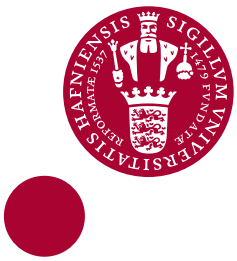

A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR

Pereira, Rui; Pereira, Vania; Gomes, Iva; Tomas Mas, Carmen; Morling, Niels; Amorim, António; Prata, Maria João; Carracedo, Angel; Gusmão, Leonor

Published in:
International Journal of Legal Medicine (Print)

DOI:
10.1007/s00414-011-0593-2

Publication date:
2012

Int J Legal Med (2012) 126:97–105
DOI 10.1007/s00414-011-0593-2

ORIGINAL ARTICLE

A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR

Rui Pereira · Vânia Pereira · Iva Gomes · Carmen Tomas · Niels Morling · António Amorim · Maria João Prata · Ángel Carracedo · Leonor Gusmão

Received: 22 April 2011 / Accepted: 16 June 2011 / Published online: 30 June 2011

Abstract Studies of human genetic variation predominantly use short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) but Insertion deletion polymorphisms (Indels) are being increasingly explored. They combine desirable characteristics of other genetic markers, especially the possibility of being analysed using short amplicon strategies, which increases the ease of analysis, contributing to justify their interest in population and forensic genetics. After the advent of autosomal and uniparental genomes (mtDNA and Y chromosome), these fields of research are also focusing on the X chromosome, given its special transmission pattern. The X chromosome markers brought new insights into the history of modern human populations and also proved useful in forensic kinship investigations, namely in deficient relationship cases and in cases where autosomes are uninformative. This work describes an X-Indel multiplex system amplifying 32 biallelic markers in one single PCR. The multiplex includes X-Indels shown to be polymorphic in the major human population groups and follows a short amplicon strategy. The set was applied in the genetic characterization of sub-Saharan African, European and East Asian population samples and revealed high forensic efficiency, as measured by the accumulated power of discrimination (0.9999990 was the lowest value in males and 0.999999999998 was the highest in females) and mean exclusion chance varied between 0.998 and 0.9996 in duos and between 0.99997 and 0.999998 in trios. Finally, a segregation analysis was performed using trio constellations of father–mother–daughters in order to address the transmission pattern and assess mutation rates of this type of markers.

Keywords Forensic genetics · Human identification · Insertion deletion polymorphism · Indel · Kinship testing · Multiplex PCR · X chromosome

Electronic supplementary material The online version of this article (doi:10.1007/s00414-011-0593-2) contains supplementary material, which is available to authorized users.

R. Pereira · V. Pereira · I. Gomes · A. Amorim · M. J. Prata · L. Gusmão (✉)
IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto,
Porto, Portugal
e-mail: lgusmao@ipatimup.pt

R. Pereira · Á. Carracedo
Institute of Forensic Sciences Luis Concheiro,
University of Santiago de Compostela,
Santiago de Compostela, Spain

V. Pereira · C. Tomas · N. Morling
Section of Forensic Genetics, Department of Forensic Medicine,
Faculty of Health Sciences, University of Copenhagen,
Copenhagen, Denmark

V. Pereira · A. Amorim · M. J. Prata
Faculty of Sciences of the University of Porto,
Porto, Portugal

Á. Carracedo
Genomic Medicine Group, CIBERER,
University of Santiago de Compostela,
Santiago de Compostela, Spain

Present Address:
I. Gomes
Institute of Legal Medicine,
University Hospital, University of Cologne,
Cologne, Germany

Introduction

Human genetic studies predominantly use short tandem repeats (STRs) or single nucleotide polymorphisms (SNPs)

to dissect patterns of genetic variation. Nevertheless, Insertion deletion polymorphisms (Indels) represent another kind of DNA variation that is being increasingly explored. After SNPs, Indels are the most abundant DNA polymorphisms [1]. Nearly 20% of all human genetic variations in the genome are Indels out of which 8% are biallelic [2]. In 2002, Weber et al. [2] reported the characterization of ~2.000 biallelic Indels and their allele frequencies in four human population groups (Europeans, Africans, Japanese and Native Americans). In a later important work, Mills et al. [3] described a map of human Indel variations with over 415,000 polymorphisms of this type.

In population and forensic genetic studies, the combination of desirable characteristics from both SNPs and STRs, especially the possibility of being analysed through short amplicons together with the ease of typing, justifies the interest in Indel polymorphisms, offering an additional tool in these fields [4]. The use of Indels is still in an early stage and not many studies have been performed, but a change of direction is becoming increasingly evident [4–10].

Concerning transmission properties, following autosomes and uniparental genomes (mtDNA and Y chromosome), the X chromosome is gaining significant importance in population and forensic genetic studies. The interesting characteristics of the X chromosome rely on its special transmission pattern: (i) it travels between both sexes in each generation, telling a different story from uniparental genomes; (ii) it has a reduced effective population size in relation to autosomes, making the X more sensitive to the effects of genetic drift and population substructure, which justifies its lower overall genetic diversity than autosomes; (iii) it only recombines in females, leading to an overall lower recombination rate in comparison with autosomes and consequent higher levels of linkage disequilibrium (LD); (iv) the hemizygous state in males allows direct access to haplotypes [11].

The combination of these features makes X chromosome markers unique tools to address general topics like the male and female migrations along the history of human populations or the differences in mutation and recombination patterns between sexes [11, 12]. Of additional interest in forensics, the X chromosome shows higher efficiency parameters than autosomes in specific kinship investigations involving mainly female offspring and it proves useful in reconstructing haplotypes in so-called deficient relationship cases [13–15]. For the abovementioned reasons, different types of X chromosome specific polymorphisms as X-STRs, X-SNPs and more recently X-Indels, have been applied in studying the genetic structure of human populations, assessing ancestry proportions in admixed populations as well as in forensic kinship investigations and identification studies [8, 9, 12, 16–19].

In summary, the increasing interest both in X chromosome and in Indels markers justifies the aim of this work that was to set up a general tool for population and forensic genetics analyses comprising a high number of X-Indels in a unique multiplex reaction using a short amplicon strategy. A set of Indel markers was selected to allow an easy access to X chromosome information, especially useful in complex kinship analysis in which autosomal markers provide limited information, and at the same time having in consideration the suitability of the method in the analysis of low quantity and/or quality DNA samples that are usual in forensic scenarios.

Materials and methods

Marker selection

Aiming to set up a simple and highly informative worldwide tool, we selected X-Indels reported to have high degree of polymorphism in the major human population groups from Africa, Europe and Asia according to the information available at dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and Marshfield Diallelic Insertion/Deletion Polymorphisms database (http://www.marshfieldclinic.org/mgs/). At present, a main drawback when working with Indels is that there is incomparably less data available in relation to SNPs. As a result, few candidates fulfilled the initial predicted criteria of previously validated X-Indels with minimum allele frequency (MAF)≥0.20 in those population groups. In order to keep a reasonable initial pool of Indels to build the multiplex reaction, we opted for a compromise solution and conceded markers with MAF≥ 0.10 in Asians to be maintained as candidates. The X-Indels included in the assay are presented in Fig. 1 and Table 1.

Development of the multiplex reaction

After obtaining a pool of candidate X-Indels fulfilling the abovementioned criteria, we proceeded with the in silico multiplex design, following essentially the same workflow as in the work of Pereira et al. [4]. At the end of the process, 32 Indels distributed along the X chromosome were successfully multiplexed in a single reaction. Thereafter, all markers were schematically arranged by expected amplicon lengths and assigned to four different fluorochromes (6FAM, VIC, NED, PET) (Applied Biosystems) in a way allowing the genotyping through a single capillary electrophoresis (CE). When necessary, a G nucleotide or a "PIGtail" [20] was added to the unlabelled primer aiming a dual function of adjusting amplicon sizes while at the same time promoting non-template adenylation of all products in order to obtain a uniform electropherogram. The unlabelled

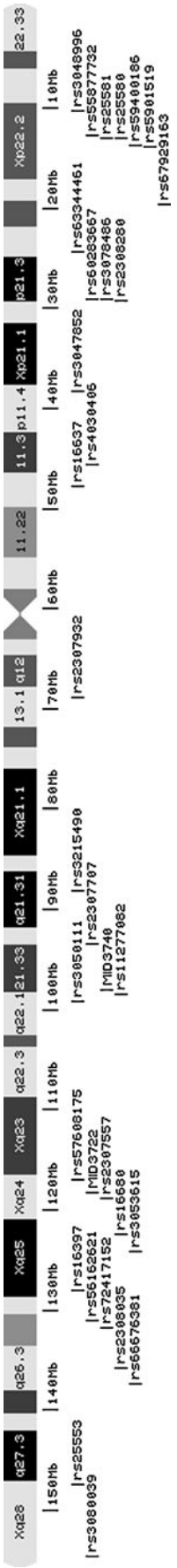

Fig. 1 Ideogram of the X chromosome indicating the location of the 32 selected Indel markers

primers of amplicons not needing mobility adjustments remained unchanged or were added a G at their 5′ end if there was none. In cases requiring small mobility adjustments, a 5′ tail GTTT or partial tails of the original sequence GTTTCTT were attached to the unlabelled primer, following the recommendations of Brownstein et al. [20] (Table 1).

Amplification and genotyping

The amplification of the 32 X-Indels was performed in a single multiplex PCR using Qiagen Multiplex PCR kit (Qiagen) at 1× Qiagen multiplex PCR master mix, 0.1 μM of all primers (sequences detailed in Table 1) except for MID3719, MID2089, MID3774 and MID3727 (0.2 μM) and 0.3–5 ng of genomic DNA in a 10 μL final reaction volume. Thermal cycling conditions consisted of an initial step at 95°C for 15 min; 30 cycles at 94°C for 30 s, 60°C for 90 s, and 72°C for 45 s; and a final extension at 72°C for 60 min.

The PCR products were prepared for subsequent analysis adding 1 μl of amplified product to 10 μl Hi-Di™ Formamide (Applied Biosystems) and 0.3 μl of internal size standard GeneScan™ 500 LIZ® (Applied Biosystems). CE and detection were performed in a 3130 Genetic Analyzer using filter set G5 and POP-7™ (Applied Biosystems), except for the Somali samples, which were typed using a 3130×l Genetic Analyzer and POP-4™ (Applied Biosystems). The resulting electropherograms were analysed and genotypes were assigned with GeneMapper v4.0 (Applied Biosystems).

Population samples

In this work we evaluated the genetic diversity of 32 X-Indels in various populations from Africa, Europe and East Asia including unrelated male and female samples from: Angola (74 males), Mozambique (42 males), Somalia (54 males and 54 females), Portugal (108 males and 108 females) and Macau (25 males and 25 females). For segregation analysis a total of 214 independent father/mother/daughter trios were also studied (106 from Somalia and 108 from Portugal).

Statistical analysis

Statistical analysis including estimation of allele and haplotype frequencies, expected heterozygosities, exact tests of Hardy–Weinberg equilibrium (HWE), pairwise F_{ST} and LD tests were assessed using Arlequin v3.5.1.2 [21]. HWE was evaluated in unrelated females from the population samples. LD analysis was performed using only male haplotypes in Angola, Mozambique and Macau. For

Table 1 X chromosome Indels and PCR primer sequences used in the multiplex

MID	rs number	Position (bp)[a]	Alleles	Forward primers[b]	Reverse primers[b]
MID2612	rs3048996	10234839	−/ATC	*g*ACCCACGGTGTTGAATTCAG	NED-CACAGCACCAGGAAAATAGC
MID3712	rs55877732	12572196	−/GAA	*gttt*AGTCTTGCTGCAATGTACCC	VIC-TTCAAGGGCAATGATGTTTG
MID357	rs25581	12912861	−/TGAGA	*g*TTTTATAGACTGTGGCCCCC	PET-GTTAGTGGTTGGATTGCTCG
MID356	rs25580	12918048	−/CTT	*gttt*CCAACTCCACGTGAGAAATG	PET-AGTCTGATGCAGTGGCAAAC
MID3703	rs59400186	13711300	−/GTTA	VIC-AGCTTCCAAGTAGTTCTGCC	*g*TTTGGCTTACTTCCTCCTCC
MID3774	rs5901519	13809000	−/ACC	*g*AAGACGGGAATTGAGTCACC	NED-TTTTTGTGCACAGGCACTCC
MID3692	rs67929163	19516252	−/CATAT	6FAM-ACATAAAAGCAAGCTTTGGC	*gtttctt*CCCGGTGTGTGAACTTTTTC
MID3716	rs63344461	24235114	−/GAG	6FAM-AAAGGGAGCATCTACTCCAG	*gtttct*AGGGCAATCCAGAATTGGAC
MID3690	rs60283667	28984076	−/TCAC	GGGCACCATATTAGGCATGT	VIC-CCCACCATCTAACCCATTTC
MID3719	rs3078486	29040938	−/TTAACT	*g*TTCTTTCTCATCTGGCACCC	VIC-CTATGAAGCCTATAGATTGG
MID2089	rs2308280	29157973	−/TTA	VIC-AATCCATTCTGGAATAAGATGTCA	*gtttctt*TCCACTCTCAGGGATTCCTTT
MID2692	rs3047852	38262701	−/ATT	*gtttctt*CAAGTTCATATGGTGTCTTGG	PET-TGCATTACACAGAGCAACTC
MID3701	rs4030406	45539201	−/ATTA	*gtttct*AGTTGGAGATGCAATGAAGC	NED-AGAGACAGGTGAATTGAGGC
MID198	rs16637	47680386	−/CAACCAAT	6FAM-CAGGCACAGGAGAGGAAGAG	*g*TCCACCCCTAGTTAAACAGC
MID1736	rs2307932	68733480	−/ATA	VIC-GTGAAAGGTGAGCTTGTCTG	*gtttct*AGGCCTTTTTGGTTAACTGG
MID3730	rs3215490	88009689	−/GACA	6FAM-AGGATCCTGACTAAGATAGC	GAATCTCTGGAAACACTTGG
MID1511	rs2307707	93392006	−/GTCT	*g*CTGCCTGGGATTTTTCCTTT	PET-CAGGGGAGAACACCCACTAA
MID3740	–	97906546	−/GT	*gtttct*ACTTGCTTTGCTTTTCCCTC	6FAM-GTACAACTGCAAGGAACRAG[c]
MID3732	rs11277082	98331815	−/ACCTCACTCA	VIC-CAGAGTCATCTATTCCCCAG	*gtttctt*CACCCATGTGGTTTCATTTC
MID3727	rs3050111	99165489	−/TT	*gtttct*GGTGGAATTCTTTTGCATGTG	PET-TTTTGGGAAGCACACTCACC
MID3754	rs57608175	116901987	−/GGTCATCACGAG	NED-TTTCACCAAGGACTTGAAAGG	*gtt*CAGCTCACACTAGGGCCTTC
MID3722	–	118156157	−/AAAGTGTACACAT	VIC-TGGCCCTTCTGAGTTCAAAC	*g*CAGTGTAATAAGGTGGGAGC
MID1361	rs2307557	118748515	-/ACA	6FAM-TCAGTCCTTAACAAGGGAGC	*gttt*GTCATTGTGAAGGCTACCTG
MID243	rs16680	122370414	-/TGT	PET-TACAGTTGGCTGCTTTTCCC	gtttcttATACGAAGATCTGTGGGAAC
MID2637	rs3053615	124135529	−/CT	*gtttct*TATGTGTCAAAATGGGAGGC	6FAM-AATCCTCAAATCACAGTGGC
MID111	rs16397	127958384	−/GTG	GAGGCAGGGAAATCAGTTAG	NED-TTGATTCCAGCTTTCCCTTT
MID3736	rs56162621	130975547	−/CT	6FAM-GGGTTAGGAGCCCCTGCT	*gtttctt*GGATGTATGACACACAACGC
MID3753	rs72417152	131760172	−/GTATAT	GCTACACCAATGGACAGATG	PET-TGTGGTGTGCATGATTTGC
MID1839	rs2308035	135695920	−/CA	*gtt*GATATCCCATAACGCCCATTT	NED-TCCTTTTGCTACGCAGACCT
MID3760	rs66676381	137369795	−/TTAAA	*g*CAAGGTTCGTACTCATTTAG	NED-AACCTAGTTCACAACCCCTG
MID329	rs25553	147393784	−/TACTCT	*g*TCTCAAAACCTTCCCATGGC	6FAM-AGAAGTTAGAGGGTGTCTGG
MID2652	rs3080039	154561961	−/TAA	PET-GCTGCTCTTTGCTTTAATTTC	*g*TATGGTAGGCACTGTGCTAA

[a] Location in the X chromosome according to UCSC Table Browser, table 131

[b] Lowercase italic letters represent nucleotide tails added to the primers

[c] R indicates a degenerate position G/A

Somalia and Portugal, the additional information on trios allowed inferring the gametic phase in daughters, and consequently to increase the number of haplotypes used in the LD analysis.

The power of discrimination (PD) in females and males and the mean exclusion chance (MEC) in trios or duos were calculated using the formulas of Desmarais et al. [22].

Results and discussion

Thirty-two short-amplicon X-Indels in one multiplex

In this work, we developed a simple X-Indel multiplex allowing the genotyping of 32 bi-allelic markers with high heterozygosities in distinct populations, performing only one PCR followed by one CE run. Furthermore, the system uses short amplicons (<160 bp) that are especially suitable for the analysis of degraded samples. Figure 2 shows an example of an electropherogram obtained with the X-Indel system for the female reference sample 9947A (Promega).

With respect to other currently existing X-Indel sets (e.g., [8, 9]), the system implemented in this work has the advantages of maximizing the multiplexing level in a single reaction, similar to that achieved by Freitas et al. [18], while at the same time employing a short amplicon strategy more adequate to challenging samples in the forensic context.

Genetic diversity in human populations

To evaluate the overall genetic diversity of the 32 X-Indels in different human populations, we studied samples from

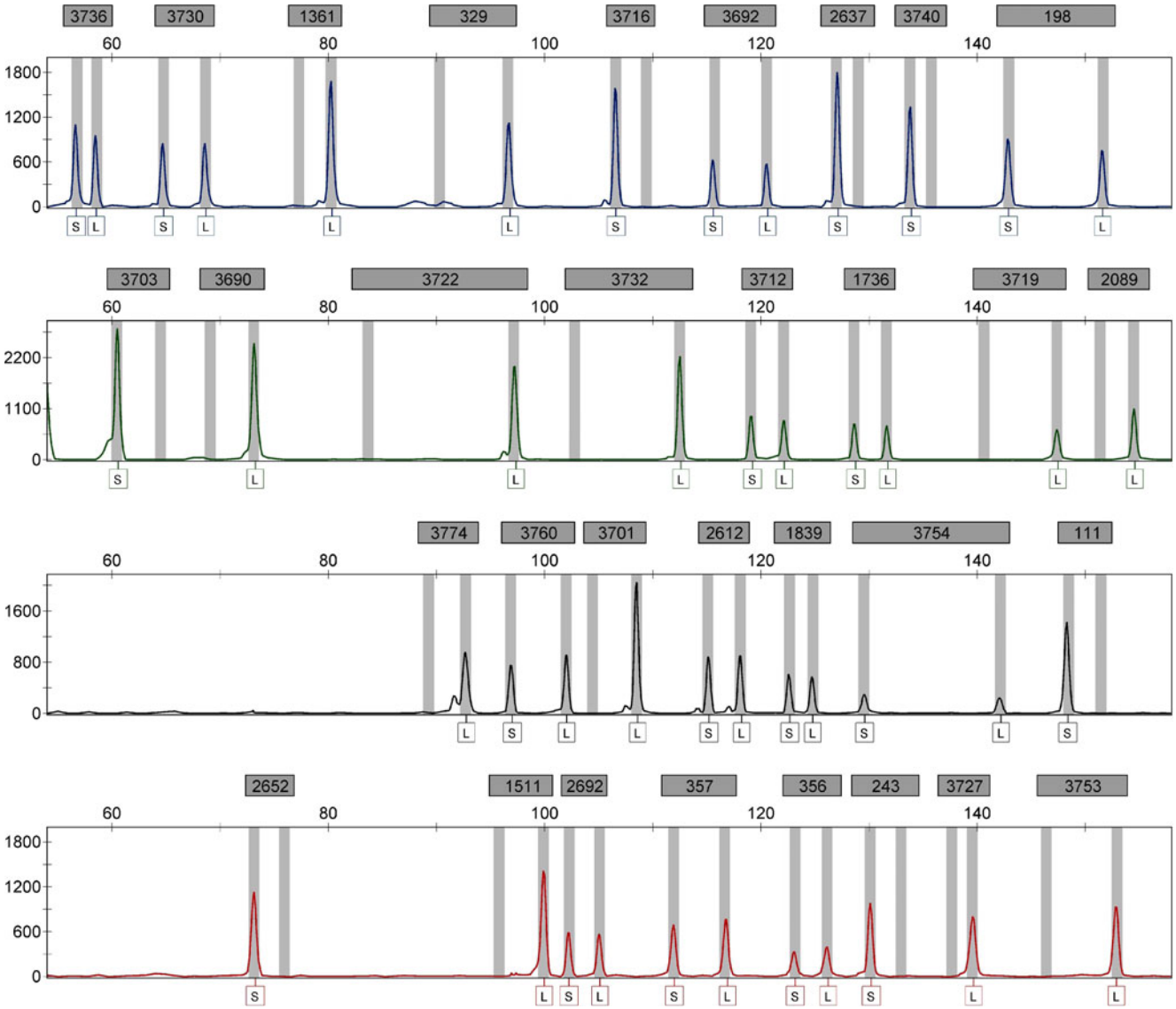

Fig. 2 Electropherogram of the female reference sample 9947A obtained with the 32 X-Indel multiplex

different continents. Table 2 presents the allele frequency distributions observed in African, European and East Asian population samples as well as the expected heterozygosity values. The allele frequencies for Somalia, Portugal and Macau were calculated after pooling the data obtained for female and male samples, since no significant differences were found between sexes ($F_{ST} \leq 0.000$; $p \geq 0.43491$) for this set of markers. The samples from Angola and Mozambique were also pooled (Ang+Moz) as they revealed to be genetically homogeneous in a preliminary analysis ($F_{ST} = 0.008$; $p = 0.06702$). As expected, all selected markers are polymorphic in the populations studied and overall they show the highest genetic diversities in Africans (mean heterozygosity was 0.435 in Somalia; 0.408 in Ang+Moz) followed by Europeans (0.388) and East Asians (0.347).

No significant deviation from HWE expectations was detected in the distribution of female genotypes in each population. The unique p value below 0.05 was found for MID329 in Somalia ($p = 0.04370$), which was not significant after Bonferroni correction for multiple tests.

Regarding the pairwise population comparisons based on the 32 X-Indels, as expected, highly significant differences

Table 2 Allele frequencies and expected heterozygosity of the 32 X-Indels in population samples from African, Europe and East Asia

MID	rs number	Short allele frequency				Expected heterozygosity				
		Somalia	Ang+Moz	Portugal	Macau	Somalia	Ang+Moz	Portugal	Macau	Total
MID3736	rs56162621	0.704	0.565	0.676	0.340	0.420	0.496	0.439	0.453	0.474
MID3730	rs3215490	0.185	0.250	0.293	0.510	0.304	0.378	0.416	0.505	0.414
MID1361	rs2307557	0.488	0.379	0.170	0.220	0.503	0.475	0.283	0.347	0.408
MID329	rs25553	0.284	0.310	0.284	0.140	0.409	0.432	0.408	0.243	0.393
MID3716	rs63344461	0.444	0.457	0.664	0.440	0.497	0.501	0.448	0.498	0.496
MID3692	rs67929163	0.315	0.448	0.256	0.060	0.434	0.499	0.382	0.114	0.398
MID2637	rs3053615	0.883	0.888	0.904	0.920	0.208	0.201	0.174	0.149	0.182
MID3740	–	0.346	0.379	0.410	0.120	0.455	0.475	0.485	0.213	0.455
MID198	rs16637	0.556	0.560	0.500	0.710	0.497	0.497	0.502	0.416	0.495
MID3703	rs59400186	0.327	0.233	0.315	0.420	0.443	0.360	0.433	0.492	0.435
MID3690	rs60283667	0.241	0.426	0.315	0.180	0.368	0.493	0.433	0.298	0.418
MID3722	–	0.302	0.328	0.364	0.050	0.425	0.444	0.465	0.096	0.420
MID3732	rs11277082	0.327	0.362	0.278	0.040	0.443	0.466	0.402	0.078	0.394
MID3712	rs55877732	0.346	0.440	0.204	0.392	0.455	0.497	0.325	0.482	0.422
MID1736	rs2307932	0.253	0.147	0.503	0.520	0.380	0.252	0.502	0.504	0.476
MID3719	rs3078486	0.302	0.357	0.165	0.404	0.425	0.463	0.276	0.487	0.386
MID2089	rs2308280	0.426	0.310	0.349	0.540	0.492	0.432	0.456	0.502	0.475
MID3774	rs5901519	0.309	0.517	0.173	0.530	0.429	0.504	0.287	0.503	0.430
MID3760	rs66676381	0.784	0.836	0.855	0.919	0.341	0.276	0.249	0.150	0.263
MID3701	rs4030406	0.562	0.586	0.460	0.410	0.495	0.489	0.498	0.489	0.501
MID2612	rs3048996	0.500	0.362	0.525	0.280	0.503	0.466	0.500	0.407	0.497
MID1839	rs2308035	0.488	0.474	0.793	0.884	0.503	0.503	0.329	0.207	0.435
MID3754	rs57608175	0.278	0.474	0.164	0.663	0.404	0.503	0.274	0.451	0.430
MID111	rs16397	0.426	0.345	0.599	0.747	0.492	0.456	0.482	0.382	0.498
MID2652	rs3080039	0.543	0.440	0.738	0.770	0.499	0.497	0.388	0.358	0.457
MID1511	rs2307707	0.389	0.190	0.707	0.390	0.478	0.310	0.416	0.481	0.501
MID2692	rs3047852	0.549	0.284	0.688	0.394	0.498	0.411	0.430	0.482	0.496
MID357	rs25581	0.259	0.095	0.306	0.810	0.386	0.173	0.426	0.311	0.444
MID356	rs25580	0.259	0.130	0.355	0.808	0.386	0.229	0.459	0.313	0.461
MID243	rs16680	0.753	0.888	0.735	0.960	0.374	0.201	0.391	0.078	0.325
MID3727	rs3050111	0.426	0.405	0.133	0.310	0.492	0.486	0.231	0.432	0.395
MID3753	rs72417152	0.395	0.103	0.133	0.110	0.481	0.187	0.231	0.198	0.302
					Mean	0.435	0.408	0.388	0.347	0.424

Ang Angola, *Moz* Mozambique

Table 3 Pairwise F_{ST} estimates between populations from Africa, Europe and East Asia obtained for the 32 X-Indel set (F_{ST} values below diagonal) and their respective p values (above diagonal)

	Somalia	Ang+Moz	Portugal	Macau
Somalia	–	$\leq 10^{-5}$	$\leq 10^{-5}$	$\leq 10^{-5}$
Ang+Moz	0.030	–	$\leq 10^{-5}$	$\leq 10^{-5}$
Portugal	0.053	0.104	–	$\leq 10^{-5}$
Macau	0.132	0.159	0.131	–

Ang Angola, *Moz* Mozambique

exist between continental populations (Table 3). A significant genetic distance was also detected between the two samples from sub-Saharan Africa, with Somalia presenting a lower distance from Europeans than Ang+Moz. The genetic distances captured by the X chromosome were higher than those obtained with autosomal Indels for the same population groups [4], except between Africans and Europeans that surprisingly showed a similar F_{ST}. Due to the reduced effective population size, higher differentiation among human populations is expected for X chromosome markers in comparison to the autosomes [11, 12]. Nevertheless, for the Indel markers included in this study, this effect was only noticed in relation with East Asians. Also, it is noticeable that the East Asian sample from Macau shows overall lower genetic diversity than Europeans or Africans, probably as a result of relaxing the marker selection criteria for this group (MAF ≥0.10), and therefore an ascertainment bias can also be contributing to the observed results.

Linkage disequilibrium

When studying markers located in the same chromosome, it is essential to perform LD analyses in order to evaluate how to properly construct genetic databases and assess forensic parameters in the reference populations.

In this study, the pairwise LD analyses revealed a significant association between MID357 and MID356 in all populations, even after Bonferroni correction for multiple testing ($p \leq 0.05/496$). This observation was expected given their very close position on the X chromosome (5.2 kb apart; see Table 1 and Fig. 1) and is in agreement with previous studies [18]. From the explained reasons, MID357 and MID356 must be treated as an haplotype when used together.

Significant p values were also observed for the pairs MID3690–MID3719 in Angola and Mozambique pooled sample, MID3719–MID2089 in Somalia and Portugal, and MID3690–MID2089 in Portugal. Interestingly, the tests performed between the loci pairs MID3690–MID3719 and MID3719–MID2089 consistently revealed low p values in all populations ($p \leq 0.02476$), indicative of a true association between these markers. In fact, MID3690–MID3719–MID2089 form a closely located cluster on the chromosome, spanning only ~170 kb, thus reinforcing the idea that they form an LD block. The present results indicate that, as for MID357–MID356, it is also recommended to treat MID3690–MID3719–MID2089 as a haplotype block and use haplotype frequencies in forensic statistical analyses. For convenience, Table S1 includes the frequencies observed for the two haplotypic blocks in the studied populations. Further studies involving different and larger samples need to be carried out to better understand the LD patterns in the different population groups.

Finally, it is worth mentioning that although no LD could be observed between other loci, some markers are closely linked and, therefore, non-independent transmission must be considered in deficiency cases with more than two generations.

Statistical parameters of forensic efficiency

The statistical parameters indicative of forensic efficiency of the X-Indel set are summarized in Table 4 and detailed in Table S2 for each marker and haplotype. The accumulated values for the set were calculated using the single locus information, except for those markers presenting association, which were treated as haplotypes (using the frequencies in Table S1).

The values obtained for PD were high in females (between 0.9999999995 in the least diverse population from Macau and 0.999999999998 in the most informative sample from Somalia) and also in males (0.9999990 to 0.99999998) taken into account their hemizygous state. The combined MEC varied between 0.99997 and 0.999998 in trios and between 0.998 and 0.9996 in duos. Overall, the accumulated PD and MEC achieved with the 32 X-Indel set are high, similar to other existent X-STR [16, 23–25], X-SNPs [17, 19] and X-Indel sets [18], therefore confirming its utility in human identification and kinship investiga-

Table 4 Statistical parameters indicative of forensic efficiency obtained for the 32 X-Indel set: accumulated power of discrimination (PD) and mean exclusion chance (MEC)

	Somalia	Ang+Moz	Portugal	Macau
PD_{Males}	0.99999998	0.99999995	0.9999997	0.9999990
$PD_{Females}$	0.999999999998	0.999999999993	0.99999999997	0.9999999995
MEC_{Duos}	0.9996	0.9995	0.9990	0.998
MEC_{Trios}	0.999998	0.999996	0.999990	0.99997

Ang Angola, *Moz* Mozambique

tions. Furthermore, due to the lower mutation rates of Indels, this set may be useful as a complementary tool to STR typing, particularly in cases showing few transmission incompatibilities with microsatellites [26, 27]. Likewise, applying simultaneously X-STRs and X-Indels to improve forensic efficiency may be a strategy to consider, provided that careful LD analyses are performed to evaluate how to deal with the data.

Segregation analysis

In order to analyse the parents/daughters segregation of the selected X-Indel markers, a total of 214 father/mother/daughter trios were studied. All markers showed the expected transmission pattern, except for MID3740 for which a second-order mother/daughter "exclusion" was found. Apparent opposed homozygosity could be explained either by mutation or a null allele, usually the result of a mutation at one of the primer binding sites. Subsequently, a new PCR was performed for these samples at a lower annealing temperature (54°C), resulting in a change of the mother's previously observed phenotype, now showing a heterozygote pattern for MID3740. In addition, sequencing analysis of the flanking region including the primer annealing sites revealed a C-to-T transition at the antepenultimate base of the 3′ end of the reverse primer.

Therefore, no mutations have been detected for the 32 studied markers in a total of 428 meiosis, corresponding to an observed mutation rate lower than 1 in 13,696 allele transmissions. Although a much larger sample is needed to improve the estimates, the present analysis allowed reaching a 95% upper bound limit of 2.19×10^{-4} for the mutation rate of these Indels, by applying Nelson's formula, $1-\alpha^{1/n}$ (for $\alpha=0.05$ and $n=13{,}696$), the same one previously used by Weir [28] to calculate the frequency of genotypes not observed in a database of size n.

Conclusions

This work describes a simple and informative X-Indel multiplex for population and forensic genetics studies. Through a single PCR and single CE it allows the genotyping of 32 bi-allelic markers distributed along the X chromosome showing high heterozygosity in distinct population groups. Furthermore, it uses short PCR fragments which are especially suitable for the analysis of degraded samples. The combination of a short amplicon strategy, simplicity of analysis and good multiplexing capacity in a single reaction constitutes one step further in relation to currently existing X-SNP and X-Indel assays.

Acknowledgements This work was partially supported through PhD grants to R.P. (SFRH/BD/30039/2006) and V.P. (SFRH/BD/70881/2010) awarded by the Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Social Fund (Human Potential Thematic Operational Programme). IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT. The assistance of Fundación Botín is also acknowledged with appreciation.

References

1. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409(6822):928–933. doi:10.1038/35057149
2. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71(4):854–862. doi:10.1086/342727
3. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16 (9):1182–1190. doi:10.1101/gr.4565806
4. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30(21):3682–3690. doi:10.1002/elps.200900274
5. Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, Shigeta R, Silva G, Patel PI, Belmont JW, Seldin MF (2005) Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. Hum Genet 118(3–4):382–392. doi:10.1007/s00439-005-0012-1
6. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005) Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1(6):e70. doi:10.1371/journal.pgen.0010070
7. Bastos-Rodrigues L, Pimenta JR, Pena SD (2006) The genetic structure of human populations studied through short insertion–deletion polymorphisms. Ann Hum Genet 70(Pt 5):658–665. doi:10.1111/j.1469-1809.2006.00287.x
8. Ribeiro-Rodrigues EM, Santos NP, Ribeiro-dos-Santos AK, Pereira R, Amorim A, Gusmão L, Zago MA, Santos SE (2009) Assessing interethnic admixture using an X-linked insertion–deletion multiplex. Am J Hum Biol 21(5):707–709. doi:10.1002/ajhb.20950
9. Edelmann J, Hering S, Augustin C, Szibor R (2009) Indel polymorphisms—An additional set of markers on the X-chromosome. Forensic Sci Int Genet Supplement Series 2 (1):510–512. doi:10.1016/j.fsigss.2009.08.148
10. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AK, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, Santos SE (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion–deletion (INDEL) ancestry-informative marker (AIM) panel. Hum Mutat 31(2):184–190. doi:10.1002/humu.21159
11. Schaffner SF (2004) The X chromosome in population genetics. Nat Rev Genet 5(1):43–51. doi:10.1038/nrg1247

12. Casto AM, Li JZ, Absher D, Myers R, Ramachandran S, Feldman MW (2010) Characterization of X-linked SNP genotypic variation in globally distributed human populations. Genome Biol 11(1): R10. doi:10.1186/gb-2010-11-1-r10
13. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117(2):67–74. doi:10.1007/s00414-002-0352-5
14. Szibor R, Plate I, Edelmann J, Hering S, Kuhlisch E, Michael M, Krause D (2003) Chromosome X haplotyping in deficiency paternity testing principles and case report. Int Congr Ser 1239:815–820. doi:10.1016/s0531-5131(02)00569-1
15. Szibor R (2007) X-chromosomal markers: past, present and future. Forensic Sci Int Genet 1(2):93–99. doi:10.1016/j.fsigen.2007.03.003
16. Gomes I, Prinz M, Pereira R, Meyers C, Mikulasovich RS, Amorim A, Carracedo A, Gusmão L (2007) Genetic analysis of three US population groups using an X-chromosomal STR decaplex. Int J Legal Med 121(3):198–203. doi:10.1007/s00414-006-0146-2
17. Tomas C, Sanchez JJ, Barbaro A, Brandt-Casadevall C, Hernandez A, Ben Dhiab M, Ramon M, Morling N (2008) X-chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic homogeneity except in North-west Africans (Moroccans). BMC Evol Biol 8:75. doi:10.1186/1471-2148-8-75
18. Freitas NS, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NP, Ribeiro-dos-Santos AK, Santos SE (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. Int J Legal Med 124(6):589–593. doi:10.1007/s00414-010-0441-9
19. Tomas C, Sanchez JJ, Castro JA, Borsting C, Morling N (2010) Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set. Transfusion (Paris) 50(10):2258–2265. doi:10.1111/j.1537-2995.2010.02696.x
20. Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20 (6):1004–1006, 1008–1010
21. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564–567. doi:10.1111/j.1755-0998.2010.02847.x
22. Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J Forensic Sci 43(5):1046–1049
23. Pereira R, Gomes I, Amorim A, Gusmão L (2007) Genetic diversity of 10 X chromosome STRs in northern Portugal. Int J Legal Med 121(3):192–197. doi:10.1007/s00414-006-0144-4
24. Gomes I, Alves C, Maxzud K, Pereira R, Prata MJ, Sánchez-Diz P, Carracedo A, Amorim A, Gusmão L (2007) Analysis of 10 X-STRs in three African populations. Forensic Sci Int Genet 1 (2):208–211. doi:10.1016/j.fsigen.2007.01.001
25. Gusmão L, Sánchez-Diz P, Alves C, Gomes I, Zarrabeitia MT, Abovich M, Atmetlla I, Bobillo C, Bravo L, Builes J, Caine L, Calvo R, Carvalho E, Carvalho M, Cicarelli R, Catelli L, Corach D, Espinoza M, Garcia O, Malaghini M, Martins J, Pinheiro F, Joao Porto M, Raimondi E, Riancho JA, Rodriguez A, Rodriguez Cardozo B, Schneider V, Silva S, Tavares C, Toscanini U, Vullo C, Whittle M, Yurrebaso I, Carracedo A, Amorim A (2009) A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. Int J Legal Med 123(3):227–234. doi:10.1007/s00414-008-0309-4
26. Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, Carracedo A, Lareu MV (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Sci Int Genet 2(3):198–204. doi:10.1016/j.fsigen.2008.02.002
27. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L (2009) Insertion/deletion polymorphisms: a multiplex assay and forensic applications. Forensic Sci Int Genet Suppl Ser 2(1):513–515. doi:10.1016/j.fsigss.2009.09.005
28. Weir BS (1992) Independence of VNTR alleles defined as fixed bins. Genetics 130(4):873–887