



REPORT

Y chromosome clues to the ancestry of Emperor Cao Cao

Chuanhao Wang¹, Shi Yan², Zheng Hou¹, Wenqing Fu¹, Momiao Xiong¹, Sheng Han¹, Li Jin^{1,2}, Hui Li¹

1.MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Department of history, Fudan University, Shanghai 200433, China; 2.CAS-MPG Partner Institute for Computational Biology, SIBS, CAS, Shanghai 200031, China

Email: LiHui.Fudan@gmail.com

Received: Nov. 22th, 2011; revised Dec. 4th, 2011; accepted Dec. 8th, 2011.

Abstract: Deep pedigrees are of great value for studying the Y chromosome evolution. However, the authenticity of the pedigree information requires careful validation. Here, we validated some deep pedigrees in China with full records of 70-100 generations spanning over 1,800 years by comparing their Y chromosomes. The present clans of these pedigrees claim to be descendants of Emperor CAO Cao (155AD–220AD). Haplotype O2-M268 is the only one that is enriched significantly in the claimed clans ($P=9.323 \times 10^{-5}$, $OR=12.72$), and therefore, is most likely to be that of the Emperor. Moreover, our analysis showed that the Y chromosome haplotype of the Emperor is different from that of his claimed ancestry of the earlier CAO aristocrats (Haplotype O3-002611). This study offers a successful showcase of the utility of genetics in studying the ancient history.

Keywords: Y chromosome, deepest pedigree, Chinese Emperor, human ancestry, stemma record

现代人类学通讯 2011年 第五卷 第107-111页 研究报告

Y 染色体揭开曹操身世之谜

王传超¹, 严实², 侯铮¹, 傅雯卿¹, 熊墨森¹, 韩昇¹, 金力^{1,2}, 李辉¹

1.复旦大学生命科学学院及历史系现代人类学教育部重点实验室, 上海 200433; 2.中科院上海生命科学研究院计算生物学研究所, 上海 200031

摘要: 大跨度家系对于研究 Y 染色体进化有着极其重要的意义, 但家系的可信度却需仔细甄别。本文中, 我们用 Y 染色体分型比对的方法确认了若干有 1800 多年历史、延续 70-100 代的大跨度家系, 这些家系宣称是魏武帝曹操后裔。单倍型 O2-M268 是唯一在宣称是曹操后裔的众多家族里频率显著升高的单倍型 ($P=9.323 \times 10^{-5}$, $OR=12.72$), 因此也极可能是曹操的 Y 染色体单倍型。我们的分析结果还显示曹操的 Y 染色体单倍型与其自称的先祖曹参的单倍型 O3-002611 并不一致。本研究是 Y 染色体和谱牒分析相结合的成功探索, 为遗传学用于历史学研究提供一个范例。

关键词: Y 染色体; 最大跨度家系; 魏武帝; 血统; 家谱

Y 染色体上的绝大部分是从父遗传且缺乏重组, 所以可以通过研究历史人物现存后代的 Y 染色体来揭示历史人物之间的父系关系[1-3]。近年来, 国际上的成功例子有美国第三任总统托马斯·杰斐逊的私生子的确认[4]、犹太教祭司的 Y 染色体单倍型的推定[5]等。实际上, Y 染色体推测历史人物的深度还可以往更古老的年代推进。通过可靠的家

谱信息可以重建跨度极大的家系, 这样的家系可用来推断其遥远祖先的 Y 染色体单倍型, 为解决历史悬案提供新途径。这样的大跨度家系对研究 Y 染色体的突变率及其进化机制也都有重要意义[6]。修著家谱是中国人的传统, 一些家谱甚至跨越 3000 年将现代人与其祖先相连, 虽然它们的真实性还有待确认。我们通过研究宣称是曹操后代的现代家

族的男性 Y 染色体, 来揭示曹操的士族出身是否属实[7], 为探索将遗传学、谱牒资料等用于进化和历史学研究提供范例。

魏武帝曹操(公元 155-220)因小说《三国演义》的风靡而成为东亚最有名的历史人物之一。陈寿的《三国志》记载曹操是西汉第二任相国曹参(公元? -190)的后代[8], 而曹操更自称其祖源可远溯至古曹国(公元前 11 世纪-公元前 487)的曹叔振铎[9], 即曹操宣称自己是皇室贵族后裔。是否具有显赫的士族出身对于曹操争夺政治利益有很重要的意义。然而曹操的祖父曹腾在东汉为宦官之首, 曹操的父亲曹嵩是曹腾的养子, 曹操的政敌袁绍在攻曹的檄文中写到“父嵩乞丐携养”[10], 说他是路边捡来的乞丐。关于曹操身世的各种说法由此流传, 且纷纷扰扰争论了近 2000 年。

曹操家族的起源争议颇大, 现存家谱却是了解曹操后裔分布情况的一条重要线索。家谱的先世传说并不十分可靠。曹氏家谱的起点, 我们选择陈寿以来言之有据的传承, 那就是始于曹参的后裔宗族, 而不是再往前的西周乃至夏朝先祖。我们课题组对上海图书馆收藏的 118 件曹氏族谱进行了全面查阅和筛选, 将家谱所载世系同历史记载相比较, 挑选出比较可靠的家谱, 其中包含自称是曹操后裔的家谱。把曹氏取样的面铺得广一些, 多取一些样本, 能够更加全面地反映出自古以来多支曹氏家族的基因状况, 获得整体的把握。家谱调查反映出, 与曹参或者曹操有关系的曹氏, 大量分布于长江流域, 与史料所见五胡十六国以后曹操家族迁居于江南的情况是吻合的。据此, 我们大致确定了几处最接近曹操后裔的居住地, 一是位于安徽、山东、河南、江苏四省的交界处, 以安徽亳州为中心, 这里是曹操政权的发源地, 也是安阳汉墓被发现的地区; 二是江东地区, 包括安徽泾县、繁昌、歙县、绩溪、浙江的金华、东阳、绍兴、余姚、萧山和江西赣州等地; 三是湖南等省沿江地区, 如湖南兴化、郴州、益阳、长沙等地。

我们用 Y 染色体上的 100 个单核苷酸位点(SNP)对全国各地曹姓 79 个家族的 280 个男性以及其他姓氏的 446 个男性进行分型。

我们所选取的 100 个 SNP 位点涵盖了最新 Y 染色体谱系树上的 O、N、C、D、Q、R、J 等东亚地区可见的所有单倍群[11]。如果同一家族的个体间 Y 染色体单倍型不同, 该家族则是由一些不同来源的单纯家族组成的复合家族, 因此实际上我们总共研究了 111 个曹姓单纯家族(图 1a)。我们也在家族成员知情同意后将大部分家谱资料影印保存。根据家谱资料, 15 个曹姓家族宣称是曹操后代, 他们中共有 6 种 Y 染色体单倍型(图 1a), 其中仅有一种单倍型是曹操的, 而其它 5 种可能是由于收养、从母姓、非亲生等原因而引入的[12]。我们需要通过分析不同单倍型在家族间的分布来确认哪个才最可能是曹操的 Y 染色体单倍型。

我们把这些曹姓家族分成三类: 宣称是曹操后代的、未宣称是曹操后代(或宣称是其他祖源的)、其他姓氏的普通对照人群。首先, 我们两两比较了每种单倍型在这三类家族间的频率差异, 在未宣称为曹操后代的家系与普通对照人群间并未发现任何单倍型有显著差异(表 1), 这表明曹姓与其他姓氏一样也是多起源的[13], 所以我们在之后的分析中把未宣称的家族和普通人群合并在一起作为对照组。非常有趣的是, O2-M268 是唯一在宣称是曹操后代的家族中显著高频出现的单倍型(Fisher 精确检验, $P=9.323 \times 10^{-5}$), 很有可能这就是曹操的 Y 染色体单倍型。某种单倍型是否是曹操的单倍型不能仅从其在宣称的家族里频率来判断, 还要考虑其在宣称的家族与对照人群的频率差异。这里我们用 OR 值去估算 O2-M268 是曹操单倍型的可能性, OR 值可判断两组变量间关系(带自信区间)。O2-M268 在宣称为曹操后代的家族和对照人群间的 OR 值为 12.72(表 1), 所以其为曹操单倍型的可能性为 92.71% ($=12.72/(12.72+1) \times 100\%$)。还有一点需注意, 有 8 个曹姓家族坚决认为不是曹操后代, 这 8 个家族的 Y 染色体单倍型也都不是 O2-M268, 这也增大了 O2-M268 为曹操单倍型的可能性。

另外, 我们也估算了反概率, 即假定所有的曹操家谱都是伪造的, 那么出现 6 个家族都是相同的 O2-M268 单倍型的巧合概率。

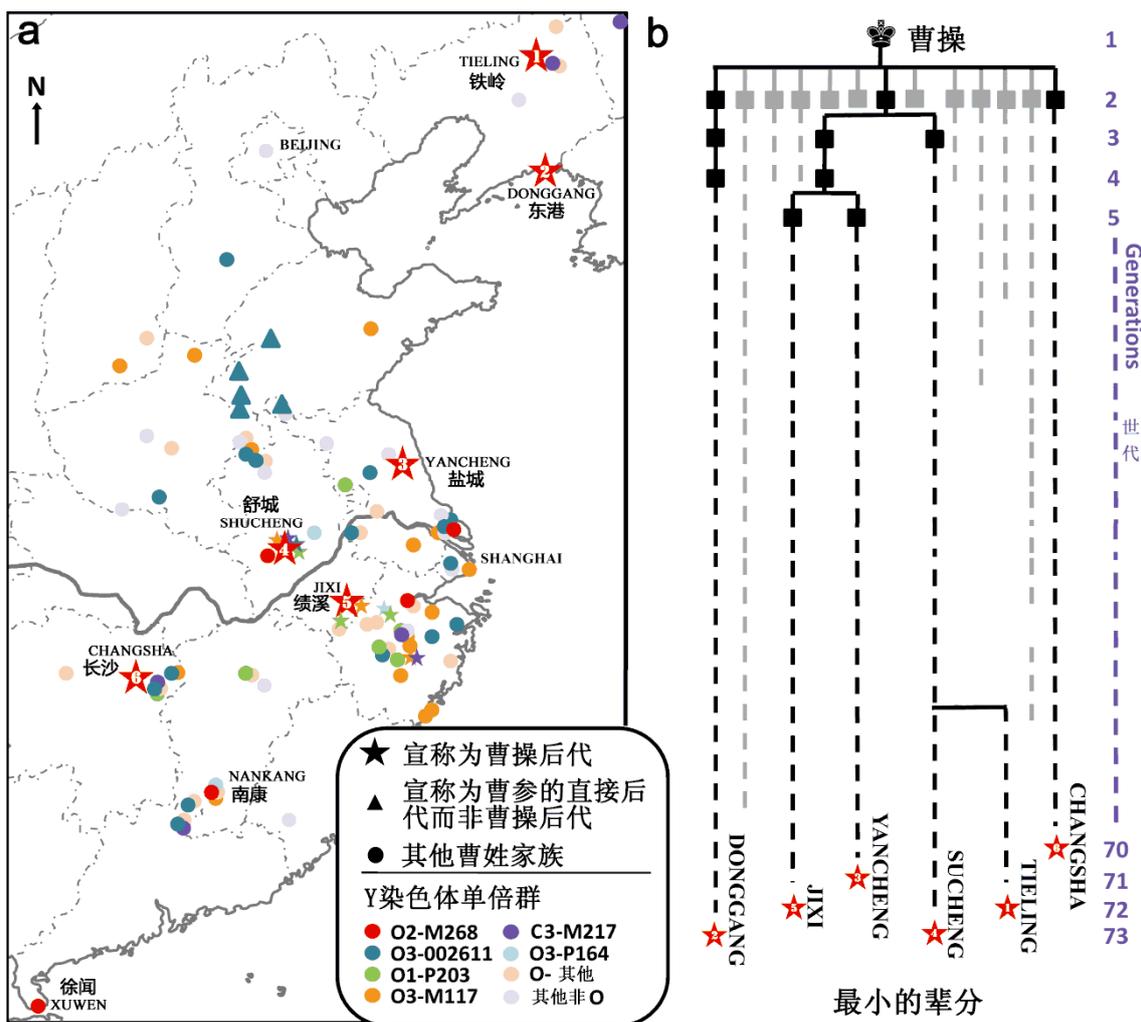


图 1 曹姓家族的 Y 染色体单倍型分布(a)和已确认的大跨度家系(b)

Fig.1. Geographical focuses and Y chromosome genotypes of the Cao clans (a) and genealogical relationship among the six related clans (b)

表 1 不同类别的曹姓家族间 Fisher 精确检验

Tab. 1. Fisher's exact tests between different clan groups

单倍型	家族的数目			两两比较的 P 值		宣称的家族和对照组间的 OR 值 [95% CI]
	宣称家族	未宣称家族	其他姓氏	未宣称家族 vs. 普通人群 ^a	宣称家族 vs. 对照组 ^b	
O2-M268	6	5	22	0.801	9.32×10⁻⁵*	12.72 [4.22-38.32]
O3-002611	1	21	79	0.384	0.952	0.32 [0.04-2.43]
O1-P203	2	6	65	0.030	0.607	1.02 [0.23-4.62]
O3-M117	3	15	67	0.876	0.408	1.40 [0.39-5.08]
C3-M217	2	5	25	1.000	0.211	2.63 [0.57-12.17]
O3-P164	1	2	13	1.000	0.358	2.51 [0.31-20.34]
Others	0	42	175	0.423	1.000	0

未宣称曹操后裔的家族和普通人群合并为对照组。*, $P < 0.01$; ^a, 双侧检验; ^b, 右侧检验。

由于这些家族在确信的历史中都没有联络，所以他们如果来源不同而基因类型相同就只能是由巧合造成的。我们知道，两个事件同时发生的巧合概率是两个事件单独发生的概率的乘积。那么我们可以这样估计曹操家谱相关家族的基因巧合概率：出现两个家族基因相同的事件计为一次巧合事件，概率就是 O2-M268 在全国汉族人中的频率约 5% [14]；出现第三个相同基因的家族则是两次巧合事件，概率为 5% 乘以 5%；那么 6 个分布在不同地域的曹姓家族同时是 O2-M268 的概率是 $(5\%)^5 = 3.125 \times 10^{-7}$ ，大约为千万分之三。也就是几乎不可能。

单倍型 O3-002611 是在其他曹姓家族中出现频率最高，也是唯一在宣称是曹参直系后裔的 5 个曹姓家族中都唯一出现的单倍型 (图 1a)。我们继续使用推断曹操 Y 染色体单倍型的方法，对不同曹姓家族进行分类去推定曹参最可能的单倍型：5 个自称是曹参后裔的 O3-002611 家族作为宣称组，79 个未宣称的 O3-002611 家族和 458 个其他单倍型的家族作为对照组。经分析，O3-002611 就是曹参最可能的单倍型 (Fisher 精确检验， $P = 1.968 \times 10^{-4}$)，可能性近于 100% (OR 值很高)。综合考虑我们的分析结果，我们认为魏武帝曹操不太可能是曹参的后裔，遗传学证据并不支持曹操自称的土族出身。

结合数据资料，我们的结果也确认了 O2-M268 的曹操后裔和 O3-002611 的曹参后裔的谱牒资料的可靠性 (图 1b)，这些谱牒将现代曹姓族人上溯 70-100 代与其 2000 多年前的祖先相连，这对于 Y 染色体的进化研究有重要意义。此次曹姓家族 Y 染色体调查是将遗传学用于古代史研究的一个成功范例，且提供了两个重要的契机，一是促成了历史学和分子生物学的深层次合作研究，建立了国内第一个以分子生物学为主要研究工具的历史人类学新学科；二是加快了人类基因调查从以民族向以家族为对象的转变。

材料与方法

样本

本项目由复旦大学生命科学学院伦理委员会审核通过，根据知情同意的原则，在全

国各地采集了 79 个曹姓家族的 280 个男性和其他姓氏的 446 个男性外周血样本，大部分谱牒资料也在家族成员知情同意后影印保存。捷瑞 GK1042 型试剂盒 (上海捷瑞® 生物工程公司) 提取 DNA。

遗传标记

本项目所选取的 100 个 SNP 位点如下：版 1 (单倍群 O 内)，M175, M119, P203, M110, M268, P31, M95, M176, M122, M324, M121, P201, M7, M134, M117, 002611, P164, L127 (rs17269396), KL1 (rs17276338)；版 2 (单倍群 O 外)，M130, P256, M1, M231, M168, M174, M45, M89, M272, M258, M242, M207, M9, M96, P125, M304, M201, M306；版 3 (单倍群 C)，P54, M105, M48, M208, M407, P33, M93, P39, P92, P53.1, M217, M38, M210, M356, P55, M347；版 4 (单倍群 D)，P47, N1, P99, M15, M125, M55, M64.2, M116.1, M151, N2, 022457；版 5 (单倍群 N)，M214, LLY22g, M128, M46/Tat, P63, P119, P105, P43, M178；版 6 (单倍群 R)，M306, M173, M124, M420, SRY10831.2, M17, M64.1, M198, M343, V88, M458, M73, M434, P312, M269, U106/M405；版 7 (单倍群 Q)，P36.2, M3, M120, MEH2, M378, N14/M265, M25, M143, M346, L53, M323。

这些位点涵盖了最新 Y 染色体谱系树上的东亚所有单倍群 [11]。基因分型采用 SNaPshot (ABI SNaPshot® 多重试剂盒) 和荧光引物 PCR 相结合的方法 [15]，PCR 产物纯化后在 ABI 3730 测序仪 (Applied Biosystems, Carlsbad, CA) 上分析。

统计分析

为确认曹操最可能的 Y 染色体单倍型，我们使用 Fisher 精确检验来分析每种单倍型在不同分类的家族中的频率差异 [16-17]，双侧显著性检验来分析分类家族间的差异，右侧检验来分析每种单倍型在宣称是曹操后裔家族中的是否显著高频出现。我们还计算了每种单倍型在宣称是曹操后裔的家族和对照家族间的 OR 值。OR 值可用来描述两组变量间的相关性，常被用在流行病学的病例-对照研究中 [18]。本项目中，OR 值是用来描述 Y 染色体单倍型和宣称的祖源之间的关系，用

N_c^H 和 N_r^H 分别代表在宣称的家族和对照家族中是某一单倍型 H 的家族的数目, T_c 和 T_r 则分别代表宣称的家族和对照家族的所有家族, 那么单倍型 H 的 OR

$$\text{值为 } OR_H = \frac{N_c^H \times (T_c - N_c^H)}{N_r^H \times (T_r - N_r^H)}.$$

致谢

感谢康栋东、李淑元、胡抗等采样员, 曹卫东、曹友平、曹祖义、曹津玮、曹典华、曹建和等家族召集人。本项目得到了复旦大学文科科研推进计划、上海市教委科技创新重点项目(11zz04)、上海市人才发展资金(2010001)的资助, 特此致谢!

参考文献:

1. Sykes B, Irven C (2000) Surnames and the Y chromosome. *Am J Hum Genet* 66:1417-1419.
2. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonn -Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26: 358-361.
3. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612.
4. Foster EA, Jobling MA, Taylor PG, Donnelly P, Knijff de P, Mieremet R, Zerjal T, Tyler-Smith C (1998) Jefferson fathered slave's last child. *Nature* 396:27-28.
5. Skorecki K, Skorecki K, Selig S, Blazer S, Bradman R, Bradman N, Waburton PJ, Ismajlowicz M, Hammer MF (1997) Y chromosomes of Jewish priests. *Nature* 385:32.
6. 王传超, 严实, 李辉(2010) 姓氏与 Y 染色体. *现代人类学通讯* 4:e5:27-34.
7. Fang A (1952) *The Chronicle of the Three Kingdoms*, vol. I. Cambridge MA: Harvard University Press.
8. 陈寿(西晋)三国志.卷一魏书一.武帝纪第一.
9. 陈寿(西晋)三国志.蒋济传.注引裴松之.
10. 陈寿(西晋)三国志.卷六魏书六.董二袁刘传第六.袁绍字本初.注引《魏氏春秋》所载袁绍讨伐曹操微文.197.
11. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830-838.
12. King TE, Jobling MA (2009) What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet* 25:351-360.
13. King TE, Ballereau SJ, Sch urer KE, Jobling MA (2006) Genetic Signatures of Coancestry within Surnames. *Curr Biol* 16:384-388.
14. Yan S, Wang CC, Li H, Li SL, Jin L; Genographic Consortium (2011) An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur J Hum Genet*. 19(9):1013-5.
15. Howard TD, Bleecker ER, Stine OC (1999) Fluorescent allele-specific PCR (FAS-PCR) improves the reliability of single nucleotide polymorphism screening. *Biotechniques* 26:380-381.
16. Lehmann EL, Romano JP (2005) *Testing Statistical Hypotheses*. 3rd ed. Springer.
17. Weir BS (1990) *Genetic Data Analysis. Methods for Discrete Population Genetic Data*. Sunderland, MA: Sinauer Associates, Inc. Publishers
18. Bland JM, Altman DG (2000) The odds ratio. *Brit Med J* 320:1468.