.328	0.02672	0.636	0.00649	2.037	0.02037
South Eastern versus North	98.000	0.02000	0.435	0.00441	1.565	0.01565
Combined versus Arnhem	98.176	0.01824	0.702	0.00710	1.121	0.01121
Combined versus North	98.344	0.01656	0.555	0.00562	1.101	0.01101
Arnhem versus North	96.017	0.03983	2.488	0.02526	1.495	0.01495 <sup>b</sup>

\* All  $p$  values are  $p < 0.00001$ , except for <sup>a</sup> $p = 0.00391$  and <sup>b</sup> $p = 0.00293$

Warnindilyakwa) or reached fixation (Duffy\*A in the Tiwi) to suggest that extreme allele frequencies in these island populations are indicative of the action of random drift. These island populations are also the most polygynous of the Arnhem Land tribes, a process that would only lead to further reduction in allele diversity. Of particular relevance to this study is that White (1997) found high sub-population heterogeneity in EAL, with dialect clusters within the Yolngu indicative of substantial diversity. This diversity was greatest among the coastal communities which occupy the most favourable habitats. White concluded that this association between linguistic diversity and genetic diversity was most probably a result of language being a barrier to gene flow. Another factor found to be strongly associated with genetic diversity was the drainage basin. Gene flow would be expected to occur more frequently between tribes along drainage systems than between people living in a different system separated by a watershed. White's (1997) research in AL suggests that the genetic diversity within those groups that the current study revealed is only a part of the total because of our sampling method, and that a much finer-scaled sampling approach will be required to identify the true level of diversity in this region.

The next most differentiated groups, North Kundjey'mi, Arnhem Guwinggu, and Arnhem Nakara, could all be regarded as WAL tribes. Combining the differentiation observed in this study with the anthropological evidence suggests grouping these tribes with WAL groups. The Ngalakan tribe is known to have prefixing language similar to other WAL tribes but lies on the South-Western edge of the Yolngu territory and it is possible there has been gene flow between these groups. Our results show the Ngalakan

tribe on the same branch as the EAL tribes (Warnindilyakwa and Yolngu), although the length of the branches suggests much differentiation between these groups.

It is clear that in AL there is clinal variation in microsatellite alleles from Melville Islands (Tiwi) off WAL to Groote Eylandt (Warnindilyakwa) in the Gulf of Carpentaria. It seems, in conclusion, that the results of this study, with blood group and protein allele frequency data and dermatoglyphic measures, all suggest that sociocultural and, particularly, linguistic, differences are important in regulating gene flow between tribes (and, for the Yolngu, even within some tribes) and thus explain the structure observed in this region. Occasional fluctuation in population sizes may also contribute to variation as a product of founder effects. For example, famine and other ecological disasters may reduce tribal sizes periodically and cause tribal fragmentation. Inter-clan fighting that occurred among the Yolngu of EAL also had a significant effect on the number of males in the group (Warner 1964). Another major factor would be diseases introduced since European contact, and possibly Macassan contact even earlier. These factors, and the resulting reduction in effective population size of a tribe,  $N_e$ , is further exacerbated by polygynous marriage systems, which in tribes such as the Tiwi, Warnindilyakwa, and Yolngu are at extreme levels.

#### Desert tribes

In Figs. 3 and 4 Central Australian (CA) tribes can be genetically distinguished as a group from Arnhem Land (AL) tribes. This finding matches conclusions from skin colour, TRC, and body-shape research (Parsons and White

**Table 4** Observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity for the 65 Aboriginal tribal populations

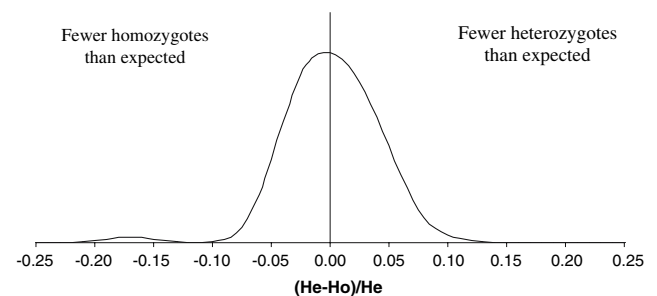
Population	$n$	$H_e$	$H_o$
Arnhem Guwinggu (pure)	17.0	0.754	0.724
Arnhem Gunwinggu (decl)	23.9	0.756	0.785
Arnhem Warnindilyakwa (pure)	11.0	0.733	0.667
Arnhem Warnindilyakwa (decl)	74.7	0.743	0.698
Arnhem Yolngu (pure)	22.1	0.753	0.748
Arnhem Yolngu (decl)	47.9	0.740	0.728
Arnhem Nakara (decl)	40.3	0.690	0.664
Arnhem Ngalakan (decl)	33.0	0.745	0.737
Desert Arrernte (pure)	116.7	0.784	0.820
Desert Arrernte (decl)	1221.5	0.769	0.752
Desert Warumungu (pure)	50.8	0.788	0.767
Desert Warumungu (decl)	142.6	0.788	0.793
Desert Jingili (decl)	23.0	0.781	0.807
Desert Luritja (decl)	11.7	0.752	0.750
Desert Pitjantjatjara (decl)	35.9	0.758	0.733
Desert Tjupany	23.0	0.792	0.797
Desert Wangkathaa	22.0	0.795	0.808
Fitzmaurice Jawoyn (pure)	34.0	0.794	0.811
Fitzmaurice Jawoyn (decl)	514.1	0.786	0.787
Fitzmaurice Murrinh'patha (pure)	19.7	0.759	0.790
Fitzmaurice Murrinh'patha (decl)	86.4	0.746	0.763
Fitzmaurice Gurindji (decl)	18.6	0.769	0.903
Fitzmaurice Malak Malak (decl)	18.0	0.754	0.759
North Kundjey'mi (decl)	55.1	0.821	0.784
North Tiwi (decl)	34.7	0.730	0.703
Gulf Binbinga (decl)	24.4	0.795	0.767
Gulf Ganggalida	21.0	0.757	0.735
Gulf Lardil	27.0	0.769	0.757
Northeast Bindal	20.0	0.811	0.833
Northeast Darumbal	33.0	0.806	0.822
Northeast Waka Waka	20.8	0.799	0.786
Rainforest Nyawaygi	23.9	0.819	0.851
Rainforest Yidinjdji	24.0	0.814	0.829
Riverine Barkindji	54.0	0.816	0.827
Riverine Kamilaroi	212.0	0.822	0.818
Riverine Kureinji	25.0	0.821	0.849
Riverine Latje	23.6	0.815	0.812
Riverine Queensland	27.9	0.795	0.793
Riverine Wailwan	137.9	0.815	0.815
Riverine Waveroo	21.0	0.834	0.857
Riverine Wiradjuri	408.0	0.818	0.806
Riverine Wongaibon	58.0	0.815	0.780
Riverine Ngarrindjeri	20.0	0.814	0.844
Southeast Awakabal	36.0	0.824	0.827
Southeast Biripi	64.0	0.810	0.828
Southeast Boonwurrung	43.9	0.809	0.810
Southeast Bundjalung	68.0	0.814	0.797
Southeast Dainggatti	60.0	0.809	0.833

**Table 4** continued

Population	$n$	$H_e$	$H_o$
Southeast Gumbainggir	81.0	0.819	0.819
Southeast Kurnai	37.2	0.814	0.803
Southeast Nganyaywana	45.0	0.810	0.778
Southeast Ngurraillam	21.0	0.809	0.841
Southeast Tharawal	128.0	0.818	0.830
Southeast Wathaurong	33.9	0.818	0.797
Southeast Woiworung	126.0	0.812	0.825
Southeast Wonnarua	24.0	0.812	0.796
Southeast Yuin	37.0	0.821	0.856
Spencer Narangga	20.0	0.797	0.828
Spencer Nukunu	39.0	0.817	0.835
Kimberley Jukun	31.0	0.818	0.824
Kimberley Nyikina	31.0	0.785	0.778
Northwest Kariyarra	28.0	0.795	0.754
Northwest Yinggarda	21.0	0.809	0.762
Torres Strait	20.8	0.813	0.791
West Cape	32.8	0.794	0.757

1973). The seven Central Desert tribes analysed in this study are not markedly differentiated from each other, plotting as two clusters in a distinct portion of the tree (Fig. 4). The only observed differentiation of Desert tribes is that two of the tribes from the North of the Desert territories (Jingili and Warumungu) plot on their own branch whereas those of the Central Desert (Arrernte, Luritja and Pitjantjatjara) and Western Desert (Wangkathaa and Tjupany) cluster on an adjacent branch.

Although blood group and dermatoglyphic data support this finding, the Arrernte may be regarded as different from other desert groups because of Birdsell's (1950) hypothesis that Arrernte came from a relatively recent migration southward from the North of Australia. The isolation of the Arrernte from the Desert tribes to their west (Pintubi and Pitjantjatjara) is shown in blood group and serum protein allele frequencies and, to a lesser extent, dermatoglyphics



**Fig. 7** Density plot for the function  $(H_e - H_o)/H_e$  for the Aboriginal tribal subpopulations

(Mader 1965; Nicholls et al. 1965; Robson and Parsons 1967). The Arrernte belong to the Pama-Nyungan linguistic family but within this family the Arrernte belong to a linguistic group different from the Pitjantjatjara and Pintubi (Walsh 1993).

There is substantial variation within Aboriginal Australia but few data are available for Aborigines from the South East and South West of Australia or from much of Central and South Queensland. These regions are the more populous areas of contemporary Australia, a factor which may bring additional drivers that could either exacerbate or diminish observed differentiation, for example more pronounced admixture with exotic populations, greater tribal fragmentation and departure from traditional marriage practices.

## Conclusion

The principal findings of this study are that the most differentiated tribal groups are located in three regions, West Arnhem Land, East Arnhem Land and Tiwi, all of which share borders with one another in the Central North of the continent. These tribal groups are most differentiated from other Aboriginal Australian tribes, especially those of the Central Desert regions, and also show marked heterogeneity from one another. These genetic findings are supportive of observations on body measurements (Macho and Freedman 1987) skin colour, and dermatoglyphic features (Parsons and White 1976), which also show vary substantially among tribes of the North (e.g. Arnhem Land) and Central Australian regions (White 1997) and, more specifically, between the Tiwi and WAL and EAL tribes.

Some findings could be enhanced by studies of lineage markers, notably Y-chromosome loci, in particular to assess the impact of polygyny among the Northern tribes and the extent of inter-tribal admixture and admixture with Caucasian or other non-indigenous populations. Previous studies have revealed much diversity among tribes of the Cape York Peninsula (Cavalli-Sforza et al. 1994); in this study we did not have a representative sample from tribes of this region and were unable to assess differentiation in this part of the continent. Notwithstanding these limitations this study has provided the most comprehensive survey of the population genetics of Aboriginal Australia. Whereas previous studies have been limited to a single sample from one tribe or community and have focussed on either the global dispersal of modern humans or Aboriginal affinities with neighbouring ethnic groups this study describes the population genetic features of widely dispersed tribal groups of Australia.

**Acknowledgments** We wish to acknowledge the support of the contributing laboratories and their Directors. They are the Forensic

Science South Australia; Forensic & Technical, Australian Federal Police; Victoria Police Forensic Services Department; Forensic Services, Queensland Police; Forensic Services, Queensland Health Scientific Services; Forensic Services, Tasmania Police; Forensic Services Group, NSW Police; Forensic Services, South Australia Police; Forensic Division, Western Australia Police; Forensic Services, Northern Territory Police; Division of Analytical Laboratories, Department of Health, NSW; Forensic Business Group, ESR, New Zealand; Forensic Science Service Tasmania; Forensic Biology, Pathwest Laboratory Medicine, WA. In addition, the research assistance of Heidi Drummond, Karah Bujtor, Jennifer Morris and Julianne Stevens in partitioning the datasets is gratefully acknowledged. This project was funded by the National Institute of Forensic Sciences, the University of Technology, Sydney, and ESR, New Zealand. We acknowledge the coordination role of Dr Tony Raymond, NIFS and Wayne Chisnall, ESR. We also thank Ms Dinusha Fernando and one anonymous reviewer whose comments greatly improved this manuscript.

## References

- ABS (1998) Experimental estimates of the Aboriginal and Torres Strait Islander population, 30 June 1991–30 June 1996. Australian Bureau of Statistics (Report No. 3230.0). Available from <http://www.abs.gov.au/auststats/abs@.nsf/lookupMF/E1B3119470F95E02CA2568A900139384> (accessed 26 June 2007)
- ABS (2004) Year Book Australia 2004. Australian Bureau of Statistics, Canberra (Publication No. 86; ABS Catalogue No. 1301.0)
- Ayub Q, Mansoor A, Ismail M, Khaliq S, Mohyuddin A, Hameed A, Mazhar K, Rehman S, Siddiqi S, Papaioannou M, Piazza A, Cavalli-Sforza LL (2003) Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites. *Am J Phys Anthropol* 122:259–268
- Balakrishnan V, Sanghvi LD, Kirk RL (1975) Genetic diversity among Aborigines. AIATSIS Research and Regional Studies No. 3. AIATSIS, Canberra
- Berndt RM, Berndt CH (1968) The World of the first Australians. Ure Smith, Sydney
- Birdsell JB (1950) Some implications of the genetical concept of race in terms of spatial analysis. *Cold Spring Harb Symp Quant Biol* 15:259–314
- Birdsell JB (1993) Microevolutionary patterns in Aboriginal Australia: a gradient analysis of clines. Oxford University Press, New York
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–457
- Brinkmann B, Junge A, Meyer E, Wiegand P (1998) Population genetic diversity in relation to microsatellite heterogeneity. *Hum Mutat* 11:135–144
- Buckleton JS (2005) Validating databases. In: Buckleton JS, Triggs CM, Walsh SJ (eds) Forensic DNA evidence interpretation. CRC, Boca Raton, pp 149–196
- Buckleton JS, Curran JM, Walsh SJ (2005) *R v Bropho*: the need for careful interpretation of independence testing in courtroom decision-making. *Aust Law J* 79:709–721
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, New Jersey
- Curran JM, Buckleton JS, Triggs CM (2003) What is the magnitude of the subpopulation effect? *Forensic Sci Int* 135:1–8

- Destro-Bisol G, Boschi I, Caglia A, Tofanelli S, Pascali V, Paoli G, Spendini G (2000) Microsatellite variation in Central Africa: an analysis of intrapopulation and interpopulation genetic diversity. *Am J Phys Anthropol* 112:319–337
- Flood J (1983) *Archaeology of the dreamtime*. Collins Publishers, Australia
- 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. *Molecular Biol Evol* 15:1108–1114
- Friedlaender JS, Gentz F, Green K, Merriwether DA (2002) A cautionary tale on ancient migration detection: mitochondrial DNA variation in Santa Cruz Islands, Solomon Islands. *Hum Biol* 74:453–471
- Guo SW, Thompson EA (1992) Performing the exact tests of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372
- Horton (creator) D (1996) *Aboriginal Australia map*. AIATSIS and Auslig, Canberra
- Huoponen K, Schurr TG, Chen Y-S, Wallace DC (2001) Mitochondrial DNA variation in an Aboriginal Australian population: Evidence for genetic isolation and regional differentiation. *Hum Immunol* 62:954–969
- Jones FL (1970) *The structure and growth of Australia's Aboriginal population*. Australian National University Press, Canberra
- Jones R (1987) Pleistocene life in the dead heart of Australia. *Nature* 328:666
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer M (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal and Y-chromosome data. *Am J Hum Genet* 66:979–988
- Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill PA, Shen P, Oefner PJ, Tommaseo-Ponzetta M, Stoneking M (2003) Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72:281–302
- 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* 68:173–190
- Kimber RG (1990) Hunter gatherer demography: the recent past in Central Australia. In: Meehan B, White NG (eds) *Hunter-gatherer demography: past and present*. Oceania Monograph 39. University of Sydney Press, Sydney, pp 160–170
- Kirk RL (1983) *Aboriginal man adapting. The human biology of Australian aborigines*. Oxford University Press, Melbourne
- Kirk RL (1989) Population genetic studies in the Pacific: red cell antigen, serum protein and enzyme systems. In: Hill AVS, Serjeantson SW (eds) *The colonization of the Pacific: a genetic trail*. Oxford University Press, New York, pp 60–119
- Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>
- Li S-L, Yamamoto T, Yoshimoto T, Uchihi R, Mizutani M, Kurimoto Y, Tokunaga K, Jin F, Katsumata Y, Saitou N (2006) Phylogenetic relationship of the populations within and around Japan using 105 short tandem repeat polymorphic loci. *Hum Genet* 118:695–707
- Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. *Am J Hum Genet* 63:613–624
- Macho G, Freedman L (1987) A reanalysis of the Andrew A Abbie morphometric data on Australian Aborigines. *Occasional Papers in Human Biology, Australian Institute of Aboriginal and Torres Strait Islander Studies* 4
- MacKnight CC (1972) Macassans and aborigines. *Oceania* 42:283–321
- Mader MK (1965) A sample of fingerprints from Tongariki in the New Hebrides. *Oceania* 35:317–320
- Merriwether DA, Friedlaender JS, Mediavilla J, Mgone C, Gentz F, Ferrell RE (1999) Mitochondrial DNA variation is an indicator of Austronesian influence in island Melanesia. *Am J Phys Anthropol* 110:243–270
- Nicholls EM, Lewis HBM, Cooper DW, Bennett JH (1965) Blood groups and serum protein differences in Central Australian Aborigines. *Am J Hum Genet* 17:293–307
- Parsons PA, White NG (1973) Genetic differentiation among Australian Aborigines with special reference to dermatoglyphics and other anthropometric traits. In: Kirk RL (eds) *The human biology of Aborigines of Cape York*. Australian Institute of Aboriginal and Torres Strait Islander Studies, Canberra, pp 81–94
- Parsons PA, White NG (1976) Variability of anthropometric traits in Australian Aborigines and adjacent populations: its bearing on the biological origin of the Australians. In: Kirk RL, Thorne AG (eds) *The biological origin of the Australians*. Australian Institute of Aboriginal and Torres Strait Islander Studies, Canberra, pp 227–243
- Perez-Miranda AM, Alfonso-Sanchez MA, Kalantar A, Garcia-Obregon S, de Pancorbo MM, Pena JA, Herrera RJ (2005) Microsatellite data support subpopulation structuring among Basques. *J Hum Genet* 50:403–414
- Redd AJ, Stoneking M (1999) Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. *Am J Hum Genet* 65:808–828
- Roberts RG, Jones R, Smith MA (1990) Thermoluminescence dating of a 50,000-year human occupation site in northern Australia. *Nature* 345:153–156
- Roberts RG, Jones R, Spooner NA, Head MJ, Murray AS, Smith MA (1994) The human colonisation of Australia: optical dates of 53,000 and 60,000 years bracket human arrival at Deaf Adder Gorge, Northern Territory. *Quater Geochronol* 13:575–583
- Robson MK, Parsons PA (1967) Fingerprint studies on four central Australian Aboriginal tribes. *Archaeol Phys Anthropol Oceania* 2:69–78
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298:2381–2385
- Schneider S, Roessli D, Excoffier L (2000) Arlequin ver.2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Shepard EM, Chow RA, Suafo'a E, Addison D, Perez-Miranda AM, Garcia-Bertrand RL, Herrera RJ (2005) Autosomal STR variation in five Austronesian populations. *Hum Biol* 77:825–851
- Simmons RT, Tindale NB, Birdsell JB (1962) A blood group genetical survey in Australian Aborigines of Bentinck, Mornington and Forsyth islands, Gulf of Carpentaria. *Am J Phys Anthropol* 20:303–320
- Smith MA (1989) The case for a resident human population in the Central Australian Ranges during full glacial aridity. *Archaeol Oceania* 24:93–105
- Stoneking M (1994) Mitochondrial DNA and human evolution. *J Bioenerg Biomembr* 26:251–259
- Stoneking M, Jorde LB, Bhatia KK, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733
- Thorne A, Grun R, Mortimer G, Spooner NA, Simpson JJ, McCulloch MT, Taylor L, Curnoe D (1999) Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *J Hum Evol* 36:591–612



- Turney CSM, Bird MI, Fifield LK, Roberts RG, Smith M, Dortch CE, Grun R, Lawson E, Ayliffe LK, Miller GH, Dortch J, Cresswell RG (2001) Early occupation at Devil's Lair, Southwestern Australia 50,000 years ago. *Quat Res* 55:3–13
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62
- van Holst Pellekaan S, Frommer M, Sved J, Boettcher B (1997) Mitochondrial D-loop diversity in Australian riverine and Australian desert Aborigines. *Electrophoresis* 18:1538–1543
- van Holst Pellekaan SM, Frommer M, Sved JA, Boettcher B (1998) Mitochondrial control-region sequence variation in Aboriginal Australians. *Am J Hum Genet* 62:435–449
- van Holst Pellekaan SM, Ingman M, Roberts-Thomson JM, Harding RM (2006) Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. *Am J Phys Anthropol* 131(2):282–294
- Vandenberg N, van Oorschot RA, Tyler-Smith C, Mitchell RJ (1999) Y-chromosome-specific microsatellite variation in Australian Aborigines. *Hum Biol* 71:915–931
- Walker A, Zorc D (1981) Austronesian loanwords in Yolngu-Matha of Northeast Arnhem Land. *Aborig Hist* 5:109–134
- Walsh M (1993) Languages and their status in Aboriginal Australia. In: Walsh M, Yallop C (eds) *Language and culture in Aboriginal Australia*. Aboriginal Studies Press, Canberra, pp 1–13
- Warner WL (1964) *A black civilization* (revised edition). Harper Row, New York
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- White JP, O'Connell JF (1982) *A prehistory of Australia, New Guinea and Sahul*. Academic, Sydney
- White NG (1997) Genes, languages and landscapes in Australia. In: McConvell P, Evans N (eds) *Archaeology and linguistics: Aboriginal Australia in global perspective*. Oxford University Press, Melbourne, pp 45–81
- Zabala Fernandez WM, Borjas-Fajardo L, Fernandez Salgado E, Castillo C, Socca L, Portillo MG, Sanchez MA, Delgado W, Morales-Machin A, Layrisse Z, Pineda Bernal L (2005) Use of short tandem repeats loci to study genetic structure of several populations from Zulia State, Venezuela. *Am J Hum Biol* 17:451–459
- Zhivotovsky LA, Rosenberg NA, Feldman MW (2003) Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* 72:1171–1186