

曲阜地区孔姓人群 17 个 Y-STR 基因座遗传多态性分析

侯伟光¹, 王传超², 蒋世洪³, 刘海东⁴, 李辉²

1. 济南铁路公安局刑事技术处, 济南 250000; 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 200438;
3. 济南铁路公安处刑事技术支队, 济南 250000; 4. 山东济宁市公安局刑事科学技术研究所, 济宁 272000

摘要:应用 AmpFLSTR® Y-filer™ PCR Amplification Kit 荧光标记复合扩增试剂盒 (ABI 公司), 对曲阜地区 1118 名孔姓男性个体血样 DNA 进行 PCR 扩增, 统计分析 17 个 Y-STR 基因座的遗传学参数。实验结果显示, 17 个基因座除 DYS385a/b 基因座检出 51 个单倍型外, 其余基因座上分别检出 4~11 种等位基因, 等位基因频率分布在 0.0009~0.8265 之间。由 17 个基因座组成的 YH 单倍型系统共检出 206 种单倍型。根据 Y-STR 单倍型推断了 Y-SNP 单倍群, 发现曲阜孔姓有 3 种高频单倍群: C3、Q1a1 和 O3, 前两者有着明显的单祖先扩散结构, 最可能是孔子类型。本实验通过对曲阜地区孔姓人群群体 17 个 Y 染色体短串联重复序列基因座遗传多态性的调查, 记录、保存孔姓人群遗传学数据。

关键词:群体遗传学; 曲阜地区; 孔姓人群; Y 染色体短串联重复序列

中图分类号: Q987; **文献标识码:** A; **文章编号:** 1000-3193(2016)01-0125-07

Genetic Diversity of Seventeen Y-STR Loci among the People with the Surname Kong from Qufu Prefecture

HOU Weiguang¹, WANG Chuanchao², JIANG Shihong², LIU Haidong³, LI Hui²

1. Criminal Technology Division of Jinan railway public security bureau, Jinan 250000;
2. MOE Key Laboratory of Contemporary Anthropology School of Life Sciences, Fudan University, Shanghai 200438;
3. Criminal technology detachment of Jinan railway public security department, Jinan 250000;
4. Jining Institute of Criminal Science and Technology in Shandong province, Jining, 272000

Abstract: Blood samples were collected from 1118 male individuals of Kong clans in the Qufu Prefecture of Shandong Province. Seventeen Y- chromosomal short tandem repeats (STR) loci were amplified in one tube by using the AmpFLSTR® Y-filer™ PCR Amplification Kit for each person, and then the frequency of alleles and genotypes of group was obtained. The results show

收稿日期: 2014-06-03; 定稿日期: 2014-10-20

基金项目: 国家自然科学基金优秀青年基金 (31222030); 教育部科学技术研究项目 (113022A)

作者简介: 侯伟光 (1981-), 男, 山东青岛人, 硕士研究生, 主要从事法医人类学及法医物证学研究。E-mail: gooddoctorhou@sohu.com

通讯作者: 李辉, 男, 教授, E-mail: lhca@fudan.edu.cn

Citation: Hou WG, Wang CC, Jiang SH, et al. Genetic diversity of seventeen Y-STR loci among the people with the surname Kong from Qufu Prefecture[J]. Acta Anthropologica Sinica, 2016, 35(1): 125-131

that, except DYS385a/b locus was detected 51 haplotypes, the rest loci were detected from 4 to 11 alleles, the allele frequency distribute from 0.0009 to 0.8265. Totally, 206 haplotypes were detected in the YH haplotype system consisting of 17 loci. According to the Y-STR haplotypes, we inferred the Y-SNP haplogroups. We found three high frequency haplogroups among the Kong clans, i.e., C3, Q1a1, and O3. Both C3 and Q1a1 exhibited a mono-ancestor diffusion structure, indicating that either of them may be the haplogroup of Confucius. By investigating the 17 Y-STR loci among the Kong clans from Qufu Prefecture, we recorded the basic date of Kong-population.

Key words: Population Genetics; Kong surname; Qufu Prefecture; Y chromosomal short tandem repeat ;

中华民族历史悠久的姓氏与具有父系遗传特点的 Y 染色体存在同源性^[1], 使 Y 染色体成为人类群体遗传学研究姓氏的一种理想模型^[2]。通过对 Y 染色体在姓氏人群的分布情况的研究, 对进一步探讨分析姓氏的起源和分化以及追踪各亚型姓氏群体的父系祖宗将提供较为可靠的和有价值的线索^[3]。曹操家族的 Y 染色体分析就是一个成功的例子^[4]。

孔子是儒家思想的创始人, 在我国被尊为圣人, 历代孔家都有社会和经济方面的特殊待遇, 孔氏谱系完善, 历史源远流长, 主要源流于历史上的姜姓、姁姓、子姓、姬姓、妫姓、蒙古族、满族及其他少数民族。一般认为, 圣人孔子出自子姓孔氏, 这支孔氏成为当代孔姓最重要的组成部分^[5]。当代孔姓人口有 270 多万, 为全国第 83 位姓氏。孔姓目前主要集中于山东省, 约占全国孔姓总人口的 45%; 山东曲阜市位于山东省西南部, 是中国古代伟大的思想家、教育家、儒家学派创始人孔子的故乡, 亦是孔姓后人繁衍栖息地。本地区孔姓人群具有一定代表性。

本研究旨在统计调研山东曲阜地区孔姓人群 17 个 Y-STR 基因座遗传多态性, 通过对基因频率的调查, 为曲阜地区孔姓人群提供基础性遗传学数据, 试探性发掘本地区孔姓人群“大众化”的位点等位基因或“代表性”表染色体单倍型。目前, 研究探讨孔姓历史渊源主要依靠古籍文献记载, 尚无对孔姓群体基因位点多态性的报道, 缺乏分子遗传学数据研究, 本研究意义性明显。随着社会、经济、文化、交通的日益发展, 姓氏融合, 基因交流亦日益频繁, 孔姓人群整体性基因特征也面临严峻挑战, 本研究及时的保存了曲阜地区孔姓人群群体性遗传数据。

1 材料与amp;方法

1.1 实验材料

样本 1118 名男性个体 FTA 卡血样样本为办案积累, 均采自山东省曲阜地区。

主要仪器设备: GeneAmp PCR System 9700 (ABI 公司); 3500xl Genetic Analyzer (ABI 公司)。

主要试剂：AmpFLSTR® Y-filer™ PCR Amplification Kit 试剂盒；POP-4 液态胶。

1.2 方法

采用 AmpFLSTR® Y-filer™ PCR Amplification Kit 荧光标记复合扩增试剂盒在 GeneAmp 9700 型扩增仪上对样本进行直接扩增，扩增体系为 10μL，其中含 PCR Reaction Mix 5.52μL，引物 3μL，AmpliTaqGold™ DNA (5U/μL) 0.48μL，模板 DNA 1ng。PCR 扩增条件：95℃ 11min，94℃ 1min，61℃ 1min，72℃ 1min，30 个循环，最后 60℃ 80min，4℃ 保存。PCR 产物在 ABI-3500xl 全自动遗传分析仪上进行毛细管电泳。使用 GeneMapper 软件通过与试剂盒里的等位基因分型标准物 (Ladder) 比较对电泳数据进行基因分型分析。

1.3 数据处理与统计分析

所得基因分型数据用直接计数法计算各基因座的等位基因与单倍型检出频率，基因多样性 (GD) 和单倍型多样性按公式 $GD = n(1 - \sum Pi^2) / (n-1)$ (Pi 为等位基因或单倍型频率，n 为样本数) 计算^[6]。

根据 Y-STR 单倍型与 Y-SNP 单倍群的对应规律^[7]，我们估算了孔姓 Y-SNP 单倍群的分布，并且用 Network 软件绘制了主要单倍群的网络结构^[8]，以分析单倍群的集中趋势。

2 结果

对曲阜地区孔姓人群 1118 例个体血样进行检验，17 个 Y-STR 基因座等位基因多态性调查结果见表 1。除 DYS385a/b 基因座检出 51 个单倍型外，其余基因座上分别检出 4~11 种等位基因，等位基因频率分布在 0.0009~0.8265 之间。经 Excell 去重功能计算，由 16 个基因座组成的 YH 单倍型系统共检出 206 种单倍型。

根据 Y-STR 单倍型与 Y-SNP 单倍群的对应关系，我们总结单倍群频率如表 2。

因为历朝历代的特殊的优待，孔子后裔在曲阜孔姓中应该占多数。曲阜孔姓出现三

表 1 曲阜地区孔姓人群 17 个 Y-STR 基因座基因频率分布及基因多样性 (n=1118)

Tab.1 Allele frequency and diversities of 17 Y-STR loci of Kong-population in the area of Qufu (n=1118)

Loci	A	n	AF	Loci	A	n	AF
DYS385a/b	10/ 12	1	0.0009		11/ 20	1	0.0009
GD=0.9664	10/ 15	1	0.0009		12	8	0.0072
	10/ 19	1	0.0009		12/ 13	5	0.0045
	11	205	0.1834		12/ 14	3	0.0027
	11/ 12	46	0.0411		12/ 15	58	0.0519
	11/ 13	2	0.0018		12/ 16	39	0.0349
	11/ 15	8	0.0072		12/ 17	18	0.0161
	11/ 16	4	0.0036		12/ 18	16	0.0143
	11/ 17	256	0.2290		12/ 19	13	0.0116
	11/ 18	20	0.0179		12/ 20	8	0.0072
	11/ 19	2	0.0018		13	12	0.0107

续表 1 曲阜地区孔姓人群 17 个 Y-STR 基因座基因频率分布及基因多样性 ($n=1118$)
Tab.1 Allele frequency and diversities of 17 Y-STR loci of Kong-population in the area of Qufu ($n=1118$)

Loci	A	n	AF	Loci	A	n	AF
	13/ 14	2	0.0018		14,	2	0.0018
	13/ 15	1	0.0009	DYS389II	27,	14	0.0125
	13/16	1	0.0009	GD= 0.6040	28,	91	0.0814
	13/ 17	4	0.0036		29,	609	0.5447
	13/ 18	18	0.0161		30,	333	0.2979
	13/ 19	13	0.0116		31,	68	0.0608
	13/20	3	0.0027		32,	3	0.0027
	13/ 21	11	0.0098	DYS438	8,	3	0.0027
	13/ 22	1	0.0009	GD= 0.4880	9,	1	0.0009
	13/ 24	2	0.0018		10,	735	0.6574
	13/ 25	1	0.0009		11,	69	0.0617
	14/ 16	2	0.0018		12,	308	0.2755
	14/ 17	5	0.0045		13,	2	0.0018
	14/ 18	12	0.0107	DYS456	13,	1	0.0009
	14/ 19	4	0.0036	GD= 0.5542	14,	29	0.0259
	14/ 20	2	0.0018		15,	497	0.4445
	14/ 21	1	0.0009		16,	555	0.4964
	15	4	0.0036		17,	35	0.0313
	15/ 19	13	0.0116		18,	1	0.0009
	15/ 20	41	0.0367	DYS458	14,	6	0.0054
	15/ 21	202	0.1807	GD= 0.4614	14.1,	1	0.0009
	15/ 22	24	0.0215		15,	38	0.0340
	15/ 23	3	0.0027		16,	800	0.7156
	16/ 19	2	0.0018		17,	130	0.1163
	16/ 20	1	0.0009		18,	120	0.1073
	16/ 21	14	0.0125		19,	19	0.0170
	19/ 20	1	0.0009		20,	3	0.0027
	9/ 11	1	0.0009		21,	1	0.0009
	9/ 16	2	0.0018	DYS437	13,	3	0.0027
DYS391	8,	6	0.0054	GD= 0.2911	14,	924	0.8265
GD=0.5404	9,	308	0.2755		15,	179	0.1601
	10,	684	0.6118		16,	12	0.0107
	11,	107	0.0957	DYS635	17,	2	0.0018
	12,	12	0.0107	GD= 0.6114	19,	44	0.0394
	13,	1	0.0009		20,	58	0.0519
DYS389I	11,	2	0.0018		21,	617	0.5519
GD= 0.6124	12,	202	0.1807		22,	72	0.0644
	13,	584	0.5224		23,	307	0.2746
	14,	320	0.2862		24,	17	0.0152
	15,	10	0.0089		25,	1	0.0009
DYS439	10,	6	0.0054	DYS448	16,	1	0.0009
GD= 0.5674	11,	616	0.5510	GD= 0.5775	17,	2	0.0018
	12,	387	0.3462		18,	33	0.0295
	13,	107	0.0957		19,	475	0.4249

续表 1 曲阜地区孔姓人群 17 个 Y-STR 基因座基因频率分布及基因多样性 ($n=1118$)Tab.1 Allele frequency and diversities of 17 Y-STR loci of Kong-population in the area of Qufu ($n=1118$)

Loci	A	n	AF	Loci	A	n	AF
	20,	546	0.4884		16,	513	0.4589
	21,	56	0.0501		17,	23	0.0206
	22,	4	0.0036		18,	4	0.0036
	23,	1	0.0009		19,	1	0.0009
Y_GATA_H4	9,	1	0.0009	DYS392	11,	530	0.4741
GD= 0.6182	10,	303	0.2710	GD= 0.6472	12,	120	0.1073
	11,	589	0.5268		13,	79	0.0707
	12,	194	0.1735		14,	373	0.3336
	13,	28	0.0250		15,	14	0.0125
	14,	3	0.0027		16,	2	0.0018
DYS447	20,	1	0.0009	DYS393	12,	219	0.1959
GD= 0.7348	21,	3	0.0027	GD= 0.6623	13,	66	0.0590
	22,	2	0.0018		14,	526	0.4705
	23,	58	0.0519		15,	305	0.2728
	24,	125	0.1118		16,	2	0.0018
	25,	173	0.1547	DYS390	20,	1	0.0009
	26,	429	0.3837	GD=0.4853	21,	1	0.0009
	27,	315	0.2818		22,	9	0.0081
	28,	11	0.0098		23,	729	0.6521
	29,	1	0.0009		24,	332	0.2970
DYS19	11,	1	0.0009		25,	39	0.0349
GD= 0.6730	13,	315	0.2818		26,	7	0.0063
	14,	54	0.0483				
	15,	207	0.1852				

GD: gene diversity; A: Allele ; AF: Allele frequency; n:number

表 2 孔姓 Y 染色体单倍群频率 (%)

Tab.2 Frequencies of the Y chromosome haplogroups of Surname Kong in Qufu

Haplogroup	C3	Q1a1	O3	O1	R	N	O2	C3c	G	J	residual
Frequency (%)	46.06	27.01	20.66	1.25	1.16	0.98	0.89	0.54	0.09	0.09	1.34

种高频单倍群，C3、Q1a1、O3。其中，O3 是汉族最常见的类型，在孔家出现频率远低于汉族中的频率 (约 60%-70%)，所以没有集中趋势。C3 和 Q1a1 都可能是孔子后裔的类型。我们进一步用 Network 分析了这两种类型的集中趋势 (图 1)。

两种单倍型都有着明显的单中心扩散结构，有着集中趋势，说明分别是从一个较近的祖先繁衍下来的。

3 讨论

孙海明^[9]等人阐述了 Y-STR 多态性在人类学中的应用后指出，利用 Y-STR 可以进行人群结构分析和谱系研究。人群的迁移和融合事件会在其遗传结构上会有所体现，迁

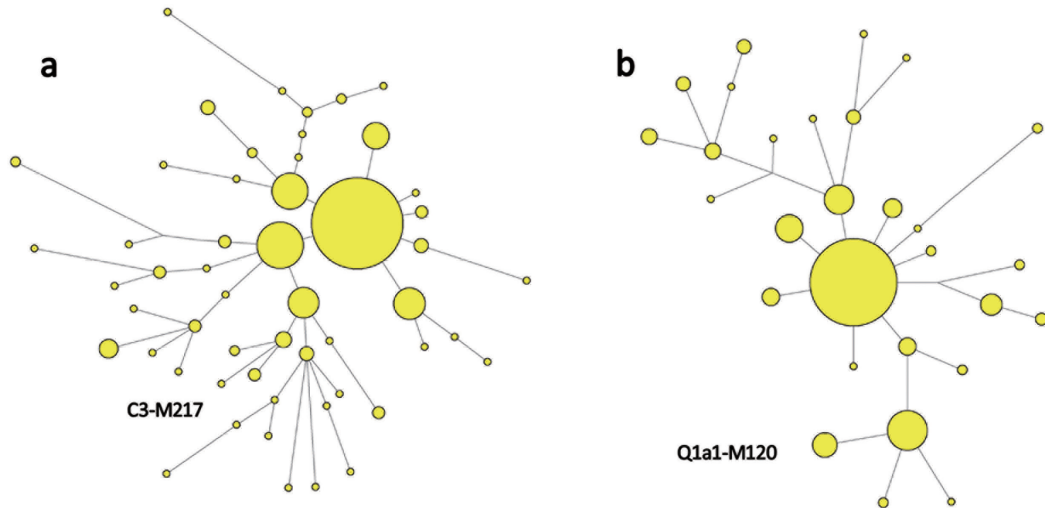


图 1 曲阜孔姓人群中两类高频单倍群的网络结构分析

Fig.1 Networks of the two high frequency Y haplogroups among the people with Kong surname from Qufu

移事件会在这些群体的遗传结构上留下相应的遗传印记^[10]。曲阜尼山是孔子出身地，曲阜地区是孔家后人最主要的生活繁衍栖息地，该地区孔姓人群具有孔姓人群的代表性。通过与其他地区孔姓人群通过与其进行 Y-STR 数据频率的对比，结合历史文献，可以推断佐证孔氏人群迁移演变过程。

本次调查 1118 例样本中，16 个基因座组成的 YH 单倍型系统共检出 206 种单倍型，重复次数最多的单倍型重复次数为 63 次，在调查人群中比例低，不具有代表性，该单倍型不能称其为代表“孔子”单倍型。但是，在单倍型重复性分析中，笔者未考虑同一父系人群 Y-STR 突变性对重复次数计算的影响，而通过表 1 各基因位点频率观察发现，某些位点的部分等位基因频率 < 0.001 ，这些等位基因不能称其为位点的多态性，只能考虑其为突变现象。如果考虑到同一家系某一单倍型某一位点的一步或二部突变，该单倍型在该人群中的重复次数将会明显增加，有可能会找到代表孔姓人群群体的典型单倍型。将表 1 中每一位点出现频率最高的等位基因组成一个新的单倍型，该单倍型在总调查人群中出现的重复次数仅为 9 次，没有明显意义性。

17 个 Y-STR 基因座组成的 YH 单倍型系统中，位点 DYS391 的等位基因 10、DYS390 的等位基因 23、DYS438 的等位基因 10、DYS458 的等位基因 16、DYS437 的等位基因 14 具有较高的等位基因频率，其频率值分别为 0.6118、0.6521、0.6574、0.7156、0.8265。参阅文献报道，DYS391 的等位基因 10、DYS390 的等位基因 23、DYS458 的等位基因 16、DYS437 的等位基因 14 在我国其他地区人群中也具有较高的频率，缺乏特异性，不能称为曲阜地区孔姓人群的“大众化”等位基因；DYS458 的等位基因 16 曲阜地区孔姓人群中具有较高频率，其他群体中频率略低，对于研究曲阜地区孔姓人群代表性等位基因具有一定意义性。

通过估算 Y-SNP 单倍群，发现曲阜孔姓有两种集中的类型，而且都是单祖先的类型，

所以这两种类型都可能是孔子类型。但是同时也存在大量其他类型，说明曲阜地区孔姓群体存在一定程度的基因融合与突变，造成了孔姓群体单倍群的多态性。笔者认为，由于历史的变迁，赐姓、改性现象的存在，某一地区某一姓氏群体存在相对较为单一的单倍型的可能性较小，某些相对封闭地区的某些群体，如某些稀有姓氏有可能存在单倍型相对单一的现象，这些群体所携带的 Y 染色体家系特征比较明显。笔者认为，即便经过历史上的变迁、基因的融合，某一姓氏群体基因座基因频率发生了改变，产生了一个新的人群群体，但是，新人群群体与旧人群群体一定会在基因频率上存在一定的相似性。因此，可以通过计算不同姓氏群体的基因频率，比较不同姓氏间的遗传距离，构建系统进化树，结合历史事件研究姓氏渊源。

由于条件限制，本研究尚存在许多不足和遗憾，孔姓历史上渊源颇丰，孔姓历史上有内孔与外孔之分，外孔原为刘姓，后改为孔姓，但本次调查取样时区分内孔姓氏与外孔姓氏存在一定困难。另外，我国其他地区亦有孔姓人群聚集区，历史文献、民间流传其与曲阜地区孔姓群体存在一定渊源，如果能够采集这些地区孔姓人群群体血样，对不同地区孔姓人群做群体遗传学分析，从而可以验证孔姓人群姓氏历史渊源。还有，距曲阜 20 余里的邹城，是孟子的故里，亦可将邹城地区孟姓人群纳入研究范围。

参考文献

- [1] 王传超, 严实, 李辉. 姓氏与 Y 染色体 [J]. 现代人类学通讯, 2010, 4: e5: 27-34
- [2] Wang CC, Li H. Inferring human history in East Asia from Y chromosomes[J]. *Investig. Genet.*, 2013, 4(1):11
- [3] Graf OM, Zlojutro M, Rubicz R, et al. Surname distributions and their association with Y-chromosome markers in the Aleutian Islands[J]. *Hum. Biol.*, 2010, 82(5-6): 745-757
- [4] Wang CC, Yan S, Hou Z, Fu W, Xiong M, Han S, Jin L, Li H. Present Y chromosomes reveal the ancestry of Emperor Cao Cao of 1,800 years ago[J]. *J. Hum. Genet.*, 2012, 57: 216-218
- [5] 张世国. 百家姓氏溯源 [M]. 北京: 中国社会科学出版社, 2001: 1-54
- [6] Hou YP, Zhang J, Li YB, et al. Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population[J]. *Forensic Sci. Int.*, 2001, 118:147-152
- [7] Wang CC, Li H. Comparison of Y-chromosomal lineage dating using either evolutionary or genealogical Y-STR mutation rates[J]. *bioRxiv*, 2014, doi: <http://dx.doi.org/10.1101/004705>
- [8] Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies[J]. *Mol. Biol. Evol.*, 1999, 16:37-48
- [9] 孙海明, 傅松滨. Y 染色体 STR 位点及其在人类学研究中的应用 [J]. *国际遗传学杂志*, 2006, 29(1): 50-53
- [10] Alonso LA, Usaquén W, et al. Y-chromosome and surname analysis of the native islanders of San Andres and Providencia[J]. *Homo: Internationale Zeitschrift für die Vergleichende Forschung am Menschen*, 2013, 64(1): 71-84