Forensic Science International: Genetics 24 (2016) 194–201

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

GlobalFiler® Express DNA amplification kit in South Africa: Extracting the past from the present

Peter Gustav Ristow, Kevin Wesley Cloete, Maria Eugenia D'Amato*

Forensic DNA Laboratory, Department of Biotechnology, University of the Western Cape, Bellville 7535, South Africa

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form 10 June 2016
Accepted 11 July 2016
Available online 13 July 2016

Keywords:
GlobalFiler®
STR
Forensics
Assignment
South Africa
Selection

ABSTRACT

In this study, the GlobalFiler® Express amplification kit was evaluated for forensic use in 541 South African individuals belonging to the Afrikaaner, amaXhosa,[1] amaZulu,[1] Asian Indian and Coloured population groups.

Allelic frequencies, genetic diversity parameters and forensic informative metrics were calculated for each population. A total of 301 alleles were observed ranging between 5 and 44.2 repeat units, 43 were rarely observed partial repeats and seven were novel. The combined match probability (CMP) ranged from 2.21×10^{-26} (Coloured) to 5.21×10^{-25} (AmaZulu), and the combined power of exclusion (CPE) 0.9999999978 (Afrikaaner) to 0.99999999979 (AmaZulu) respectively. No significant departures from Hardy-Weinberg equilibrium (HWE) were observed after Bonferroni correction.

Strong evidence of genetic structure was detected using the coancestry coefficient θ, Analysis of Molecular Variance (AMOVA) and an unsupervised Bayesian clustering method (STRUCTURE). The efficiency of assignment of individuals to population groups was evaluated by applying likelihood ratios with WHICHRUN, and the individual ancestral membership probabilities inferred by STRUCTURE. Likelihood ratios performed the best in the assignment of individuals to population groups. Signs of positive selection were detected for TH01 and D13S317 and purifying/balancing selection for locus SE33. These three loci also displayed the largest informativeness for assignment (I_n) values.

The results of this study supports the use of the GlobalFiler® STR profiling kit for forensic applications in South Africa with the additional capability to predict ethnicity or continental origin of a random sample.

1. Introduction

South Africa is located at the southernmost point of the African continent and the ancestries of its inhabitants have diverse geographic origins. The autochthonous population groups of South Africa are the click-speaking Khoisan, and the Bantus (black Africans). Skeletal evidence indicates the presence of the Khoisan in South Africa dating between 2000 and 4500 years ago [1–3]. Presently, the Khoisan populations constitute a minority of the total South African population. The Bantu-speaking groups which constitute 79% of the total population [4], entered South Africa via the Eastern coastal routes prior to the fifth century [5]. Both groups have historically been subject to racial killing by indigenous and foreign populations. In 1652, Dutch Europeans moored on the shores of South Africa to set up a refreshment settlement for the Dutch East India Company. Following the Dutch colonists, inhabitants from several other European countries (the majority of British, French and German descent) settled in South Africa and their descendants today constitute 8.9% of the overall South African inhabitants [4]. The Afrikaaners are Southern African, descendants of Dutch settlers who speak the Dutch-derived language: Afrikaans. The onset of colonisation resulted in the trade of slaves from Central Africa, Eastern Africa and South East Asia [6].

Abbreviations: AIMs, ancestry informative markers; AMOVA, analysis of molecular variance analysis; BIC, Bayesian inference of clusters; CMP, combined match probability; CPE, combined power of exclusion; CPI, combined paternity index; DC, discrimination capacity; FDR, false discovery rate; He, expected heterozygosity; HGDP, human genome diversity project; Ho, observed heterozygosity; HWE, Hardy Weinberg equilibrium; I_n, informativeness for assignment; InDels, insertions and deletions; $Log_{10}(L)$, logarithm of likelihood ratio; PE, power of exclusion; PIC, polymorphic information content; RMP, random match probability; STR, short tandem repeats; TPI, typical paternity index; R_{ST}, *R*-statistic; F_{ST}, *F*-statistic.

* Corresponding author at: Forensic DNA Laboratory, Department of Biotechnology, University of the Western Cape, Bellville 7535, South Africa.
E-mail addresses: medamato@uwc.ac.za, medamato@gmail.com (M.E. D'Amato).

[1] Prefixes ama- and isi- are used within Bantu populations to indicate population group and language respectively.

http://dx.doi.org/10.1016/j.fsigen.2016.07.007
1872-4973/

Commencing in 1860 and continuing until 1902, the local slave force in Durban was augmented by the influx of indentured labourers shipped from India [7] which at that time was under British rule. The South African Indian population presently constitute 2.5% of South Africa's inhabitants [4]. The Coloured populations are the resultant progeny of several centuries of admixture between indigenous South African and immigrant population groups [8–10], and constitute 8.9% of the South African population [4]. The admixture present within the Coloured subpopulations is not homogenous; however, they display a large European ancestral contribution with varying contributions from Khoisan, Bantu and Asian Indian ancestries [11,12].

The AmpFlSTR® Identifiler® Plus (Thermo Fisher Scientific) autosomal STR genotyping kit is routinely used in forensics and paternity casework in South Africa [13–15]. This kit displays a low random match probability for the Coloured, Bantu, Khoisan populations from South Africa [13,16]. This study however utilises the GlobalFiler® kit which implements a 6-dye system to scrutinise 21 autosomal STRs and 3 gender specific markers (D3S1358; vWA; D16S539; CSF1PO; TPOX; D8S1179; D21S11; D18S51; D2S441; D19S433; TH01; FGA; D22S1045; D5S818; D13S317; D7S820; SE33; D10S1248; D1S1656; D12S391; D2S1338; Y InDel; DYS391; and Amelogenin). This kit showed a high level of discrimination within American [17], Japanese [18], Mexican [19], Southern Portuguese [20], and the United Arab Emirates [21] populations.

In this study, the GlobalFiler® Express DNA amplification kit was evaluated for use in forensic applications in South Africa. The ascertainment bias in the selection of STR loci for forensic applications implies that lower genetic differentiation is expected than with randomly selected STRs [22], however Algee-Hewitt et al. [23] demonstrated otherwise. The main contributors to genetic differentiation between populations using forensic markers are thus expected to be historical events and demographic processes leading to genetic drift. The evaluation of population structure and ancestry information with STR forensic markers has been extensively reported [22–29]. Given the complexity of the ethnic composition in South Africa, the population structure was evaluated using different methods: summary statistics (coancestry coefficient θ, R_{ST} and AMOVA) and STRUCTURE.

Furthermore, we investigated the possibility of loci being subjected to selection processes using the traditional and hierarchical FDIST2 methods and a Bayesian approach. The final component of the study was to investigate the possibility of the assignment of individuals to population groups and this was evaluated using the log likelihood ratios (Log_{10}(L)) of population probabilities with WHICHRUN, as well as from the ancestral components with STRUCTURE. The implications of our results for forensic identification in South Africa are discussed.

2. Materials and methods

2.1. Samples and DNA purification

The sampling procedure was approved by the Ethics Committee of the University of the Western Cape (10/3/39). Buccal swabs were collected from 541 consenting individuals belonging to five South African population groups, namely Afrikaaner (n = 106), Asian Indian (n = 102), Cape Coloured (n = 113), amaXhosa (n = 120) or amaZulu (n = 100). The Cape Coloured population group is referred to as "Coloured" throughout. Genomic DNA was extracted from cotton swabs using a modified salting out method [30]. The quality and quantity of the extracted DNA was estimated using a Nanodrop ND-2000 spectrophotometer (Thermo Fisher Scientific).

2.2. Amplification and genotyping

Amplification of genomic DNA (2 ng) was performed as per the GlobalFiler® Express DNA amplification kit's protocol [31] at half reaction volume and cycled in a Veriti thermocycler (Thermo Fisher Scientific). Liz 600 internal size standard (Thermo Fisher Scientific) was run alongside amplified products, resolved on an ABI 3500 (Thermo Fisher Scientific), and raw data was captured with 3500 Series Data Collection Software 2 (Thermo Fisher Scientific). STR genotyping was completed using GeneMapper® IDX v1.4 software (Thermo Fisher Scientific). The above work was conducted as per the quality control guidelines outlined by Schneider [32]. The individuals displaying tri-allelic patterns or rare variants were subject to reamplification for confirmation and tri-alleles were excluded from further statistical analysis.

2.3. Statistical analysis

2.3.1. Genetic diversity and forensic parameters

Genetic diversity parameters including allele frequency, observed (Ho) and expected (He) heterozygosity were estimated using Gstudio [33], in R v3.1.3 (http://www.r-project.org/), and Hardy-Weinberg equilibrium using Arlequin v3.5 [34]. Bonferroni correction [35] was applied to the probability of HWE to determine significant deviations. Forensic metrics, namely: random match probability (RMP), combined match probability, discrimination capacity (DC), polymorphic information content (PIC) [36], typical paternity index (TPI), combined paternity index (CPI), power of exclusion (PE), and combined power of exclusion (CPE) were calculated using Gstudio [33], in R v3.1.3 with an in-house script.

2.3.2. Population structure

To evaluate the presence of genetic structure between population groups, several statistical methods were implemented. Coancestry between populations were calculated implementing Weir and Cockerham's θ [37], at a 5% significance level over 10 000 permutations as implemented in Genetix v4.05 [38]. Hierarchical AMOVA was performed to quantify the partition of genetic variation of the South African population groups placed into four groups (Afrikaaner, Asian Indian, Coloured and Bantu which is constituted by the amaXhosa and amaZulu groups) with Arlequin v3.5 [34], using R_{ST} like distance matrix and running 50 000 permutations.

The Bayesian clustering method in STRUCTURE v.2.3.4 [39], was applied in an unsupervised manner implementing *admixture* and *correlated allele frequencies* parameters, with 600 000 repeats after a burn-in of 100 000. Three iterations for each K from 1 to 6 were run and the optimal number for K was estimated using Evanno's ΔK method [40]. The iterations for each K were merged using CLUMPP [41], and the ancestral components plotted with Distruct [42], using the online package Clumpak [43].

2.3.3. Ancestry informative loci

The contribution of each locus to population differentiation was evaluated with the frequentist method of Rosenberg I_n [44], calculated using Infocalc v1.1 [45], excluding the admixed Coloured population.

2.3.4. Investigation of loci under processes of selection

To investigate whether any loci were under processes of selection, we applied the coalescence-based approach to identify F_{ST} outliers [46], implemented in Lositan [47], and in an hierarchical manner [48], as implemented in Arlequin v3.5 [34]. Additionally, a Bayesian approach was used as implemented in BayeScan v2.1 [49]. The population structure used for all selection tests were determined by Beaumont & Nichols' suggestion for the

FDIST2 method [46], and the hierarchical equivalent version by Excoffier et al. [48]. The criteria stipulate using a large number of markers (≥20), with large heterozygosity values and large sample sizes from well-defined population groups. Therefore, only three population groups (Afrikaaner, Asian Indian and Bantu) were used for the selection tests.

Analysis with Lositan was conducted as suggested by Antao et al. [47], calculating a neutral F_{ST} by removing probable non-neutral loci after calculating an initial mean F_{ST} over 1 000 000 iterations, assuming a stepwise-mutation model. The calculated mean neutral F_{ST} was then used in a second run, including all loci to identify loci deviating from neutrality expectations, using a 0.99 confidence interval and a false discovery rate (FDR) of 0.1. The hierarchical island model of Excoffier et al. [48], was implemented assuming 20 groups, 200 demes over 20 000 simulations and loci were considered under selection when the R_{st} value displayed a p-value below 5%. The Bayesian approach to detect loci departing from neutrality was implemented in BayeScan v2.1 [49]. BayeScan estimates the locus-specific (α) and population-specific (β) components of F_{ST}, the probability of α significantly deviating from zero, the posterior odds (PO) as the ratio of posterior probabilities between models (assuming selection, and neutrality), and the q-values per locus as the minimum FDR at which this locus may become significant. The program was run using equal weighting for each of the three population groups. Model parameters were tuned using 20 pilot runs with 5000 steps. Following this, a burn-in of 5000 iterations was followed by 100 000 iterations with sample sizes of 5000 and ten thinning intervals. The prior odds to neutral model were set to ten and the Log_{10}(PO) values were evaluated at a 1% FDR.

2.3.5. Assignment

Individual assignments were evaluated by applying a Log likelihood approach and a cluster inference method assuming two (African or non-African) or three (European, Asian Indian or African) population groups. For the first evaluation, all individuals were jackknifed and likelihoods of their genotype probabilities were estimated given the allele frequencies for the assumed scenarios of two or three population groups (see above) using Log_{10}(L) of assignment ratios as in [50]. A cut off value of 0.602 ($Log_{10}4$) was chosen as FDR. For the cluster analysis method, criteria similar to that of Phillips et al. [24,51] was implemented whereby an assignment cut-off of 0.7 for cluster membership probability was implemented. Individuals were considered "correctly assigned" when the assignment corresponded with the self-declared population group (Log_{10}(L) > 0.602; cluster membership probability > 0.7). The error rate of assignment was calculated as the fraction of individuals whose assigned population differed from the self-declared population. Individuals not matching these criteria were considered unassigned (admixed).

3. Results and discussion

3.1. Rare variant, and off-ladder alleles

Alleles falling within virtual bins were classified as "rare variants", or when no bin was present they were classified as "off-ladder" variants. A total of 43 rare variant alleles and seven novel off-ladder alleles were observed for GlobalFiler®'s allelic ladder (BIN set v1.2) in this study (Supplementary Table 1). The high frequency of the observed rare variant alleles in the indigenous South African population groups is likely a result of the dearth of genetic population data available for South Africa for these loci. The observed novel off-ladder alleles were all partial repeats and observed in the Afrikaaner (D22S1045 allele 13.2), Coloured (D1S1656 allele 16.1) and Bantu (SE33 alleles 6.2, 21.3, 22.3 and 24.1 and D8S1179 allele 14.3) population groups. The allelic frequencies of all loci for all populations are shown in Supplementary Table 2A and individual population groups in Supplementary Table 2B–F.

3.2. Tri-allelic patterns

Tri-allelic patterns have been intensively researched and the various patterns have been well described [52–54]. Presently two patterns are known: type 1 (three imbalanced alleles) and type 2 (one imbalanced allele peak). Tri-allelic genotypes have also been identified to be of significant importance in paternity as well as forensic cases [53,55,56]. Tri-allelic genotypes, while a rare variant, do occur frequently, with 388 variants reported for autosomal STRs as of 09/11/2015 in the NIST STR database (http://www.cstl.nist.gov/biotech/strbase/tri_tab.htm) with TPOX (19 variants) and FGA (40 variants) displaying the most observations and several hundred profiles from previous studies [15,53,57,58]. The genotypes of six individuals who displayed tri-allelic patterns for loci D1S1656, TPOX and vWA are shown in Table 1.

Type 2 patterns were the only variants observed in this study (Supplementary Fig. 1). The tri-allelic patterning of the TPOX locus is the best characterised and studied with the extra allele hypothesised as being a translocation of allele 10 onto the X-Chromosome [15,53]. All individuals displaying tri-allelic patterns in this study presented this allele. Frequencies of tri-alleles, for the locus TPOX, below 0.006 were observed in non-African regions [53–55,57,59]. Greater frequencies (0.004–0.045) have been observed in African populations [15,60–62], and the Dominican republic [58]. Interestingly, Muro et al. [63], and Takeshita et al. [64], reported no tri-allelic variants for the Ovambo population in Namibia. A frequency increment by two orders of magnitude was reported for African-Americans [54], suggesting this pattern might occur in higher frequency in Western Africa since it is the predominant ancestry of African Americans [12].

The frequency of tri-alleles for the locus TPOX in the current study is in agreement with previous studies [51,56–58], with all observed tri-alleles originating in the indigenous South African populations: 0.9% and 1.5% for the Coloured and Bantu population groups respectively. Recent studies [15,53,58] have shown that due to the TPOX translocation onto the X-chromosome, males will transmit an extra allele to their daughters, while tri-allelic females have a 50% probability to transmit their extra X-linked allele to their progeny. Lane et al. [15], also suggested that the translocation of the extra TPOX allele with the X-chromosome occurred prior to the Bantu expansion based on the frequencies of tri-allelic patterns observed between the Western and Eastern Bantu populations. Therefore, taking into account the method of transmittance and estimated time of occurrence [15] we hypothesise that the driving force behind the high frequency of tri-allelic TPOX genotypes in African populations is the cultural practice of polygamy.

Table 1
Tri-allelic patterns observed in South African populations. The number of tri-alleles observed per locus, the pattern observed and population group.

Locus	N° observed	Alleles	Type	Population
D1S1656	1	13, 14, 15	II	Coloured
TPOX	4	9, 10, 11	II	Coloured
		8, 10, 11	II	AmaXhosa
		6, 8, 10	II	AmaZulu
		6, 9, 10	II	AmaZulu
vWA	1	14, 17, 20	II	Asian Indian

3.3. Genetic diversity and forensic statistic metrics

The genetic diversity parameters for all five populations tested are shown in Supplementary Table 3. No locus showed significant deviation from Hardy-Weinberg equilibrium after Bonferroni correction [35] (Supplementary Table 3). The computed forensic metrics are shown in Supplementary Table 4. The least and most polymorphic loci were TPOX and SE33 displaying PIC values of 0.56 and 0.94 respectively. The probability of obtaining a random match between individuals (CMP) ranged between 2.21×10^{-26} (Coloured), and 5.212×10^{-25} (amaZulu), and the CPE ranged from 0.9999999978 (Afrikaaner) to 0.99999999979 (amaZulu). The large CPE and small CMP values supports the use of GlobalFiler® in individual identification within South African populations. In comparison with Identifiler® Plus, GlobalFiler® achieves four and seven larger orders of magnitude for CPE and CMP respectively [13,14,16,20]. Additionally, GlobalFiler® displayed large likelihoods for CPI ranging from 3.41×10^{8} (Afrikaaner) to 5.3×10^{9} (AmaZulu) indicating that this kit will be of value in paternity cases.

3.4. Population structure

In this section, we investigated the degree of differentiation between populations. To investigate the genetic heterogeneity between populations, pairwise θ values for coancestry were calculated and were found to be significant for all populations at the 5% level except between the amaXhosa and amaZulu groups (Table 2).

This coancestry result can be explained by the two groups belonging, linguistically, to the same Nguni subgroup of the Bantoid macrogroup [65], and they are also geographically in close proximity. Genetic similarity between the amaXhosa and amaZulu has also previously been identified using both Profiler plus™ [28], and Investigator DIPplex® kits [66] using the same individuals in the latter. These two Nguni population groups were therefore grouped into one large group named "Bantu" for hierarchical AMOVA between the Afrikaaner, Asian Indian, Coloured and Bantu population groups. AMOVA indicated that 93.36% of the variation is contained within populations, 6.3% among groups, and 0.34% between populations within groups (Table 3). The presence of significant population structure between the four groups was also indicated by a large R_{ST} value (0.0664, p-value = 0.00000).

The results of STRUCTURE (Fig. 1), display the genetic substructure between the Afrikaaner, Asian Indian, Coloured and Bantu populations and also provides evidence of admixture within the Coloured population (Fig. 1) as previously shown by [11,12,66]. The optimal K was determined to be three using the Evanno method indicating that each cluster was representative of one of three continental groups (European, Asian Indian or African). Individuals similarly appear to be grouped as African or non-African when K = 2 (Fig. 1). The similarities between the amaXhosa and amaZulu populations are shown in all Ks > 1 with the majority of individuals from both population groups getting assigned to the same cluster(s).

The major ancestral contributor to the Coloured population, in this study, was observed as originating from Asian Indian (K = 3) and non-African (K = 2) ancestries (Supplementary Fig. 2). The Coloured population has previously been shown to exhibit high levels of admixture with the main contributors differing between studies which used SNPs [11], microsatellites and InDels [12], and InDels [66].

Table 2
Pairwise population θ values for coancestry. Significant p-values (<0.05) are indicated with an asterisk (*).

Population	Coloured	AmaXhosa	Afrikaaner	Asian Indian	AmaZulu
Coloured	0	0.01194*	0.0089*	0.00617*	0.0131*
AmaXhosa		0	0.02914*	0.02958*	0.0001
Afrikaaner			0	0.01732*	0.02867*
Asian Indian				0	0.02761*
AmaZulu					0

3.5. Ancestry informative loci

Several parameters can be implemented to identify loci which drive differentiation between populations such as absolute allele frequency differences (δ), F statistic (F_{ST}), Fisher Information Curvature Criterion (*FIC*), Shannon Information Content (*SIC*) and Informativeness for assignment (I_n). Ding et al. [67] showed that I_n and F_{ST} are highly correlated, and provided conservative values for the best performance of the above-mentioned parameters for binary markers. I_n was therefore implemented in this evaluation of ancestry informative STR loci. Rosenberg et al. [44] and Listman et al. [68] state that dinucleotide STRs are the most informative due to their stability. However, GlobalFiler® or any other forensic kit do not contain dinucleotide repeats. The most informative (I_n) locus was also the most polymorphic locus SE33 in this study and [69]. The 5 most informative markers identified in this study were SE33, D1S1656, D21S11, TH01 and D8S1179 (Supplementary Table 5). The largest I_n was 0.168 and suggests the individual markers are not optimal for inference of ancestry.

3.6. Selection

For the analysis of selective pressure over these loci, we applied different methods based on different assumptions. FDIST2 [46], is based on the identification of outliers from the simulated null distribution of F_{ST} given an infinite island model of population structure, with demes of equal sizes and migration rates. Excoffier et al. [48], improved the FDIST2 method by incorporating a hierarchical island model, because of the trend of hierarchical population structure and bottlenecks to render false positives. The Bayesian method in BayeScan models F_{ST} allowing for locus and population effect, permits unequal gene flow between demes differentiated from a common ancestral population and is robust under different demographic scenarios and small sample sizes [49]. The performance of these methods has been evaluated using simulation studies [70], showing higher Type I and Type II error rates with the hierarchical FDIST2 method, and the lowest Type I error with BayeScan and FDIST2.

It is worth noting that the results of these tests may not indicate a direct effect on the microsatellites but a hitchhiking effect from the genomic regions they are linked to. Discrepant results among different methods might originate from deviations from the

Table 3
AMOVA results of four South African population groups (Afrikaaner, Asian Indian, Coloured and Bantu).

Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	Fixation indices (p-values)
Among groups	3	492304.6	581.732 Va	6.3	R_{CT} 0.06302 (0.10003)
Among populations within groups	1	15273.753	30.95761 Vb	0.34	R_{SC} 0.00358 (0.07544)
Within populations	1065	9179092.253	8618.865 Vc	93.36	R_{ST} 0.06637 (0.00000)
Total	1069	9686670.606	9231.55558		

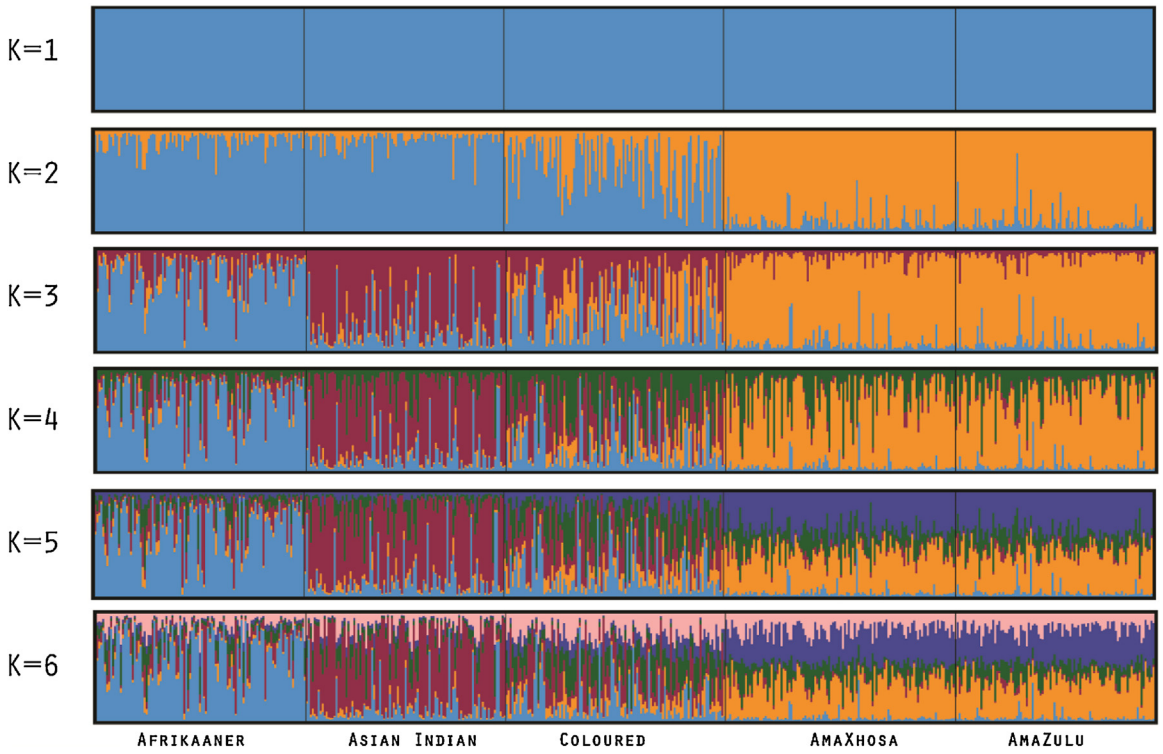

Fig. 1. STRUCTURE results for GlobalFiler® genotypic data from South African population groups (Afrikaaner, Asian Indian, Coloured, amaXhosa and amaZulu). The individual ancestral components were obtained by merging results of three iterations for K = 1–6.

assumptions these methods are based on, or limited power to detect low levels of selection [70]. Consistent results among the methods lend support to the hypothesis of the loci being under selective pressure.

Our results are compatible with the hypothesis that locus TH01 (classic outlier test) and D13S317 (hierarchical test) are affected by positive selection (Supplementary Fig. 3A and B). The locus TH01 is located within intron 1 of the human tyrosine hydroxylase gene on chromosome 11 (11p15.5), the protein of which is responsible for the conversion of tyrosine to L-Dopamine. The locus D13S317 is located on the long arm of chromosome 13 (13q31.1) and is located > 100 kb from the nearest gene coding region. The Bayesian method displayed a $\text{Log}_{10}(\text{PO}) > 2$ with a negative α value for locus SE33 which is indicative of balancing/purifying selection (Supplementary Fig. 3C). Locus SE33 is located on the long arm of chromosome 6 (6q14) and codes a β-actin related pseudogene which is believed to be functionless [71]. At the less stringent FDR of 5%, loci vWA and FGA showed "strong" and "very strong" signs of purifying/balancing selection respectively based on the scales of Jeffreys [72] and Foll [73]. The loci vWA and FGA are found in introns of functional genes. A summary of all three methods for all loci is shown in Supplementary Table 6. It is notable that the three loci with the lowest F_{ST}, closer to the 'balancing selection' zone defined by Lositan, are identified by BayeScan as being under purifying/balancing selection, whereas TH01 shows the highest averaged posterior F_{ST}. As expected, the loci identified here as being under positive selection processes are among the loci with highest ancestry information content, discussed in the previous section.

3.7. Assignment

The assignment of individuals to population groups is important for the identification of human remains in mass disasters [24,74], identification of possible criminal suspects [24,29], biobanking [75] and for evaluation of the weight of evidence [76,77]. The validity of the assignment of individuals to population groups using GlobalFiler® was evaluated using likelihood ratios in WHICHRUN, and ancestral proportions in STRUCTURE (Supplementary Table 7). The assignment tests were investigated assuming the assignment to either two (African and non-Africa in Supplementary Table 7A) or three (European, Asian Indian and African in Supplementary Table 7B) population groups.

For the assignment tests assuming two population groups (Supplementary Table 7A), both methods displayed large rates of correct assignment (99.52–100%), with STRUCTURE being the most efficient in detecting admixture (unassigned individuals). A similar trend was observed for the analyses assuming three population groups (Supplementary Table 7B), with both methods showing large rates of correct assignment. No cross-assignment was observed between African and European population groups (Supplementary Table 7B) and WHICHRUN was also the most efficient in distinguishing individuals with Indian ancestry from those with European ancestry. The similarities between the two methods are observed for both K = 2 and 3 as shown in Fig. 2A and B respectively.

The negligible error rate of assignment (0.48%) when assuming three population groups was due to a single Bantu individual being assigned to the Asian Indian population. This individual self-identified as an amaZulu who was originally from the Kwa-Zulu Natal province, which historically had a large influx of Asian Indian slaves. This type of cross-assignment error was also observed in the work of Phillips et al. [51] with forensic STR markers, and Londin et al. [75] with ancestry-informative STRs. Hefke et al. [66] showed similar population substructure and assignment using STRUCTURE with larger cross-assignment error rates (<4%) between African and non-African population groups when assuming three population groups. The error rate of assignment in this study, however low, may exist due to the recent admixture between the population groups, or more likely due to the suboptimal efficiency of these genetic markers for population assignment. This is because these

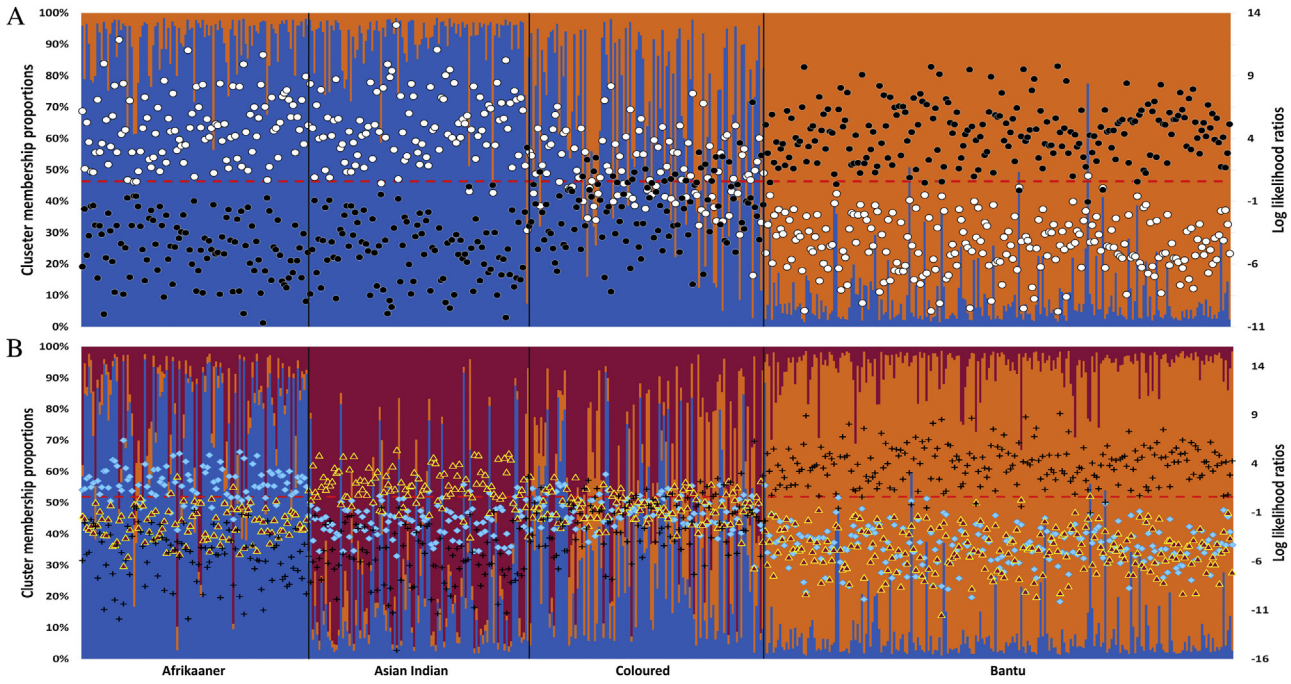

Fig. 2. A. The assignment ratios of individuals to groups using likelihood ratios, assuming two and three population groups, is plotted against cluster membership probabilities from STRUCTURE for K = 2 (A) and 3 (B) respectively. STRUCTURE assigned ancestral components to clusters with either non-African (Blue bars) or African (Orange bars) ancestry. The $Log_{10}(L)$ assignment ratios between non-African/African (white dots) and Africa/non-African (black dots) were calculated by jackknifing individuals from non-African (Afrikaaner, Asian Indian) and African population groups. The Coloured individuals were assigned to either of the two population groups using likelihood ratios. The assignment of an individual to a population group was implemented at a 4 times more likely cut-off value (red dashed line). Individuals with assignment ratios below $Log_{10}(L) < 0.602$ (red dashed line) were deemed admixed. B. The assignment ratios of individuals to groups using likelihood ratios, assuming three population groups, is plotted against the cluster membership probabilities from STRUCTURE for K = 3. STRUCTURE assigned ancestral components to clusters with either European (Blue bars), Asian Indian (Maroon bars) or African (Orange bars) ancestry. $Log_{10}(L)$ ratios (European/other – light Blue diamond, Asian Indian/other – Yellow/maroon triangles and Bantu/other – black plus) were calculated by jackknifing individuals from Afrikaaner, Asian Indian and Bantu population groups. The Coloured individuals were assigned to either of the three population groups using likelihood ratios. The assignment of an individual to a population group was implemented at a 4 times more likely cut-off value (red dashed line). Individuals with all assignment ratios below the red dashed lines ($Log_{10}(L) < 0.602$) were not assigned and deemed admixed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

markers have not been selected for their ancestry information content, as observed by Barnholtz-Sloan et al. [78].

In summary, the advantages of using WHICHRUN over STRUCTURE for population assignment are: higher rates of assignment, lower error rates and a greatly reduced computational time.

4. Conclusion

This study has thoroughly evaluated the GlobalFiler® Express DNA amplification kit for use in forensics applications within South African. The observation of several off-ladder alleles in a small subset of a majority population of South Africa underlines the need for further investigation into African polymorphisms. The kit showed increased levels of discrimination capacity and lower probability of random matches for African populations than with the Identifiler® Plus STR kit. In this study, GlobalFiler® also showed large paternity index values for South African populations indicating that the kit is a valuable asset in paternity testing.

In addition to the above, significant genetic substructure between South African populations was also identified. This differentiation appears to be driven by selection processes on highly informative markers. This observation can be beneficial in forensics investigations as the possibility of population group prediction is possible by evaluating the ancestry informative content of the autosomal STR markers. The large correct assignment rate and almost negligible errors when assuming two population groups (<0.5%) makes GlobalFiler® an important tool for forensic analysts. Furthermore, several of the loci containing the highest ancestry informative content, produce amplicons below 220 bp. In cases of degraded human remains or mass disasters, the profiles obtained with this kit could be used for group prediction applying a simple likelihood ratio method. GlobalFiler® is therefore highly recommended for human identification and paternity testing within South Africa.

Conflict of interests

The authors declare that there are no conflicts of interest. This work followed the guidelines for publication of population data as requested by the journal [79,80].

Acknowledgements

We would like to extend our sincere thanks and appreciation to Applied Biosystems (Thermo Fisher Scientific, South Africa) for the generous donation of the GlobalFiler® Express DNA amplification kits used in this study. Additionally, we would also like to acknowledge the National Research Foundation of South Africa (NRF) (Grant UID: 88640) for funding Peter Ristow during this study, as well as NRF Incentive Funding for Rated Researchers (IFRR) and UWC Research grants to MED for running costs. Finally, we would like to thank Jacolien Volschenk for proof-reading this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.07.007.

References

[1] A.G. Morris, H. Anja, E.K.F. Chan, A.B. Smith, V.M. Hayes, First ancient mitochondrial human genome from a pre-pastoralist Southern African, 6 (2010) 1–15. 10.1093/gbe/evu202.

[2] J. Orton, Rescue excavations at Diaz Street midden, Saldanha Bay, South Africa, Azania Archaeol. Res. Africa. 44 (2009) 107–120, doi:http://dx.doi.org/10.1080/00671990902795822.

[3] J. Orton, J. Compton, A reworked mid-Holocene lithic assemblage at Dunefield midden 1, Elands Bay, South Africa, South Afr. Archaeol. Bull. 61 (2006) 90–95.

[4] Statistics South Africa (2015). http://www.statssa.gov.za/newsletters/Fieldworker_January_2011.pdf (accessed 01.04.15).

[5] T.N. Huffman, General history of Africa, Gen. Hist. Africa, vol. 3, UNESCO general history of Africa (abridged) (v. 3) Press, University of California, California, 1992, pp. 664–680.

[6] F. Welsh, A History of South Africa, Harper Collins, London, 2000.

[7] S. Bhana, Indentured Indians in Natal, 1860–1902: a study based on ship's lists, University of Kwa-Zulu Natal, 1987. http://scnc.ukzn.ac.za/doc/SHIP/Bhana_Indentured_Indians_Natal_Study_based_Ships_list.pdf.

[8] S. Balson, Children of the Mist: the lost tribe of South Africa, first ed., Interactive Presentations Pty Ltd, Australia, 2007.

[9] K. Schoeman, The Griqua captaincy of Philippolis, 1826-1861, first ed., Protea Book House, Pretoria, 2002.

[10] J.M. Greeff, Deconstructing Jaco: genetic heritage of an Afrikaner, Ann. Hum. Genet. 71 (2007) 674–688, doi:http://dx.doi.org/10.1111/j.1469-1809.2007.00363.x.

[11] D.C. Petersen, O. Libiger, E.A. Tindall, R.A. Hardie, L.I. Hannick, R.H. Glashoff, et al., Complex patterns of genomic admixture within Southern Africa, PLoS Genet. 9 (2013) 10–13, doi:http://dx.doi.org/10.1371/journal.pgen.1003309.

[12] S.A. Tishkoff, F.A. Reed, F.R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, et al., The genetic structure and history of Africans and African Americans, Science 324 (2009) 1035–1044, doi:http://dx.doi.org/10.1126/science.1172257.

[13] A. Lucassen, K. Ehlers, P.J. Grobler, A.L. Shezi, Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa, Int. J. Legal Med. 128 (2014) 275–276, doi:http://dx.doi.org/10.1007/s00414-013-0898-4.

[14] A.B. Lane, STR null alleles complicate parentage testing in South Africa, South Afr. Med. J. 103 (2013) 1004–1008, doi:http://dx.doi.org/10.7196/SAMJ.7067.

[15] A.B. Lane, The nature of tri-allelic TPOX genotypes in African populations, Forensic Sci. Int. Genet. 2 (2008) 134–137, doi:http://dx.doi.org/10.1016/j.fsigen.2007.10.051.

[16] C.M. Schlebusch, H. Soodyall, M. Jakobsson, Genetic variation of 15 autosomal STR loci in various populations from Southern Africa, Forensic Sci. Int. Genet. 6 (2012) e20–e21, doi:http://dx.doi.org/10.1016/j.fsigen.2010.12.013.

[17] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. population data for 29 autosomal STR loci, Forensic Sci. Int. Genet. 7 (2013) e82–e83, doi:http://dx.doi.org/10.1016/j.fsigen.2012.12.004.

[18] K. Fujii, Y. Iwashima, T. Kitayama, H. Nakahara, N. Mizuno, K. Sekiguchi, Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population, Leg. Med. (Tokyo) 16 (2015) 234–237, doi:http://dx.doi.org/10.1016/j.legalmed.2014.03.007.

[19] B. Ramos-González, J.A. Aguilar-Velázquez, M. de, L. Chávez- Briones, J.R. Delgado-Chavarría, E. Alfaro-Lopez, H. Rangel-Villalobos, Population data of 24 STRs in Mexican-Mestizo population from Monterrey, Nuevo Leon (Northeast, Mexico) based on Powerplex Fusion and GlobalFiler kits, Forensic Sci. Int. Genet. 21 (2015) 15–17, doi:http://dx.doi.org/10.1016/j.fsigen.2015.12.004.

[20] C. Almeida, T. Ribeiro, A.R. Oliveira, M.J. Porto, J. Costa Santos, D. Dias, et al., Population data of the GlobalFiler® Express loci in South Portuguese population, Forensic Sci. Int. Genet. 19 (2015) 39–41, doi:http://dx.doi.org/10.1016/j.fsigen.2015.06.001.

[21] O. Ali Alhmoudi, R.J. Jones, G.K. Tay, H. Alsafar, S. Hadi, Population genetics data for 21 autosomal STR loci for United Arab Emirates (UAE) population using next generation multiplex STR kit, Forensic Sci. Int. Genet. 19 (2016) 190–191, doi:http://dx.doi.org/10.1016/j.fsigen.2015.07.009.

[22] N.M. Silva, L. Pereira, E.S. Poloni, M. Currat, Human neutral genetic variation and forensic STR data, PLoS One 7 (2012) e49666, doi:http://dx.doi.org/10.1371/journal.pone.0049666.

[23] B.F.B. Algee-Hewitt, M.D. Edge, J. Kim, J.Z. Li, N.A. Rosenberg, Individual identifiability predicts population identifiability in forensic microsatellite markers, Curr. Biol. (2016) 1–8, doi:http://dx.doi.org/10.1016/j.cub.2016.01.065.

[24] C. Phillips, L. Prieto, M. Fondevila, A. Salas, A. Gómez-Tato, J. Álvarez-Dios, et al., Ancestry analysis in the 11-M Madrid bomb attack investigation, PLoS One 4 (2009) e6583, doi:http://dx.doi.org/10.1371/journal.pone.0006583.

[25] C. Phillips, M. Gelabert-Besada, L. Fernandez-Formoso, M. García-Magariños, C. Santos, M. Fondevila, et al., New turns from old STaRs: enhancing the capabilities of forensic short tandem repeat analysis, Electrophoresis (2014) 3173–3187, doi:http://dx.doi.org/10.1002/elps.201400095.

[26] L. Pereira, F. Alshamali, R. Andreassen, R. Ballard, W. Chantratita, N.S. Cho, et al., PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile, Int. J. Legal Med. 125 (2011) 629–636, doi:http://dx.doi.org/10.1007/s00414-010-0472-2.

[27] M. Graydon, F. Cholette, L.K. Ng, Inferring ethnicity using 15 autosomal STR loci-comparisons among populations of similar and distinctly different physical traits, Forensic Sci. Int. Genet. 3 (2009) 251–254, doi:http://dx.doi.org/10.1016/j.fsigen.2009.03.002.

[28] A.B. Lane, H. Soodyall, S. Arndt, M.E. Ratshikhopha, E. Jonker, C. Freeman, et al., Genetic substructure in South African Bantu-speakers: evidence from autosomal DNA and Y-chromosome studies, Am. J. Phys. Anthropol. 119 (2002) 175–185, doi:http://dx.doi.org/10.1002/ajpa.10097.

[29] X. Fosella, F. Marroni, S. Manzoni, A. Verzeletti, F. De Ferrari, N. Cerri, et al., Assigning individuals to ethnic groups based on 13 STR loci, Int. Congr. Ser. 1261 (2004) 59–61, doi:http://dx.doi.org/10.1016/S0531-5131(03)01699-6.

[30] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, Nucleic Acids Res. 16 (1988) 1215, doi:http://dx.doi.org/10.1093/nar/16.3.1215.

[31] L.T. Corporation, GlobalFiler™ Express PCR amplification kit user guide, Life Technol. (2012) Publication number 4476609, Revision A..

[32] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238–243, doi:http://dx.doi.org/10.1016/j.forsciint.2006.06.067.

[33] R.J. Dyer, Analyses and functions related to the spatial analysis of genetic marker data (2014) https://github.com/dyerlab/gstudio.

[34] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47–50, doi:http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x.

[35] B.S. Weir (Ed.), Genetic data analysis. Methods for discrete population genetic data, first ed., Sinauer Associates, Sunderland, Mass, 1990.

[36] D. Botstein, R.L. White, M. Skolnick, R.W. Davis, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, Am. J. Hum. Genet. 32 (1980) 314–331.

[37] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution (N. Y). 38 (1984) 1358–1370.

[38] K. Belkhir, P., Borsa, L., Chikhi, N., Raufaste, F. Bonhomme, GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations (1996).

[39] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.

[40] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620, doi:http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x.

[41] M. Jakobsson, N.A. Rosenberg, CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, Bioinformatics 23 (2007) 1801–1806, doi:http://dx.doi.org/10.1093/bioinformatics/btm233.

[42] N.A. Rosenberg, Distruct: a program for the graphical display of population structure, Mol. Ecol. Notes 4 (2004) 137–138.

[43] 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 (2015) 1179–1191, doi:http://dx.doi.org/10.1111/1755-0998.12387.

[44] N.A. Rosenberg, L.M. Li, R. Ward, J.K. Pritchard, Informativeness of genetic markers for inference of ancestry, Am. J. Hum. Genet. 73 (2003) 1402–1422, doi:http://dx.doi.org/10.1086/380416.

[45] N.A. Rosenberg, Algorithms for selecting informative marker panels for population assignment, J. Comput. Biol. 12 (2005) 1183–1201, doi:http://dx.doi.org/10.1089/cmb.2005.12.1183.

[46] M.A. Beaumont, R.A. Nichols, Evaluating loci for use in the genetic analysis of population structure, 263 (1996) 1619–1626.

[47] T. Antao, A. Lopes, R.J. Lopes, A. Beja-Pereira, G. Luikart, LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method, BMC Bioinf. 9 (2008) 323, doi:http://dx.doi.org/10.1186/1471-2105-9-323.

[48] L. Excoffier, T. Hofer, M. Foll, Detecting loci under selection in a hierarchically structured population, Hered. 103 (2009) 285–298, doi:http://dx.doi.org/10.1038/hdy.2009.74.

[49] M. Foll, O. Gaggiotti, A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective, Genetics 180 (2008) 977–993, doi:http://dx.doi.org/10.1534/genetics.108.092221.

[50] M.A. Banks, W. Eichert, WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data, J. Hered. 1 (2000) 87–89.

[51] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magariños, L. Porras, T. Tvedebrink, J. Amigo, et al., Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169, doi:http://dx.doi.org/10.1016/j.fsigen.2010.02.003.

[52] T.M. Clayton, J.L. Guest, A.J. Urquhart, P.D. Gill, A genetic basis for anomalous band patterns encountered during DNA STR profiling, J. Forensic Sci. 49 (2004) 1207–1214, doi:http://dx.doi.org/10.1520/JFS2003145.

[53] J.B. Picanço, P.E. Raimann, C.H.A.S. Da Motta, R. Rodenbusch, L. Gusmão, C.S. Alho, Identification of the third/extra allele for forensic application in cases with TPOX tri-allelic pattern, Forensic Sci. Int. Genet. 16 (2015) 88–93, doi:http://dx.doi.org/10.1016/j.fsigen.2014.11.016.

[54] C.A. Crouse, S. Rogers, E. Amiott, S. Gibson, A. Masibay, Analysis and interpretation of short tandem repeat microvariants and three-banded allele patterns using multiple allele detection systems, J. Forensic Sci. 44 (1999) 87–94.

[55] M. Lukka, G. Tasa, P. Ellonen, K. Moilanen, V. Vassiljev, I. Ulmanen, Triallelic patterns in STR loci used for paternity analysis: evidence for a duplication in

chromosome 2 containing the TPOX STR locus, Forensic Sci. Int. 164 (2006) 3–9, doi:http://dx.doi.org/10.1016/j.forsciint.2005.11.006.
[56] L.-F. Wang, Y. Yang, X.-N. Zhang, X.-L. Quan, Y.-M. Wu, Tri-allelic pattern of short tandem repeats identifies the murderer among identical twins and suggests an embryonic mutational origin, Forensic Sci. Int. Genet. 16 (2015) 239–245, doi:http://dx.doi.org/10.1016/j.fsigen.2015.01.010.
[57] R.L.M. Huel, L. Basić, K. Madacki-Todorović, L. Smajlović, I. Eminović, I. Berbić, et al., Variant alleles triallelic patterns, and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia, Croat. Med. J. 48 (2007) 494–502.
[58] V. Díaz, P. Rivas, A. Carracedo, The presence of tri-allelic TPOX genotypes in Dominican Population, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 371–372, doi:http://dx.doi.org/10.1016/j.fsigss.2009.09.021.
[59] C. Fridman, P.C.C. Dos Santos, P. Kohler, C.F. Garcia, L.F. Lopez, E. Massad, et al., Brazilian population profile of 15 STR markers, Forensic Sci. Int. Genet. 2 (2008) 2–5, doi:http://dx.doi.org/10.1016/j.fsigen.2007.09.003.
[60] C. Alves, L. Gusmão, A. Damasceno, B. Soares, A. Amorim, Contribution for an African autosomic STR database (AmpF/STR Identifiler and Powerplex 16 System) and a report on genotypic variations, Forensic Sci. Int. 139 (2004) 201–205, doi:http://dx.doi.org/10.1016/j.forsciint.2003.11.002.
[61] A.C. Semo, Estudo genético-populacional dos principais grupos de Moçambique-aplicação forense, Universidade Da Beira Interior, 2013.
[62] W. Charoenchote, W. Charoenchote AmpFlSTR Identifiler STR allele frequencies and PowerPlex Y-STR haplotype frequencies of the Mero population of Northern Tanzania, University of California, 2004.
[63] T. Muro, J. Fujihara, S. Imamura, H. Nakamura, T. Yasuda, H. Takeshita, Allele frequencies for 15 STR loci in Ovambo population using AmpFlSTR® Identifiler kit, Leg. Med. 10 (2008) 157–159, doi:http://dx.doi.org/10.1016/j.legalmed.2007.10.003.
[64] H. Takeshita, E. Meyer, B. Brinkmann, The STR loci HumTPO and HumLPL: population genetic data in eight populations, Int. J. Legal Med. 110 (1997) 331–333, doi:http://dx.doi.org/10.1007/s004140050099.
[65] L.M. Paul, G.F. Simons, C.D. Fennig, Ethnologue Languages of the World, (2015). (accessed 09.08.15.) https://www.ethnologue.com/subgroups/s-0.
[66] G. Hefke, S. Davison, M.E. D'Amato, Forensic performance of Investigator DIPplex indels genotyping kit in native, immigrant and admixed populations in South Africa, Electrophoresis 36 (2015) 3018–3025.
[67] L. Ding, H. Wiener, T. Abebe, M. Altaye, R.C. Go, C. Kercsmar, et al., Comparison of measures of marker informativeness for ancestry and admixture mapping, BMC Genomics 12 (2011) 622, doi:http://dx.doi.org/10.1186/1471-2164-12-622.
[68] J.B. Listman, R.T. Malison, A. Sughondhabirom, B.-Z. Yang, R.L. Raaum, N. Thavichachart, et al., Demographic changes and marker properties affect detection of human population differentiation, BMC Genet. 8 (2007) 21, doi:http://dx.doi.org/10.1186/1471-2156-8-21.
[69] J.M. Butler, C.R. Hill, M.C. Kline, D.L. Duewer, C.J. Sprecher, R.S. McLaren, et al., The single most polymorphic STR Locus: SE33 performance in U.S. populations, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 23–24, doi:http://dx.doi.org/10.1016/j.fsigss.2009.08.173.
[70] S.R. Narum, J.E. Hess, Comparison of FST outlier tests for SNP loci under selection, Mol. Ecol. Resour. 11 (2011) 184–194, doi:http://dx.doi.org/10.1111/j.1755-0998.2011.02987.x.
[71] M. Moos, D. Gallwitz, Structure of a human β-actin-related pseudogene which lacks intervening sequences, Nucleic Acids Res. 10 (1982) 7843–7849, doi:http://dx.doi.org/10.1093/nar/10.23.7843.
[72] H. Jeffreys, The theory of probability, third ed., Oxford University Press, 1961.
[73] M. Foll, BayeScan v2.0 User Manual, (2010) http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.0_manual.pdf.
[74] C.H. Brenner, Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities, Forensic Sci. Int. 157 (2006) 172–180, doi:http://dx.doi.org/10.1016/j.forsciint.2005.11.003.
[75] E.R. Londin, M.A. Keller, C. Maista, G. Smith, L.A. Mamounas, R. Zhang, et al., CoAIMs: a cost-effective panel of ancestry informative markers for determining continental origins, PLoS One 5 (2010) e13443, doi:http://dx.doi.org/10.1371/journal.pone.0013443.
[76] C.D. Steele, D.J. Balding, Choice of population database for forensic DNA profile analysis, Sci. Justice 54 (2014) 487–493, doi:http://dx.doi.org/10.1016/j.scijus.2014.10.004.
[77] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic Sci. Int. Genet. 18 (2015) 49–65, doi:http://dx.doi.org/10.1016/j.fsigen.2015.05.012.
[78] J.S. Barnholtz-Sloan, C.L. Pfaff, R. Chakraborty, J.C. Long, Informativeness of the CODIS STR loci for admixture analysis, J. Forensic Sci. 50 (2005) 1322–1326 (JFS2003404).
[79] Á. Carracedo, J.M. Butler, L. Gusmão, W. Parson, L. Roewer, P.M. Schneider, Publication of population data for forensic purposes, Forensic Sci. Int. Genet. 4 (2010) 145–147, doi:http://dx.doi.org/10.1016/j.fsigen.2010.02.001.
[80] Á. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al., New guidelines for the publication of genetic population data, Forensic Sci Int Genet 7 (2013) 217–220, doi:http://dx.doi.org/10.1016/j.fsigen.2013.01.001.