

# 茶洞话群体的 Y 染色体遗传结构 及其父系起源研究

邓琼英<sup>1</sup>, 王晓庆<sup>1</sup>, 王传超<sup>2</sup>, 李辉<sup>2</sup>

1. 广西医科大学人体解剖学教研室, 南宁 530021;

2. 复旦大学生命科学学院现代人类学教育部重点实验室, 上海 200433

**摘要:** 我国广西的桂东北地区大约有 20,000 人使用茶洞话, 该群体的族源问题一直存在争议。本文为调查茶洞话群体的 Y 染色体遗传结构, 探讨其父系起源, 对临桂县使用茶洞话的 21 名无关男性个体的 Y-STR 和 Y-SNP 进行了检测分型, 并对该群体与周边民族的遗传关系进行了研究分析。结果显示: 茶洞话群体的 17 个 Y-STR 位点具有丰富的遗传多态性, 适用于群体遗传学和法医学研究; Y 染色体高频单倍群为 O2\*-P31 和 O2a1\*-M95, 表明茶洞话群体具有显著的百越民族系统侬傣族群的遗传背景; N-J 树和主成分分析显示茶洞话群体与仫佬族的父系遗传关系较之与毛南族和汉族更亲近。本研究结果为茶洞话群体的族源研究提供了遗传学证据。

**关键词:** 茶洞话群体; Y-STR; Y-SNP; 遗传结构; 父系起源

中图分类号: Q986; 文献标识码: A; 文章编号: 1000-3193(2013)01-0000-07

广西桂东北地区大约有 20000 人使用一种不同于周边民族的新语言, 当地其他族群称之为“茶洞话”, 其使用范围主要集中在临桂县茶洞乡的大部分地区, 以及临桂县两江镇的山口村一带和永福县龙江乡的部分村寨。李锦芳教授<sup>[1]</sup> 研究认为茶洞话不同于周边的汉语、壮语和瑶语, 它应是侬台语族侬水语支中一支新发现的语言。梁金荣博士也在《临桂县志》<sup>[2]</sup> 中记载茶洞话的系属尚未明确, 初步研究认为其与壮侬语族侬水语支的仫佬语比较接近。而王楠楠等<sup>[3]</sup> 通过将茶洞话与侬台语族各语支语言进行了语音和词汇等的比较, 发现在侬水语支内部, 茶洞话与毛南语的关系词比例最高 (68.5%), 其次是与仫佬语 (60.4%), 因此认为茶洞话与毛南语关系最密切, 但与仫佬语的关系也较接近。目前, 使用茶洞话群体的民族成分多被定为汉族, 少部分为壮族, 但这一群体并不认可自己的这一身份, 他们根据自己的族谱、祠堂碑刻和已有的文献记载, 认为其祖先是在明初洪武年间从广西庆远府南丹洲充军到茶洞来镇压当地的桂北壮瑶起义。因此, 他们曾多次向广西壮族自治区人民政府、广西壮族自治区民族委员会递交申请, 要求认定为毛南族 (迁出地的原民族), 但至今这一请求尚未有结果<sup>[3]</sup>。综上所述, 目前学术界对茶洞话的系属尚无明

收稿日期: ; 定稿日期:

基金项目: 广西自然科学基金资助项目 (2011GXNSFA018240); 国家自然科学基金资助项目 (31160222)

第一作者: 邓琼英, 女, 广西博白人, 主要研究方向: 体质人类学、群体遗传学和人体组学, E-mail: yingzideng@sina.com

确的结论, 所做的研究也仅局限于史料和语言等方面, 而这些材料或研究的因素, 往往容易受到客观条件的限制, 难以获得直接证据。

人类 DNA 标记, 具有稳定的世代遗传的特点, 并按照一定的频率发生突变, 可为研究群体的传承、迁徙和融合的历史提供线索, 因此被广泛地应用到人类群体的起源、进化和遗传关系的研究<sup>[4-6]</sup>。其中 Y 染色体非重组区 (Non-recombination portion of the Y chromosome, NRY) 存在两种遗传标记——短串联重复序列 (Short tandem repeat, STR) 和单核苷酸多态 (Single nucleotide polymorphism, SNP), 前者突变率高, 多态性丰富, 后者突变率稳定, 具有群体特异性<sup>[7-8]</sup>。由于 NRY 区严格遵循父系遗传、不重组等特点, 能清晰地记录群体的历史, 因此在家系识别、群体父系遗传学及法医父权鉴定方面有重要的应用价值<sup>[9]</sup>。

本文通过对茶洞话群体的 Y 染色体非重组区的遗传结构进行研究分析, 探索茶洞话群体与周边民族的遗传关系, 从分子人类学的角度为茶洞话群体的父系起源和迁徙提供了遗传学证据。

## 1 材料和方法

### 1.1 材料

在临桂县茶洞乡随机选取无亲缘关系的健康男性个体 21 人, 追溯其 3 代以上均居住在茶洞乡, 均讲茶洞话, 根据“知情同意”的原则, 抽取静脉血 2mL, 枸橼酸钠抗凝, -20℃ 保存, 采用天根试剂盒提取 DNA。

### 1.2 方法

#### 1.2.1 Y-STR 的检测分析

采用 Y-filer 试剂盒 (ABI 公司) 对 21 例样本的 17 个 Y-STR 位点 (*DYS19*、*DYS389I*、*DYS389II*、*DYS390*、*DYS391*、*DYS392*、*DYS393*、*DYS437*、*DYS438*、*DYS439*、*DYS448*、*DYS456*、*DYS458*、*DYS635*、*GATA\_H4*、*DYS385a* 和 *DYS385b*) 进行一次 PCR 复合扩增。PCR 产物纯化后在 ABI 3730 测序仪 (Appliedbiosystems, Carlsbad, CA) 上进行毛细管电泳, 用 GeneMapper ID 3.1 软件对电泳产物进行比对分型。

#### 1.2.2 Y-SNP 的检测分析

采用 SNaPshot (ABI 公司) 和荧光引物 PCR 相结合的方法, 对最新 Y 染色体谱系树上东亚所能见到的 40 个 Y-SNP 位点的单倍群<sup>[10-11]</sup> 进行分型。选取的 SNP 位点主要有单倍群 Corset: M130, P256, M1, M231, M168, M174, M45, M89, M272, M258, M242, M207, M9, M96, P125, M304, M201, M306; 单倍群 O: M175, M119, P203, M110, M268, P31, M95, M176, M122, M324, M121, P201, M7, M134, M117, 002611, P164, L127 (rs17269396), KL1 (rs17276338), 各位点的扩增引物和 PCR 引物均是本实验室自主设计的。Peakscanner 软件读取实验结果。

#### 1.2.3 数据处理

17 个 Y-STR 基因座的等位基因频率与单倍型频率用直接计数法计算, 单倍型多样性 HD(Haplotype diversity) 以及基因多样性 GD(Gene diversity) =  $n(1 - \sum P_i^2) / (n-1)$  ( $P_i$  为单倍型频率或等位基因频率,  $n$  为样本数)<sup>[12]</sup>。用 Arlequin3.5 软件<sup>[13]</sup> 计算茶洞话群体与其他群体的 *Fst* 遗传距离和平均基因多样性值, 再根据 *Fst* 值, 用 MEGA4.0 软件<sup>[14]</sup> 绘制 Neighbor-Joining 系统进化树。以文献 [10-11] 为标准将所有 Y-SNP 分型并统计各单倍群频率。

## 2 结果

### 2.1 茶洞话群体 17 个 Y-STR 基因座的遗传多态性

在茶洞话群体 21 例样本的 17 个 Y-STR 基因座上共发现了 72 个等位基因, 等位基因频率分布在 0.0476 ~ 0.8095, 最高频的是 *DYS391* 的等位基因 10; 基因多样性 GD 值为 0.3381 ~ 0.9571, 平均 GD 为 0.6438±0.1749, 多样性最高的是 *DYS385a/b*, 最低的为 *DYS391* (表 1)。共检出 20 种单倍型, 单倍型多样性 HD 值为 0.9952。

### 2.2 茶洞话群体 Y-SNP 的单倍群分型及频率分布

按照 YCC<sup>[10,11]</sup> 的分型方法, 21 例样本可分为 10 个单倍群, 其中高频单倍群有 O2\*-P31、O2a1\*-M95 和 C\*-M130, 分别占 19.05%、14.29% 和 14.29%, 其他单倍群频率均不足 10% (表 2)。

### 2.3 茶洞话群体与周边民族 Y-STR 的遗传距离及 N-J 系统进化树

选取广西境内茶洞话群体周边的壮族<sup>[15]</sup>、瑶族<sup>[15]</sup>、彝族<sup>[15]</sup>、京族<sup>[15]</sup>、仫佬族<sup>[16]</sup> 和侗族 (上海复旦大学生命科学院现代人类学教育部重点实验室未发表数据) 作为比较研究对象, 由于缺少毛南族和汉族 17 个 Y-STR 的原始实验数据, 故本文未能选取毛南族和汉族进行比较分析。考虑到 *DYS385a/b* 易发生重组, 所以只根据这 6 个民族和茶洞话群体的 15 个 Y-STR 位点的原始数据, 计算他们之间的遗传距离, 并依据遗传距离构建 N-J 系统进化树 (表 3; 图 1)。遗传距离显示: 茶洞话群体与仫佬族的距离最近; N-J 树显示: 茶洞话群体先与壮、瑶族聚类, 再与侗族、仫佬族相聚类。

### 2.4 茶洞话群体与周边民族 Y-SNP 的主成分分析

为弥补 Y-STR 研究数据的不足, 进一步探清茶洞话群体与周边民族的遗传关系, 选取已发表的 17 个群体的 Y-SNP 的频率数据<sup>[16-20]</sup> 和茶洞话群体的实验数据一起进行主成分分析, 第一主成分和第二主成分合计解释了 45.7% 的总方差。以第一主成分和第二主成分构建散点图 (图 2), 结果可见山东汉族、广东汉族和浙江汉族距离较近, 进一步证实了南北汉族父系起源的一致性, 这与文波<sup>[19]</sup> 等的研究结论相同, 但是广西的平话汉族除外, 在散点图中平话汉族与南方原住居民聚在一起, 与甘瑞静<sup>[18]</sup> 等的研究一致, 即广西平话汉族是中国南北汉族父系遗传同源性的例外; 本研究中的茶洞话群体与南方侗傣群体和部分平话汉族聚在一起, 尤其与仫佬族、壮族和侗族距离较近, 与毛南族距离较远, 但与苗瑶语系的苗族、瑶族的关系也较近。

表 1 茶洞话群体 17 个 Y-STR 的等位基因频率及基因多样性 (N=21)  
**Tab.1 Frequencies and genetic diversity of 17 Y-STR in Chadong (N=21)**

基因座 Locus	等位基因 Allele	频率 Frequency	基因多样性 GD	基因座 Locus	等位基因 Allele	频率 Frequency	基因多样性 GD		
DYS19	14	0.2381	0.5667	DYS437	14	0.7143	0.4286		
	15	0.6190			15	0.2857			
	16	0.1429		DYS439	11	0.5238	0.6048		
DYS389I	12	0.3810	12		0.3810				
	13	0.2857	13		0.0476				
	14	0.3333	14		0.0476				
DYS389II	27	0.0952	0.8333	DYS458	15	0.2857	0.8000		
	28	0.2381			16	0.0952			
	29	0.2857			17	0.2857			
	30	0.1905			18	0.2381			
	31	0.0476			19	0.0952			
	32	0.1429			DYS393	12		0.3333	0.6238
	DYS438	10				0.7143		13	
11		0.2857	14	0.5238					
DYS448		18	0.3810	0.7190	DYS390	21	0.0952	0.8143	
	19	0.2381	22			0.1429			
	20	0.3333	23			0.2381			
	21	0.0476	24			0.2857			
	DYS456	14	0.1429			0.4095	DYS391		25
15		0.7619	10	0.8095					
16		0.0952	11	0.1429					
GATA_H4	10	0.1429	0.6714	DYS385a/b	12	0.0476	0.9571		
	11	0.4762			10/16	0.0476			
	12	0.3333			11/12	0.0476			
	13	0.0476			11/18	0.1429			
DYS635	19	0.0952	0.8524	12/16	0.0952				
	20	0.1905		12/17	0.0476				
	21	0.0476		13/17	0.0476				
	22	0.1429		13/18	0.0476				
	23	0.2857		13/19	0.0952				
	24	0.1905		14/14	0.0476				
	26	0.0476		14/15	0.1429				
	DYS392	11		0.3333	0.7095	14/19	0.0476		
12		0.0952	15/15	0.0952					
13		0.4286	15/18	0.0476					
14		0.1429	18/18	0.0476					

表 2 茶洞话群体 Y 染色体单倍群频率分布 (%) (N=21)  
**Tab.2 Frequencies of Y chromosome haplogroups in Chadong(%) (N=21)**

单倍群Haplogroup	C*	C3*	D1	N*	O2*	O2a1*	O2a1a	O3a1c	O3a2c1*	O3a2c1a
SNP	M130	M217	M15	M231	P31	M95	M88	2611	M134	M117
频率Frequency (%)	14.29	9.52	9.52	4.76	19.05	14.29	4.76	4.76	9.52	9.52

表 3 茶洞话群体与周边民族 Y-STR 的 *Fst* 遗传距离矩阵Tab.3 The *Fst* genetic distance among 7 populations based on Y-STR

茶洞话群体	仫佬族	壮族	瑶族	侗族	彝族	京族
茶洞话群体	\	+	+	+	+	+
仫佬族	0.06	\	+	+	+	+
壮族	0.08	0.04	\	+	+	+
瑶族	0.17	0.11	0.08	\	+	+
侗族	0.18	0.12	0.20	0.27	\	+
彝族	0.07	0.02	0.03	0.10	0.14	\
京族	0.06	0.03	0.04	0.09	0.16	0.03

注: 对称轴上是 P 值, 下是 *Fst* 值; “+”:  $P < 0.05$ 。The data on the left bottom is *Fst* distance and that on the right upper is P value

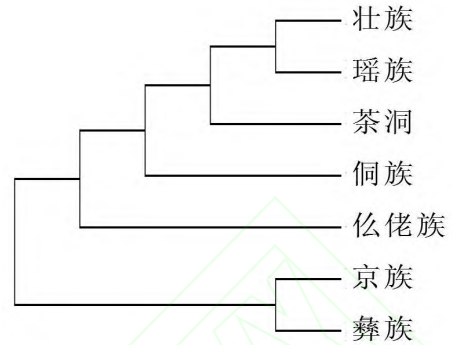


图 1 茶洞话群体与周边民族 Y-STR 的 N-J 树  
Fig.1 N-J tree of 7 populations with *Fst* genetic distances based on 15 Y-STRs

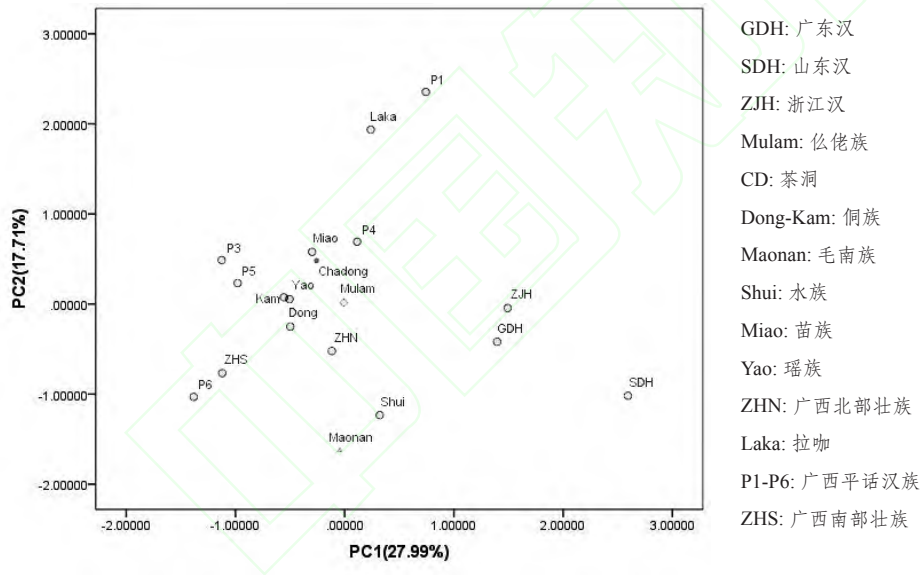


图 2 茶洞话群体与周边民族 Y-SNP 的主成分分析  
Fig 2 PCA plot based on Y-SNP haplogroups frequencies of 18 populations

## 3 讨论

### 3.1 茶洞话群体的 Y 染色体遗传结构

茶洞话群体 Y 染色体的遗传多态性结果显示: 17 个 Y-STR 基因座上共发现了 72 个等位基因, 频率分布在 0.0476~0.8095, 最高频率的等位基因是位于 *DYS391* 基因座的等位基因 10; 基因多样性 GD 在 0.3381~0.9571, 平均 GD 为 0.6438, 大于 0.5, 多样性最高的是 *DYS385a/b*。17 个 Y-STR 基因座共构成 20 种单倍型, 单倍型多样性为 0.9952, 可见这 17 个 Y-STR 在茶洞话群体中具有丰富的遗传多样性, 可适用于茶洞话群体的个体识别、父系

起源和父权鉴定等研究。

### 3.2 茶洞话群体与周边民族的遗传关系及其父系起源初探

目前,茶洞话群体的民族成分主要被定为汉族,而已有的史料和语言学等方面的研究结果却显示其可能是毛南族或仡佬族,茶洞话群体到底在父系遗传上最接近哪个民族呢?目前尚未有分子遗传学方面的文献报道,而史料和语言学方面的研究容易受到客观条件的限制,难以获得直接的证据。

近年来,随着分子生物学技术的迅猛发展,Y染色体的STR和SNP以其高度的遗传多态性、群体特异性和遗传稳定性等特点成为追踪父系起源的最理想遗传标记。迄今,中国许多群体的Y染色体单倍群已相继被界定,某些民族表现出显著的单倍群特异性,例如汉族的父系特征单倍群是O3a1c-002611、O3a2c1\*-M134和O3a2c1a-M117<sup>[21]</sup>,而O1-M119和O2-M95则是侗傣族群的高频单倍群<sup>[17]</sup>。本文所研究的茶洞话群体在民族成分上虽然多数被定为汉族,但其O3a1c-002611、O3a2c1\*-M134和O3a2c1a-M117这三种汉族特征的单倍群类型只占了23.81%,而主要单倍群为O2a1\*-M95(包含O2a1a-M88)和O2\*-P31,两者的频率达到了38.1%,表明茶洞话群体在父系遗传结构方面更接近侗傣族群,但近23.81%的O3存在也说明茶洞话群体与周边汉族有频繁的基因交流。为更确切的分析茶洞话群体的父系起源,我们利用15个Y-STR位点计算了遗传距离并绘制N-J树,结果显示茶洞话群体与侗傣族群中的壮族、侗族和仡佬族有较近的亲缘关系;此外,基于Y-SNP单倍群频率进行了主成分分析,结果表明茶洞话群体与仡佬族的遗传关系较之与毛南族和平话汉族更为亲近,而与其余各地汉族的距离较远。值得关注的是N-J树和主成分散点图均显示茶洞话群体与苗瑶语系中瑶族的关系也较近,估计历史上茶洞话群体的祖先在镇压了壮瑶起义后就留居在桂北一带,与当地的瑶族间存在过基因交流。综合以上多种分析的结果,可以得出以下结论:茶洞话群体的父系遗传结构具有典型的百越民族系统中侗傣族群的特征,与仡佬族的亲缘关系较之与毛南族和汉族更亲近。

### 参考文献

- [1] 李锦芳. 茶洞语概况. 民族语文, 2001(1): 67-80.
- [2] 临桂县志编纂委员会. 临桂县志. 北京: 方志出版社, 1996. 10: 145.
- [3] 王楠楠. 茶洞话研究 [D]. 南宁: 广西大学, 2010.
- [4] Cavalli-Sforza L, Menozzi P, Piazza A. The history and geography of human genes. Princeton: Princeton University Press, 1994, 373-374.
- [5] 肖春杰, 杜若甫, Cavalli-Sforza L, et al. 中国人群基因频率主成分分析. 中国科学, C辑, 2000, 30(4): 434-442.
- [6] 文波, 石宏, 任玲, 等. Y染色体、线粒体DNA多态性与云南宁波摩梭人的族源研究. 中国科学, C辑, 2003, 33(4): 375-384.
- [7] Jobling MA, Pandya A, Tyler-Smith C, et al. The Y chromosome in forensic analysis and paternity testing. Int J Legal Med, 1997, 110(3): 118-124.
- [8] Su B, Xiao JH, Underhill P, et al. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. Am J Hum Genet, 1999, 65(6): 1718-1724.
- [9] 王传超, 严实, 李辉. 姓氏与Y染色体. 现代人类学通讯, 2010(4): 27-34.
- [10] Karafet TM, Mendez FL, Meilerman MB, et al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res, 2008, 18(5): 830-838.
- [11] Yan S, Wang C-C, Li H, Li S-L, Jin L; Geographic Consortium. 2011. An updated tree of Y-chromosome Haplogroup O and

- revised phylogenetic positions of mutations P164 and PK4. *European Journal of Human Genetics* 19:1013-5.
- [12] Nei M. Phylogenetic analysis in molecular evolutionary genetics. *Annu Rev Genet*, 1996, 30(1):371-403.
- [13] Excoffier L, Lischer HE. Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*, 2010, 10(3): 564-567.
- [14] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*, 2007, 24(8):1596-1599.
- [15] 冯冬亮, 刘长晖, 梁祚仁, 刘超. 广西 4 个少数民族 17 个 Y-STR 基因座的多态性分析. *遗传*, 2009(9): 921-935.
- [16] 王晓庆, 王传超, 邓琼英, 李辉. 广西仫佬族 Y 染色体和 mtDNA 的遗传结构分析. *遗传*, 2013, 35 (2): 168-174.
- [17] Li H, Wen B, Chen SJ, Su B, Pramoongjago P, Liu Y, Pan S, Qin Z, Liu W, Cheng X, Yang N, Li X, Tran D, Lu D, Hsu MT, Deka R, Marzuki S, Tan CC, Jin L. Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evolutionary Biology*, 2008(8): 146.
- [18] Gan RJ, Pan SL, Mustavich LF, Qin ZD, Cai XY, Qian J, Liu CW, Peng JH, Li SL, Xu JS, Jin L, Li H\*: the Genographic Consortium. Pinghua Population as an Exception of Han Chinese's Coherent Genetic Structure. *J Hum Genet* 2008(53):303-313.
- [19] Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Deka R, Su B, Chakraborty R, Jin L. Genetic evidence supports demic diffusion of Han culture. *Nature* 2004(431):302-305.
- [20] Lu Yan, Cai Xiaoyun, Li Hui. Genetic affinity between the Hmong-Mien and Mon-Khmer populations. The 2nd International Meeting of Linguistic Evolution and Genetic Evolution, Shanghai, September 16-18, 2011. Section 5: In the flanking region of East Asia.
- [21] Shi H, Dong YL, Wen B, et al. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet*, 2005, 77(3): 408-419.

## Genetic structure of Y chromosome and paternal origin of the Population speaking Chadong in Guangxi, China

DENG Qiongying<sup>1</sup>, WANG Xiaoqing<sup>1</sup>, WANG Chuanchao<sup>2</sup>, LI Hui<sup>2</sup>

1. Department of Anatomy, Guangxi Medical University, Nanning 530021;

2. MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433

**Abstract:** There are about 20,000 individuals speaking a special dialect of Chadong living in Northeastern Guangxi. The origin of the population has been a controversial research subject. In order to study the paternal origin of Chadong people, this research aims to investigate the genetic structure of Y chromosome by studying Y-STR and Y-SNP from 21 male Chadong individuals (unrelated) in Lingui County and by comparing the data of Chadong people with those of the surrounding populations. The multiplex detection of 17 Y-STR loci revealed a highly polymorphic genetic distribution, showing powerful potential for the loci to be used for population genetics and forensic research. High frequencies of haplogroup O2\*-P31 and O2a1\*-M95 suggested that Chadong people have prominent Daic genetic background while N-J tree and principal component analysis indicated that Chadong people are closer to Mulam than to Maonan and Han peoples. These results provide genetic evidence for the origin of Chadong people.

**Keywords:** Chadong people; Y-STR; Y-SNP; genetic structure; paternal origin