Forensic Science International: Genetics 20 (2016) 71–80

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Pacifiplex: an ancestry-informative SNP panel centred on Australia and the Pacific region

Carla Santos[a,*], Christopher Phillips[a], Manuel Fondevila[a], Runa Daniel[b], Roland A.H. van Oorschot[b], Esteban G. Burchard[c], Moses S. Schanfield[d], Luis Souto[e], Jolame Uacyisrael[f], Marc Via[g], Ángel Carracedo[a,h], Maria V. Lareu[a]

[a] Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Spain
[b] Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Victoria, Australia
[c] University of California San Francisco, San Francisco, California, USA
[d] Department of Forensic Science, George Washington University, Mount Vernon College Campus, Washington, USA
[e] Department of Biology, University of Aveiro, Aveiro, Portugal
[f] Fiji Police Forensic Biology and DNA Laboratory, Nasova, Suva, Fiji
[g] Universitat de Barcelona, Barcelona, Spain
[h] Galician Foundation of Genomic Medicine (SERGAS), CIBERER (University of Santiago de Compostela), Sanatiago de Compostela, Spain

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form 22 September 2015
Accepted 6 October 2015
Available online 20 October 2015

Keywords:
Forensic assay
Ancestry informative
Single nucleotide polymorphisms
Australia
Pacific region

ABSTRACT

The analysis of human population variation is an area of considerable interest in the forensic, medical genetics and anthropological fields. Several forensic single nucleotide polymorphism (SNP) assays provide ancestry-informative genotypes in sensitive tests designed to work with limited DNA samples, including a 34-SNP multiplex differentiating African, European and East Asian ancestries. Although assays capable of differentiating Oceanian ancestry at a global scale have become available, this study describes markers compiled specifically for differentiation of Oceanian populations. A sensitive multiplex assay, termed *Pacifiplex*, was developed and optimized in a small-scale test applicable to forensic analyses. The *Pacifiplex* assay comprises 29 ancestry-informative marker SNPs (AIM-SNPs) selected to complement the *34-plex* test, that in a combined set distinguish Africans, Europeans, East Asians and Oceanians. Nine Pacific region study populations were genotyped with both SNP assays, then compared to four reference population groups from the HGDP-CEPH human diversity panel. *STRUCTURE* analyses estimated population cluster membership proportions that aligned with the patterns of variation suggested for each study population's currently inferred demographic histories. Aboriginal Taiwanese and Philippine samples indicated high East Asian ancestry components, Papua New Guinean and Aboriginal Australians samples were predominantly Oceanian, while other populations displayed cluster patterns explained by the distribution of divergence amongst Melanesians, Polynesians and Micronesians. Genotype data from *Pacifiplex* and *34-plex* tests is particularly well suited to analysis of Australian Aboriginal populations and when combined with Y and mitochondrial DNA variation will provide a powerful set of markers for ancestry inference applied to modern Australian demographic profiles. On a broader geographic scale, *Pacifiplex* adds highly informative data for inferring the ancestry of individuals from Oceanian populations. The sensitivity of *Pacifiplex* enabled successful genotyping of population samples from 50-year-old serum samples obtained from several Oceanian regions that would otherwise be unlikely to produce useful population data. This indicates tests primarily developed for forensic ancestry analysis also provide an important contribution to studies of populations where useful samples are in limited supply.

1. Introduction

Forensic DNA analysis is mainly utilised to identify the donor of an evidential sample by comparing the obtained short tandem repeat (STR) profile with a known reference sample. However, over the last ten years there has been increasing interest in the use of

* Corresponding author at: Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Rúa San Francisco, s/n, 15782 Santiago de Compostela, A Coruña, Spain. Fax: +34 981 56 31 00.
E-mail address: carla.arcanjo.santos@gmail.com (C. Santos).

http://dx.doi.org/10.1016/j.fsigen.2015.10.003
1872-4973/

single nucleotide polymorphisms (SNP) and, in particular, their application to forensic DNA-based intelligence tests used to predict biogeographic ancestry or externally visible characteristics. Such intelligence can be used to generate investigative leads when STR profiling is unsuccessful (a common occurrence in the analysis of highly degraded or low template DNA), no DNA database matches are obtained and in the absence of reliable eyewitness testimony [1]. SNPs are highly abundant in the genome and are well characterized in publicly available databases as a result of global GWAS efforts, which assist in the selection of appropriate markers in populations of interest. SNPs have two other characteristics making them suitable for forensic DNA analysis: (i) the capacity to be amplified in very short fragments (commonly less than 100 bp), critical for successful genotyping of highly degraded samples [2–4]; (ii) base substitutions are stably inherited markers as they have low overall mutation rates, 1.2×10^{-8} per nucleotide per generation [5,6], advantageous for complex relationship testing with ambiguous STR genotypes [7,8]. A small proportion of SNP variants have emerged as particularly informative for ancestry, as they have well differentiated allele frequency distributions that can help distinguish populations. When selecting suitable ancestry informative markers, the degree of divergence between populations and the number of populations that a test seeks to differentiate both have a bearing on the selection process. Using the delta (δ) allele frequency differential metric to identify the most differentiated SNPs centres on pairwise population comparisons [9]. To predict when a SNP is informative for several populations, Rosenberg's informativeness for assignment, *In* [10], or the closely related Shannon's Divergence metric [11] are more informative and appropriate diversity measures, given the multiplex size constraints of forensic SNP sets. Although mitochondrial DNA (mtDNA) and Y-chromosome markers routinely demonstrate informative geographic differentiation, they are less suited to analyzing individuals with admixed parentage and genomic backgrounds (i.e., when recent gene flow has occurred), leading to possible classification errors [12]. For this reason, autosomal SNPs are much less prone to the risk of misinterpretation that can occur when finding an atypical patri- or matri-lineage, unrepresentative of the overall ancestry of an individual [13–15]. Therefore, forensic analysis has adopted a range of autosomal SNP sets that can be used as a stand-alone tool or to provide additional ancestry data complementary to uniparental variation [11,14,16–21].

Studies of worldwide population groups using hundreds of markers generally identify genetic clusters that tend to correspond to a broadly continental distribution of the tested samples [17,22]. Although large sets of randomly selected SNPs enable differentiation of these groups, the constraints of forensic analysis of scant DNA traces requires selection of the most differentiated markers in much smaller panels. Similar conclusions of global differentiation were recently obtained using such small-scale sets [21,23]. However, there remains a gap in the development of a forensically optimised ancestry informative test for the Pacific region.

The genetic patterns of human population variation arose from a series of sequential migrations and bottleneck events. Around 50–70 thousand years ago (KYA), modern humans migrated out of Africa into Asia following a southern coastal route [24–26]. These migrant populations underwent population bottlenecks followed by rapid expansions, creating founder effects that reduced their diversity [27]. The Pacific region was settled in a particularly complex way, reflected in much of the current genetic diversity of this area. Evidence from Lake Mungo suggests modern humans reached Australia 50 KYA [28,29] and were related to individuals first occupying Asia 62-75 KYA, as revealed by Aboriginal Australian whole genome data [29]. Furthermore, Aboriginal Australians are most closely related to Highlanders from Papua New Guinea (PNG) [29]. This is an expected pattern if Australia and PNG are considered parts of a previous single Sahul landmass existing 50 KYA [29]. In contrast, the eastern side of the Pacific region (also known as Remote Oceania) was only settled in the last 3500 years [30]. More detailed overviews of modern human migration into the Pacific region and its current genetic diversity are given in two recent comprehensive reviews [31,32]. To date, most genetic evidence supports the origin of Pacific Islanders in the west followed by a slow migration eastwards. From mtDNA and Y-marker studies, distinct lineage distribution differences are evident in Asia and Near Oceania. Uniparental lineage diversity reduces from west to east going towards Remote Oceania, while lineages present in Remote Oceania can be traced back to both Asia and Near Oceania [31,33,34]. Evidence of sex-bias also exists, with high numbers of Melanesian Y-marker and Asian mtDNA variation in Polynesia [35]. However, Duggan et al. recently indicated that reconstruction of migration history in the Pacific region is far from straightforward [33]. Oceanian population variation is further complicated by indications of a recent relationship between the Indian subcontinent and some Aboriginal Australian populations [36,37]. Lastly, a small number of studies using autosomal markers [38–40] indicate low genetic diversity in Remote Oceania and high diversity in Melanesia, with Polynesian variation closer to East Asia than Melanesia.

This study describes a new multiplex of 29 ancestry-informative SNPs (AIM-SNPs) termed *Pacifiplex* that is focused on Oceanian population variation. The panel comprises 27 autosomal and two X-chromosome SNPs selected to be complementary to an existing 34-SNP ancestry panel [41], so that the combined markers allow ancestry inferences to be made comparing African, European, East Asian and Oceanian origins. Unlike previously mentioned SNP sets [21,23], *Pacifiplex* is a fully optimized and ready-to-use forensic assay that can be analysed using readily available capillary electrophoresis (CE) technology. Although *Pacifiplex* focuses on autosomal variability in Oceania, it complements analysis of Pacific region mtDNA variability [42,43]. We assessed the capability of *Pacifiplex* to infer ancestry using the HGDP-CEPH human diversity sample panel and then compared 1000 Genomes test populations and nine novel Oceanian study populations genotyped for the combined 62 AIM-SNPs.

2. Materials and methods

2.1. Ethical approval for use of samples from Oceania and East Asia

Oceanian study population samples comprised: Northern Territory (NT) Aboriginal Australian individuals obtained by Victoria Police Forensic Services Department; samples from Papua New Guinea, Easter Island, Truk Lagoon, Aboriginal Taiwanese and Western Australia (WA) Aboriginal individuals obtained by George Washington University (GWU); Philippine island samples obtained by University of California San Francisco; East Timor samples obtained by University of Aveiro working in collaboration with National University of East Timor, Dili, East Timor; and Fijian island samples obtained by Fiji Police Forensic Biology and DNA Laboratory. Written informed consent was obtained from donors regarding the use of samples for the characterisation of population variation, with three exceptions These were the GWU samples collected in the fifties plus Fijian and NT Aboriginal Australians, both collected as part of routine forensic casework, with approval given by each institute for use of samples for forensically relevant research purposes. The NT Aboriginal Australian samples have been used in other population studies [42,44]. In all cases, samples were anonymised at the point of collection. George Washington University, University of California, San Francisco and University of Aveiro-National University of East Timor had Institutional Review