Research Article   **Open Access**

Analysis of 30 INDEL Polymorphic Markers in the Panamanian Population: Gene Admixture Estimates, Population Structure and Forensic Parameters

Ramos CW[1], **Castro-Pérez E**[2*], **Molina-Jirón C**[3] and **Trejos DE**[4]

[1]Department of Genetics and Molecular Biology, College of Natural Science and Technology, University of Panama, Panama

[2]Department of Dermatology, University of Wisconsin-Madison School of Medicine and Public Health & William S. Middleton Memorial Hospital, USA

[3]Department of Biochemistry, College of Natural Sciences and Technology, University of Panama, Panama

[4]Laboratory of Biomolecular Analysis, Legal Medicine and Forensic Science Institute, Public Ministry, Panama

*Corresponding author: Castro-Pérez E, Department of Dermatology, University of Wisconsin-Madison, Madison, School of Medicine and Public Health & William S. Middleton Memorial Hospital, USA, Tel: 608261-1878; E-mail: ecastroperez@dermatology.wisc.edu

Abstract

Short tandem repeat (STR) loci have been successfully employed for forensic genetic, paternity and anthropological analyses in the admixed and Amerindian populations of Panamanian for several years. Nevertheless, reports indicate that the use of STRs might be limited in cases involving degraded DNA samples due to their PCR amplicon size and in paternity because of their relatively high mutation rates. Therefore, as a complement to STRs, markers with higher PCR efficiency, such as insertion/deletion (INDEL), have been developed. However, the genetic polymorphisms and distribution of INDELs in the Panamanian population was unknown. Using the Investigator® DIPplex kit (Qiagen), we report here for the first time the genetic profile of 30 INDEL markers in the Panamanian population. Gene admixture estimates and population structure indicated that the Panamanian population is differentially admixed highly polymorphic. Interestingly, admixture estimates where highly similar to our previous report using STR indicating Amerindian, African and European gene contributions. INDELs showed three gene clusters in these proportions: 0.46 (cluster 1), 0.24 (cluster 2) and 0.30 (cluster 3) versus 0.51 (Amerindian), 0.24 (African) and 0.25 (European) for STRs. We also found that, both, INDELs and STRs indicated that ancestral gene distribution is heterogeneous among the provinces, across the country. Furthermore, allele frequency variation showed that all loci accomplished Hardy-Weinberg expectations and display high diversity as indicated by the average of observed heterozygosity (0.4477), expected heterozygosity (0.4553) and high similarity with reference U.S. and world populations. Additionally, forensic statistic parameters showed a strong significance in the combined power of discrimination (0.9999999999987) and the combined matching probability (1.27×10^{-12}) but relatively lower significance in the combined power of exclusion (0.992655332). The forensic parameters calculated indicate that INDEL markers can be affectively applicable to the Panamanian population for forensic uses but should be complemented with additional markers, such as STRs for paternity analyses.

Keywords: Insertion/Deletion; INDEL markers; Short tandem repeats; STR loci; Gene admixture; Forensic parameters

Introduction

Short tandem repeats (STRs) genotyping is one of the most extensively used systems of molecular markers for human forensic identification worldwide [1,2]. STRs have been successfully employed for forensic and paternity analyses in the Panamanian population for more than 15 years. In addition, we have used STR markers to characterize Amerindian Ngobe and Embera populations of Panama (also located in Costa Rica and Colombia, respectively) [3]. More recently, we determined the ancestry and tri-hybrid genetic structure with admixture estimates of the population of Panama using STR loci [4]. These markers have shown high polymorphism and genetic variation suitable for forensic and paternity analyses across the country's provinces. Nevertheless, several reports indicate that the use of STRs might potentially be limited in some cases due to their PCR amplicon size and relatively high mutation rates [2,5,6].

In particular, analysis involving paraffin-embedded samples, ancient DNA and degraded DNA, all commonly found in forensic casework [7,8]. Therefore, as a complement to STRs, additional markers with higher PCR efficiency in degraded DNA samples have been developed including single nucleotide polymorphisms (SNPs) and insertion/deletion (INDELs) polymorphic markers [9,10]. Unfortunately, SNP-based genotyping systems require more sophisticated instrumentation and more laborious sample processing chemistry that make it incompatible with STR genotyping technology currently available in most forensic laboratories [6,11]. In contrast, INDELs combine the best features of both SNPs and STRs by using reduced size of PCR amplicon ranges, lower binary loci mutation rates but direct capillary electrophoresis platform for genotyping as STR loci [6,7,11]. Hence, the implementation of INDEL markers has been proposed as a complement for forensic analysis in the Panamanian population together with STRs. However, the allelic polymorphisms and distribution of gene variation for INDEls across the country is unknown.

The Investigator® DIPplex kit (Qiagen) contain a 30 biallelic panel of INDEL markers plus amelogenin previously shown of valuable forensic

application in several world populations [12,13]. Here we analyzed 332 unrelated subjects distributed across the country's provinces with the Investigator® DIPplex panel kit. Importantly, our previous work using STRs showed that the Panamanian population displays a tri-hybrid-Mestizo admixture composed of Amerindian, African and European genes [4]. Additionally, several studies have shown the utility of INDEL markers for gene admixture estimations in multiple populations [8,14-16].

Therefore, to further extend our admixture studies with STRs, we determined here the genetic structure of the Mestizo population of Panama using INDEL markers and estimated the gene admixture proportions. Moreover, we calculated allelic frequencies and diversity parameters of each locus; and estimated statistical parameters of forensic and paternity importance. Results showed that polymorphic genetic structure and admixture estimates using INDEL markers are consistent with previous report using STR loci. Additionally, forensic statistic parameters showed a strong significance in the combined power of discrimination (0.9999999999987) and the combined matching probability (1.27×10^{-12}) but lower in the combined power of exclusion (0.992655332). The data generated indicate that INDEL markers might be applicable to the Panamanian population for most forensic uses but should be complemented with additional markers such as STRs for paternity and kinship analyses.

Materials and Methods

Selection of mestizos, blood collection and DNA extraction

We selected 332 unrelated individuals from Mestizo's origins half male and half female with both parents and grandparents born in Panama. The number of individuals per province for sampling was proportionally determined according to the percentage of contribution of each province to the total country's population as reported by the Government's National statistics Figure 1. This model was chosen because it would have the advantage of allowing determining a good representation for the total country's population. The disadvantage of this model is that for regional calculation of population statistics, some provinces would have relatively low number of samples. However, the random collection of DNA from unrelated individuals and unbiased markers could be sufficient to get a general view of genes at each province. Samples were primarily collected in urban areas were most of the Mestizo (admixed) populations are concentrated. Conversely, areas with low amounts of Mestizo population, such as Amerindian reservations, known as Comarcas, were not considered for sampling.

Venous blood samples (10 mL) were obtained by phlebotomy and were collected into tubes containing EDTA anticoagulant buffer. Tubes were gently mixed to ensure that the EDTA and blood were adequately mixed. After extraction, blood samples were immediately processed or stored at 4˚C, for a maximum of 1-4 days. Approximately 125 µL of blood were placed onto the center of each of four circles of commercially prepared FTA® cards (Whatman® FTA® card technology). An FTA-blood punch was used to cut a small piece of the FTA card-containing blood, and the small piece was processed for DNA extraction according to the procedure specified by the DNA IQ™ Reference Sample Kit for Maxwell® 16 (Promega® Corporation, Madison WI, USA). DNA samples were eluted in 300 µL of elution buffer; their concentration determined and later used for PCR amplification.

Figure 1: Map of Panama's Provinces showing the locations of samples collections. The number of samples collected per province representing the percentage of contribution to the country's population is Bocas Del Toro n=9 (2.8%) Coclé n=23 (7%), Colon n=25 (7.4%), Chiriqui n=48 (14%), Darien n=5 (1.4%), Herrera n=14 (4%), Los Santos n=12 (3.4%), Panama n=174 (53%), Veraguas n=24 (7%).

PCR amplification and genotyping of INDEL's alleles

PCR Amplification of 30 INDEL markers were performed in multiplex reaction with a GeneAmp® PCR System 9700 thermal cycler (Applied BiosystemsTM) according to the Investigator® DIPplex kit (Qiagen) user manual. Allelic genotypes of INDEL's multiplex PCR were determined by capillary electrophoresis using a ABI PRISM 3500 genetic analyzer according to Investigator® DIPplex kit (Qiagen) protocols. Alleles' sizes were identified by co-migration comparison with the DNA size standard BTO 550 (Qiagen®) and using the data Collection Software version 1.0 and the GeneMapper® ID-X Software version 1.2 (Thermo Fisher Scientific Inc.) as indicated the Investigator® DIPplex kit manufacturer.

Gene admixture estimates and population forensic parameters

The software STRUCTURE [17,18] version 2.3.4 (2012) was used to estimate the admixture proportions and population structure using 5,000 burn-in period and 50,000 MCMC reps after burn-in. The number of populations assumed was from one to eight and the log likelihood for each K was calculated using the Structure Harvester software (2012) implementing the Evanno method [19]. Allelic frequencies for insertions (Dip+) and deletions (Dip-) and population parameters were calculated using PopGene [20], version 1.32. Observed and expected heterozygosity were determined using Nei's algorithm [21] and the likelihood ratio (G-square) test was performed to estimate Hardy-Weinberg expectations per locus. The MS Excel workbook template PowerStatsV12 (Promega Corporation, Madison WI, USA) was used to calculate forensic and paternity parameters matching probability (MP), power of discrimination (PD) power of exclusion (PE) typical paternity index (TPI) and polymorphic information content.

Results

Gene admixture estimates

After multiple Structure's Bayesian permutations, analyses showed that Ks with the highest log likelihood L(K) probability estimated were

for K2 (11 assays), K3 (12 assays) and k4 (9 assays) corresponding to average probabilities of -1287.01, -12944.04 and -13013.91, respectively (Figure 2).
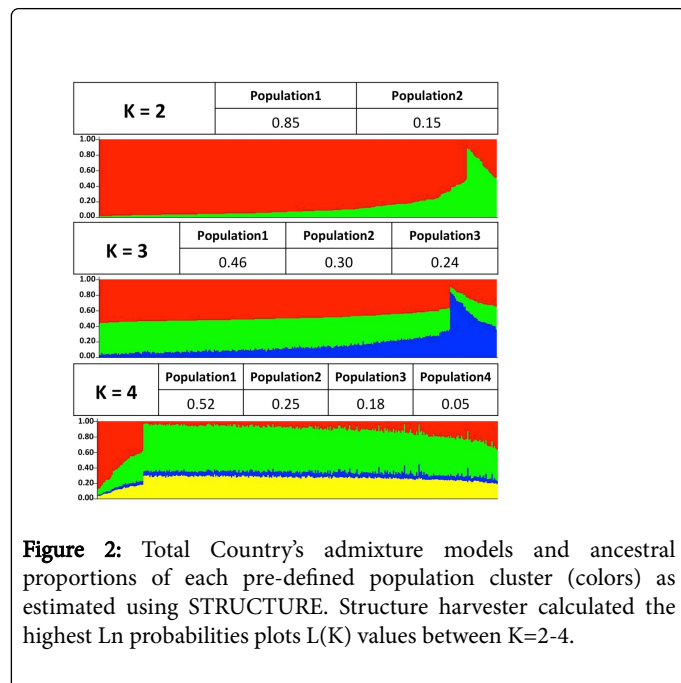

Figure 2: Total Country's admixture models and ancestral proportions of each pre-defined population cluster (colors) as estimated using STRUCTURE. Structure harvester calculated the highest Ln probabilities plots L(K) values between K=2-4.

K2 showed admixture proportions of 0.85 and 0.15; K3 admixture proportions were 0.46, 0.24 and 0.30; whereas K4 showed admixture estimates of 0.52, 0.25, 0.18 and 0.05. Unfortunately, there was no INDEL's data available from ancestral populations; therefore, to further evaluate these results we compared the admixture proportions generated using INDELs with our previous work using STR markers [4].

PROVINCE	INDELs[*]			STRs[**]		
	Cluster 1	Cluster 2	Cluster 3	Amerindian	African	European
Total Country	0.46	0.24	0.30	0.51	0.24	0.25
Bocas del T.	0.18	0.14	0.69	n.d	n.d	n.d
Cocle	0.51	0.26	0.23	0.70	0.10	0.20
Colon	0.30	0.54	0.16	0.34	0.47	0.19
Chiriqui	0.61	0.19	0.20	0.64	0.15	0.21
Darien	0.41	0.35	0.24	n.d	n.d	n.d
Herrera	0.52	0.37	0.11	0.40	0.20	0.40
Los Santos	0.48	0.41	0.11	0.33	0.25	0.42
Panama	0.61	0.24	0.14	0.50	0.27	0.23
Veraguas	0.59	0.16	0.25	0.50	0.11	0.39
[*]Present Study; [**]Castro-Perez et al. 2016; n.d=not determined.						

Table 1: Gene admixture estimates per province, the total country and comparison of INDELs markers with STR loci.

Importantly, our STR studies involved the use of ancestral data from Amerindian (Ngobe-Chibchan), Western African (Angola and Guiné-Bissau) and European (Spain) populations and supported a tri-hybrid admixture model (K3). Additionally, the tri-hybrid admixture model is highly consistent with ethno-historic [22,23] and genetic data from classical [24], mitochondrial [25] and Y-chromosome [26] DNA markers. We therefore compared K3's admixture proportions per province and total country from INDELs analysis with our previous admixture STR data. Interestingly, we found that admixture estimates from both, INDELs and STRs (Table 1) were very similar for the total country 0.46 (cluster 1), 0.24 (cluster 2) and 0.30 (cluster 3) for INDELs; versus 0.51 (Amerindian), 0.24 (African) and 0.25 (European) for STRs. Furthermore, admixture proportions among provinces with INDELs also displayed overall similar patterns but with some variations compared to STRs (Table 1). This data indicate that both INDELs and STRs display comparable gene admixture proportions for the Panamanian population.

Genetic diversity and population structure

Admixture proportion among the provinces determined by INDEL markers showed that the Panamanian population is heterogeneously admixed and highly polymorphic (Figure 3). Similar to STR loci, INDEL's analysis generated a diverse distribution of genes across the provinces that resemble Amerindian, African and European derived ancestral genes (Table 1). Triangle plot (Figure 3) represent the gene pool of the general Panamanian population displaying three genetic clusters that behave similar to STR's parental populations: Amerindian (cluster 1), African (cluster 2) and European (cluster 3).

Dots within the triangle represent the individuals of each province, who were clustered and spatially distributed according to their individual admixture affinity to each ancestral population. Structure bar plot (Figure 3, bottom) displays the differential proportions of gene ancestry per individuals and demonstrates the polymorphic genetic variation among provinces. INDELs demonstrated that Chiriqui and Panama provinces display the highest proportion of Amerindian-like (cluster 1) genes, both with 0.61, followed by Veraguas (0.59), Herrera (0.52) and Cocle (0.51). STRs also showed that the same provinces showed the highest Amerindian gene admixture but in different proportions (Table 1).

Moreover, African-like (cluster 2) admixture determined by INDELs showed that, as expected, Colon province display the highest African gene ancestry (0.54), followed by Los Santos (0.41) and Herrera (0.37). STR loci also showed the highest proportion of African genes in Colon (0.47), but with some variation in the consecutive provinces. In STR studies, Colon was followed by Panama province with 0.27, Los Santos (0.25) and Herrera (0.20). INDELs also indicated that European-like (cluster 3) genes show relatively lower contribution to the gene pool in all provinces, ranging from 0.11 for Herrera and Los Santos, to 0.25 for Veraguas. An exception to European genes was Bocas, showing an unusual 0.69. Comparison with STRs also indicated relatively lower levels of European genes contribution among the provinces ranging from 0.19 for Colon to 0.42 for Los Santos. Taken together, this data demonstrate that both INDELs and STRs display similar gene admixture patterns and genetic structure across the country's provinces.
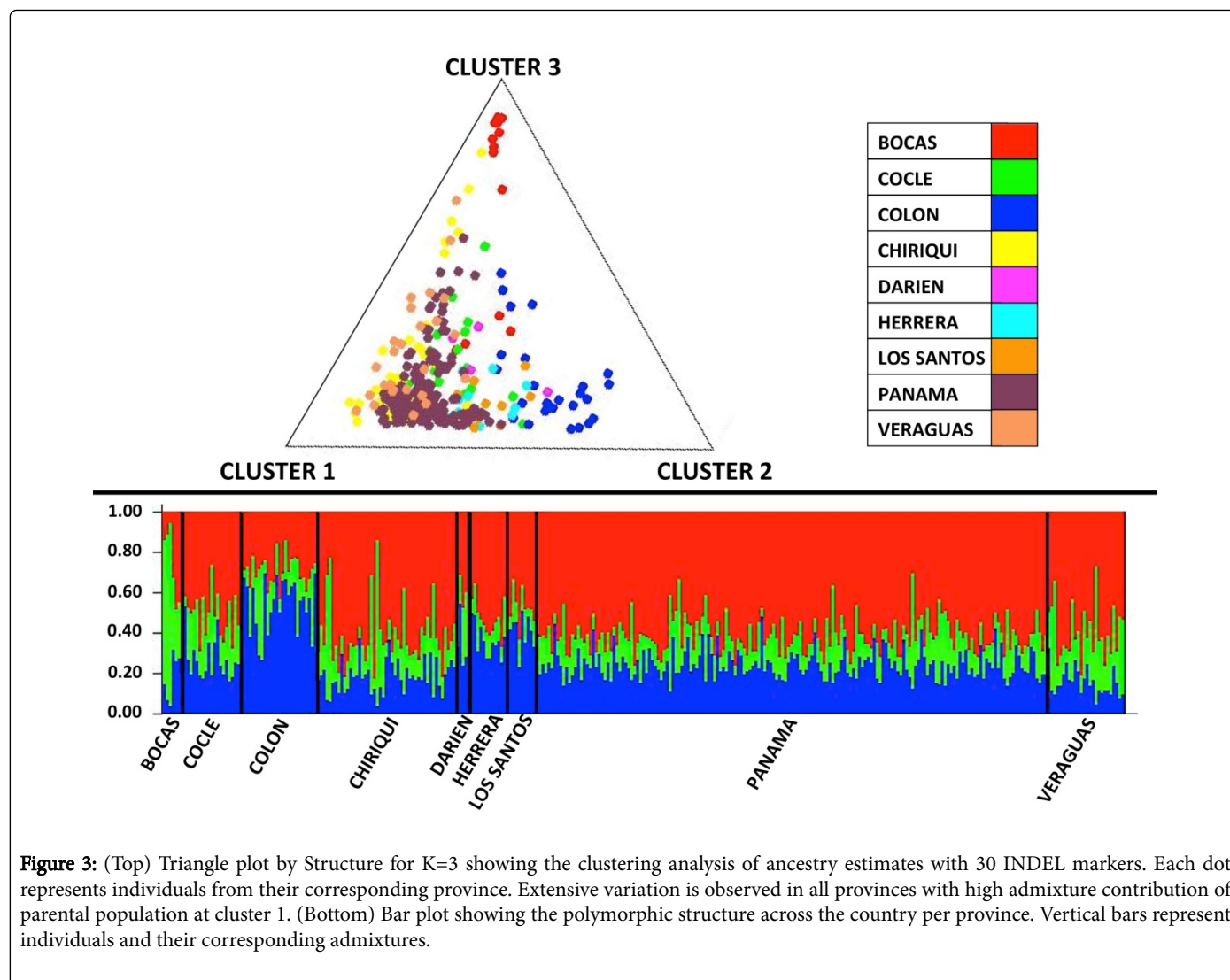
Figure 3: (Top) Triangle plot by Structure for K=3 showing the clustering analysis of ancestry estimates with 30 INDEL markers. Each dot represents individuals from their corresponding province. Extensive variation is observed in all provinces with high admixture contribution of parental population at cluster 1. (Bottom) Bar plot showing the polymorphic structure across the country per province. Vertical bars represent individuals and their corresponding admixtures.

Allele frequency variation

Allelic frequencies of 30 INDELs insertions (Dip+) and deletions (Dip-) from the total mestizo population of Panama are shown in Table 2. Frequencies of insertions varied between 0.2024 (locus HLD58) and 0.8030 (locus HLD64) with an average of 0.48349; whereas frequencies for deletions ranged between 0.1970 (locus HLD64) and 0.7976 (locus HLD58) and 0.51651 as average. Observed and expected heterozygosity and Hardy-Weinberg equilibrium values are also shown (Table 2). Observed heterozygosity (Ho) ranged from 0.2961 (locus HLD58) to 0.5287 (locus HLD136) with an average of 0.4477. Expected heterozygosity (He) showed values ranging from 0.3175 (locus HLD64) to 0.4997 (locus HLD131) with an average of 0.4553, which is close to the average of Ho (0.4477). However, when comparing the observed with the expected heterozygosity for each marker, we found that 5 loci (HLD 131, HLD 114, HLD 122, HLD 81 and HLD 128) display low Ho, contrasting with He, ranging from 0.003 (locus HLD131) to 0.0487 (HLD128), thus suggesting a deficiency of heterozygosity. However, after Bonferroni correction (p=0.0017) all loci showed HWE [27]. This data indicate that INDEL markers variation is highly polymorphic in the Panamanian population.

LOCUS	ID	Allele Dip-	Allele Dip +	Ho	He	HWE
HLD77	rs1611048	0.6485	0.3515	0.4773	0.4571	0.3764
HLD45	rs2307959	0.3973	0.6027	0.4743	0.4789	0.8404
HLD131	rs1611001	0.4879	0.5121	0.4018	0.4997	0.003
HLD70	rs2307652	0.2613	0.7387	0.3776	0.3861	0.6707
HLD6	rs1610905	0.4517	0.5483	0.5106	0.4953	0.5944
HLD111	rs1305047	0.6843	0.3157	0.3958	0.4321	0.1195
HLD58	rs1610937	0.7976	0.2024	0.2961	0.3229	0.1233
HLD56	rs2308292	0.3833	0.6167	0.4592	0.4729	0.5407
HLD118	rs16438	0.6888	0.3112	0.429	0.4287	0.9885
HLD92	rs17174476	0.6692	0.3308	0.4441	0.4428	0.9775
HLD93	rs2307570	0.4486	0.5514	0.4985	0.4947	0.9117
HLD99	rs2308163	0.4486	0.5514	0.4683	0.4947	0.3173

HLD88	rs8190570	0.4018	0.5982	0.5015	0.4807	0.4476
HLD101	rs2307433	0.4834	0.5166	0.5136	0.4994	0.6257
HLD67	rs1305056	0.4743	0.5257	0.5196	0.4987	0.4611
HLD83	rs2308072	0.6409	0.3591	0.4532	0.4605	0.7084
HLD114	rs2307581	0.5303	0.4697	0.4441	0.4982	0.0394
HLD48	rs28369942	0.2455	0.7545	0.3776	0.3712	0.8144
HLD124	rs6481	0.6576	0.3424	0.429	0.4497	0.4031
HLD122	rs8178524	0.6348	0.3652	0.4018	0.4647	0.0163
HLD125	rs16388	0.7379	0.2621	0.4169	0.3875	0.1923
HLD64	rs1610935	0.197	0.803	0.2991	0.3175	0.2534
HLD81	rs17879936	0.3167	0.6833	0.3807	0.4332	0.0217
HLD136	rs16363	0.4409	0.5591	0.5287	0.4927	0.1782
HLD133	rs2067235	0.6341	0.3659	0.4894	0.4639	0.3674
HLD97	rs17238892	0.5136	0.4864	0.4955	0.4996	0.8147
HLD40	rs2307956	0.5854	0.4146	0.4894	0.4852	0.9513
HLD128	rs2307924	0.596	0.404	0.429	0.4807	0.0487
HLD39	rs17878444	0.6037	0.3963	0.5045	0.4776	0.3093
HLD84	rs3081400	0.4343	0.5657	0.5257	0.4916	0.3067

Ho=Observed Heterozygosity, He=Expected Heterozygosity, HWE=Hardy-Weinberg Equilibrium.

Table 2: Allele and genotypic frequencies of 30 INDEL markers in the panamanian population.

Forensic genetic population parameters

The following (Table 3) shows the values calculated for INDEL's forensic and paternity parameters (all combined loci) per province and the country; and per loci for the total population. We found that the average for matching probability (MP) was 0.4032, ranging from the most informative marker for this parameter HLD64 with 0.519 to locus HLD131, the least informative with 0.341. However, the combined matching probability (CMP) for the country (1.27×10^{-12}) and most of the provinces was relatively high. Among the provinces, CMP varied starting from Darien (1.37219×10^{-07}) and Bocas Del Toro (4.17645×10^{-09}), to Panama province with 9.17764×10^{-13}. Colon (8.34379×10^{-12}) and Chiriqui (4.56524×10^{-12}) also showed relatively high CMP, followed by Los Santos (2.13145×10^{-11}), Herrera (2.10797×10^{-11}), Veraguas (1.76484×10^{-11}) and Cocle (1.69054×10^{-11}). Furthermore, the power of discrimination (PD) showed an average of 0.597, ranging from the most informative marker for this parameter (locus HLD131) with 0.659 to the least informative loci (locus HLD64) with 0.481. However, like the CMP, the combined power of discrimination (CPD) was also relatively high in the country and most provinces. For the country, CPD was 0.9999999999987 and for the provinces varied from 0.99999986 for Darien and 0.9999999999991 for Panama province. Additionally, the power of exclusion (PE) displayed an average of 0.15, and ranged from the most informative marker for this parameter (locus HLD136) with 0.215, to the least informative loci HLD58 and HLD64, both with 0.062. In

contrast to the CMP and the CPE, the combined power of exclusion (CPE) showed a relatively lower significance for the country (0.992655332) and provinces ranging from 0.991455029 for Panama province to 0.999564716 for Herrera. An exception was Bocas Del Toro showing a CPE of 1 that might be related to its relatively small sample size as mentioned in methods. The typical paternity index (TPI) per loci showed an average of 0.915, ranging from 0.710 (locus HLD58) to 1.065 (locus HLD136). The polymorphic information content (PIC) showed an average of 0.3504 and varied from 0.266 (locus HLD64) to 0.375 (loci HLD131, HLD101 and HLD97). Most PIC values were close to 0.375, except 5 loci (HLD64, HLD58, HLD48, HLD70 and HLD125) that showed values lower than the average. Overall, the combined statistics of 30 INDELs for forensic and paternity parameters demonstrate that most calculations are highly significant for these markers' application for forensic genetic in the Panamanian population.

PROVINCE	CMP	CPD	CPE		
Bocas del T.	4,18E-04	0.999999996	0.994436451		
Cocle	1,69E-06	0.99999999998	0.997620088		
Colon	8,34E-07	0.999999999992	0.996414523		
Chiriqui	4,57E-07	0.999999999995	0.995164563		
Darien	1,37E-02	0.99999986	1		
Herrera	2,11E-06	0.99999999998	0.999564716		
Los santos	2,13E-06	0.99999999997	0.998165382		
Panama	9,18E-08	0.9999999999991	0.991455029		
Veraguas	1,76E-06	0.99999999998	0.998657597		
Total Country	1.27 × 10⁻¹²	0.9999999999987	0.992655332		
LOCUS	**MP**	**PD**	**PE**	**TPI**	**PIC**
HLD77	0.409	0.591	0.170	0.959	0.352
HLD45	0.384	0.616	0.166	0.951	0.364
HLD131	0.341	0.659	0.115	0.836	0.375
HLD70	0.450	0.550	0.101	0.803	0.312
HLD6	0.385	0.615	0.197	1.022	0.373
HLD111	0.407	0.593	0.111	0.828	0.339
HLD58	0.513	0.487	0.062	0.710	0.271
HLD56	0.384	0.616	0.153	0.922	0.361
HLD118	0.416	0.584	0.133	0.876	0.339
HLD92	0.409	0.591	0.143	0.899	0.345
HLD93	0.380	0.620	0.186	0.997	0.372
HLD99	0.367	0.633	0.162	0.943	0.372
HLD88	0.395	0.605	0.189	1.003	0.365
HLD101	0.383	0.617	0.200	1.028	0.375
HLD67	0.387	0.613	0.205	1.041	0.374

HLD83	0.394	0.606	0.148	0.912	0.354
HLD114	0.353	0.647	0.142	0.897	0.374
HLD48	0.466	0.534	0.100	0.801	0.302
HLD124	0.397	0.603	0.133	0.878	0.349
HLD122	0.377	0.623	0.116	0.838	0.356
HLD125	0.457	0.543	0.123	0.855	0.312
HLD64	0.519	0.481	0.062	0.711	0.266
HLD81	0.404	0.596	0.102	0.805	0.339
HLD136	0.399	0.601	0.215	1.065	0.371
HLD133	0.405	0.595	0.177	0.976	0.356
HLD97	0.372	0.628	0.182	0.988	0.375
HLD40	0.384	0.616	0.177	0.976	0.368
HLD128	0.366	0.634	0.133	0.877	0.366
HLD39	0.400	0.600	0.193	1.012	0.364
HLD84	0.394	0.606	0.205	1.041	0.370

Table 3: Combined forensic genetic and paternity statistical parameters per province and the total country; and per loci for the country.

Discussion

INDEL's gene admixture variation is comparable to STR's

Structure computations showed that the highest log likelihood L(K) probability of admixture models were for K2, K3 and K4 (Figure 2). However, multiple reports have previously shown that K2 and K4 models are poorly supported genetically and ethno historically. In contrast, the tri-hybrid (K3) model (Figure 2, middle) is the best supported as demonstrated by us with STR loci (4) and classical markers (24) but also by other research groups using mtDNA [25] and Y-chromosome analysis (26). Moreover, Ethno-historical data strongly support that the Panamanian population is the result of an admixture process involving Amerindians, African and European ancestors (22, 23). Therefore, further INDEL's analyses were all performed based on a tri-hybrid admixture model (K3) showing that the general population is composed of three ancestral populations in the following proportions 0.46, 0.24 and 0.30 (Figure 2, middle). Interestingly, admixture proportions and gene distribution are highly consistent with our previous report using STR loci (Table 1). Furthermore, admixture proportions are polymorphic across the country's provinces as evidenced in Table 1 and in the Structure bar plot (Figure 3, bottom). These similarities with STRs suggest that even though in the absence of ancestral data for INDELs, the observed proportions highly correspond to Amerindian (cluster 1), African (cluster 2) and European (cluster 3) ancestry. Consistent with this notion, the triangle plot (Figure3, top) depicts three gene clusters resembling three parental populations corresponding to Amerindian (cluster 1), African (cluster 2) and European (cluster 3). Importantly, although the INDEL's ancestral proportions and distribution within provinces display some variation compared to STRs, these differences might be related to some sources of variation including lack of INDEL's ancestral data, relatively small sample size in some provinces or simply differences between

both methods. Another source of variation might be the fact that, samples used for INDELs are different from those used in the previous study with STRs. DNA samples (332 subjects, half male/female) used for INDEL analyses represent a totally new sampling, and therefore we did not expect exactly the same results when compared them with our STR studies (650 subjects, half male/female). However, it is exciting that analyses still demonstrate high similarity using both methods and different samplings. This reproducibility of methods and samplings is very useful for multiple applications of these markers including anthropology and biomedicine but particularly in paternity and forensic genetic casework.

Allele frequencies are highly similar to U.S. Africans and U.S. Hispanics

Gene admixture distribution and population structure demonstrated that INDEL markers are highly consistent with STRs in the Panamanian population. However, how these INDELs in the Panamanian population were compared to other world populations was unknown. We consider very important to compare our allele frequency data (Table 2) with U.S. Caucasian-American, U.S. African-American, U.S. Hispanic-American and Asians, reference populations where INDEL markers are already in use for paternity and forensic casework [12]. We found that allele frequency of 11 loci from the Panamanian population were highly similar to U.S. African-Americans (loci HLD77, HLD70, HLD111, HLD58, HLD118, HLD92, HLD93, HLD122, HLD125, HLD64 and HLD133). Additionally, 9 loci from Panamanians (loci HLD131, HLD99, HLD88, HLD101, HLD67, HLD114, HLD136, HLD97 and HLD39) were very similar to U.S. Hispanic-American population whereas HLD48 locus was similar to both U.S. African-American and U.S. Hispanic-American populations. Two loci showed mixed allele frequency similarities with U.S. Hispanic-Americans and U.S. Caucasian-Americans (loci HLD56 and HLD40), while locus HLD45 displayed mixed allele similarity with U.S. Hispanic-Americans and Asians. Allele frequency for locus HLD83 was only similar to Asians whereas locus HLD84's allele frequency was similar to all reference populations, except to Asians. Interestingly, four loci (HLD6, HLD124, HLD81 and HLD128) were not comparable to the U.S. reference populations, thus showing unique frequency patterns for the Panamanian population (Table 2). Importantly, heterozygosity values (Ho and He) also displayed high diversity among the loci. The observed heterozygosity for these loci varied between 0.2961 (locus HLD58) to 0.5287 (locus HLD136), with an average of 0.4477. The expected heterozygosity varied between 0.3175 (HLD64) to 0.4997 (HLD131) with an average of 0.4553, which is close to the theoretical expected value for bi-allelic markers (0.500). In general, the range of allele variation and heterozygosity, together with allelic polymorphism similarities with reference populations indicates high genetic diversity of these loci for forensic use in the Panamanian population.

Polymorphic INDEL markers are effective complement for STRs

Forensic and paternity parameters indicated that when combined, all 30 INDEL markers are suitable for most forensic application in the Panamanian population (Table 3). The combined matching probability (CMP) for Panamanians was 1.27×10^{-12}, which is acceptable for forensic use, similar to other world populations currently using INDELs for forensic casework. For instance, CMP for Chinese populations range between $1.8^{-3}.17 \times 10^{-11}$ [2], whereas U.S. reference

populations were relatively higher in U.S. Caucasian-Americans with 3.65×10^{-13}. However, U.S. Hispanic-Americans (2.12×10^{-12}) and U.S. African-American (1.43×10^{-11}) [11] display CMP highly similar to the Panamanian population. Moreover, Brazil (3.4×10^{-13}) [28], which is also an admixed population like Panama's, showed higher significance for the CMP than Panamanians. Furthermore, the combined power of discrimination for Panamanian's (CPD) was 0.9999999999987, which is similar to those reported in admixed populations from Brazil (0.9999999999997) [28] and Mexico (0.99999999) [7]. In addition, the combined power of exclusion (CPE) (Table 3) for Panamanians was 0.992655332, which is relatively lower than other populations including all U.S. reference populations, all displaying 0.999999999 [11] and also lower than Brazil with 0.997250806 [28]. However, the CPE of Panamanian mestizos is higher than Somalia (0.9860) [29] and Chinese (0.988) [2]. The typical paternity index (TPI), displayed values between 0.710 (locus HLD58) and 1.065 (locus HLD136), with an average of 0.9150, which is relatively lower, when compared to STRs. The polymorphic information content (PIC) ranged between 0.266 (locus HLD64) to 0.375. Altogether, this data indicate that in the Panamanian population, INDEL markers are suitable for most forensic genetic applications such as those involving samples from sources with DNA degradation. Nevertheless, INDELs should be complemented with STRs or other markers for paternity and kinship. Although DNA degradation is usually less common in samples used for paternity tests, they might be susceptible to mutations more common in STR loci than INDEL markers.

Conclusion

The results obtained in this research provide positive evidence for the use of INDEL markers in the Panamanian population for forensic genetic, paternity and anthropological applications. The forensic statistics parameters of INDELs are also comparable to those reported for U.S. reference and world populations. Furthermore, gene admixture estimates using both, INDELs and STRs demonstrate that both markers are highly powerful to detect highly similar gene ancestry variation across the country. Altogether, this data suggest that INDEL markers can be affectively applicable to the Panamanian population for most forensic applications but complementing them with additional markers is recommendable for paternity analyses.

Acknowledgement

References

1. Shewale JG, Liu RH (2013) Forensic DNA analysis: Current practices and emerging technologies: CRC press, Taylor and Francis Group, USA.

2. Wang Z, Zhang S, Zhao S, Hu Z, Sun K, et al. (2014) Population genetics of 30 insertion-deletion polymorphisms in two chinese populations using Qiagen Investigator® DIPplex kit. Forensic Sci Int Genet 11: e12-e14.

3. Castro EA, Trejos DE, Berovides-Alvarez V, Arias TD, Ramos CW, et al. (2007) Genetic polymorphism and forensic parameters of nine short tandem repeat loci in Ngöbé and Emberá Amerindians of Panama. Hum Biol 79: 563-577.

4. Castro-Pérez E, Trejos DE, Hrbek T, Setaluri V, Ramos CW, et al, (2016) Genetic ancestry of the panamanian population: Polymorphic structure, chibchan amerindian genes; and biological perspectives on diseases. Inter J Biol Anthropology 9: 1-14.

5. Torres SR, Uehara CJ, Sutter-Latorre AF, de Almeida BS, Sauerbier TS, et al. (2014) Population genetic data and forensic parameters of 30 autosomal InDel markers in santa catarina state population, Southern Brazil. Mol Biol Rep 41: 5429-5433.

6. TZ Gao, LB Yun, W He, Y Gu, YP Hou, et al. (2015) The application of multi-InDel as supplementary in paternity cases with STR mutation. Forensic Sci Int Genet 5: e218-e219.

7. Martínez-Cortés G, García-Aceves M, Favela-Mendoza AF, Muñoz-Valle JF, Velarde-Felix JS, et al. (2016) Forensic parameters of the Investigator DIPplex kit (Qiagen) in six Mexican populations. Int J Legal Med 130: 683-685.

8. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, et al. (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One 7: e29684.

9. Weber JL, David D, Heil J, Fan Y, Zhao C, et al. (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71: 854-862.

10. da Costa Francez PA, Rodrigues EM, de Velasco AM, dos Santos SE (2012) Insertion-deletion polymorphisms-utilization on forensic analysis. Int J Legal 126: 491-496.

11. LaRue BL, Ge J, King JL, Budowle B (2012) A validation study of the Qiagen investigator DIPplex® kit; An INDEL-based assay for human identification. Int J Legal Med 126: 533-540.

12. Fondevila M, Phillips C, Santos C, Pereira R, Gusmão L, et al. (2014) Forensic performance of two insertion-deletion marker assays. Int J Legal Med 126: 725-737.

13. Investigator DIPplex Handbook (2014) QIAGEN.

14. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmão L, et al. (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. Hum Mutat 31: 184-190.

15. Resque RL, Freitas NoS, Rodrigues EM, Guerreiro JF, Santos NP, et al. (2010) Estimates of interethnic admixture in the Brazilian population using a panel of 24 X-linked insertion/deletion markers. Am J Hum Biol 22: 849-852.

16. Saloum de Neves MF, Pereira R, Vianna R, Rodolfo Beuttenmüller de AA, Leite Góes Gitaí D, et al. (2013) Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. PLoS One 8: e75145.

17. Falush D, Stephens M, Pritchard JK (2000) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

18. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

19. Earl DA, vonHoldt BM (2012) Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

20. Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX, et al. (1997) PopGene, the user-friendly shareware for population genetic analysis, molecular biology and biotechnology centre.

21. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

22. Jaén-Suárez O (1978) The population of the isthmus of panama Xvi century to the Xx: Impresora de La Nación, Spain.

23. Jaén-Suárez O (1998) The population of the isthmus of panama, study of geo-history, Agencia Española de Cooperación Internacional, Spain.

24. Arias TD, Castro E, Ruiz E, Barrantes R, Jorge-Nebert L, et al. (2002) Racial mix of the panamanian population. Rev Med Panama 27: 5-17.

25. Perego UA, Lancioni H, Tribaldos M, Angerhofer N, Ekins JE, et al. (2012) Decrypting the mitochondrial gene pool of modern Panamanians. PLoS One 7: e38337.

26. Grugni V, Battaglia V, Perego UA, Raveane A, Lancioni H, et al. (2015) Exploring the Y chromosomal ancestry of modern panamanians. PLoS One 10: e0144223.

27. Carvalho A, Pinheiro MF (2013) Population data of 30 insertion/delection polymorphisms from a sample taken in the north of Portugal. Int J Legal Med 127: 65-67.

28. Ferreira Palha TJB, Ribeiro Rodrigues EM, Cavalcante GC, Marrero A, de Souza IR, et al. (2015) Population genetic analysis of insertion-deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. Forensic Sci Int Genet 19: 10-14.

29. Neuvonen AM, Palo JU, Hedman M, Sajantila A (2012) Discrimination power of Investigator DIPplex loci in finnish and somali populations. Forensic Sci Int Genet 6: e99-e102.