Flinders
UNIVERSITY

Title: Analysis of the South Australian Aboriginal population using the Global AIMs Nano ancestry test

Authors: Catherine Hopkins, Duncan Taylor, Kelly Hill, Julianne Henry

PII: S1872-4973(18)30662-8
DOI: https://doi.org/10.1016/j.fsigen.2019.03.020
Reference: FSIGEN 2079

To appear in: *Forensic Science International: Genetics*

Received date: 27 December 2018
Revised date: 19 March 2019
Accepted date: 23 March 2019

Please cite this article as: Hopkins C, Taylor D, Hill K, Henry J, Analysis of the South Australian Aboriginal population using the Global AIMs Nano ancestry test, *Forensic Science International: Genetics* (2019), https://doi.org/10.1016/j.fsigen.2019.03.020

TITLE:

Analysis of the South Australian Aboriginal population using the Global AIMs Nano ancestry test

AUTHORS:

Catherine Hopkins[1], Duncan Taylor[1,2], Kelly Hill[1], Julianne Henry[1,2]

1. College of Science and Engineering, Flinders University, GPO Box 2100 Adelaide SA, Australia 5001

2. Forensic Science SA, GPO box 2790, Adelaide, South Australia 5001

CORRESPONDING AUTHOR:

Duncan A. Taylor, PhD
Forensic Science SA
GPO Box 2790
Adelaide
South Australia 5001
Phone: +61-8 8226 7700
Fax: +61-8 8226 7777
Email: Duncan.Taylor@sa.gov.au

HIGHLIGHTS:

- 97 Australian Aboriginals were profiled using 31 ancestry –informative SNP markers
- Australian Aboriginality was based on self-declaration
- The Aboriginal group overlapped with European and Oceanic groups
- There was limited ability to distinguish Aboriginal ancestry at the regional level
- There was limited success in distinguishing Aboriginals from worldwide populations

ABSTRACT:

We investigate the ability of the 31 SNP loci in the Global AIMs Nano set to distinguish self-declared Australian Aboriginal individuals from European, Oceanic, African, Native American and East Asian populations. Human evolution suggests that Australian Aboriginal individuals came to Australia approximately 50 000 years ago, during the time it made up part of Sahul. Since then the colonisation of Australia by Europeans has meant significant admixture within the Australian Aboriginal population. These two events present themselves in our study with the Aboriginal population creating a continuous genetic cline between the Oceanic and European groups. We also assigned the Aboriginal individuals into their traditional regional groups to determine whether there was any ability to distinguish these from each other. We found similar results to studies using other markers, namely that the more remote regions (that have been less affected by admixture) diverged from the rest. Overall, we found the ability of the GNano system to differentiate self-declared Australian Aboriginal individuals was reasonable but had limitations that need to be recognised if these assignments are applied to unknown individuals.

KEY WORDS:

Global AIMs Nano; Australian Aboriginal; SNP profile; Ancestry Informative Markers.

1. INTRODUCTION

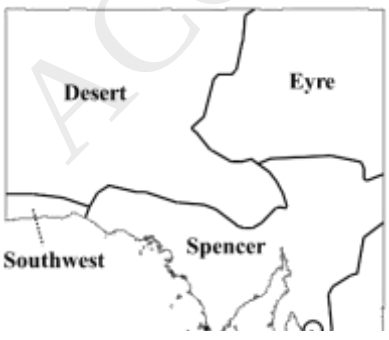
Australian Aborigines are believed to be one of the first settlements to occur after the "out of Africa" expansion [1]. Whilst some controversy still surrounds the mode of dispersal of modern humans, there is genetic and archaeological evidence to support that the landmass Sahul, made up of the Australian continent and Papua New Guinea, was settled approximately 47,000 years ago [1-3]. Studies of mitochondrial and Y chromosome DNA demonstrate the strong link between Aboriginal Australians and Oceanians, such as New Guineans [1, 2]. However, haplogroups unique to the non-admixed Aboriginal population also demonstrate their long-standing genetic isolation after the separation of Sahul and prior to European settlement [4].

There are no accurate estimates of the population of Australia before European settlement in 1788 [5] but immediately prior to European settlement it is possible that the Aboriginal population was between 300 000 and 1 million [5, 6]. A large proportion of these Australian Aborigines were displaced from their traditional lands and forced to assimilate with western culture [7, 8]. Mixed marriages between Aboriginal women and European men were common

[7] and according to the 1954 Census, only around 26,000 full-blood Aborigines remained due to inter-breeding. Regardless of the genetic extent of Aboriginal-European admixture, current census data shows that these individuals (who currently make up ~3% of the Australian population) identify themselves as Aboriginal [8]. Therefore, it is expected that self-declared Aborigines will exhibit genetic ancestries of variable Oceanic and European composition.

Aboriginal Australia was divided into eighteen cultural regions by Horton [9] based upon cultural, language and trade boundaries. Whilst these regions represent the general location of large groups of people, many of these regions were composed of smaller tribes. Four of these Aboriginal regions (Desert, Eyre, Spencer and Riverine) and an urbanised population in the capital city of Adelaide, fall within the state of South Australia (Figure 1, left). Phylogenetic analyses of South Australian self-declared Aboriginal and European individuals using autosomal STRs [10, 11], Y-chromosome STRs [12, 13] and Y-chromosome SNPs [4, 14], while acting on different areas of the DNA with different modes of inheritance, speak to a common picture of contemporary Aboriginal people (Figure 1). Those in remote areas, such as the Desert region, have a less admixed genetic history and so appear more genetically divergent from Europeans. In contrast, individuals in regions where European settlement was highest, such as the Riverine region and Adelaide, were the most admixed.

Ancestry-informative markers (AIMs) based on single-nucleotide polymorphisms (SNPs) are widely used to study human evolution and migration but are now becoming more widely applied in a forensic context to provide biogeographical ancestry (BGA) assignment of an unknown individual when no other investigative leads are available. Numerous commercial and non-commercial AIMs panels are available for this purpose (see review article [15]) and the depth of the assignment (whether it occurs at a continental or population level) depends upon the nature and number of markers used. The use of massively-parallel sequencing (MPS) to analyse large BGA SNP panels, often in combination with other forensically-relevant markers, is gaining momentum. While very large amounts of information can be generated, equipment and consumables are expensive, and analysis and storage of the data can be complex. A simpler and less expensive alternative to MPS is analysis of smaller AIMs panels using a SNaPshot assay which allows laboratories to implement BGA testing on their current CE platforms.



Autosomal-STR	Y-STR

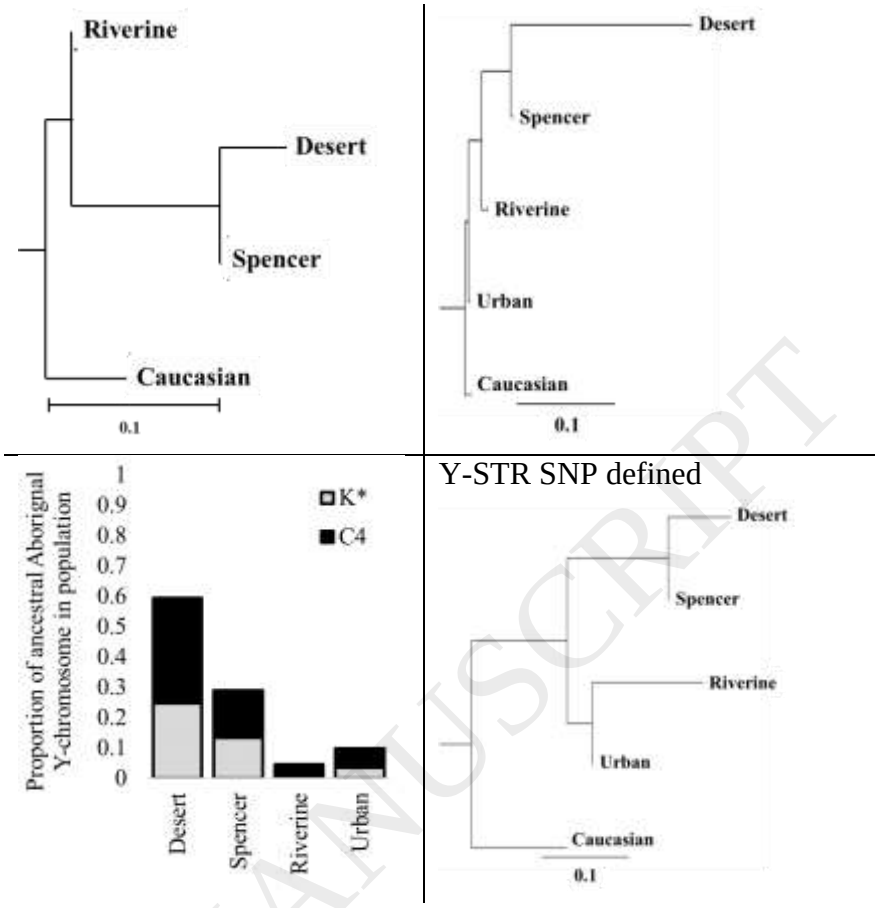


Figure 1: (left) Map of South Australia showing traditional Aboriginal regional groups [9], (top middle) Neighbour-joining tree based on nine autosomal STR loci, Nei genetic distances, and self-declaration of ethnicity (based on data from [16]), (top right) Neighbour-joining tree based on 17 Y-STR loci using data from [14], Nei genetic distances and self-declaration of ethnicity, (bottom middle) Proportion of confirmed Aboriginal Y-chromosomes (as indicated by the K and C4 haplogroups) in the self-declared South Australian population as an indication of paternal European admixture using data from [14], and (bottom right) Neighbour-joining tree based on 17-Y STR loci for SNP confirmed ancestral Y-chromosomes using data from [14].*

The Global AIMs Nano ancestry test (hereafter GNano) is a 31-plex SNaPshot assay developed by de la Puente et. al [17] which enables differentiation between five continental populations; African, European, East Asian, Native American and Oceanian. Assignment of an individual with unknown ancestry into one of these continental groups is achieved by comparing the resultant SNP typings to the associated reference databases. As is common of any population genetic assignment, if a person originates from a population where reference data does not exist (or there is significant admixture), the assignment will be made as best it can against existing data. GNano was recently investigated by our laboratory as a possible addition to our DNA testing armoury with the aim of providing additional leads to police investigators. Part of this investigation examined the performance of the test for the three

main populations seen in South Australia, namely European, East Asian and Aboriginal. As discussed above, GNano reference data sets exist for the European and East Asian populations. Due to difficulty in obtaining DNA samples from Australian Aborigines for the purpose of genetic studies, there is currently no published reference population data for AIMs, including those in GNano. Therefore, the performance of GNano for predicting the ancestry of self-declared Aboriginal people was unknown prior to our study but it was expected that this population would exhibit some form of continuous genetic cline between Oceanic and European groups. This paper describes our findings in relation to the ancestry prediction of GNano for 97 South Australian Aboriginal individuals both in the presence and absence of a self-declared Aboriginal reference data set. It also evaluates the ability of GNano to designate Aboriginal people to individual regional groups.

2. METHODS

2.1 Samples

The use of 97 self-declared Australian Aboriginal DNA casework samples was approved under the South Australian Criminal Law (Forensic Procedures) Act 2007 [18]. All samples were de-identified. The Research and Development Committee at Forensic Science SA and the Human Research Ethics Committee at Flinders University of South Australia both determined that, as a result, the informed consent of the donors was not required. Because informed consent was not possible to attain for these samples (they had been previously de-identified according to the Act), we believe their ethical use is implied under clause 32 of the World Medical Association Declaration of Helsinki. Assignment of each individual to a region used location of residence or location of offence for which the individual was arrested (if the location of residence was not available), as per [10, 14]. This manner of region assignment had limitations (namely the fact that the residence or crime of an individual is not necessarily in the same geographical location as their ancestors) but has shown to be an approximation that is fit for purpose. Table 1 shows the breakdown of the dataset into regional grouping for our study.

Table 1: Number of self-declared Australian Aboriginal individuals included in the study

Region	Samples
Desert	23
Eyre	3
Riverine	20
Spencer	26
Urban	25
TOTAL	97

2.2 DNA extraction and SNP genotyping

DNA was extracted from FTA paper using the Promega DNA-IQ system (Promega, Madison, WI, USA) and quantified using the Quantifiler® Trio DNA quantification kit (ThermoFisher Scientific, Waltham, MA, USA). SNPs were amplified using GNano primer set and methods as described in [17] on a a GeneAmp® 9700 thermocycler (ThermoFisher Scientific) and separated using a 3500*xl* Genetic Analyser (ThermoFisher Scientific). DNA profiles were visualised using Genemapper ID-X v1.4 (ThermoFisher Scientific) using an analytical threshold of 50 relative fluorescence units (rfu) and homozygous thresholds that were locus specific (based on internal validations, data not shown). Only complete profiles were included for analysis in this study.

2.3 Statistical analysis

Ancestry prediction of the GNano assay under different conditions (as described below) was evaluated using the online tool Snipper (http://mathgene.usc.es/snipper/). Analyses did not assume Hardy Weinberg Equilibrium. The Aboriginal population was included in the dataset as either a group of individuals of unknown origin (and hence being assigned to one of the existing GNano reference populations) or as a reference population itself. When included as a reference population the ability to assign Aboriginal individuals into the self-declared Aboriginal reference group was trialled by leave one out cross validation i.e. each target individual was removed from the dataset (so that the reference Aboriginal population did not include that target individual) and the target individual was analysed as being of unknown ethnicity.

Population structures were analysed using STRUCTURE v.2.3.4 [19]. The length of Burn-in and the number of MCMC repetitions after Burn-in were both set at 100,000 for 5 iterations per *K* (the number of assumed populations within a dataset). Correlated allele frequencies were employed. For all analyses, the default populations provided with the GNano publication were designated as references (popflag = 1) and the Aboriginal Australian individuals were designated as of unknown origin (popflag = 0). CLUMPAK [20] (http://clumpak.tau.ac.il) was used to align the runs completed in STRUCTURE.

2.4 Phylogenetic analyses

Phylogenetic analyses (to produce phylogenetic trees seen in Figure 1) were carried out either using Genetic Data Analyser [21] or R [22] with addon ‘ape’ [23].

3. RESULTS

3.1 Discrimination power of GNano loci for Australian Aborigines

The divergence output calculated in Snipper is expressed in Jensen-Shannon Divergence values (in natural log form) [24]. Rosenberg’s informativeness for assignment (I_n) [25] unit is

used more widely and the Jensen-Shannon Divergence were converted to I_n by multiplying by $ln(2)$. The I_n indicates the amount of information that the multiallelic markers provide about an individual's ancestry. The larger the I_n value between populations the more difference and therefore divergence there is between the allelic markers. I_n was determined for each locus in the GNano kit for the Aboriginal population as a whole compared to all remaining non-Aboriginal groups considered together as a single population (Figure 2).

When comparing the informativeness of the loci between the Australian Aboriginal group and the GNano reference populations, there were no loci where the highest level of informativeness was for the Australian Aboriginal group. Given the Oceanic ancestry of the Australian Aboriginal people it is not surprising that the most informative loci for the Australian Aboriginal group were from this region (Figure 2). The construction of the GNano system (and specifically the selection of loci) was carried out to separate European, East Asian, African, Native American and Oceanic groups (and notably not Australian Aboriginals) and this is reflected in the fact that the highest value for I_n in the Australian Aboriginal group was just below 0.25, compared to values ranging up to approximately 0.7 in other populations [17].

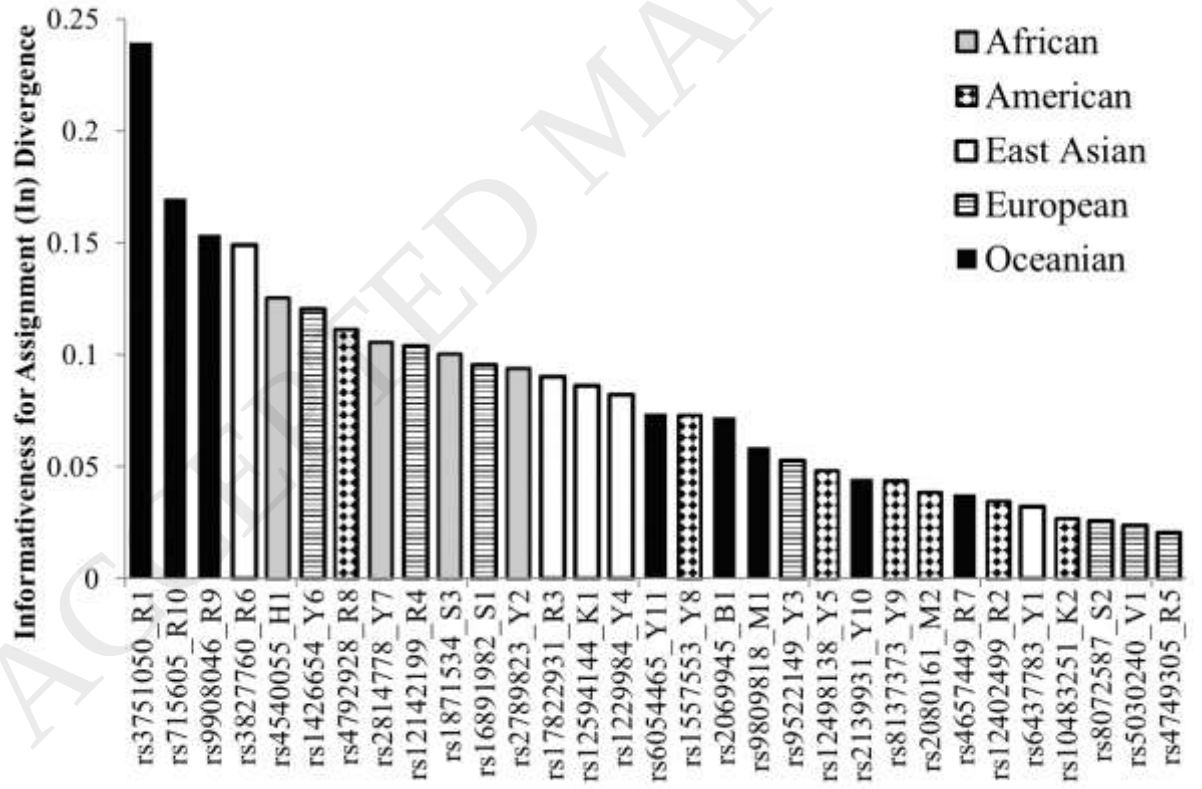


Figure 2. Population specific divergence of the biallelic GNano loci for the Aboriginal population compared to the five GNano default populations considered together as a single non-Aboriginal population. The GNano default reference population for which the SNP is most informative is represented by the patterning of the columns.

The accumulated divergence (i.e. taking into account the divergence of each locus from Figure 2) comparing the Australian Aboriginal group as a whole to the GNano default reference populations is shown in Figure 3. Note that Figure 3 is created by comparing the Australian Aboriginal group to each other population separately. The order of loci is by greatest to smallest divergence for each population comparison (and so are different to the order of loci in Figure 2). We have included the raw data that led to Figures 2 and 3 as supplementary material. Again, given the ancestral connection of the Australian Aboriginal people with the Oceanic group, and the contemporary connection to the European group it was expected that the lowest divergence was seen between the Australian Aboriginal and these two groups. Between the Australian Aboriginal group and the Oceanic and European groups, Figure 3 shows that the accumulated divergence is only significantly contributed to by approximately 15 of the loci.

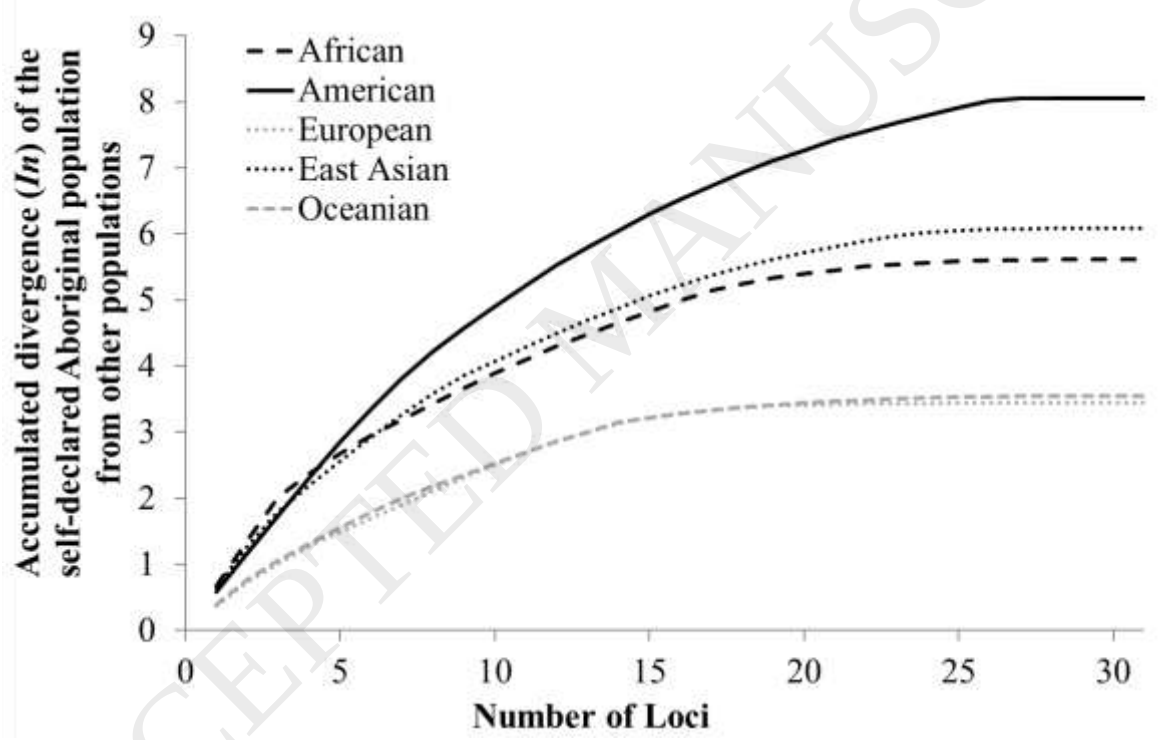


Figure 3. Accumulated pairwise divergence for Aboriginal population against each GNano reference population

As well as considering the Australian Aboriginal group as a whole (as in Figures 2 and 3), we also investigated the ability to distinguish the individual regional groups. Table 2 shows the population specific divergence (PSD) values for each of the regional groups for each locus, compared to the GNano reference populations treated as a single 'non-Aboriginal' group. Highlighted in Table 2 are the regions for which each locus shows the highest PSD value, out of the Australian Aboriginal regional groups. As may be expected from the higher instance of

ancestral Aboriginal genetics, the Desert and Eyre groups possessed the greatest number of loci with high divergence values. Total accumulated divergence values (shown in the bottom row of Table 2) are lower than the values seen in Figure 3 (comparing the Australian Aboriginal group as a whole to GNano default populations).

Table 2: Population specific divergence (PSD), as I_n, of the Aboriginal regional groups compared separately against the five GNano reference populations (treated as a single 'non-Aboriginal' group). Highlighted PSD values indicate population where SNP loci are of highest divergence (out of the regional Aboriginal groups). Numerical values have been reduced to two significant figures for display. Note that the Eyre category is based on only three individuals and so caution should be used when interpreting these results.

SNP	Desert	Eyre	Riverine	Spencer	Urban
rs9908046	0.44	0.30	0.08	0.12	0.13
rs2139931	0.17	0.03	0.04	0.02	0.04
rs3827760	0.15	0.04	0.14	0.12	0.12
rs5030240	0.12	0.12	0.05	0.04	0.03
rs4657449	0.12	0.08	0.09	0.06	0.04
rs8137373	0.11	0.07	0.07	0.03	0.03
rs4749305	0.09	0.03	0.09	0.02	0.07
rs2080161	0.08	0.06	0.07	0.08	0.03
rs12498138	0.06	0.00	0.05	0.05	0.05
rs3751050	0.39	0.48	0.16	0.23	0.20
rs715605	0.45	0.45	0.15	0.15	0.09
rs6054465	0.27	0.37	0.05	0.08	0.01
rs4540055	0.18	0.25	0.10	0.15	0.17
rs9809818	0.12	0.20	0.11	0.08	0.05
rs12594144	0.09	0.19	0.04	0.08	0.14
rs10483251	0.05	0.15	0.05	0.03	0.01
rs6437783	0.06	0.11	0.06	0.06	0.06
rs12142199	0.04	0.00	0.29	0.11	0.17
rs16891982	0.05	0.03	0.18	0.14	0.16
rs2069945	0.13	0.23	0.14	0.08	0.03
rs17822931	0.11	0.02	0.11	0.11	0.06
rs12402499	0.04	0.00	0.05	0.04	0.04
rs4792928	0.11	0.02	0.09	0.13	0.13
rs1229984	0.05	0.00	0.05	0.06	0.05
rs1426654	0.08	0.02	0.21	0.16	0.24
rs9522149	0.02	0.06	0.13	0.03	0.14
rs8072587	0.04	0.08	0.05	0.07	0.14
rs2814778	0.07	0.05	0.11	0.07	0.12
rs1871534	0.07	0.04	0.10	0.07	0.11
rs2789823	0.06	0.02	0.06	0.06	0.09

rs1557553	0.09	0.01	0.06	0.07	0.09
Total Divergence	**3.89**	**3.52**	**3.06**	**2.59**	**2.82**

3.2 Principal component analyses

Principal component analyses (PCA) was carried out including all GNano reference populations and the Australian Aboriginal individuals, broken up into regional groups (Figure 4). The placement of the Aboriginal Australian individuals, separated into regional groups, showed that they did not cluster close enough within regions (or separate enough from other regions) to be able to be classified into their own region. There are however, some patterns emerging from this regional separation that can be explained by admixture, isolation of certain groups and the migration of humans into Australia. The individuals from the Desert and Eyre populations appeared to cluster together and toward the Oceanian population (apart from one Desert individual who is clearly placed in the European cluster). This is expected due to the isolation of these two regions from the more coastal areas of Australia such as the Riverine and Spencer regions. Samples from the urban areas of Australia are seen closer to the European population. This is also the same for the Riverine samples. There is some overlap seen between the Riverine and Urban samples and the European population. No individuals from these two regional groups are observed within the Oceanian population. Samples from the Spencer region are seen throughout the entire Aboriginal cluster, spread from the European population to the Oceanian population.

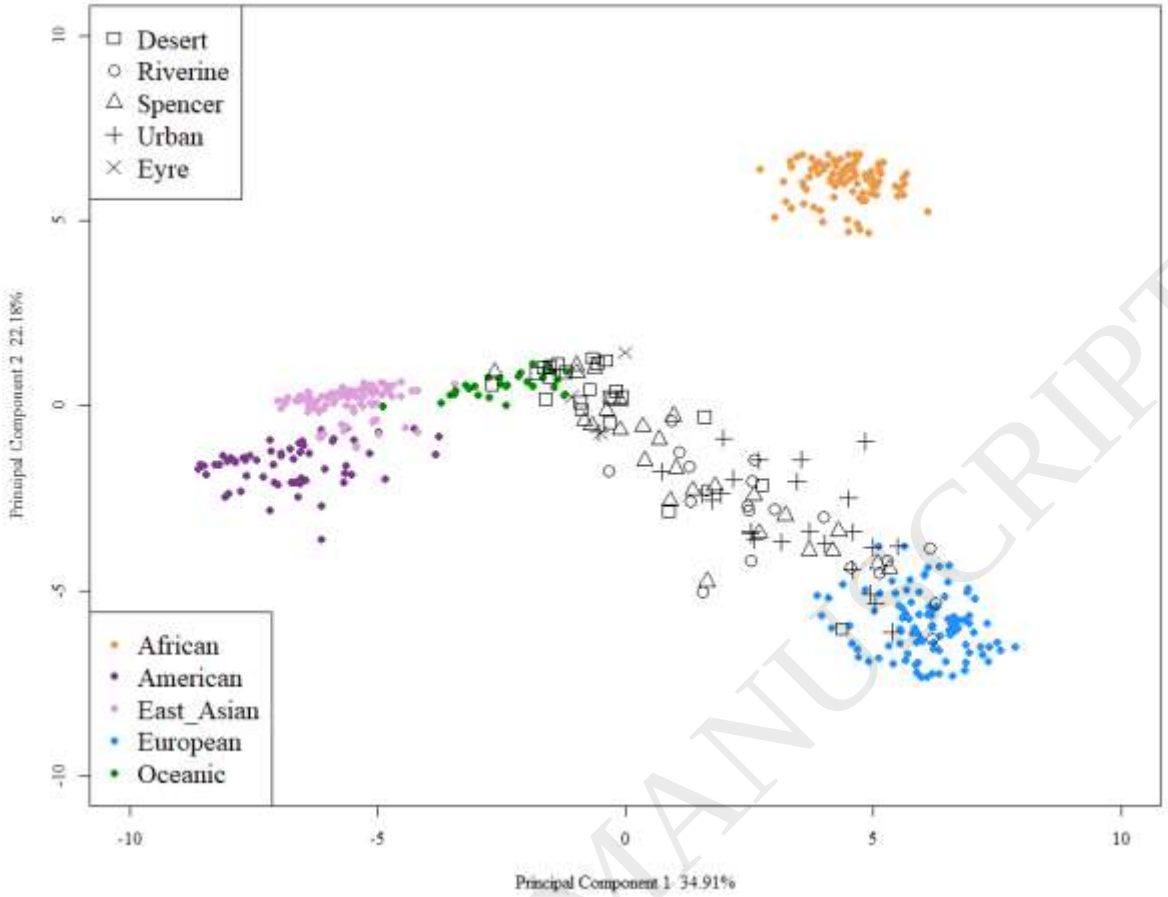


Figure 4. GNano reference populations and Aboriginal regional groups compared with PCA analysis.

3.3 Population structure

Many of the patterns seen throughout the previous analyses can be explained by European admixture within the Aboriginal DNA samples as well as the similarities shared with the Oceanian population due to them being the most recent known common ancestor. The ability to observe the level of admixture present within each individual's DNA in comparison to the GNano reference populations allows for further information to be gained regarding the migration into Australia and the effect European colonization had on the Aboriginal population. To this end STRUCTURE was used to assign individuals to populations based on their SNP genotypes and those of the entire analysis population [26].

Figure 5 shows the results of multiple STRUCTURE runs combined using CLUMPAK, when the Aboriginal population was either considered as a single group, or as five groups (based on their tribal region). The number of populations was trialled as either five or six, but not any higher. When using five populations the Australian Aboriginal group was considered an

admixed group of Oceanic and European ancestry, which is consistent with other analyses conducted (and reported in previous sections). The number of populations trialled was not taken beyond six, because at $K = 6$, the default reference populations of the GNano started to split into separate populations. This may suggest that there is more variation between the individuals within the GNano default reference populations than there is between the Australian Aboriginal and European and Oceanic groups.

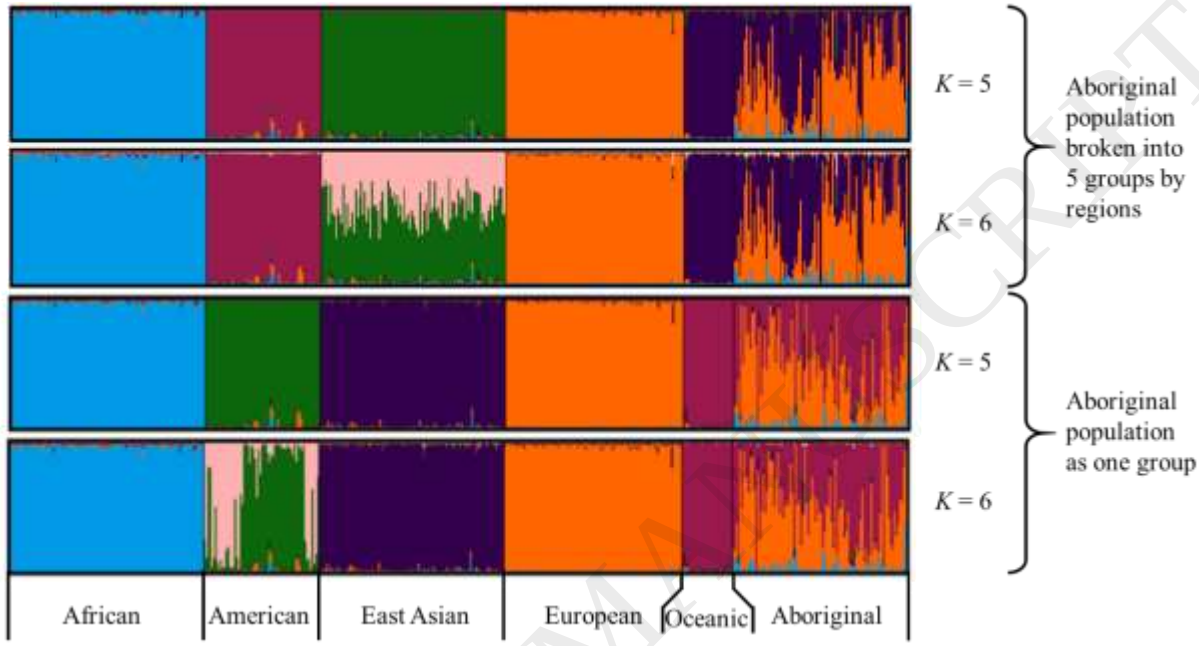


Figure 5. CLUMPAK analysis of Australian Aboriginal data with GNano reference populations, where Aboriginal individuals are either separated into regional groups (top) or considered as a single group (bottom).

Figure 6 shows the STRUCTURE analysis run on only the Australian Aboriginal data. This was done to see if the absence of the other five populations would allow for separation of these regions into their own 'populations'. The K value was ranged from two to five (the maximum based on the number of regions used in the analysis). Initially run in STRUCTURE, the data appeared quite varied between the five iterations. CLUMPAK was used to align these runs. As there were no reference populations included in this analysis, it was not able to be determined what populations contributed to the admixture seen in the Aboriginal individuals, however based on the similarities seen between Figure 5 the orange and blue colouring seem to align with the European and Oceanian populations respectively.

At all values of *K* (Figure 6), differences can be seen between the more remote regional groups (Desert and Eyre) and the Spencer, Riverine and Urban groups. The increase in the *K* value beyond two, did not produce any further separation of regional groups. The Desert and Eyre region samples produced similar admixture levels at every number of assumed populations. This is also the same between the Spencer, Riverine and Urban samples. These

samples contained much more ‘orange’ admixture, whereas the Desert and Eyre were predominantly blue.

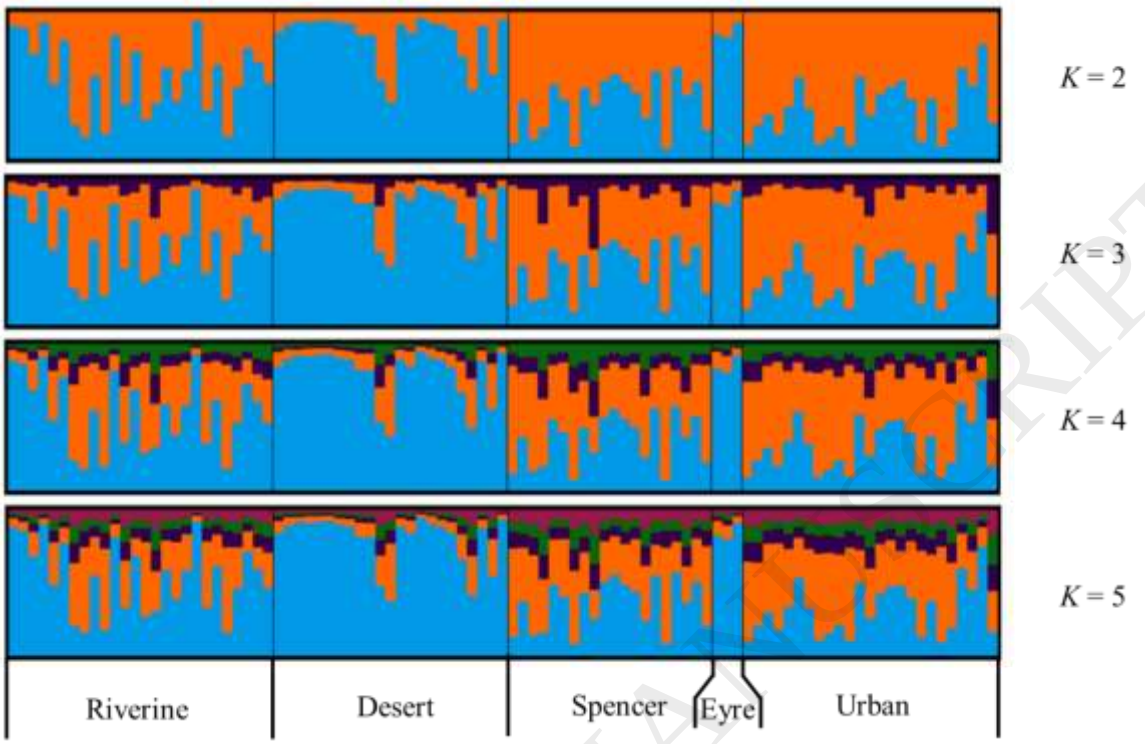


Figure 6. CLUMPAK analysis of Australian Aboriginal data separated into regional groups and trialled as two to five separate populations.

3.4 Population assignment

The ultimate aim of this study was to determine whether the GNano system would be suitable to identify ancestral origins of individuals from South Australia. A large proportion of the population in South Australia is of European descent (approximately 95%), and smaller proportions are Asian (approximately 3%) and Australian Aboriginal (approximately 2%) [27]. The European and Asian populations exist as reference populations in GNano, however the performance of assigning self-declared Aboriginal individuals into an Aboriginal reference population is unknown.

Figure 7 shows the results of assignments of European, Oceanic and self-declared Australian Aboriginal individuals into each of these three groups either in the absence of a self-declared Australian Aboriginal reference population, or in the presence of such a population. As seen in Figure 7, both prior and post the addition of the self-declared Australian Aboriginal reference population, all European individuals are assigned to the European group and Oceanic individuals are assigned into the Oceanic group. However, due to the overlap between the self-declared Aboriginal individuals with both these two groups (most clearly

seen in the PCA plot in Figure 4) the strength of the assignment, displayed as a log(likelihood), decreases in both instances.

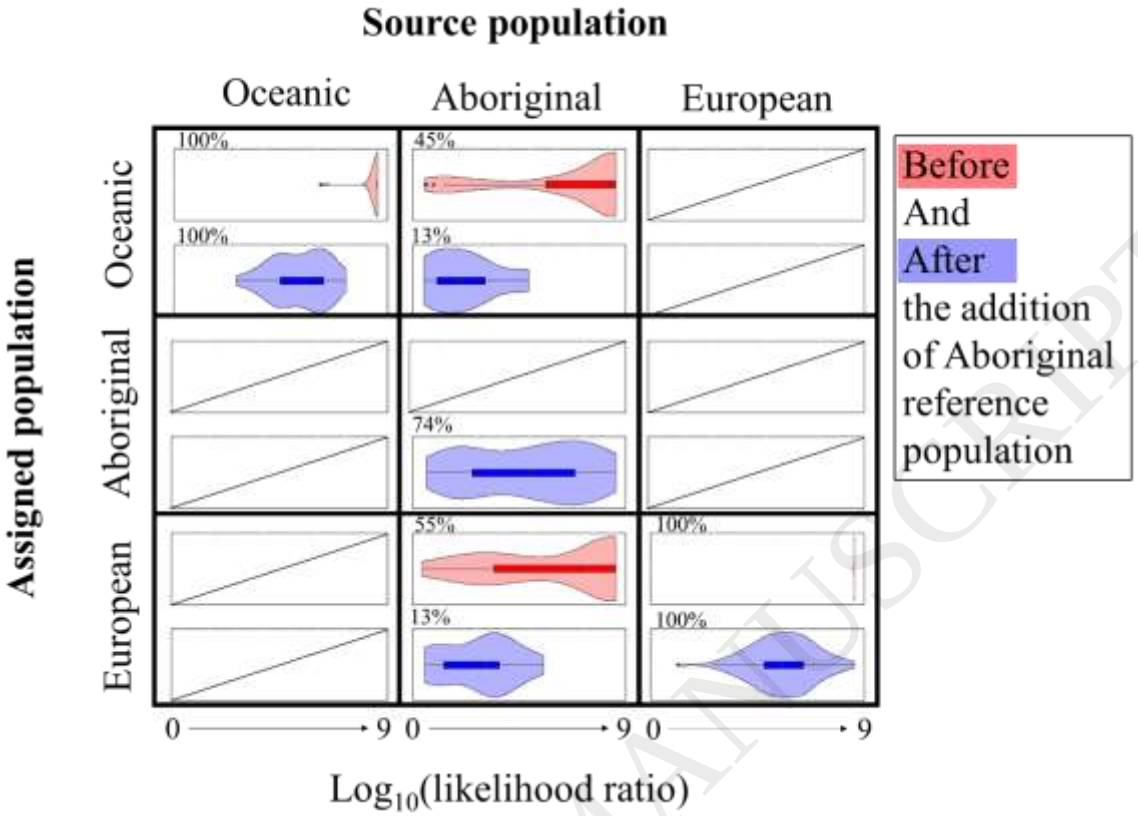


Figure 7: Assignments of individuals in the Oceanic, European and Aboriginal groups by Snipper into these three populations before (pink) and after (blue) the addition of a self-declared Aboriginal reference population. Violin plots show the distribution of log_{10}(likelihood ratio) values and the percentage above each box represents the percentage of individuals that were assigned in that category. A box with a diagonal line represents a category that has no assignments.

For the assignment of the individuals in the self-declared Australian Aboriginal group, prior to the addition of this group as a reference, the individuals are assigned (with approximately equal occurrence) into either the European or Oceanic groups, with a range of strengths. After the addition of the self-declared Australian Aboriginal group as a reference, the majority of self-declared Australian Aboriginal individuals are assigned into their own group (remembering that this is being determined by leave one out cross validation), with a much lower assignment into either the European or Oceanic groups, and a lower likelihood when this mis-assignment does occur.

4. DISCUSSION

The SNPs present in the GNano system were chosen for their ability to separate the default reference populations (European, East Asian, African, Native American and Oceanian), and did not have data on the Australia Aboriginal group. Australian Aboriginals diverged from other Oceanic groups approximately 50 000 years ago and so the Oceanic group within the GNano reference populations is expected to be the most closely related. More recently the Australian Aboriginal group has undergone extreme admixture from (predominantly) European colonisation of Australia. In our study we trialled 97 self-declared Australian Aboriginal individuals using the 31 GNano loci. Given the ancient and modern genetic history of the Australian Aboriginal people, and the design of the GNano system, it is not surprising that the ability to distinguish Australian Aboriginals is limited.

All analyses in our study show a similar trend, which is that those Australian Aboriginal individuals that are from more remote areas of Australia (the Desert and Eyre regions shown in Figure 1), have a history that is less subject to European admixture and tend to cluster with the Oceanic group. Note that while the de-identification of the individuals in this study means the family histories are unknown, speaking in general terms, individuals who are in more coastal and urbanised areas tend to have a history of greater admixture and will appear either as clustering with European individuals (if the individual family history has a predominantly European heritage) or somewhere between Oceanic and European. This is consistent with studies on other autosomal and Y-chromosome markers (as shown in Figure 1).

When the self-declared Australian Aboriginal group is used as a reference population there is some ability to identify a self-declared Aboriginal individual from other default reference populations in the GNano system. Approximately three quarters of individuals are so assigned, with the remaining quarter spread across the Oceanic or European groups. It is the Aboriginal individuals that have partially admixed family histories (again, speaking in generality and not from knowledge of the family histories of the 97 profiled individuals in our study) that are most clearly distinguishable, as they fall neither in the Oceanic, or European clusters. However, it is important to recognise that we do not know how the assignment performs in the presence of other admixed individuals e.g. if a self-declared Aboriginal reference population is added, it is quite likely that an individual who has one Polynesian and one European parent would have very strong assignment into the Aboriginal group.

The other consideration is how the addition of a reference population for self-declared Australian Aboriginal individuals affects assignments of other individuals into their correct populations. As shown in Figure 7, all assignments of Europeans were into the European group and Oceanians were into the Oceanic group, however there was a drop in the strength of the assignments. One possibility for assignment may be to model the distributions of likelihoods and use this value as a type of score that could then be used to assign probabilities for being a self-declared Australian Aboriginal (noting that the use of scores have been

shown to have limitations [28]). The need for some additional consideration (to the likelihood value that is produced by Snipper for assignments) comes from the fact that self-declaration and genetic composition, while expected to broadly align, are not always the same i.e. in the case of Australian Aboriginal individuals it comes down to the difference between identifying someone who is ancestrally genetically Aboriginal and someone who identifies as Aboriginal (which the Australian Government recognises as being more than a proportion of ancestrally genetic heritage).

Given this we feel that the assignment of an individual as either 'Australian Aboriginal' or not, is not the best manner to interpret results or express the findings. A dichotomous decision is likely to cause interpretative issues and is not necessary. A better treatment of the problem is to consider the results of analyses using some form of multidimensional scaling (e.g. PCA) and STRUCTURE to guide investigators as to the range of potential ancestral makeup the individual might possess and allow the investigator to incorporate that with other (non-genetic) information they may have.

5. CONCLUSION

We have explored the performance of the 31 GNano SNP loci to distinguish self-declared Australian Aboriginal individuals from other continental population groups. The inability to achieve strong segregation of the Aboriginal population with this test is unsurprising given the absence of SNP markers specific for this population. In fact, this limitation applies to all currently available SNP-based ancestry tests and whether such markers exist is currently unknown.

We make the final note that while there may be some ability to identify the (predominant) genetic heritage of an individual, this does not necessarily translate to a phenotype or social affiliation. Whilst we plan to undertake a study to link the GNano results with the physical appearance of an individual, the potential for discordance between the two must be recognised. Regardless of the limitation of the test and the results, we feel the GNano system can provide potentially important investigative information on individuals from South Australia, which takes into account the genetic make-up of the region.

SUPPLEMENARY MATERIAL:

97 Australian Aboriginal GNano SNP profiles with Regional information in Snipper format

REFERENCES:

[1] N. Nagle, K.N. Ballantyne, M. van Oven, C. Tyler-Smith, Y. Xue, S. Wilcox, L. Wilcox, R. Turkalov, R.A. van Oorschot, S. van Holst Pellekaan, T.G. Schurr, P. McAllister, L. Williams, M. Kayser, R.J. Mitchell, Mitochondrial DNA diversity of present-day Aboriginal Australians and implications for human evolution in Oceania, Journal of human genetics 62(3) (2017) 343-353.

[2] I. Pugach, F. Delfin, E. Gunnarsdóttir, M. Kayser, M. Stoneking, Genome-wide data substantiate Holocene gene flow from India to Australia, Proceedings of the National Academy of Sciences 110(5) (2013) 1803.

[3] A.-S. Malaspinas, M.C. Westaway, C. Muller, V.C. Sousa, O. Lao, I. Alves, A. Bergström, G. Athanasiadis, J.Y. Cheng, J.E. Crawford, T.H. Heupink, E. Macholdt, S. Peischl, S. Rasmussen, S. Schiffels, S. Subramanian, J.L. Wright, A. Albrechtsen, C. Barbieri, I. Dupanloup, A. Eriksson, A. Margaryan, I. Moltke, I. Pugach, T.S. Korneliussen, I.P. Levkivskyi, J.V. Moreno-Mayar, S. Ni, F. Racimo, M. Sikora, Y. Xue, F.A. Aghakhanian, N. Brucato, S. Brunak, P.F. Campos, W. Clark, S. Ellingvåg, G. Fourmile, P. Gerbault, D. Injie, G. Koki, M. Leavesley, B. Logan, A. Lynch, E.A. Matisoo-Smith, P.J. McAllister, A.J. Mentzer, M. Metspalu, A.B. Migliano, L. Murgha, M.E. Phipps, W. Pomat, D. Reynolds, F.-X. Ricaut, P. Siba, M.G. Thomas, T. Wales, C.M.r. Wall, S.J. Oppenheimer, C. Tyler-Smith, R. Durbin, J. Dortch, A. Manica, M.H. Schierup, R.A. Foley, M.M. Lahr, C. Bowern, J.D. Wall, T. Mailund, M. Stoneking, R. Nielsen, M.S. Sandhu, L. Excoffier, D.M. Lambert, E. Willerslev, A genomic history of Aboriginal Australia, Nature 538 (2016) 207.

[4] N. Nagle, K.N. Ballantyne, M. van Oven, C. Tyler-Smith, Y. Xue, D. Taylor, S. Wilcox, L. Wilcox, R. Turkalov, R.A. van Oorschot, P. McAllister, L. Williams, M. Kayser, R.J. Mitchell, Antiquity and diversity of aboriginal Australian Y-chromosomes, Am J Phys Anthropol 159(3) (2016) 367-81.

[5] Australian Bureau of Statistics. 1301.0 - Year Book Australia, 2002. <http://www.abs.gov.au/ausstats/abs@.nsf/94713ad445ff1425ca25682000192af2/bfc28642d31c215cca256b350010b3f4!OpenDocument>, 2002 (accessed 21-02-2019.2019).

[6] A.H. Chisholm, The Australian encyclopaedia, [2nd ed.] ed., Angus & Robertson, Sydney, 1958.

[7] G. Blyton, J. Ramsland, Mixed-race unions and Indigenous demography in the Hunter Valley of New South Wales, 1788-1850, Journal of the Royal Australian Historical Society 98(1) (2012) 125-148.

[8] Australian Museum. Indigenous Australia Timeline - 1500 to 1900. <https://australianmuseum.net.au/indigenous-australia-timeline-1500-to-1900>, 2018 (accessed 21-02-2019.2019).

[9] D. Horton, Aboriginal Australia Map, AIATSIS and Auslig, Canberra, ACT, 1996.

[10] S.J. Walsh, J. Buckleton, Autosomal microsatellite allele frequencies for 15 regionally defined Aboriginal Australian population datasets, Forensic Science International 168 (2007) e29-e42.

[11] J. Buckleton, J.-A. Bright, D. Taylor, Forensic DNA Evidence Interpretation, second edition, CRC Press, Boca Raton, Florida, 2016.

[12] D. Taylor, J. Henry, Haplotype data for 16 Y-chromosome STR loci in Aboriginal and Caucasian populations in South Australia, Forensic Science International: Genetics 6 (2012) e187-e188.

[13] T.E. Collins, R. Ottens, K.N. Ballantyne, N. Nagle, J. Henry, D. Taylor, M.G. Gardner, A.J. Fitch, A. Goodman, R.A. van Oorschot, R.J. Mitchell, A. Linacre, Characterisation of novel and rare Y-chromosome short tandem repeat alleles in self-declared South Australian Aboriginal database, Int J Legal Med 128(1) (2014) 27-31.

[14] D. Taylor, N. Nagle, K.N. Ballantyne, R.A.H. van Oorschot, S. Wilcox, J. Henry, R. Turakulov, R.J. Mitchell, An investigation of admixture in an Australian Aboriginal Y-chromosome STR database, Forensic Science International: Genetics 6 (2012) 532-538.

[15] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic science international. Genetics 18 (2015) 49-65.

[16] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery, B. Weir, Population-specific FST values for forensic STR markers: A worldwide survey., Forensic Science International: Genetics 23 (2016) 126-133.

[17] M.d.l. Puente, C. Santos, M. Fondevila, L. Manzo, The-EUROFORGEN-NoE-Consortium, Á. Carracedo, M.V. Lareu, C. Phillips, The Global AIMs Nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs, Forensic Science International: Genetics 22 (2016) 81-88.
[18] South Australian Criminal Law Forensic Procedures Act. <https://www.legislation.sa.gov.au/LZ/C/A/CRIMINAL%20LAW%20(FORENSIC%20PROCEDURES)%20ACT%202007.aspx>, (accessed 30-11-2018.2018).
[19] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of Population Structure Using Multilocus Genotype Data, Genetics 155 (2000) 945-959.
[20] N.M. Kopelman, J. Mayzel, M. Jakobsson, N.A. Rosenberg, I. Mayrose, Clumpak: a program for identifying clustering modes and packaging population structure inferences across K, Mol Ecol Resour 15(5) (2015) 1179-91.
[21] P.O. Lewis, D. Zaykin, Genetic Data Analysis: Computer program for the analysis of allelic data, 2001.
[22] M. Plummer, rjags: Bayesian graphical models using MCMC, 2012.
[23] E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language, Bioinformatics 20(2) (2004) 289-290.
[24] I. Grosse, P. Bernaola-Galvan, P. Carpena, R. Roman-Roldan, J. Oliver, H.E. Stanley, Analysis of symbolic sequences using the Jensen-Shannon divergence, Physical review. E, Statistical, nonlinear, and soft matter physics 65(4 Pt 1) (2002) 041905.
[25] N.A. Rosenberg, L.M. Li, R. Ward, J.K. Pritchard, Informativeness of genetic markers for inference of ancestry, Am J Hum Genet 73(6) (2003) 1402-22.
[26] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M.V. Lareu, An overview of STRUCTURE: applications, parameter settings, and supporting software, Front Genet 4 (2013) 98.
[27] The Australian Bureau of Statistics. <http://www.abs.gov.au/census>, 2018 (accessed 30-11-2018.2018).
[28] A. Bolck, H. Ni, M. Lopatka, Evaluating score- and feature-based likelihood ratio models for multivariate continuous data: applied to forensic MDMA comparison, Law, Probability and Risk 14(3) (2015) 243-266.