

· 技术与应用 ·

广西6个民族Y-STR网络图分析及其法医学意义

肖月¹, 邓盼², 常开创³, 马泉⁴, 钱恩芳^{4,5}, 余建华⁶, 程宝文⁷, 李彩霞⁴, 江丽⁴

(1. 广东中一司法鉴定中心, 广东 深圳 518000; 2. 株洲市公安局, 湖南 株洲 412007; 3. 昆明市公安局, 云南 昆明 650034; 4. 公安部物证鉴定中心 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; 5. 贵州医科大学法医学院, 贵州 贵阳 550004; 6. 昭通市公安局, 云南 昭通 657000; 7. 云南警官学院, 云南 昆明 650221)

摘要: 目的 探究Y-SNP和Y-STR遗传标记信息在不同民族中的遗传结构分布及其法医学应用。方法 采用SNaPshot微测序技术对广西汉族、广西京族、广西苗族、广西瑶族、广西壮族和广西侗族6个民族439名男性个体12个Y-SNP位点多态性进行检测, 采用DNATyper™ Y26试剂盒对26个Y-STR基因座进行复合扩增, 3130xl基因分析仪进行基因分型。对同一Y-SNP单倍群下不同个体Y-STR单倍型使用Network 5.0绘制网络图并分析。结果 12个Y-SNP位点在6个民族中检出6个单倍群, 26个Y-STR基因座检出362个单倍型, 单倍型多样性为0.996 6。C单倍群中, 广西瑶族、广西壮族和广西侗族样本聚集在不同簇上; O1单倍群中, 广西壮族、广西苗族和广西京族样本相对独立并分别聚集一簇; O2单倍群中部分广西苗族和广西瑶族样本聚集一簇。结论 基于同一Y-SNP单倍群下Y-STR网络图分析, 能初步对部分民族进行区分, 为相关地区利用犯罪现场遗留男性生物检材的检验推断族群来源。

关键词: 法医遗传学; 多态性; 单核苷酸; Y染色体; 短串联重复; 单倍型; 广西

中图分类号: DF795.2 文献标志码: A doi: 10.12116/j.issn.1004-5619.2019.03.010

文章编号: 1004-5619(2019)03-0314-05

Network Analysis of Y-STR in Six Ethnic Populations in Guangxi and Its Forensic Significances

XIAO Yue¹, DENG Pan², CHANG Kai-chuang³, MA Quan⁴, QIAN En-fang^{4,5}, YU Jian-hua⁶, CHENG Bao-wen⁷, LI Cai-xia⁴, JIANG Li⁴

(1. Guangdong Zhongyi Forensic Science Center, Shenzhen 518000, Guangdong Province, China; 2. Zhuzhou Municipal Public Security Bureau, Zhuzhou 412007, Hunan Province, China; 3. Kunming Municipal Public Security Bureau, Kunming 650034, China; 4. Key Laboratory of Forensic Genetics, National Engineering Laboratory for Forensic Science, Beijing Engineering Research Center of Crime Scene Evidence Examination, Institute of Forensic Science, Ministry of Public Security, PRC, Beijing 100038, China; 5. Institute of Forensic Medicine, Guizhou Medical University, Guiyang 550004, China; 6. Zhaotong Municipal Public Security Bureau, Zhaotong 657000, Yunnan Province, China; 7. Yunnan Police Officer Academy, Kunming 650221, China)

Abstract: **Objective** To explore the distribution of genetic structure of Y-SNP and Y-STR genetic markers in different ethnic groups and its application in forensic science. **Methods** SNaPshot minisequencing was used to detect the polymorphisms of 12 Y-SNP loci in 439 males from 6 ethnic groups, including Guangxi Han, Guangxi Jing, Guangxi Miao, Guangxi Yao, Guangxi Zhuang and Guangxi Dong. DNATyper™ Y26 kit was used to multiplex-amplify 26 Y-STR loci. The PCR products were analyzed by 3130xl genetic analyzer. The network analysis of Y-STR haplotype under the same Y-SNP haplogroup was analyzed by Network 5.0 software. **Results** Six haplogroups defined by 12 Y-SNP loci were detected in 6 ethnic groups, and 362 haplotypes were detected in 26 Y-STR loci. The haplotype diversity was 0.996 6. In the C haplogroup, the samples from Guangxi Yao, Guangxi Zhuang and Guangxi Dong were clustered on different branches; in the O1 haplogroup, those from Guangxi Zhuang, Guangxi Miao and Guangxi Jing were relatively independent and clustered separately; in the O2 haplogroup, some samples from Guangxi Miao and Guangxi Yao were gathered in a cluster. **Conclusion** Based on the Y-STR network analysis of samples with identical haplogroup of Y-SNP, some ethnic groups can be preliminarily distinguished, which could be used to infer male suspects' ethnic group through detecting their genetic markers left in the crime scene.

Keywords: forensic genetics; polymorphism, single nucleotide; Y chromosome; short tandem repeat; haplotype; Guangxi

基金项目: 国家重点研发计划资助项目(2017YFC0803501); 国家自然科学基金资助项目(81772027); 公安部技术研究计划资助项目(2015JSYJC47, NO2016JSYJA05); 国家科技资源共享服务平台计划资助项目(YCZYPT[2017]01-3); 公安部物证鉴定中心基本科研业务费专项资助项目(2017JB027, 2016JB039, 2017JB025)

作者简介: 肖月(1992—), 女, 硕士研究生, 主要从事法医物证学鉴定和研究; E-mail: 625287645@qq.com

通信作者: 李彩霞, 女, 博士, 主任法医师, 主要从事法医物证学研究及鉴定; E-mail: licaixia@tsinghua.org.cn

通信作者: 江丽, 女, 博士, 副主任法医师, 主要从事法医物证学研究及鉴定; E-mail: jl@mail.bnu.edu.cn

目前,我国多省市正广泛开展Y染色体短串联重复(Y-chromosome short tandem repeat,Y-STR)数据库建设^[1-2],但偏远地区及多民族混居地区建库较困难,因此如何利用现场遗留男性犯罪嫌疑人生物检材的遗传标记来推测其种族、地域来源一直是法医学检验的难点。本研究在对广西6个民族12个Y染色体单核苷酸多态性(Y-chromosome single nucleotide polymorphism,Y-SNP)位点检测基础上,对不同人群同一Y-SNP单倍群下的26个Y-STR基因座分别进行检测,并构建网络图分析,探索不同民族之间的差异。

1 材料与方法

1.1 样本及DNA提取

在志愿者知情同意的原则下,采集广西地区6个民族共439份男性静脉血样本,其中广西汉族65名、广西京族78名、广西苗族86名、广西瑶族73名、广西壮族84名、广西侗族53名。被采集血样的人员追溯其3代以上均居住在广西地区且是本族人员。采用

QIAamp® DNA Blood Midi试剂盒(德国Qiagen公司)进行DNA提取,提取方法按照说明书进行。DNA用NanoDrop 2000c分光光度计(美国Thermo Scientific公司)进行定量。

1.2 Y-SNP的检测分析

依据国际遗传谱系学会(International Society of Genetic Genealogy,ISOGG;<https://isogg.org/tree>)2017年版的谱系树及相关文献^[3-4]选择12个在中国人群中分布频率较高的Y-SNP位点:C-M130、D-M174、F-M89、G-M201、J-M304、K-M9、N-M231、O-M175、Q-M242、R-M207、O1-F265、O2-CTS10736。各位点的扩增引物和延伸引物通过Primer Premier 5.0设计,位点及引物序列信息见表1。采用SNaPshot微测序技术对12个Y-SNP位点进行复合扩增及检测,应用Mastercycler® Pro梯度PCR仪(德国Eppendorf公司)进行复合扩增,采用3130xl基因分析仪(美国AB公司)对扩增产物进行检测,使用GeneMapper ID v3.2软件(美国Thermo Fisher Scientific公司)对数据结果进行分型。

表1 12个Y-SNP位点及其引物序列信息

Y-SNP 位点	单倍群	PCR 扩增引物序列(5'→3')	单碱基延伸引物序列(5'→3')	引物长度/ bp
M130	C	F-TCTCCTCCCTTTCTTTCT R-TTAGCCACTGCTCTGTTT	(CT) ₆ TGCAGGGCAATAAACCTTGGATTTC	37
M174	D	F-TATGATGCACTGTGCGTTCT R-GTGCAATAAACACTATAAGT	(CT) ₈ TCTCTGAATACCTTCTGGAGTGCCC	41
M89	F	F-TGAGGTGCCATGAAAGTG R-AGGCTGGCACACTTTGGGTCCAGGA	(CT) ₁₀ CTGCAACTCAGGCAAAGTGAGAGAT	45
M201	G	F-TATAGTAATTAGTTTCTCAGATCT R-ACGTATCTGAGGTTCAAATCC	(CT) ₁₅ ATCCAGTATCAACTGAGG	48
M304	J	F-AGTTTGTAACAAACAGTATGTGGGA R-GTCTTATACCAAAATATCACCAG	(CT) ₁₂ TGTTCAATTTGAAAGTAACTTGTGA	49
M9	K	F-ACTGCAAAGAAACGGCCTAA R-AGCGCTACCTTACTTACATAAC	(CT) ₁₃ TGTCTAAATTAAAGAAAAATAAAGAG	52
M231	N	F-ATTTACTGTTTCTACTGCTTTC R-TGGCCAGAGTCTTTCACA	(CT) ₈ AACATTTACTGTTTCTACTGCTTTC	41
M175	O	F-CACATGCCTTCTCACTTC R-ACTTTGTCCAATGCTGAA	(CT) ₁₃ TTTGTTTCTGTTTCATTCTTGAGAAG	51
M242	Q	F-AAGCTTTCTACGGCATAG R-AACAACCTCTGAAGCGGTG	(CT) ₁₅ TTGTGCAAAAAGGTGACCAAGGTGCT	56
M207	R	F-TACAACATATGGGGCAAAT R-CTGAAGGAAAAGTGAGT	(CT) ₁₉ AGTCAAGCAAGAAATTTA	56
F265	O1	F-CCAGGAAGTGGTAGAATC R-GCAGCCTAGAGCTATTTT	(AT) ₂₀ AGGTTTAAGTAGCACAGTTTATTCA	65
CTS10736	O2	F-ACAAAGCCAAAGCATATC R-AGAAAAGCACCTACTGAG	(AT) ₂₂ AACGTAAACATAAAGAGAGAAGGTG	69

1.3 Y-STR的检测分析

DNATyper™ Y26 PCR试剂盒为公安部物证鉴定中心研发,其中不包含高突变率基因座。对26个Y-STR基因座用Mastercycler® Pro梯度PCR仪进行复合扩增,扩增产物用3130xl基因分析仪进行检测,使用GeneMapper ID v3.2软件对数据结果进行STR分型。

1.4 统计学分析

用直接计数法计算各个单倍型及单倍群频率。单倍型多样性(haplotype diversity, HD)计算公式为:

$$HD=[n(1-\sum P_i^2)]/(n-1)。(1)$$

式中, n 为观察到的单倍型个数, P_i 为单倍型频率。根据Median-joining方法^[5],对样本分布频率较高、同一Y-SNP单倍群下的Y-STR单倍型使用Network 5.0(英国Fluxus Technology公司)构建网络图(多拷贝Y-STR基因座被排除在外),权重值^[6-7]根据Y-STR基因座突变率^[8]及基因多样性(多样性计算所需数据来源于文献[9])综合评估^[6-8],突变率越高,基因多样性越好的基因座权重值越低(表2)。

表2 Y-STR基因座突变率及6个民族基因多样性

基因座	权重	突变率	民族					
			广西侗族	广西瑶族	广西京族	广西汉族	广西苗族	广西壮族
DYS444	1	0.003	0.783	0.560	0.688	0.722	0.553	0.644
DYS458	1	0.006	0.830	0.490	0.740	0.817	0.831	0.762
DYS635	1	0.004	0.826	0.585	0.781	0.738	0.815	0.762
DYS447	1	0.002	0.737	0.586	0.837	0.814	0.760	0.807
DYS391	2	0.002	0.109	0.080	0.480	0.400	0.516	0.308
DYS393	2	0.001	0.714	0.248	0.668	0.654	0.707	0.628
GATA-H4	2	0.003	0.589	0.557	0.618	0.538	0.686	0.653
DYS460	2	0.001	0.718	0.505	0.595	0.675	0.668	0.521
DYS389 II	2	0.004	0.774	0.435	0.703	0.717	0.491	0.594
DYS19	2	0.002	0.729	0.520	0.712	0.712	0.671	0.485
DYS439	3	0.005	0.656	0.263	0.636	0.663	0.564	0.632
DYS390	3	0.002	0.627	0.557	0.500	0.703	0.610	0.494
DYS448	3	0.002	0.716	0.372	0.616	0.755	0.566	0.561
DYS508	4	0.004	0.496	0.524	0.498	0.474	0.659	0.380
DYS533	4	0.004	0.545	0.294	0.593	0.557	0.578	0.566
DYS389 I	4	0.003	0.597	0.339	0.662	0.618	0.538	0.488
DYS522	4	0.003	0.688	0.501	0.566	0.734	0.405	0.474
DYS617	5	0.004	0.293	0.486	0.329	0.508	0.424	0.114
DYS456	5	0.004	0.354	0.180	0.391	0.584	0.434	0.311
DYS437	5	0.001	0.506	0.390	0.405	0.513	0.427	0.397
DYS392	7	0.001	0.619	0.531	0.392	0.591	0.153	0.476
DYS438	8	0.001	0.538	0.054	0.267	0.311	0.212	0.320

注:突变率来源于YHRD数据库(https://yhrd.org/pages/resources/mutation_rates)及文献[8]

2 结 果

2.1 Y-SNP单倍群频率分布及Y-STR单倍型分型结果

根据ISOGG 2017年版谱系树命名方法,12个Y-SNP位点在6个民族439名样本中检出了C、D、J、N、O1、O2共6个单倍群,其频率及频数分布结果见表3。除了广西瑶族群体在C单倍群高频分布外,其他民族单倍群主要分布在O1和O2单倍群,D、J、N单倍群也存在少量的分布。其中广西汉族、广西京族、广西壮族、广西苗族、广西侗族在O1单倍群高频分布,广西汉族、广西苗族、广西侗族在O2单倍群分布频率也较高。

26个Y-STR基因座在439个广西地区男性中检出362个单倍型,单倍型多样性为0.996 6,在广西汉、京、苗、瑶、壮、侗族单倍型多样性分别为0.997 6、0.994 2、0.983 2、0.933 9、0.986 3、0.997 6。

2.2 C、O1、O2单倍群下构建Y-STR网络图

6个民族分别在C、O1、O2单倍群分布频率较高,对不同民族同一Y-SNP单倍群下22个Y-STR单倍型绘制网络图(DYS385a/b、DYS527a/b未纳入),在C单倍群(图1A)中,广西瑶族群体有44名,且呈典型的星拓状结构,广西壮族、广西侗族、广西京族、广西汉族分布较少,各个民族相对独立聚集一簇。O1单倍群人群(图1B)分布较广,广西壮族、广西京族、广西苗族以及广西汉族群体较多,其中部分广西京族个体

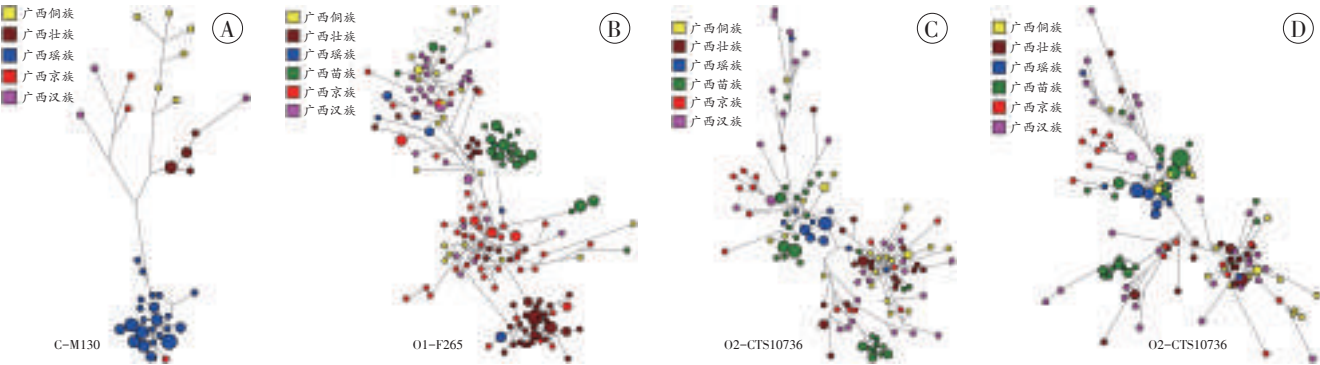
Y-STR单倍型与广西壮族聚集一簇且处在网络图边缘;2名广西汉族个体和1名广西侗族个体及1名广西京族个体共享Y-STR单倍型;1名广西瑶族个体与1名广西壮族个体共享Y-STR单倍型,但两个个体在*DYS527a/b*基因座上存在差异,分别为21-26、23-26。O2单倍群(图1C)在各个民族均有分布,广西汉族、广

西瑶族、广西侗族群体较多,部分广西苗族和广西瑶族聚集一簇,广西壮族、广西京族分布较少。基于AmpF ℓ STRTM YfilerTM PCR扩增试剂盒(美国AB公司)中的15个Y-STR单倍型构建网络图(*DYS385a/b*未纳入)(图1D)时,其共享单倍型个体比22个Y-STR单倍型多且单倍型间的分支较少。

表3 6个Y-SNP单倍群在广西6个民族中的频率及频数分布

群体	C	D	J	N	O1	O2
广西汉族	0.046(3)	-	-	0.046(3)	0.462(30)	0.446(29)
广西京族	0.038(3)	0.064(5)	0.013(1)	0.013(1)	0.679(53)	0.192(15)
广西苗族	-	-	-	-	0.523(45)	0.477(41)
广西瑶族	0.603(44)	-	-	-	0.137(10)	0.260(19)
广西壮族	0.083(7)	-	-	-	0.655(55)	0.262(22)
广西侗族	0.113(6)	0.019(1)	-	0.038(2)	0.415(22)	0.415(22)

注:“-”表示无分布;括号中为频数



A: 基于C单倍群下22个Y-STR网络图;B: 基于O1单倍群下22个Y-STR网络图;C: 基于O2单倍群下22个Y-STR网络图;D: 基于O2单倍群下15个Y-STR网络图

图1 广西6个民族基于C、O1、O2单倍群下的Y-STR网络图

3 讨论

如何利用现场遗留犯罪嫌疑人生物检材的遗传标记来推测其种族、地域来源一直是法医学检验的难点和重点。近年来,在汉族聚集的省市,通过对现场遗留男性生物检材的Y-STR分型检验进行男性家系排查侦破了许多重大现案及积案^[10],而在中原文化影响较小、多民族混居地区的省份很难建设Y-STR数据库并进行此类排查。Y染色体非重组区的Y-SNP突变率低,具有种族特异性^[11-12],如本研究中6个民族的Y-SNP单倍群主要以O1、O2和C单倍群为主,其中广西瑶族的C单倍群频率最高,而其他群体分别在O1或O2单倍群高频分布。因此,仅利用现场遗留男性生物检材的Y-SNP单倍群分型很难区分其种族来源。Y-SNP单倍群划分可以揭示群体间的大尺度遗传差异,对不同人群相同单倍群下的Y-STR单倍型进行分析,则可以进一步描述群体间的小尺度遗传结构的差别。

2015年,OLOFSSON等^[6]通过对3个群体15个Y-STR基因座的网络图分析发现不同的群体可聚在不同的分支上。本研究的DNATyperTM Y26试剂盒包括了AmpF ℓ STRTM YfilerTM PCR扩增试剂盒中的17个基因座,还增加了9个(*DYS444*、*DYS447*、*DYS460*、*DYS508*、*DYS522*、*DYS533*、*DYS617*、*DYS527a/b*)针对中国人群遗传特征的Y-STR基因座^[13-14],在439个广西地区男性样本中共检出362个单倍型,HD为0.9966。对同一Y-SNP单倍群下不同人群22个Y-STR单倍型绘制的网络图上,C单倍群中,广西瑶族、广西壮族和广西侗族相对独立并分别聚集在不同簇上,能较好地对不同民族进行区分;O1单倍群中,除部分广西京族个体样本与壮族群体样本混聚在一起外,广西壮族、广西苗族和广西京族相对独立并分别聚集一簇,其中2名广西汉族个体和1名广西侗族个体、1名广西京族个体Y-STR分型相同而聚在一起;O2单倍群各民族聚集特点不明显,本研究仅发现部分广西苗族和广西瑶族样本相对聚集一簇,网络图较复杂,这可

能与O2单倍群在东亚人群中频率分布高且包含亚群较多^[4, 15-16]、本研究选取的Y-STR基因座及各基因座在网络图中的权重值等因素有关。下一步随着检测样本的增加,基于更精细O2单倍群亚群及更多的Y-STR基因座构建的网络图能更好地区分不同起源的民族。如本研究中O2单倍群下基于15个Y-STR单倍型构建网络图时,其共享单倍型个体比22个Y-STR单倍型多且单倍型间的分支较少,增加基因座后,网络图能更好地区分不同族群。

在常规Y-STR基因座检验的基础上,通过12个Y-SNP复合体系检测并对C、O1和O2单倍群下Y-STR数据进行了网络图分析,C、O1单倍群能初步对不同民族进行区分,为此类地区利用犯罪现场遗留男性生物检材的检验推断其族群来源,进而为案件的侦破提供方向。

参考文献:

- [1] 葛建业,严江伟,谢群,等. 中国Y-STR数据库建设相关问题探讨[J]. 法医学杂志,2013,29(3):212-215,221.
- [2] 陈振乾,黄书琴. 郑州市Y-STR DNA数据库建设及应用的调查研究[J]. 中国人民公安大学学报(自然科学版),2017,23(1):15-19.
- [3] SHI H, ZHONG H, PENG Y, et al. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations[J]. BMC Biol,2008,6:45.
- [4] SHI H, DONG Y L, WEN B, et al. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122[J]. Am J Hum Genet, 2005,77(3):408-419.
- [5] BANDELT H J, FORSTER P, ROHL A. Median-joining networks for inferring intraspecific phylogenies[J]. Mol Biol Evol,1999,16(1):37-48.
- [6] OLOFSSON J K, MOGENSEN H S, BUCHARD A, et al. Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit[J]. Forensic Sci Int Genet,2015,16:232-236.
- [7] ULLAH I, OLOFSSON J K, MARGARYAN A, et al. High Y-chromosomal differentiation among ethnic groups of Dir and Swat Districts, Pakistan[J]. Ann Hum Genet,2017,81(6):234-248.
- [8] 王新杰,许欣,黄磊,等. 山东汉族人群63个Y-STR基因座突变观察及法医学应用[J]. 刑事技术,2016,41(5):424-428.
- [9] 邓盼,江丽,马泉,等. 广西6个民族24个Y-STR基因座遗传多态性及群体遗传结构分析[J]. 刑事技术,2018,43(3):198-201.
- [10] YAO H, WEN S, TONG X, et al. Y chromosomal clue successfully facilitated the arrest of Baiyin serial killer[J]. Sci Bull,2016,61(22):1715-1717.
- [11] SU B, XIAO C, DEKA R, et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas[J]. Hum Genet,2000,107(6):582-590.
- [12] KAYSER M. Forensic use of Y-chromosome DNA: A general overview[J]. Hum Genet, 2017, 136(5): 621-635.
- [13] 刘亚举,李效阳,岳俊涛,等. 河南汉族人群22个Y-STR基因座遗传多态性[J]. 中国法医学杂志,2013,28(6):502-503.
- [14] 姜先华,贾菲,沈红缨,等. 复合扩增20个Y-STR基因座及其法医学的应用[J]. 中国法医学杂志,2013,28(1):18-22.
- [15] YAN S, WANG C C, LI H, et al. An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4[J]. Eur J Hum Genet,2011,19(9):1013-1015.
- [16] YAN S, WANG C C, ZHENG H X, et al. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers[J]. PLoS One, 2014, 9(8): e105691.

(收稿日期:2017-12-25)

(本文编辑:张素华)