

武汉汉族群体 23 个 Y 染色体双等位基因标记遗传多态性研究

黄代新, 杨庆恩, 尹 慧, 翟仙敦, 杨荣芝

(华中科技大学同济医学院法医学系, 武汉 430030)

摘要: 筛选汉族群体中具有多态性的 Y 染色体双等位基因标记并获取其群体遗传学数据。采用片段长度差异等位基因特异性 PCR 和 PAGE 技术对武汉地区 160 名男性汉族无关个体的 23 个 Y 染色体双等位基因标记 (M7, M9, M50, M88, M89, M95, M111, M117, M119, M121, M122, M134, M159, M164, M175, M214, LINE1, MSY2, RPS4 Y₇₁₁, SRY₊₄₆₅, IMS-JST164520, IMS-JST021354 和 IMS-JST003305) 进行分型。除 M50、M159 和 M164 外, 其余 20 个标记在武汉汉族群体中均具有遗传多态性, 其基因多样性 (GD) 范围为 0.0126~0.4855, 共检出 35 种不同单体群组合 (Hg1~35), 单体群多样性 (HD) 为 0.9471。表明 20 个 Y 染色体双等位基因标记组成的单体群具有较高的遗传多样性, 在法医学应用和群体进化研究中具有较高的实用价值。

关键词: Y 染色体; 双等位基因标记; 遗传多态性; 单体群

中图分类号: Q347

文献标识码: A

文章编号: 0253-9772(2006)07-0791-08

Genetic Polymorphism of 23 Y Chromosome Biallelic Markers in Wuhan Han Population

HUANG Dai-Xin, YANG Qing-En, Yin Hui, ZHAI Xian-Dun, YANG Rong-Zhi

(Faculty of Forensic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China)

Abstract: To search polymorphic Y chromosome biallelic markers in Chinese Han population, and obtain their population genetic data. Genotyping of 23 biallelic markers on human Y chromosome (M7, M9, M50, M88, M89, M95, M111, M117, M119, M121, M122, M134, M159, M164, M175, M214, LINE1, MSY2, RPS4 Y₇₁₁, SRY₊₄₆₅, IMS-JST164520, IMS-JST021354 and IMS-JST003305) were carried out in a sample of 160 unrelated Chinese male individuals living in Wuhan using fragment length discrepant allele specific PCR (FLDAS-PCR) and PAGE technique. In all 23 biallelic markers, genetic polymorphism were identified for 20 loci in Wuhan Han population except for M50, M159 and M164, and the ranges of gene diversity (GD) were 0.0126~0.4855. A total of 35 different haplogroups (Hg1~35) were observed and the haplogroup diversity (HD) was 0.9471. The haplogroups formed by 20 biallelic markers are highly polymorphic, and can be used in forensic science and population evolution studies.

Key words: Y chromosome; biallelic marker; genetic polymorphism; haplogroup

人类 Y 染色体除拟常染区外, 绝大部分为特异性非重组区, 在减数分裂中不发生重组, 以单倍体的形式呈父系遗传特征并为男性所特有。因此 Y-DNA 序列携带男性进化史的信息, 被广泛应用于人

类的起源和进化研究。此外, Y 染色体遗传标记的研究对某些特殊的亲权鉴定和个人识别案, 如涉及父子、兄弟、叔侄和祖孙鉴定、混合斑鉴定以及犯罪现场斑痕的群体来源推断等具有独特的应用价值。

虽然 Y-STR 标记多态性程度较高并得以广泛应用,但近年来,作为第 3 代遗传标记的双等位基因标记系统(biallelic markers)——包括单核苷酸多态性(single nucleotide polymorphisms, SNPs)和小片段插入/缺失(insertion/deletions, indels),因其具有分布广泛、遗传稳定、突变率低及易于分析等优点而逐渐成为研究热点。为了筛选在汉族群体中具有多态性的 Y 染色体双等位基因标记并获取其群体遗传学数据,本文应用片段长度差异等位基因特异性 PCR(fragment length discrepant allele specific PCR, FLDAS-PCR)技术^[1]对武汉汉族群体 160 名男性无关个体 23 个 Y 染色体双等位基因标记进行了分型研究,现报道如下。

1 材料和方法

1.1 样品

武汉地区 160 份汉族男性无关健康个体及 5 份女性 EDTA 抗凝血由武汉同济医院血库提供。用 Chelex-100 法提取基因组 DNA^[2]。

1.2 分型方法的选择及引物

本文选择的 23 个 Y 染色体双等位基因标记包

括 16 个 SNPs (M7, M9, M50, M88, M89, M95, M119, M122, M159, M164, M214, RPS4Y₇₁₁, SRY₊₄₆₅, IMS-JST164520, IMS-JST021354 和 IMS-JST003305) 和 7 个 indels (M111, M117, M121, M134, M175, LINE1 和 MSY2)。所有的 SNPs 均采用 FLDAS-PCR 技术分型;7 个 indels 中, M134 为单碱基缺失(G/del G),两个等位基因相差仅 1 bp,普通聚丙烯酰胺凝胶电泳难以区分,故也采用 FLDAS-PCR 策略分型,其余皆采用普通 PCR 技术分型。

除 SRY₊₄₆₅、LINE1 和 MSY2 分别参照 Naito E 等^[3]、Santos F R 等^[4]和 Bao W 等^[5]报道的引物外,其余 20 个标记的引物根据 GenBank 上的序列资料利用 Primer 3 软件自行设计,由上海生工生物工程技术服务有限公司合成。每套引物(除 6 个 indels 外)均由 2 条相差 4~5 bp 的特异性引物和 1 条反向公用引物组成,其中 2 条特异性引物分别在其 3' 端第 3 或第 4 位碱基处引入不同碱基错配,部分引物在较长等位基因特异性引物的 5' 端不同碱基处引入 1~3 个错配碱基,以提高特异性(表 1)。

表 1 23 个 Y 染色体双等位基因标记的引物序列、PCR 反应条件及产物大小

Table 1 The primer sequences, PCR condition and product size for 23 Y chromosome biallelic Markers

基因座 Locus	引物 Primer	序列(5'→3') Sequence(5'→3')	终浓度(μmol/L) Final concentration (μmol/L)	复性温度(℃) Annealing temperature(℃)	产物大小(bp) Product size(bp) (genotype)
M7*	M 7-F	tggattagtcacttagacctac	0.14	61	176 (G)
	M 7-G	ttgtagttgagtactgttctt gttc	0.10		172 (C)
	M 7-C	agttgagttactgttcttc atg	0.14		
M9	M 9-G	aaaga ttcggcctaagatggtga ttg	0.14	60	197 (G)
	M 9-C	aacggcctaagatggttg t atc	0.20		192 (C)
	M 9-R	cctcatttttgaagctcgtg	0.20		
M50*	M 50-F	tgaagggtgcagggacagac	0.12	61	150 (T)
	M 50-T	tccaggaaaggcctctggtaa caca	0.12		145 (C)
	M 50-C	gaaagggtcctctggtaa t cg	0.12		
M88	M 88-A	cattctattctgcttcttc agca	0.12	61	156 (A)
	M 88-G	cttattctgcttctctt c cg	0.14		152 (G)
	M 88-R	ggtcttccatgtgatggtcag	0.16		

续表 1

基因座 Locus	引物 Primer	序列(5'→3') Sequence(5'→3')	终浓度 Final concentration ($\mu\text{mol/L}$)	复性温度($^{\circ}\text{C}$) Annealing temperature($^{\circ}\text{C}$)	产物大小(bp) Product size(bp) (genotype)
M89	M 89-F	gaattccaagctggtagtctt	0.16	61	142 (C)
	M 89-G	gctgc ttctcaggcaaatgagag ttg	0.10		137 (T)
	M 89-A	aactcaggcaaatgaga cata	0.16		
M95	M 95-F	tgaacccactttcacaaca	0.18	61	160 (T)
	M 95-T	g tatgaaaggctaagccat gcaa	0.10		155 (C)
	M 95-C	gaaaggctaagccatc gag	0.16		
M111	M 111-F	agaacaagtctgttttcaca	0.20	61	83 (TT)
	M 111-R	gctgtgacacttctttgtgt	0.20		81 (del TT)
M117	M 117-F	tcaccaaggaatgcacatc	0.20	61	187 (ATCT)
	M 117-R	gcatctccctaatagttgga	0.20		183 (del ATCT)
M119	M 119-G	gttattccaattcagcataca cgcg	0.12	61	171 (C)
	M 119-T	tccaattcagcatacag cct	0.20		166 (A)
	M 119-R	agggaaatgccaaggtaaatg	0.20		
M121	M 121-F	ttcaacatcgctaaacacacg	0.20	60	198 (AGAAA)
	M 121-R	agggggattttgttaacg	0.20		193 (del AGAAA)
M122	M 122-T	gaaaagcaattgagataactaat gat	0.14	60	123 (T)
	M 122-C	gcaattgagataactaat acac	0.20		118 (C)
	M 122-R	caggcgatgctgatatgcta	0.20		
M134	M 134-A	gagag taacttttgatccccacc tata	0.14	60	172 (del G)
	M 134-C	atacttttgatccccacca ttc	0.20		168 (G)
	M 134-R	ggggctttctgaagcaaata	0.20		
M159	M159-C	gttttattatgatgcaagccct tac	0.12	60	120 (C)
	M159-A	attatgatgcaagccctaaa	0.20		115(A)
	M159-R	ttctgtgttccttgctcca	0.20		
M164	M164-T	atatgtgccaggcatcaagca gct	0.10	61	173 (T)
	M164-C	tgccaggcatcaagc tccc	0.14		168 (C)
	M164-R	aaaatggctctggagcaaatcc	0.14		
M175	M175-F	cccaaatcaactcaactccag	0.14	62	221 (TTCTC)
	M175-R	tgcagcatttcagttagcc	0.14		216 (del TTCTC)
M214	M214-F	aggctgattttgctgctgat	0.20	60	112 (T)
	M214-T	ctgtgagacactgtctgaaaac taca	0.10		108 (C)
	M214-C	gagacactgtctgaaaaca tgcg	0.16		
LINE1*	LINE1-F	gcacaatgtgcacatgtacccta	0.20	61	120~140
	LINE1-R	tgatgtgtcattcatctcatatat	0.20		(LINE1+)
MSY2	MSY2-F	ctggctttggggacgtagta	0.20	62	499 (4 repeats)
	MSY2-R	ctgcgctttatgagcatacg	0.20		400 (3 repeats)

续表 1

基因座 Locus	引物 Primer	序列(5'→3') Sequence(5'→3')	终浓度 Final concentration ($\mu\text{mol/L}$)	复性温度($^{\circ}\text{C}$) Annealing temperature($^{\circ}\text{C}$)	产物大小(bp) Product size(bp) (genotype)
RPS4 Y ₇₁₁	RPS4 Y ₇₁₁ -T	t <u>aga</u> cggcaataaaccttggatt <u>act</u>	0.20	59	189 (T)
	RPS4 Y ₇₁₁ -C	ggcaataaaccttggat <u>atcc</u>	0.30		184 (C)
	RPS4 Y ₇₁₁ -R	cacaagggggaaaaaacac	0.30		
SRY ₊₄₆₅	SRY ₊₄₆₅ -F	tatcgacctctgcggaagg	0.20	60	104(T)
	SRY ₊₄₆₅ -T	aa <u>ggag</u> ttgtccagttgcac <u>atca</u>	0.16		99(C)
	SRY ₊₄₆₅ -C	ggtgtccagttgcactcg	0.20		
IMS-JST 164520*	JST164520-T	atccaactaatgccaccaccaa <u>cgt</u>	0.10	61	123 (T)
	JST164520-C	actaatgccaccacca <u>tggc</u>	0.12		118 (C)
	JST164520-R	gaagacgtttcttgccctga	0.12		
IMS-JST 021354	JST021354-C	atc <u>acgg</u> ttaagtcattgatct <u>gagc</u>	0.10	60	133 (C)
	JST021354-T	ggttaagtcattgatctc <u>tgt</u>	0.10		128 (T)
	JST021354-R	actaaccccaaatccaagg	0.12		
IMS-JST 003305	JST003305-A	tcatcagctctgtgactggita <u>tta</u>	0.12	60	127 (A)
	JST003305-G	agctctgtgactggtt <u>t</u> <u>tatg</u>	0.20		122 (G)
	JST003305-R	tgtaaacacaaagccagccta	0.20		

注:引物序列中带下划线的碱基为引入的错配碱基,带星号基因座执行热启动 PCR。

Note: The underlined bases in primer sequences denote the noncomplementary nucleotides. The asterisk denotes hot start reactions were performed.

1.3 PCR 扩增及产物检测

反应体系 10 μL , 内含 100 $\mu\text{mol/L}$ dNTPs, 50 mmol/L KCl, 10 mmol/L Tris-HCl (pH 8.3), 1.5 mmol/L MgCl_2 , 0.5 U Taq DNA 聚合酶, 5~20 ng 模板 DNA, 优化后的各引物浓度见表 1。PCR 循环参数为 95 $^{\circ}\text{C}$ 2 min, 然后 94 $^{\circ}\text{C}$ 30 s, 最佳复性温度(表 1) 30 s, 72 $^{\circ}\text{C}$ 35 s, 共 32 个循环, 再 72 $^{\circ}\text{C}$ 延伸 5 min。其中, M7、M50、LINE1 和 IMS-JST164520 4 个标记采用热启动 PCR, 于反应温度达到 90 $^{\circ}\text{C}$ 后再加入 Taq DNA 聚合酶。扩增产物经非变性聚丙烯酰胺凝胶(T=6%, C=5%)电泳后, 银染显带^[6]。

1.4 测序

选择 6 个 Y-SNPs (M9, M89, M95, M119, M214 和 IMS-JST003305) 不同等位基因的 PCR 产物纯化后, 送上海申友公司测序。

1.5 统计分析

用直接计数法计算双等位基因标记的等位基因和单体群频率。基因多样性(gene diversity, GD)和单体群多样性(haplogroup diversity, HD)按公式 h

$= n(1 - \sum x_i^2) / (n - 1)$ 计算, 其中 n 为样本数, x_i 为等位基因或单体群频率。

2 结果

2.1 23 个 Y 染色体双等位基因标记分型结果及基因频率分布

23 个 Y 染色体双等位基因标记中, 有 20 个在武汉汉族群体检出多态性, 仅 M50、M159 和 M164 3 个标记无多态性。20 个多态性标记除 LINE1 和 MSY2 外, 每个标记均能检出大小相差 2~5 bp 的 2 个不同等位基因(表 1, 图 1), 每个男性个体均为单一谱带, 女性样品无扩增产物。选择用于测序的 6 个标记其分型结果与测序结果一致。23 个标记的等位基因频率分布见表 2。

2.2 20 个多态性标记构成的单体群频率分布

20 个多态性双等位基因标记在武汉汉族群体中共观察到 35 种不同单体群组合(Hg1~35), 单体群频率分布及其多样性值见表 3。

表 2 23 个 Y 染色体双等位基因标记的基因多样性

Table 2 Gene diversity of 23 Y chromosome biallelic markers (n=160)

基因座 Locus	等位基因 Alleles	频率 Frequency	基因多样性 Gene diversity
M7	C/G	0.9625/0.0375	0.0726
M9	G/C	0.9062/0.0938	0.1711
M50	T/C	1.0000/0.0000	0.0000
M88	A/G	0.9875/0.0125	0.0248
M89	T/C	0.9187/0.0813	0.1503
M95	C/T	0.9375/0.0625	0.1179
M111	TT/del TT	0.9875/0.0125	0.0248
M117	ATCT/del ATCT	0.8437/0.1563	0.2654
M119	A/C	0.8437/0.1563	0.2654
M121	AGAAA/del AGAAA	0.9875/0.0125	0.0248
M122	C/T	0.6000/0.4000	0.4830
M134	G/del G	0.7125/0.2875	0.4123
M159	A/C	1.0000/0.0000	0.0000
M164	T/C	1.0000/0.0000	0.0000
M175	TTCTC/del TTCTC	0.1438/0.8562	0.2477
M214	C/T	0.8937/0.1063	0.1912
LINE1	+ / -	0.1563/0.8437	0.2654
MSY2	3repeats/4repeats	0.1563/0.8437	0.2654
RPS4 Y ₇₁₁	C/T	0.9250/0.0750	0.1396
SRY + ₄₆₅	C/T	0.9937/0.0063	0.0126
IMS-JST164520	T/C	0.5937/0.4063	0.4855
IMS-JST021354	T/C	0.6312/0.3688	0.4685
IMS-JST003305	A/G	0.9187/0.0813	0.1503

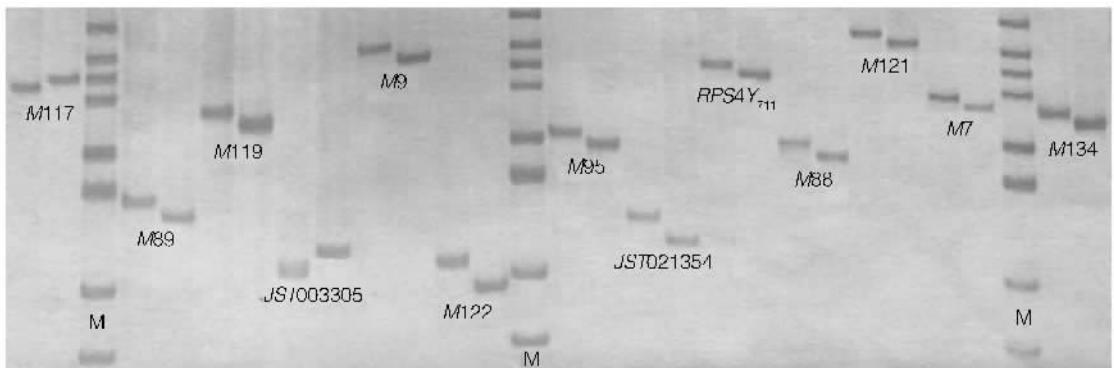


图 1 部分 Y 染色体双等位基因标记电泳分型图

M: pBR322/ Msp I 分子量标准。

Fig.1 The gel patterns of some Y chromosome biallelic markers

M: pBR322/ Msp I Marker.

表 3 35 种单体群频率分布及其多样性值 (n = 160)

Table 3 Frequencies of 35 haplogroups and haplogroup diversity (n = 160)

单体群 Haplogroup	基因型组合 Arrangement of genotypes	例数 No.	频率 Frequency
1	C/G/A/T/C/TT/del ATCT/T/AGAAA/C/del G/del TTCTC/C/-/4/C/C/T/C/A	20	0.1250
2	C/G/A/T/C/TT/ATCT/G/AGAAA/T/G/del TTCTC/C/-/3/C/C/T/T/A	15	0.0937
3	C/G/A/T/C/TT/ATCT/T/AGAAA/C/del G/del TTCTC/C/-/4/C/C/T/C/A	14	0.0875
4	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/+4/C/C/C/T/A	13	0.0812
5	C/G/A/T/C/TT/ATCT/G/AGAAA/T/G/del TTCTC/C/-/3/C/C/C/T/A	10	0.0625
6	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/+4/C/C/T/T/A	8	0.0500
7	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/-/4/C/C/T/T/A	7	0.0437
8	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/-/4/C/C/C/T/A	7	0.0437
9	C/G/A/T/C/TT/ATCT/T/AGAAA/C/del G/del TTCTC/C/-/4/C/C/C/C/A	7	0.0437
10	C/C/A/C/C/TT/ATCT/T/AGAAA/T/G/TTCTC/T/-/4/T/C/T/T/G	7	0.0437
11	C/G/A/T/T/TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/C/T/T/A	5	0.0312
12	C/C/A/C/C/TT/ATCT/T/AGAAA/T/G/TTCTC/T/-/4/T/C/C/T/G	5	0.0312
13	C/G/A/T/C/TT/del ATCT/T/AGAAA/C/del G/del TTCTC/C/-/4/C/C/C/C/A	5	0.0312
14	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/TTCTC/C/-/4/C/C/T/T/A	4	0.0250
15	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/C/T/T/A	3	0.0187
16	C/G/A/T/T/TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/C/C/T/A	3	0.0187
17	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/-/4/C/C/T/C/A	3	0.0187
18	G/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/-/4/C/C/C/C/A	3	0.0187
19	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/TTCTC/C/-/4/C/C/C/T/A	2	0.0125
20	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/C/C/T/A	2	0.0125
21	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/-/4/C/C/C/C/A	2	0.0125
22	C/C/A/T/C/TT/ATCT/T/AGAAA/T/G/TTCTC/T/-/4/C/C/T/T/A	2	0.0125
23	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/T/C/T/A	1	0.0063
24	G/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/+4/C/C/C/C/A	1	0.0063
25	C/G/A/T/C/TT/ATCT/T/del AGAAA/C/G/del TTCTC/C/-/4/C/C/C/T/A	1	0.0063
26	C/G/G/T/T/del TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/C/C/T/A	1	0.0063
27	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/TTCTC/T/-/4/C/C/T/T/A	1	0.0063
28	C/G/A/T/C/TT/ATCT/T/AGAAA/T/G/TTCTC/T/-/4/C/C/C/T/A	1	0.0063
29	G/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/+4/C/C/T/C/A	1	0.0063
30	C/G/G/T/T/del TT/ATCT/T/AGAAA/T/G/del TTCTC/C/-/4/C/C/T/T/A	1	0.0063
31	C/C/A/C/C/TT/ATCT/T/AGAAA/T/G/TTCTC/T/-/4/C/C/T/T/G	1	0.0063
32	C/G/A/T/C/TT/ATCT/T/del AGAAA/C/G/del TTCTC/C/-/4/C/C/T/T/A	1	0.0063
33	G/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/-/4/C/C/T/C/A	1	0.0063
34	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/+4/C/C/T/C/A	1	0.0063
35	C/G/A/T/C/TT/ATCT/T/AGAAA/C/G/del TTCTC/C/+4/C/C/C/C/A	1	0.0063
合计 Total		160	1.0000

单体群多样性 Haplogroup diversity = 0.9471

注:基因型组合的双等位基因标记排列顺序为 M7/M9/M88/M89/M95/M111/M117/M119/M121/M122/M134/M175/M214/LINE1/MSY2/RPS4Y₇₁₁/SRY₊₄₆₅/IMS-JST164520/IMS-JST021354/IMS-JST003305。

Note: The order of biallelic markers in arrangement of genotypes was M7/M9/M88/M89/M95/M111/M117/M119/M121/M122/M134/M175/M214/LINE1/MSY2/RPS4Y₇₁₁/SRY₊₄₆₅/IMS-JST164520/IMS-JST021354/IMS-JST003305.

3 讨论

3.1 FLDAS-PCR 分型方法学

等位基因特异性 PCR (allele specific PCR, AS-PCR) 是一种广泛应用于已知点突变检测的传统技术,但存在需分管扩增、分道电泳、实验条件要求严格等缺点。为了简化操作、提高特异性,部分学者在 AS-PCR 基础上先后建立了诱变分离 PCR (mutagenically separated PCR, MS-PCR)^[7] 和扩增产物长度多态性 (amplified product length polymorphism, APLP)^[8] 技术。其共同点是在等位基因特异性引物的 3' 端第 2、3 或 4 位碱基引入 1 个错配碱基,以提高特异性,并在其中 1 条引物的 5' 端加上不同长度的碱基使 2 个等位基因形成长度差异,由此基因分型可在同一反应管内扩增并于相同泳道分离。不同之处是 MS-PCR 将长引物中和短引物 5' 末端两位碱基对应的碱基引入错配,防止短扩增产物以长扩增产物为模板继续延伸,以此进一步提高特异性。事实证明,MS-PCR 和 APLP 是非常简便有效的突变检测技术。综合 MS-PCR 和 APLP 法设计原理,我们直观地命名为 FLDAS-PCR^[1]。

本文根据 GenBank 里的序列信息,利用 Primer 3 软件按上述策略自行设计了分型 15 个 Y-SNPs (文献报道的 *SRY*₊₄₆₅ 引物也是依此原理设计) 和 1 个 indels *M134* 的引物序列 (表 1)。用这些引物分型效果良好,除 3 个未检出多态性的标记外,其余每个基因座均检出长度相差 4~5 bp 的 2 个等位基因 (图 1),选择用于测序的 6 个标记其分型结果与测序结果一致,且女性样本没有扩增产物,证明本文所设计的分型引物是可靠的。另外,设计的用于检测 5 个 indels 的 5 对普通引物 (另 2 个 indels *LINE1* 和 *MSY2* 采用文献报道引物) 也获得了良好的分型效果。

3.2 汉族群体 20 个双等位基因标记多态性及部分标记不同群体之间的比较

双等位基因标记也称为二态性标记 (binary polymorphisms) 或单次突变事件多态性 (unique event polymorphisms, UEPs)。通常,根据双等位基因确定的不同的染色体,称为单体群 (haplogroups), 而由 STR 所确定的不同的染色体称为单倍型 (haplotypes)。由双等位基因和 STR 共同确定的不同染色体称为谱系 (lineage)^[9]。双等位基因标记突变率低,不易受重组和回复突变的影响,由其构成

的单体群保持完整,是进化事件的忠实记录者,可以用于人类的起源和迁移研究。

本文调查的 23 个双等位基因标记中,有 20 个在武汉汉族群体中具有遗传多态性,其基因频率分布见表 2。值得注意的是 *LINE1*, 该标记为一种逆转录子插入 (retroposon insertion), Santos 等^[4] 在研究中发现 *LINE1*(+) PCR 产物条带较宽且边缘模糊,推测可能是因为 PCR 过程中 *Taq* DNA 聚合酶在跨越 poly (A) 尾时滑动所致。同时,他们还发现 *LINE1*(+) 扩增产物的大小在少数个体发生变化。本文在实验中也观察到类似现象 (图 2)。从表 2 可以看出, *LINE1*(+) 在武汉汉族群体中的频率为 15.63%, 这一结果与以前文献报道的 *LINE1*(+) 在汉族群体中具有较高频率是一致的^[4,10]。此外, *SRY*₊₄₆₅ 的等位基因频率分布在中国人和日本人之间具有显著性差异: 本文调查的中国汉族 T 等位基因频率为 0.63%, 而日本人中则高达 25%~31%^[3,10,11]。

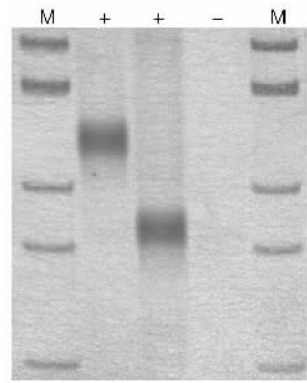


图 2 *LINE1* 电泳分型图

M: pBR322/*Msp* I 分子量标准; +: *LINE1*(+); -: *LINE1*(-)。

Fig. 2 The gel patterns of *LINE1* locus

M: pBR322/*Msp* I Marker; +: *LINE1*(+); -: *LINE1*(-)。

20 个多态性标记对应的基因多样性范围从 0.0126 到 0.4855, 其中以 *IMS-JST164520* 的 GD 值最高, *M122* 次之, *SRY*₊₄₆₅ 最低 (表 2)。由于 Y 染色体上的遗传标记间存在连锁, 故实际检出的单倍体数量要远远低于理论上可能组合的数量^[12]。本研究在 160 个男性个体中共检出 35 种不同单体群, 其中 Hg1 频率最高, 为 0.1250, Hg23~Hg35 频率最低, 为 0.0063。35 种单体群多样性为 0.9471。此外, 20 个标记中, *M89* 和 *IMS-JST003305* 2 个标记

总是紧密联系在一起,只出现 TA 和 CG 组合,未发现 TG 和 CA 组合。

表 4 为部分双等位基因标记在不同汉族群体中的基因频率分布。从表中可以发现, M9 和 M119

在 3 个南方汉族群体中基因频率分布相似,而 M89、M111、M122 和 M134 四个标记在几个汉族群体间频率分布差异较大,提示其存在明显的地域性差异,可作为群体迁徙研究的候选标记。

表 4 不同汉族群体部分双等位基因标记的基因频率分布

Table 4 Allele frequency distribution of a few biallelic markers in different Han population

汉族群体 Han population	M9G	M89T	M111del TT	M119C	M122C	M134del G	SRY ₊₄₆₅ T
武汉 Wuhan (n=160)	0.9062	0.9187	0.0125	0.1563	0.6000	0.2875	0.0063
福建 Fujian (n=79) ^[12]	0.9620	0.7625	0.1139	0.1392	0.1646	0.4557	—
四川 Sichuan (n=56) ^[12]	0.9643	0.1852	0.0536	0.1818	0.1964	0.0893	—
北京 Beijing (n=69) ^[10]	0.7536	—	—	—	—	—	0.0580

注:“—”代表未检测。

Note: ‘—’ denotes not detected.

参考文献(References):

- [1] HUANG Dai-Xin, YANG Qing-En, ZHAO Gui-Sen. A simple and rapid modified-new method for SNP typing by fragment length discrepant allele specific PCR. *J Forensic Med*, 2005, 21(1): 11~14.
黄代新, 杨庆恩, 赵贵森. 片段长度差异等位基因特异性 PCR——一种改良的 SNP 分型新方法. 法医学杂志, 2005, 21(1): 11~14.
- [2] Walsh P S, Metzger D A, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, 1991, 10: 506~513.
- [3] Naito E, Umetsu K, Yuasa I, Dewa K, Sumi H, Yamanouchi H. A novel dimorphism in the human SRY gene: usefulness in human migration studies. *Int J Legal Med*, 2001, 114 (4-5): 274~277.
- [4] Santos F R, Pandya A, Kayser M, Mitchell RJ, Liu A, Singh L, Destro-Bisol G, Novelletto A, Qamar R, Mehdi SQ, Adhikari R, de Knijff P, Tyler-Smith C. A polymorphic L1 retroposon insertion in the centromere of the human Y chromosome. *Hum Mol Genet*, 2000, 9: 421~430.
- [5] Bao W, Zhu S, Pandya A, Zerjal T, Xu J, Shu Q, Du R, Yang H, Tyler-Smith C. MSY2: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations. *Gene*, 2000, 244: 29~33.
- [6] Bassam B J, Caetano-Anolles G, Gresshoff P M. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem*, 1991, 196 (1): 80~83.
- [7] Rust S, Funke H, Assmann G. Mutagenically separated PCR (MS-PCR): a highly specific one step procedure for easy mutation detection. *Nucleic Acids Res*, 1993, 21 (16): 3623~3629.
- [8] Watanabe G, Umetsu K, Yuasa I, Suzuki T. Amplified product length polymorphism (APLP): a novel strategy for genotyping the ABO blood group. *Hum Genet*, 1997, 99 (1): 34~37.
- [9] Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*, 2002, 12 (2): 339~348.
- [10] Jin H J, Kwak K D, Hammer M F, Nakahori Y, Shinka T, Lee J W, Jin F, Jia X, Tyler-Smith C, Kim W. Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans. *Hum Genet*, 2003, 114 (1): 27~35.
- [11] Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Yoshitome K, Miyaishi S, Ishizu H. Typing of Y chromosome single nucleotide polymorphisms in a Japanese population by a multiplexed single nucleotide primer extension reaction. *Leg Med*, 2002, 4: 202~206.
- [12] YU Min, ZHANG Yong-Li, CHEN Feng, XUE Ya-Li, YU Yang, MA Lin-Lin, HUANG Xiao-Yi, LIU An, SHI Rong-Qian, LÜ Fu-Qu, HUANG Cheng-Bin, ZHANG Gui-Yin, LI Pu, FU Song-Bin. Y chromosome variations and haplogroups from 15 biallelic markers in six Chinese populations. *Acta Genetica Sinica*, 2002, 29 (4): 283~289.
于敏, 张咏莉, 陈峰, 薛雅丽, 于畅, 马琳琳, 黄小义, 刘岸, 史榕茜, 吕芙蕖, 黄承滨, 张贵寅, 李璞, 傅松滨. 中国 6 个人群中 Y 染色体 15 个双等位基因标记变异频率分布及单体系群分析. 遗传学报, 2002, 29 (4): 283~289.