



单管直扩38重祖先信息InDels体系的构建及其在微流控芯片系统的应用*

韩俊萍^{1,2)} 赵蕾¹⁾ 王庆国³⁾ 江丽¹⁾ 谢何鑫¹⁾庄斌⁴⁾ 赵丽健⁴⁾ 李唐松²⁾ 吴瑾²⁾ 李彩霞^{1)**}

¹⁾ 公安部物证鉴定中心法医遗传学公安部重点实验室, 北京市现场物证检验工程技术研究中心, 现场物证溯源技术国家工程实验室, 北京 100038;

²⁾ 北京市公安局朝阳分局, 北京 100025; ³⁾ 山东第一医科大学, 济南 250000;

⁴⁾ 北京博奥晶典生物技术有限公司, 北京 101111)

摘要 本文基于插入-缺失多态性 (insertion-deletion polymorphism, InDel) 遗传标记, 从相关文献和 dbSNP 库中筛选出 38 个在东亚、欧洲、非洲人群中具有等位基因频率差异的祖先来源 InDel 位点, 同时整合 Amelogenin 位点和 Y 染色体 STR (DYS439) 位点, 以辅助未知样本的性别鉴定, 采用 PCR-CE 技术构建了可以区分三大洲际人群的单管直接扩增复合检测体系, 该体系可与微流控芯片系统相结合, 1.7 h 内完成 DNA 样本自动化扩增和电泳分型. 利用公共数据库 1000Genomes 中 16 个人群 1 607 名个体的分型数据对筛选的 38 个 InDel 位点进行初步评估; 同时对构建的 38-plex InDels 体系的灵敏度、准确性、直接扩增能力进行验证评估; 检测 5 种不同人群的 779 份样本和 215 份唾液卡、血卡的直接扩增样本, 采用聚类分析和主成分分析方法, 评价体系的推断准确性. 结果表明该 38-plex InDels 复合检测体系分型准确, 灵敏度达 157 pg, 能够对唾液卡和血卡直接扩增, 可以区分三大洲际人群及其混合人群, 能够准确推断待测样本的族群来源、估算祖先成分, 且通过集成细胞裂解步骤将有望 2 h 实现“样本进-结果出”式快速自动化 InDel 分型.

关键词 法医物证学, 插入-缺失多态性, 祖先信息位点, 微流控芯片

中图分类号 R89, D919.2

DOI: 10.16476/j.pibb.2020.0020

在人类基因组中存在一些遗传标记, 其等位基因频率在不同人群中的分布有明显差异^[1], 在法医分析中选取一组这样的遗传标记位点可以对某人群/个体的遗传成分构成进行分析, 对个体族群来源进行推断^[2-3]. 插入/缺失多态性 (insertion-deletion polymorphism, InDel) 在基因组中含量广泛, 每 7.2 Kb 就至少有一个 InDel 位点^[4], 其结合了 SNP 和 STR 的优势: a. 突变率较低且相当稳定^[5]; b. 表现为二等位基因, 本质上仍属于长度多态性遗传标记, 可通过常规 PCR-CE 实现检测, 与现有法医基因分型系统相兼容; c. 易于从短片段中获得分型, 适于降解 DNA 样本的分析; d. InDel 在距离较远的群体之间能显示出等位基因频率分布差异^[6-7]. 基于上述特点, InDel 位点不仅可以用于法医学个体识别, 也可以用于生物地理祖先来源推

断, 因而受到了国内外法医研究者的更多关注, 相继建立了多种 InDel 分型体系. 在个体识别方面, 早在 2002 年, Weber 等^[8] 发现并定义了 2 000 个人类二等位基因 InDel 标记. 2006 年, Mills 等^[4] 完成了一个具有里程碑意义的研究——绘制出 InDel 变异草图, 通过后续深入研究^[9], 使这些标记成为法医学应用的焦点. 2009 年, Pereira 等^[10] 选取 38 个常染色体 InDel 标记, 构建一套用于东亚、欧洲以及非洲三大群体法医学个体识别的检测体系, 位点的随机匹配概率达 10^{-15} ~ 10^{-14} , 之后不少学者利

* 国家自然科学基金 (81772027) 和中央级公益性科研院所基本科研业务费专项资金计划 (2018JB037) 资助项目.

** 通讯联系人.

Tel: 010-63262392, E-mail: licaixia@tsinghua.org.cn

收稿日期: 2020-01-14, 接受日期: 2020-06-10

用该体系开展了一系列相关研究^[11-14]. 国内学者赵书民等^[15]、李成涛等^[16]建立了适于中国民族的InDel个体识别检测体系, 如InDel_typer30复合扩增体系, 用于中国汉族、回族、维吾尔族、蒙古族、藏族等5个民族DNA鉴定, 位点的随机匹配概率达 10^{-11} . 在生物地理族群来源推断方面, 早在2006年, Bastos-Rodrigues等^[17]筛选了40个二等位基因InDel标记, 研究了世界人群HGDP-CEPH多态性, 结果表明该40个位点能将世界范围内群体多态性划分为5个遗传结构(美洲、非洲、东亚、大洋洲及欧洲-中东-中亚), 与其主要地理区域分布相一致. 2012年, Pereira等^[18]筛选了46个祖先信息InDel位点, 构建一个用于区分非洲、欧洲、东亚及美洲人群及评估混合人群的遗传成分. Zaumsegel等^[19]建立了包含21个InDel位点, 可用于三大洲际人群(东亚、欧洲、非洲)族群来源推断的检测体系. 除了区分洲际人群, 将上述两个体系共计67个位点组合使用, 能够实现对亚人群—中南亚人群的准确区分^[20]. 目前, 已有研究将物理位置接近的2~3个InDel位点当作一个位点, 即multi-InDel标记, 也使得InDel标记对人群结构区分、祖先来源推断^[21-22]等遗传分析的效率进一步提高^[23].

本研究主要针对非洲、东亚和欧洲人群的遗传结构特点, 筛选出38个祖先信息InDel位点, 同时整合Amelogenin位点和Y染色体STR(DYS439)位点, 基于PCR-CE构建了38-plex InDels检测体系, 对该体系的准确性、灵敏度、直接扩增能力进行验证, 通过1000Genomes数据库中的参考群体及本实验收集不同群体样本的检测, 评估该体系对三大洲际人群的区分效能. 上述研究均是基于传统实验室操作, 为了实现快速一体化InDel检测, 本研究又将该体系与微流控芯片系统集成, 以自动化完成DNA样本扩增和电泳分型, 提高InDel分型效率.

1 材料与方法

1.1 实验样本

本研究选取公共数据库1000Genomes中的16个人群1607份样本, 检测样本包括5个人群779份样本和用于直接扩增的215份唾液卡和血卡, 其中3个人群615份和215份直接扩增样本来源于国家科技资源共享服务平台计划项目(YCZYPT

[2017] 01-3); 2个人群164份DNA样本来源于北德克萨斯大学, 详细信息见表1. 人类基因组DNA标准品9947A、9948和K562(初始浓度为10 mg/L), 购买自Promega公司. 所有样本对象均签署知情同意书. 本研究已通过公安部物证鉴定中心伦理委员会的伦理审查.

1.2 方法

1.2.1 DNA提取和定量

采用QIAamp® DNA Blood Mini Kit试剂盒(Qiagen公司, 德国)提取上述测试样本DNA, 采用NanoDrop 2000c Spectrophotometer(Thermo公司, 美国)进行DNA定量, 以去离子灭菌水稀释浓度至1~2 mg/L备用. 将DNA标准品9947A和9948进行梯度稀释, 使用Quantifiler™人类DNA定量试剂盒(Life Technologies公司, 美国)进行定量, 用于灵敏度测试.

1.2.2 InDel位点筛选

参考相关文献报道^[18-19, 24]和dbSNP库, 通过分析1000Genomes库中与非洲(AFR)、东亚(EAS)、欧洲(EUR)人群表型相关的InDels位点, 根据以下原则筛选构建复合体系的38个InDel位点: a. 位于常染色体的二等位基因位点; b. 等位基因长度2~15 bp; c. 两个位点之间的最小物理距离>5 Mb, 所选位点不连锁; d. 所选位点符合Hardy-Weinberg平衡; e. 三大群体两两人群间的等位基因频率差异值 $\delta > 0.25$; f. 位点组合可以进行单管复合PCR.

1.2.3 InDel位点区分效果评价

以1000Genomes数据库中16个人群(1607份样本)作为参考人群, 对所筛选InDel位点的区分效果进行评估. 采用以下方法进行评价: a. 使用种族推断软件DAA v1.0^[25](DNA Ancestry Analyzer, DAA)计算样本的群体匹配概率(population assignment match probability, AMP)和似然比(likelihood ratio, LR); b. 使用DAA v1.0软件计算人群祖先成分比例, 使用CLUMPAK对STRUCTURE结果进行可视化^[26]; c. 采用Rv3.2.3进行主成分分析(principal component analysis, PCA), 并用R程序包ggplot2画出种族归类图^[27-28]; d. 综合分析AMP值和祖先成分进行最终判断^[29-30], 当 $LR < 100$ 时^[31], AMP的前两位人群均不排除. 同时与1000Genomes所提供的样本来源信息进行比较.

Table 1 Sample information¹⁾

World Region	Population	Abbr	Sample Size (N)	Source
Africa (AFR)	African Caribbean in Barbados	ACB	96	1000genome
	Esan in Nigeria	ESN	99	1000genome
	Gambian in Western Division, The Gambia	GWD	113	1000genome
	Luhya in Webuye, Kenya	LWK	99	1000genome
	Mende in Sierra Leone	MSL	85	1000genome
	Yoruba in Ibadan, Nigeria	YRI	108	1000genome
	<i>African American</i>	<i>AFA</i>	<i>90</i>	<i>University of north Texas</i>
East Asia (EAS)	Chinese Dai in Xishuangbanna, China	CDX	93	1000genome
	Han Chinese in Beijing, China	CHB	103	1000genome
	Southern Han Chinese, China	CHS	105	1000genome
	Japanese in Tokyo, Japan	JPT	104	1000genome
	Kinh in Ho Chi Minh City, Vietnam	KHV	99	1000genome
	<i>Han Chinese</i>	<i>THC</i>	<i>56</i>	<i>NICGR</i>
	<i>Hui in China</i>	<i>HUX</i>	<i>333</i>	<i>NICGR</i>
Europe (EUR)	Utah residents with Northern and Western European ancestry	CEU	99	1000genome
	Finnish in Finland	FIN	99	1000genome
	British in England and Scotland	GBR	91	1000genome
	Iberian populations in Spain	IBS	107	1000genome
	Toscani in Italy	TSI	107	1000genome
	<i>Caucasian in Europe</i>	<i>CAU</i>	<i>74</i>	<i>University of north Texas</i>
<i>CentralAsia (CAS)</i>	<i>Uyghur in Xinjiang, China</i>	<i>CUX</i>	<i>226</i>	<i>NICGR</i>

¹⁾ Five test populations are highlighted in bold and italic, others are reference populations of 1000Genomes.

1.2.4 引物设计及合成

采用Primer premier 5.0 设计出所筛选的InDel位点以及DYS439基因座和Amelogenin性别基因座的PCR引物, 设置扩增子长度范围50~200 bp, 最佳Tm值范围55°C~65°C, 最佳GC含量50%~60%, 利用NCBI的Primer-BLAST工具进行引物特异性验证, 分别采用FAM、HEX、TAMRA和ROX进行荧光素标记. 所有引物均由上海生工生物有限公司合成和标记.

1.2.5 InDel多重复合扩增检测体系

先以9948为模板将设计的引物进行单位点PCR, 再调整各引物浓度和MgCl₂浓度, 设置退火温度58°C, 各位点的复合扩增效率最佳. 建立的快速扩增总体系为10 μl, 内含2.5×NiHi S9 PCR master mix 4 μl, MgCl₂ (25 mmol/L) 0.2 μl, 引物混合物1 μl, 模板DNA 1 μl (浓度1~2 mg/L). 使用Eppendorf Master cycler nexus PCR扩增, 热循环参数为: 95°C 10 min; 95°C 15 s, 58°C 1 min, 共28个循环; 62°C 30 min, 25°C保温.

取1 μl PCR产物与9 μl甲酰胺、0.5 μl Typer 500内标混匀, 95°C 3 min后立即冰浴5 min. 毛细管电泳在ABI 3130 XL型遗传分析仪, 使用Pop7凝胶, 36 cm毛细管进行电泳分离. 通过GeneMapper®ID v3.2软件进行分析判型, 得到每个位点的遗传参数, 按照软件说明书, 制作适用于本研究的Panel文件、Bin文件及method文件. 等位基因峰高阈值设置为50RFU (相对荧光强度). 在电泳图谱中, 每个位点的InDel分型结果以“+/-”表示.

1.3 人群样本测试和个体族群来源推断

对收集的来自5个人群的779份测试样本以及12份未知来源个体进行分型检测, 分型结果位点丢失数≤3个为有效分型. 以1000Genomes中16个群体的基因型数据作为参考数据库进行分析, 利用PHYLIP v 3.69构建系统发育树. 对个体进行族群来源推断时, 综合群体匹配概率的似然比和种族成分进行判断^[29, 32].

1.4 体系性能的验证评估

1.4.1 准确性验证

从测试样本中随机抽取5份, 对其38个InDel位点进行DNA一代测序, 以验证体系分型结果的准确性.

1.4.2 灵敏度测试

选择9947A、9948两种DNA标准品进行梯度稀释: 10 ng、5 ng、2.5 ng、1.25 ng、625 pg、313 pg、157 pg、78.5 pg、39.3 pg, 用构建的38-plex InDels体系扩增和分型检验, 以测试体系的灵敏度.

1.4.3 直接扩增能力测试

采用0.5 mm直径打孔器从待测的唾液卡、血卡上取样, 将0.5 mm圆片放于10 μ l PCR反应液中, 直接扩增的PCR反应体系和热循环参数均按照1.2.5进行设置.

1.5 微流控芯片系统的测试

本研究采用的集成式微流控芯片卡盒装置如图1所示, 该卡盒由上层塑料材质的提取-扩增芯片 (DEA芯片)、底层玻璃材质的毛细管电泳芯片 (CAE芯片) 以及用于连接上下芯片的电极针固定器构成. 芯片卡盒和一体化芯片检控仪器的详细信息可参照文献 [33]. 实验操作步骤如下: 配制45 μ l PCR Mix, 包含4.5 μ l 引物混合物, 18 μ l 2.5 \times NiHi S9 PCR master mix, 4.5 μ l DNA模板 (1 mg/L) 以及18 μ l 去离子水. 用移液器将45 μ l PCR Mix加入DEA芯片的样本槽, 将DEA芯片插入卡盒并固定电极针; 将组装好的集成式芯片卡盒放于检控仪上, 运行控制程序, 由仪器自动完成所有操作. 芯片PCR热循环参数同1.2.5. DNA模板采用K562和本实验室收集的DNA样本 (HJP).

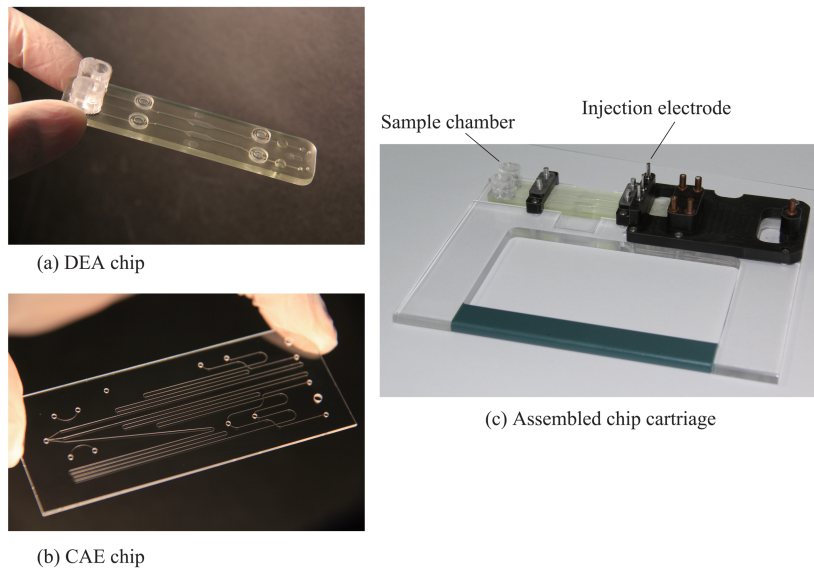


Fig. 1 The structure of the integrated chip cartridge consists of an upper DEA chip (a), lower CAE chip (b), and an electrode holder with three injection electrodes to connect the two chips in a plastic cartridge (c)

2 结 果

2.1 InDels位点筛选结果

本研究最终筛选出38个InDel位点 (详细信息见表2), 等位基因长度 ≤ 18 bp, 此外, 增加1个Y-STR基因座 (DYS439) 和Amelogenin性别基因座, 以辅助未知样本的性别判断. 如表3所示, 所有位点都符合Hardy-Weinberg平衡, P 值范围为0.147~1, 所有位点处于连锁平衡 ($r^2 < 0.2$). F_{st} 值范围0.085~0.572, 每一位点的等位基因频率在两

两人群间具有较大差异, 其中9个位点区分非洲人群 (即这些位点在非洲/欧洲人群以及非洲/东亚人群间具有较大的 δ_p 值); 14个位点可用作区分东亚人群 (即这些位点在非洲/东亚人群以及欧洲/东亚人群间具有较大的 δ_p 值); 15个位点区分欧洲人群 (即这些位点在非洲/欧洲人群以及东亚/欧洲人群间具有较大的 δ_p 值). PSD值在欧洲人群中为3.75, 明显低于平均值4.37, 而在东亚人群中相对较高, 为4.72.

Table 2 Information of the 38 InDels

rs number	Position (bp) ¹⁾	Chromosome	Insertion (+) /Deletion (-) ²⁾	Expected amplicon length	Tag number
rs3054057	86010537	15	-/TGTT	53.39–57.24	F1
rs112634351	35701898	22	-/GAGTG	65.51–71.07	F2
rs1160852	137345857	6	-/AA	84.3–86.98	F3
rs145415095	50955581	18	CT/-	99.89–101.94	F4
rs2308036	65207011	15	-/CC	113.06–115	F5
rs10656283	82545540	5	-/AATC/AGTC	156.44–160.53	F6
rs35851958	61190685	8	-/GTGA/TTGA	168.38–172.43	F7
rs2308101	51245594	20	TA/-	175.29–177.23	F8
rs16416	25365169	15	ACAG/-	180.87–184.8	F9
rs1610878	58763163	16	CTGT/-	194.72–197.9	F10
rs140847	12617324	9	AACG/-	58.18–62.56	H1
rs143825911	170720014	6	AAATGGC/-	64.79–72.18	H2
rs72173922	84924926	12	AT/-	79.85–81.92	H3
rs5900168	118346559	9	AT/-	88.32–90.29	H4
rs34921138	123774810	4	CTC/-	96.7–99.89	H5
rs111759013	25278464	20	-/GCGGG/GTGGG	108.97–114	H6
rs5877023	67176759	6	-/TCCAATAG/TCCCATAG	132.1–140.29	H7
rs35633537	7814342	5	GAG/-	147.8–150.83	H8
rs548753180	84476378	6	-/ATGG	161.91–166.04	H9
rs17879936	95047150	7	-/ACAATGCTTAC	183.97–195.12	H10
rs3045215	234740917	1	-/AGTTATAAT	58.93–67.64	T1
rs34122827	63778777	13	ACTT/-	70.79–74.53	T2
rs16715	61393657	17	-/AAGCTC	79–85	T3
rs5789229	1668636	11	-/TTTC	95.52–99.31	T4
rs74748892	128317272	5	-/TAG	112.19–115.17	T5
rs1160953	115626076	5	-/TAGA/GAGA	124.3–128.27	T6
rs66850318	11715430	6	CTC/-	146.82–150.12	T7
rs68050185	23619105	21	ATTTC/-	155.17–160.22	T8
rs112109748	45526063	15	CTCCAAATATAATTT/-	170.2–184.81	T9
rs71425754	52868424	15	TG/-	181.71–184.32	T10
rs2307832	55590787	1	-/TT	196.38–198.14	T11
rs2308067	127291541	7	-AA	85.63–87.97	R1
rs16711	20082419	17	-/TAGGAAGAAA	122.2–132.34	R2
rs3071043	8815887	10	AG/	143.54–145.78	R3
rs3991155	19401861	21	TT/-	161.77–164.04	R4
rs3083306	112712192	3	TCTT/-	165.88–169.69	R5
rs67625332	41359305	14	TA/-	192.88–194.85	R6
rs1610963	112247083	5	-/TTAGTTAT	199.35–207.35	R7

¹⁾ The physical location of the InDel marker on the chromosome, the Human Genome Database version is GRCh37.p13; ²⁾ In the electropherograms, the short allele is designated as “-” and the long allele is designated as “+”.

Table 3 The P value for HWE, F_{st} value, δ_p , and PSD values of the 38-plex InDels

Rs number	HWE (p)	δ_p value			Population-specific divergence (PSD)			F_{st}	Ancestry-specific
		EAS/EUR	EAS/AFR	EUR/AFR	EAS	EUR	AFR		
<i>rs2308036</i>	0.637	0.604	0.451	0.153	0.195	0.102	0.016	0.264	EAS
<i>rs1610878</i>	0.6	0.376	0.526	0.15	0.147	0.006	0.084	0.202	EAS
<i>rs140847</i>	0.989	0.586	0.47	0.116	0.201	0.095	0.025	0.269	EAS
<i>rs17879936</i>	0.147	0.337	0.245	0.092	0.065	0.028	0.005	0.085	EAS
<i>rs3045215</i>	0.386	0.522	0.619	0.097	0.244	0.027	0.11	0.323	EAS
<i>rs34122827</i>	0.591	0.448	0.273	0.175	0.151	0.072	0.002	0.176	EAS
<i>rs2308067</i>	0.84	0.368	0.679	0.311	0.215	0	0.25	0.355	EAS
<i>rs74748892</i>	0.997	0.53	0.549	0.019	0.262	0.069	0.105	0.368	EAS
<i>rs143825911</i>	0.939	0.695	0.823	0.128	0.455	0.051	0.216	0.565	EAS
<i>rs111759013</i>	0.973	0.487	0.653	0.166	0.269	0.013	0.119	0.309	EAS
<i>rs5789229</i>	0.922	0.352	0.367	0.015	0.104	0.017	0.026	0.117	EAS
<i>rs68050185</i>	0.476	0.346	0.449	0.103	0.115	0.007	0.053	0.147	EAS
<i>rs3991155</i>	0.972	0.475	0.477	0.002	0.168	0.033	0.039	0.205	EAS
<i>rs548753180</i>	0.707	0.311	0.351	0.04	0.093	0.015	0.038	0.135	EAS
<i>rs3054057</i>	0.869	0.471	0.013	0.484	0.08	0.247	0.105	0.358	EUR
<i>rs16711</i>	0.999	0.423	0.006	0.417	0.08	0.206	0.075	0.302	EUR
<i>rs10656283</i>	0.77	0.582	0.088	0.494	0.078	0.202	0.029	0.267	EUR
<i>rs66850318</i>	0.635	0.57	0	0.57	0.048	0.235	0.057	0.285	EUR
<i>rs112109748</i>	0.965	0.471	0.002	0.473	0.037	0.159	0.043	0.212	EUR
<i>rs16416</i>	0.553	0.388	0.029	0.359	0.029	0.198	0.02	0.137	EUR
<i>rs1160953</i>	0.855	0.289	0.21	0.499	0	0.113	0.094	0.176	EUR
<i>rs2308101</i>	0.984	0.356	0.096	0.452	0.011	0.13	0.071	0.189	EUR
<i>rs145415095</i>	0.763	0.319	0.056	0.375	0.011	0.092	0.04	0.132	EUR
<i>rs72173922</i>	0.632	0.417	0.036	0.453	0.021	0.135	0.04	0.168	EUR
<i>rs5900168</i>	0.958	0.402	0.015	0.417	0.028	0.129	0.043	0.188	EUR
<i>rs3071043</i>	0.813	0.438	0.064	0.502	0.026	0.174	0.076	0.248	EUR
<i>rs71425754</i>	0.778	0.679	0.081	0.598	0.114	0.301	0.053	0.387	EUR
<i>rs3083306</i>	0.503	0.4	0.183	0.583	0.005	0.176	0.113	0.244	EUR
<i>rs67625332</i>	0.712	0.428	0.029	0.457	0.026	0.141	0.046	0.192	EUR
<i>rs1160852</i>	0.16	0.008	0.692	0.684	0.212	0.182	0.446	0.572	AFR
<i>rs2307832</i>	0.53	0.148	0.685	0.537	0.173	0.038	0.276	0.383	AFR
<i>rs1610963</i>	0.61	0.121	0.624	0.503	0.112	0.033	0.285	0.321	AFR
<i>rs34921138</i>	0.983	0.185	0.803	0.618	0.215	0.044	0.383	0.488	AFR
<i>rs35633537</i>	0.856	0.01	0.511	0.501	0.151	0.119	0.286	0.385	AFR
<i>rs112634351</i>	0.739	0.146	0.409	0.555	0.017	0.094	0.183	0.234	AFR
<i>rs35851958</i>	0.988	0.049	0.394	0.345	0.041	0.019	0.112	0.146	AFR
<i>rs16715</i>	0.977	0.234	0.832	0.598	0.257	0.032	0.382	0.503	AFR
<i>rs5877023</i>	0.964	0.248	0.775	0.527	0.261	0.021	0.308	0.445	AFR
Cumulative values					4.717	3.751	4.649		

2.2 InDel位点组合的人群评估结果

基于1000Genomes中16个群体1607份样本的基因型数据,对本文筛选的38个InDel位点进行STRUCTURE分析和PCA分析,以评估其区分效果.结果如图2所示,当 $K=3$ 时,16个人群被聚类

成3个大的群体,即非洲人群(ACB、ESN、MAG、LWK、MSL和YRI,主成分超过0.91)、东亚人群(CDX、CHB、CHS、JPT和KHV,主成分超过0.98)、欧洲人群(CEU、FIN、GBR、IBS和TSI,主成分超过0.96).

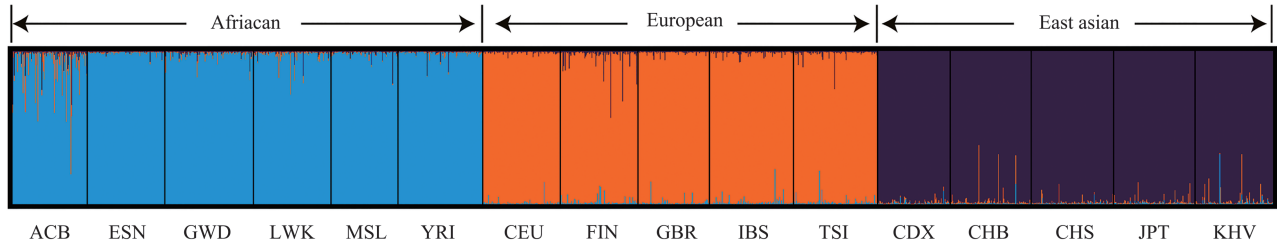


Fig. 2 Analysis of 38-plex InDels in the 16 population from 1000Genomes using STRUCTURE (K=3, Iteration=15)

采用 Rv3.2.3 基于群体等位基因频率对 1000Genomes 数据库中 16 个人群进行主成分分析 (图 3) . 16 个人群分为较为明显的 3 个部分, 非洲人群 (ACB、ESN、MAG、LWK、MSL 和 YRI)、

东亚人群 (CDX、CHB、CHS、JPT 和 KHV) 以及欧洲人群 (CEU、FIN、GBR、IBS 和 TSI) .

2.3 测试人群的检测结果

采用构建的 38-plex InDels 复合检测体系, 检测 5 个人群的 779 份样本, 共获得 29 580 个 Indels 位点的分型, 检出率为 99.9%, 图 4 为其中 1 份测试样本的分型图, 可以看出分型完整, 同种荧光不同位点间的峰高均衡性良好.

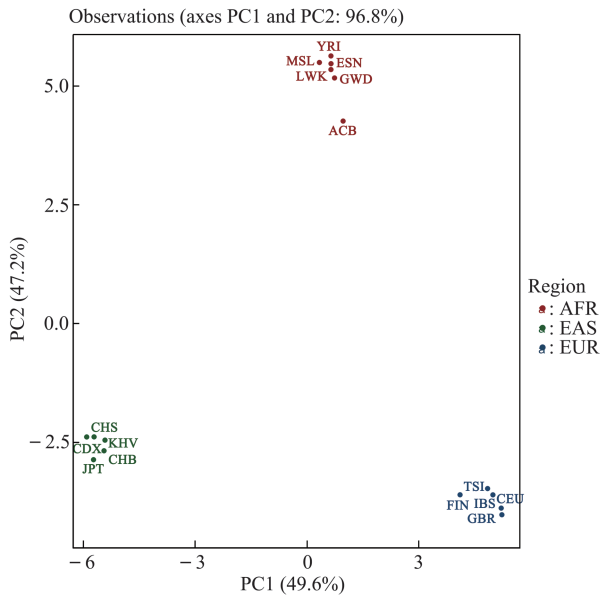


Fig. 3 Principal component analysis of 16 populations based on allele frequencies of 38-plex InDels

基于 1000Genomes 中 16 个人群的各位点等位基因频率, 计算待测样本的群体随机匹配概率和似然比, 结合祖先成分对其族群来源进行推断, AFA 人群除了 1 个样本来源于欧洲人群, 不排除来源于非洲人群外, 其余个体均被归入非洲人群; CAU 人群个体均被归入欧洲人群; THC 人群所有个体均被归入东亚人群; HUX 人群中有 5 个来源于东亚不排除来源于欧洲, 4 个来源于欧洲不排除来源于东亚, 其余样本均被归入东亚人群; CUX 人群中, 除了 11 份样本归入东亚和 16 份样本归入到欧洲人群外, 其余样本均被归类到不排除的混合人群中 (表 4) .

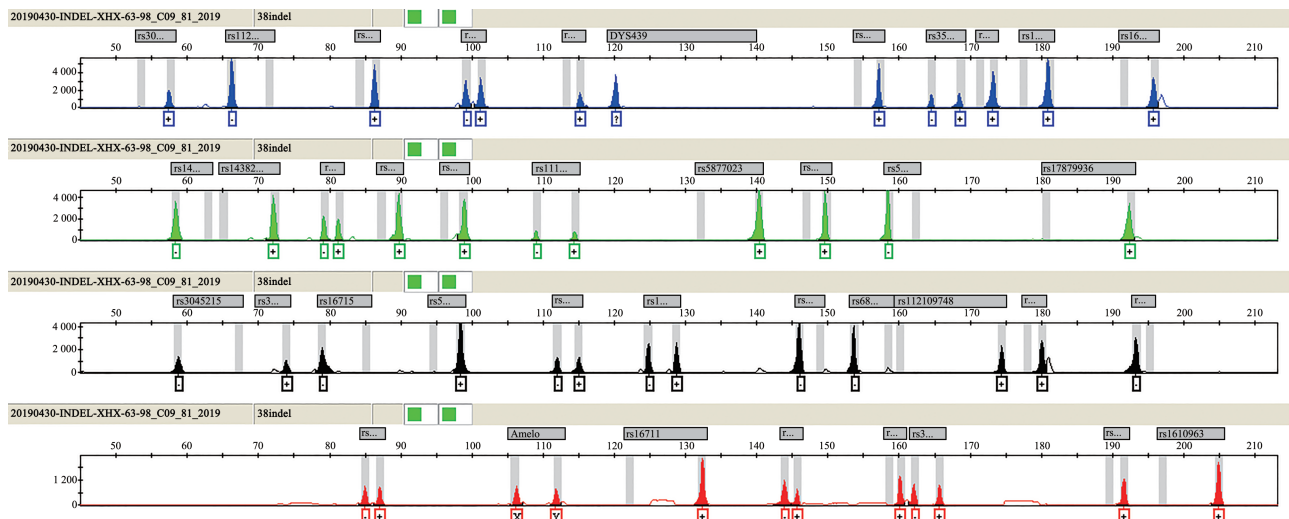


Fig. 4 Electropherogram of a 38-plex InDels profile obtained from a testing DNA sample

Table 4 The inferred ancestry components and the accuracy of sample assignment for five test populations

Population	Sample size	Accuracy of sample assignment (%)				The distribution of three ancestry components		
		AFR	EUR	EAS	inclusive	EAS	EUR	AFR
AFA	90	98.89%			1.11%	0.037±0.045	0.073±0.103	0.890±0.110
CAU	74		100%			0.018±0.024	0.970±0.029	0.013±0.013
THC	56			100%		0.980±0.0199	0.013±0.017	0.008±0.008
HUX	333			97.30%	2.70%	0.934±0.078	0.045±0.071	0.021±0.031
CUX	226		7.08%	4.87%	88.05%	0.490±0.216	0.481±0.218	0.029±0.044

将1000Genomes数据库的16个人群和检测的5个人群共计2386份样本进行STRUCTURE聚类分析, 当K=3时, 上述人群样本主要聚成4类: 非洲人群、欧洲人群、欧洲-东亚混合人群、东亚人群

(图5). 5个人群中非洲(AFA)、欧洲(CAU)和东亚(THC、HUX)均根据其祖先主成分大于0.9正确地聚类, 而CUX人群则表现出明显的欧亚混合成分, 即48.1%/49%.

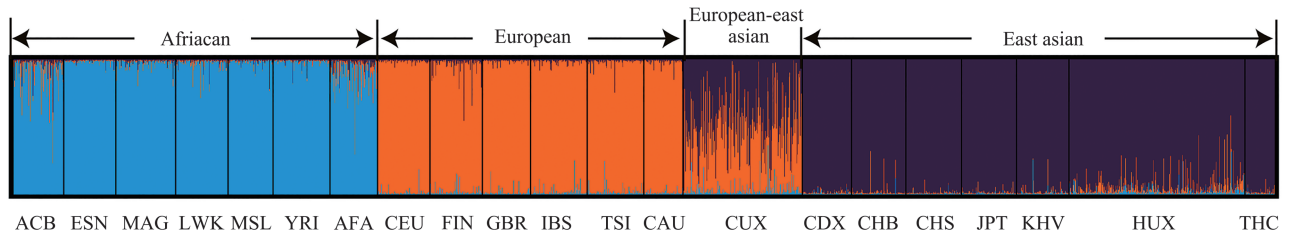


Fig. 5 Analysis of 38-plex InDels in the 21 population using STRUCTURE

以上21个人群的PCA分析结果(图6)显示, 主成分1和主成分2合计占96.4%, 21个人群的散点形成了3个大的聚类. 左中为东亚人群, 右上为非洲人群, 右下为欧洲人群, 混合人群CUX在欧洲-东亚之间.

基于38个InDel在上述21个人群中的等位基因频率构建系统发育树(图7), 非洲人群(ACB、ESN、MAG、LWK、MSL、YRI、AFA)聚为一支, 处于系统发育树的根部; 欧洲人群(CEU、

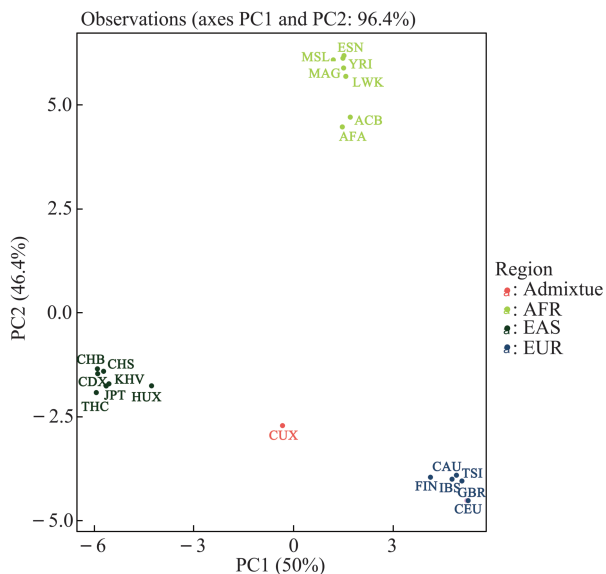


Fig. 6 Principal Component Analysis of 21 populations based on allele frequencies of the 38 InDels

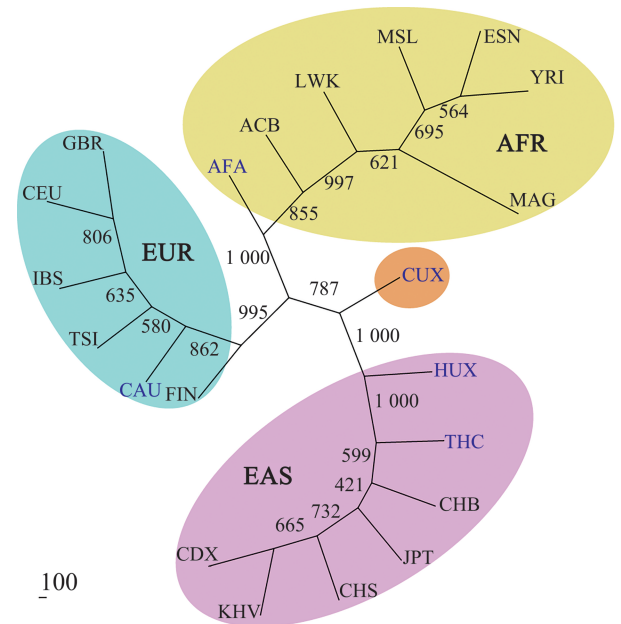


Fig. 7 Phylogenetic trees based on the allelic frequency of 21 populations

FIN、GBR、IBS、TSI、CAU) 聚为一支; 东亚人群 (CDX、CHB、CHS、JPT、KHV、THC、HUX) 聚为一支; 混合人群CUX的一支位于欧洲和东亚人群之间.

2.4 InDels多重复合PCR体系性能验证结果

2.4.1 准确性验证

随机抽取的5份样品, 经比对DNA一代测序结果与本体检测体系检测结果分型完全一致, 证明本文构建的体系对38个InDel位点的基因分型结果准确性可靠.

2.4.2 灵敏度测试

采用本文38-plex InDels体系对稀释的9947A、9948 DNA标准品进行检验, 以等位基因峰高≥50 RFU为标准 (图8). 结果显示, 最佳DNA模板量在313 pg~1.25 ng之间, 当DNA量低于157 pg时开始出现等位基因丢失, 如rs16711和rs67625332位点丢失 (图8方框所示); 当DNA量高于2.5 ng时, 等位基因峰荧光信号过强, 产生渗透而干扰结果分析, 如rs5877023、rs548753180、rs16416位点 (图8箭头所示).

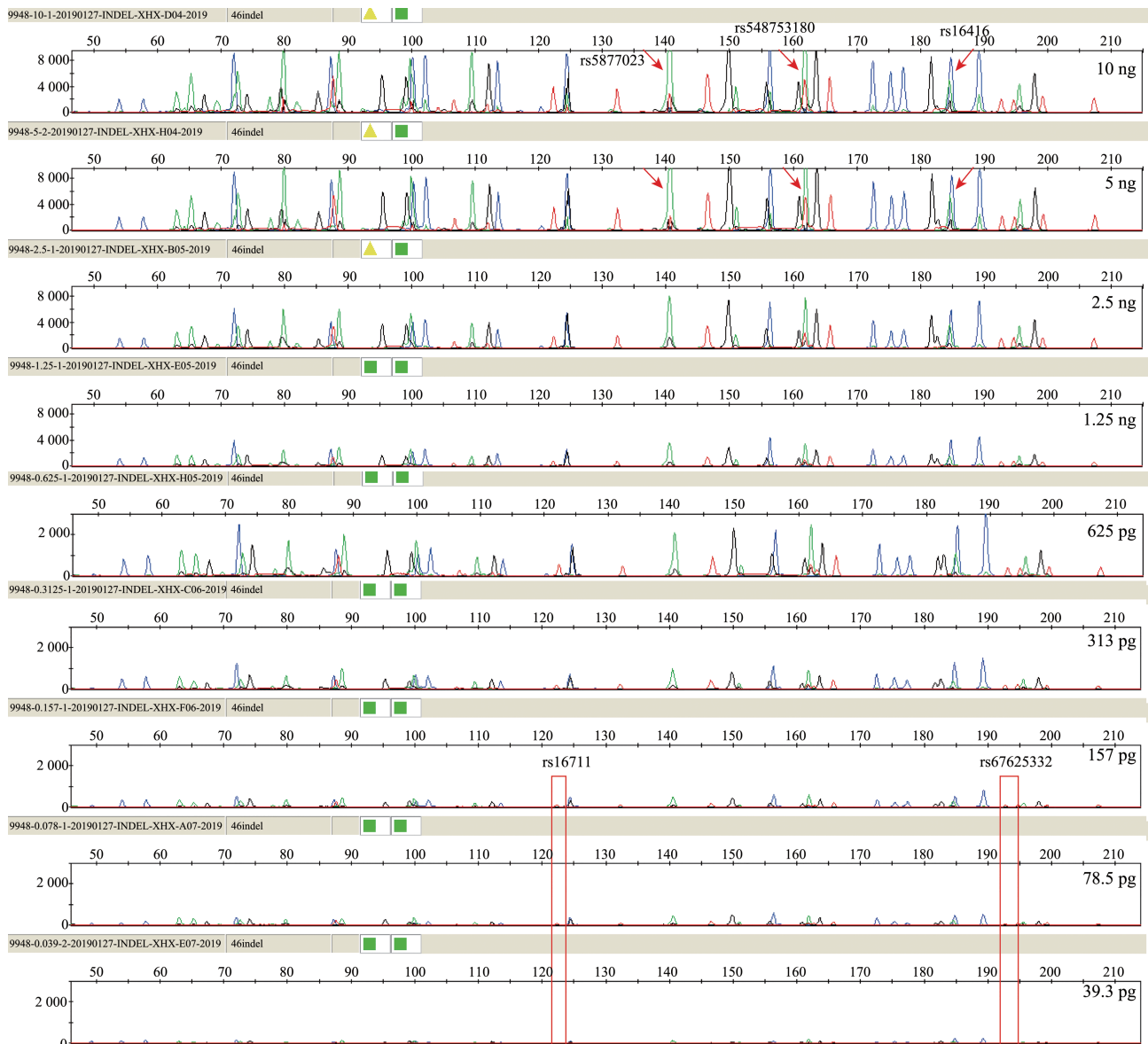


Fig. 8 Electropherograms from the sensitivity study of 9948 standard human genomic DNA

2.4.3 直接扩增能力

采用本文38-plex InDels体系对215份唾液卡和血卡进行直接PCR扩增, 结果显示160份血卡可以

获得完整分型, 未出现等位基因丢失; 55份唾液卡中有3份出现扩增失败或未获得有效分型. 分析其原因可能是口腔细胞转移至唾液卡上的数量较少

Table 5 Statistic of sample assignment for 215 direct amplification samples¹⁾

Sample type	Population	Sample size	Accuracy of match probability (%)			
			AFR		EUR	
			EUR	EAS-EUR	EAS	EAS
blood sample	Nigeria in African	3	100			
	European	6	100			
	Han-Chinese	127			100	
	Xinjiang in China	24			100	
saliva sample	Southern Chinese Han	55	100			

¹⁾ Three saliva samples that failed to provide results of direct PCR were not included in the table.

或者某些个体因饮食等原因导致唾液中存在某些抑制扩增的物质^[34]. 利用16个族群的参考数据库, 对直扩获得的212份样本分型结果进行个体族群来源推断, 结果如表5所示, 所有样本的推断结果均与已知样本信息一致.

2.5 微流控芯片系统的测试结果

如图9所示, DNA样本在微流控芯片系统上完成了PCR扩增和电泳分离检测, 总时间为1.7 h. 2份测试样本均获得完整的38-plex InDels分型图, 其结果与传统实验室方法相同. 利用16个族群的参考数据库, 对该2份样本结果进行个体族群来源推断, 结果显示DNA标准品K562来源于欧洲人群, 本实验室样本HJP来源于东亚人群.

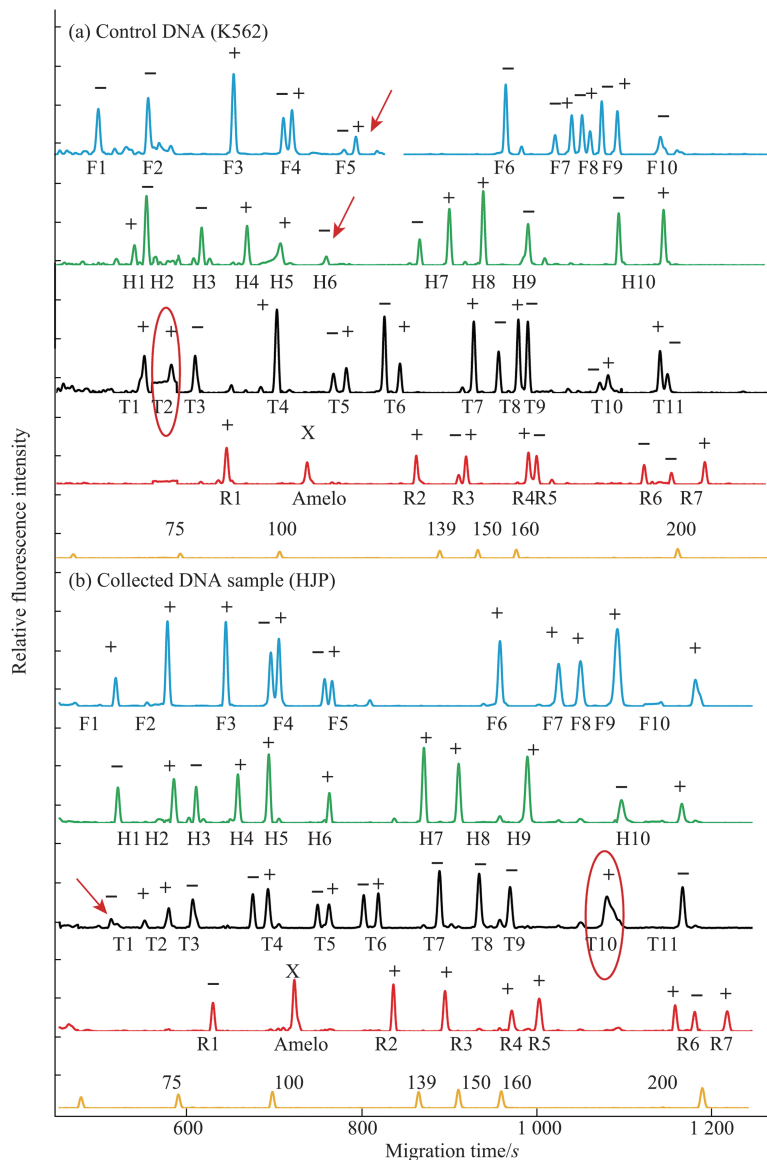


Fig. 9 Electropherograms obtained for K562 and HJP samples (markers are identified by tag number)

3 讨 论

3.1 位点筛选

本研究主要是针对非洲、欧洲、东亚人群，共筛选出 38 个用于族群推断的 InDels 位点，为使这组位点在不同人群之间具有相对平衡的区分效能，其中选取 9 个非洲人群特异性位点，15 个欧洲人群特异性位点，14 个东亚人群特异性位点。由于非洲/东亚、非洲/欧洲人群间的差异较为明显，选取非洲特异性位点较少，而欧洲与东亚人群间的差异性较小，所以选取的特异性位点数相对较多，PSD 值在欧洲人群中为 3.75，低于平均值 4.37，而在东亚人群中相对较高，为 4.72。通过对 1000Genomes 中 16 个人群进行 STRUCTURE (图 2) 和 PCA 分析 (图 3)，所有人群被分成了三大部分 ($K=3$ 时)，主成分 1 (PC1) 和主成分 2 (PC2) 解释了 96.8% 的差异，分布集中的人群为来自同一洲际祖先来源。为了进一步提高平衡度，未来需要增加欧洲人群特异位点。

3.2 体系构建

本研究采用常规 PCR-CE 检测分型技术，通过对引物浓度配比、扩增条件等初步优化，构建了 38-plex InDels 复合检测体系，38 个位点等位基因均可明显判断分型 (图 4)，但仍存在部分位点均衡性差和非特异峰问题，需要后续继续优化。同时，验证了该 38-plex InDels 体系具有较高灵敏度和准确性，与本研究组前期建立的 27-plex SNP 检测体系^[29-30]类似，均可实现对非洲、欧洲、东亚及混合人群遗传成分的区分，而本体系的优势在于采用 PCR-CE 技术，在 3 h 内即可完成样本检测，耗时短、检测成本低等。与其他 InDel 族群推断体系^[18-19]相比，本体系所需试剂均为国产，降低了检测成本，同时可以直接扩增，操作简单。由于与常规 STR 检验方法相同，也易于在国内法医 DNA 实验室推广应用。

3.3 测试样本人群区分效果

采用该 38-plex InDels 检测体系对 5 个人群 779 份样本进行检测，然后与 1000Genomes 的 16 个群体合并进行分析。STRUCTURE 结果 (图 5) 显示该体系能对非洲、欧洲、东亚、欧亚混合人群分别聚类 and 区分，系统发育树 (图 7) 显示同一个洲际祖先来源人群 (椭圆形区域覆盖部分) 内部的距离较近，洲际人群之间的距离较远。CAU 为欧洲高加索人群，所有个体均归入欧洲人群，欧洲成分占

97%；AFA 是非裔美国人群，虽然有一些欧洲白人成分 (7.3%)，但其基本的构成还是以非洲黑人成分 (89%) 为主，均与之前的研究报道一致^[35]。汉族 (THC) 所有个体均归入东亚人群，东亚成分占 98%。回族 (HUX) 绝大多数个体归入东亚人群，东亚成分占 93.4%，有 4 份样本得到不排除东亚的结论，根据文献记载^[36-37] 回族萌芽于唐代，来自阿拉伯、小亚细亚的蕃客、胡商经丝绸之路进入中国；元朝时，大批中亚、波斯和阿拉伯的士兵、工匠被蒙带入中国；到清朝，从西北各省区迁居到新疆，与汉族及当地少数民族混杂居住、不断融合，回族与我国国内及东亚群体均表现出较近的遗传距离 (图 7)，这与 Y-STR^[38-39]、其他 InDel 研究^[40] 相一致。维吾尔族 (CUX) 的欧亚混合成分分别为 48.1%/49%，该结果与 27-plex SNP 研究结果类似^[29]。PCA 分析 (图 6) 也可以看出，38 个 InDel 位点把 21 个人群明显的区分成三大部分，东亚、非洲、欧洲分布相对较集中，而 CUX 位于东亚人群和欧洲人群之间，属于欧亚混合人群，根据人群随机匹配概率结果 (表 4) 有个别个体被归入东亚或欧洲人群，分析原因可能是种族和民族的概念不完全一致，前者主要从生物学意义上定义，后者主要从文化意义上定义，而且不同人群之间的融合是一个动态过程，没有绝然的界值，因而混合人群的遗传结构相对复杂^[41-42]，推断结果有可能与户籍填报的民族信息出现差别。因此，在实际应用中，要综合考虑祖先成分等其他指标综合研判。

3.4 微流控芯片体系的应用

采用微流控芯片技术进行 STR 快速分型已有较多报道^[43-45]，尚未见应用于族群推断领域。本研究基于前期建立的微流控芯片系统^[33, 46]，将该 38-plex InDels 体系集成于微流控芯片上，利用芯片检测仪自动完成对 DNA 样本的 PCR 扩增和电泳分离检测。虽然芯片电泳得到的 InDel 分型图存在如等位基因不均衡，个别位点峰值较低 (如图 9 箭头所示) 以及出现异常信号峰 (如图 9 中椭圆所示) 等缺陷，但该实验结果已初步显示，利用微流控芯片系统能够实现集成式、自动化 InDel 检测，与传统实验室方法相比，简化了操作步骤，同时降低污染风险^[46]。下一步将尝试把细胞裂解和 DNA 提取步骤整合到芯片系统中，同时进一步缩短时间，以实现 2 h 内完成对口腔拭子、血样等样本全集式检验。

综上，本研究建立了 38 个 InDel 位点的复合检

测体系, 该体系可以有效进行东亚、欧洲、非洲三大人群以及混合人群的遗传结构分析和个体的祖先来源推断, 同时该体系基于PCR-CE技术, 具有较高灵敏度和准确性, 可以直扩检测, 也可以集成于微流控芯片系统, 快速的进行未知DNA样本的族群来源推断, 为案件提供侦查线索.

参 考 文 献

- [1] Phillips C, Salas A, Sanchez J J, *et al.* Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet*, 2007, **1**(3-4): 273-280
- [2] Halder I, Shriver M, Thomas M, *et al.* A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat*, 2008, **29**(5): 648-658
- [3] 李彩霞, 贾竟, 魏以梁, 等. 30个祖先信息位点的筛选及应用. *遗传*, 2014, **36**(8): 779-785
Li C X, Jia J, Wei Y L, *et al.* *Hereditas*, 2014, **36**(8): 779-785
- [4] Mills R E, Luttig C T, Larkins C E, *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res*, 2006, **16**(9): 1182-1190
- [5] Nachman MW, Crowell S L. Estimate of the mutation rate per nucleotide in humans. *Genetics*, 2000, **156**(1): 297-304.
- [6] 刘海渤, 孙启凡, 陈健刚, 等. 27重SNP种族推断体系准确性验证研究. *中国法医学杂志*, 2015, **30**(6): 563-566
Liu H B, Sun Q F, Chen J G, *et al.* *Chin J Forensic Med*, 2015, **30**(6): 563-566
- [7] Yang N, Li H Z, Criswell L A, *et al.* 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*, 2005, **118**(3-4): 382-392
- [8] Weber J L, David D, Heil J, *et al.* Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet*, 2002, **71**(4): 854-862
- [9] Mills R E, Pittard W S, Mullaney J M, *et al.* Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res*, 2011, **21**(6): 830-839
- [10] Pereira R, Phillips C, Alves C, *et al.* A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, 2009, **30**(21): 3682-3690
- [11] Pinto N, Magalhães M, Conde-Sousa E, *et al.* Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. *Forensic Sci Int Genet*, 2013, **7**(1): 16-21
- [12] Romanini C, Catelli M L, Borosky A, *et al.* Typing short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. *Forensic Sci Int Genet*, 2012, **6**(4): 469-476
- [13] Cardoso S, Sevillano R, Gamarra D, *et al.* Population genetic data of 38 insertion-deletion markers in six populations of the northern fringe of the Iberian Peninsula. *Forensic Sci Int Genet*, 2017, **27**: 175-179
- [14] Ferragut J F, Pereira R, Castro J A, *et al.* Genetic diversity of 38 insertion - deletion polymorphisms in Jewish populations. *Forensic Sci Int Genet*, 2016, **21**: 1-4
- [15] 赵书民, 张素华, 李成涛. InDel_typer30: 用于中国5个主要民族DNA鉴定的多重PCR系统. *法医学杂志*, 2010, **26**(5): 343-346
Zhao S M, Zhang S H, Li C T, *et al.* *J Forensic Med*, 2010, **26**(5): 343-346
- [16] Li C T, Zhang S H, Li L, *et al.* Selection of 29 highly informative InDel markers for human identification and paternity analysis in Chinese Han population by the SNPlex genotyping system. *Mol Biol Rep*, 2012, **39**(3): 3143-3152
- [17] Bastos-Rodrigues L, Pimenta J R, Pena S D J. The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann Hum Genet*, 2006, **70**(5): 658-665
- [18] Pereira R, Phillips C, Pinto N, *et al.* Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *Plos One*, 2012, **7**(1): e29684
- [19] Zaumsegel D, Rothschild M A, Schneider P M. A 21 marker insertion deletion polymorphism panel to study biogeographic ancestry. *Forensic Sci Int Genet*, 2013, **7**(2): 305-312
- [20] Zaumsegel D, Pereira R, Rothschild M A, *et al.* Comparative analysis of two indel-based ancestry informative multiplex PCR typing kits. *Forensic Sci Int Genet Suppl*, 2013, **4**(1): e21-e22
- [21] Sun K, Ye Y, Luo T, *et al.* Multi-InDel analysis for ancestry inference of sub-populations in China. *Sci Rep*, 2016, **6**: 39797
- [22] Sun K, Yun L, Zhang C, *et al.* Evaluation of 12 Multi-InDel markers for forensic ancestry prediction in Asian populations. *Forensic Sci Int Genet*, 2019, **43**: 102155
- [23] 杨洁, 赫佳, 王丹碧, 等. InDel标记的研究和应用进展. *生物多样性*, 2016, **24**(2): 237-243
Yang J, He Jia, Wang D B, *et al.* *Biodiversity Sci*, 2016, **24**(2): 237-243
- [24] Santos N P, Ribeiro-Rodrigues E M, Ribeiro-Dos-Santos A K, *et al.* Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat*, 2010, **31**(2): 184-190
- [25] 刘京, 李盛, 江丽, 等. 对于未知来源个体进行族群推断的自动分析系统. *生命科学研究*, 2018, **22**(1): 3-7
Liu J, Li S, Jiang L, *et al.* *life Sci Res*, 2018, **22**(1): 3-7
- [26] Kopelman N M, Mayzel J, Jakobsson M, *et al.* Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol Ecol Resour*, 2015, **15**(5): 1179-1191
- [27] Team RDC. R: a language and environment for statistical. *Computing*, 2011, **14**: 12-21
- [28] H W. ggplot2: elegant graphics for data analysis. Incorporated: Springer Publishing Company, 2009, 180-185
- [29] Wei Y L, Wei L, Zhao L, *et al.* A single-tube 27-plex SNP assay for estimating individual ancestry and admixture from three continents. *Int J Legal Med*, 2016, **130**(1): 27-37
- [30] Jiang L, Sun Q F, Ma Q, *et al.* Optimization and validation of

- analysis method based on 27-plex SNP panel for ancestry inference. *Yi Chuan*, 2017, **39**(2): 166-173
- [31] Santos C, Fondevila M, Ballard D, *et al.* Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: results of a collaborative EDNAP exercise. *Forensic Sci Int Genet*, 2015, **19**: 56-67
- [32] Jia J, Wei Y L, Qin C J, *et al.* Developing a novel panel of genome-wide ancestry informative markers for bio-geographical ancestry estimates. *Forensic Sci Int Genet*, 2014, **8**(1): 187-194
- [33] Han J P, Gan W P, Zhuang B, *et al.* A fully integrated microchip system for automated forensic short tandem repeat analysis. *Analyst*, 2017, **142**(11): 2004-2012
- [34] Park S J, Kim J Y, Yang Y G, *et al.* Direct STR amplification from whole blood and blood- or saliva-spotted FTA without DNA purification. *J Forensic Sci*, 2008, **53**(2): 335-341
- [35] Kidd K K, Speed W C, Pakstis A J, *et al.* Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet*, 2014, **10**: 23-32
- [36] 张静. 近现代新疆回族的来源. *新疆地方志*, 2013, **2**: 56-58
Zhang J. *Local Chronicles of Xinjiang*. 2013, **2**: 56-58
- [37] 李玲. 回族形成的过程及其特点. *新疆社科论坛*, 1991, **1**: 60-61
Li L. *Xinjiang Social Sciences Forum*. 1991, **1**: 60-61
- [38] 赵琪. Y染色体STR基因座在回族群体间的遗传学研究及法医学应用初探. 江苏: 苏州大学, 2017: 56-62
Zhao Q. *Jiangsu: Soochow university*, 2017: 56-62
- [39] 雷强. 三个民族27个Y-STR基因座遗传多态性研究. 云南: 昆明医科大学, 2017: 33-39
Lei Q. *Yunnan: Kunming Medical University*, 2017: 33-39
- [40] 解通. 35个InDel位点荧光复合扩增检测体系的构建、验证及新疆三个少数民族遗传背景的探索研究. 广州: 南方医科大学, 2019: 76-86
Xie T. *Guangzhou: Southern Medical University*, 2019: 76-86
- [41] Li C X, Pakstis A J, Jiang L, *et al.* A panel of 74 AISNPs: improved ancestry inference within Eastern Asia. *Forensic Sci Int Genet*, 2016, **23**: 101-110
- [42] Xu S, Huang W, Qian J, *et al.* Analysis of genomic admixture in Uyghur and its implication in mapping strategy. *Am J Hum Genet*, 2008, **82**(4): 883-894
- [43] Liu P, Li X J, Greenspoon S A, *et al.* Integrated DNA purification, PCR, sample cleanup, and capillary electrophoresis microchip for forensic human identification. *Lab Chip*, 2011, **11**(6): 1041-1048
- [44] Tan E, Turingan R S, Hogan C, *et al.* Fully integrated, fully automated generation of short tandem repeat profiles. *Investig Genet*, 2013, **4**: 1-16
- [45] Andrew J. Hopwood C H, Yang J N, *et al.* Integrated microfluidic system for rapid forensic DNA analysis: sample collection to DNA profile. *Anal Chem*, 2010, **82**: 6991-6999
- [46] Zhuang B, Han J P, Xiang G X, *et al.* A fully integrated and automated microsystem for rapid pharmacogenetic typing of multiple warfarin-related single-nucleotide polymorphisms. *Lab Chip*, 2016, **16**(1): 86-95

A Single-tube Direct Amplification 38-plex Ancestry Informative InDels Assay and Its Integration With a Microfluidic System*

HAN Jun-Ping^{1,2}, ZHAO Lei¹, WANG Qing-Guo³, JIANG Li¹, XIE He-Xin¹, ZHUANG Bin⁴,
ZHAO Li-Jian⁴, LI Tang-Song², WU Jin², LI Cai-Xia¹**

¹Key Laboratory of Forensic Genetics, Beijing Engineering Research Center of Crime Scene Evidence Examination,

National Engineering Laboratory for Forensic Science, Institute of Forensic Science, Beijing 100038, China;

²Technology Department of Chaoyang Sub-bureau, Beijing Public Security Bureau, Beijing, 100025, China;

³Shandong First Medical University, Jinan 250000, China;

⁴Beijing CapitalBio Technology Ltd. Co., Beijing 101111, China)

Abstract In this paper, 38 ancestry informative insertion-deletion (InDel) polymorphism markers, with allele frequency differences between three major continental population groups (East Asian, European, and African) were selected from the related literatures and dbSNP database. The Amelogenin locus and a Y chromosomal STR (DYS439) locus were integrated to aid in the sex determination of an unknown sample. All markers were designed and amplified in a single-tube and performed direct amplification assay based on the PCR-CE technology for inferring ancestry of three continental populations. This assay was integrated with a microfluidic system for automatic amplification and InDel genotyping using DNA samples in 1.7 h. The genotypes of 1 607 individuals obtained from the 1000Genomes Phase III panel were used to assess the differentiation capacity of this 38-plex InDels assay. Then, this assay was validated by sensitivity, genotyping accuracy, and direct PCR ability. 779 samples from 5 populations and 215 direct amplification from saliva and blood samples were genotyped by this assay, then using the structure cluster analysis and principal component analysis to infer ancestral origins and evaluate the accuracy of the system. The results indicate that the 38-plex InDels assay can provide an accurate genotyping, with a sensitivity of 157 pg, and enable the direct amplification of DNA from saliva and blood on filter paper without sample purification. In addition, this assay is sufficient to distinguish between three continental populations and the admixture of Eurasian populations, and also accurately infer the ancestry origins of testing DNA samples. The integration of cell lysis into the microfluidic system is possible to yield 'sample-in-profile-out' results of InDel genotyping in 2 h.

Key words forensic biological evidence, insertion-deletion polymorphisms, ancestry informative markers, microfluidic chip

DOI: 10.16476/j.pibb.2020.0020

* This work was supported by grants from The National Natural Science Foundation of China (81772027) and The Basic Research Project (2018JB037).

** Corresponding author.

Tel: 86-10-63262392, E-mail: licaixia@tsinghua.org.cn

Received: January 14, 2020 Accepted: June 10, 2020