

DOI: 10.16359/j.cnki.cn11-1963/q.2017.0081

重构司马光家族基因家谱

杜盼新, 文少卿, 王凌翔, 韦兰海, 蒙海亮, 熊建雪, 包若瑜, 韩昇, 李辉*

复旦大学生命科学院现代人类学教育部重点实验室, 上海 200433

摘要: 司马光家族延续了上千年, 家谱记载较为完整, 为谱牒学、历史学、遗传学等人文和自然学科的跨学科研究提供了较好的材料。本研究对 11 个声称司马光后代的家族进行了 Y-STR 分型, 结果表明有 5 个家族的 STR 单倍型彼此之间十分接近, 同属下游单倍群 O1a1a1a1a1a-F492+,F656-。因此, 我们推断司马光家族的父系遗传类型极有可能属于此单倍群。同时, 我们使用 BATWING 法逐层计算所有支系的最近共祖时间, 其结果与根据谱牒资料构建的家族谱系图非常吻合。本研究为从遗传学角度研究现代家族的父系谱系和重构家族谱牒材料提供了参考, 并将有助于进一步研究汉代史学家司马迁家族以及西晋司马王室的源流。

关键词: 司马光; 基因家谱; Y 染色体; 父系遗传谱系; SNaPshot

中图分类号: Q987; 文献标识码: A; 文章编号: 1000-3193(2019)02-0304-13

Reconstructing the genetic stemma of Sima Guang family

DU Panxin, WEN Shaoqing, WANG Lingxiang, WEI Lanhai, MENG Hailiang,
XIONG Jianxue, BAO Ruoyu, HAN Sheng, LI Hui*

MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Department of History, Fudan University, Shanghai 200433

Abstract: Genetic family trees, especially those of celebrities, have become the research focuses of genetic genealogy in recent years. Among them, the history of Sima Guang's family tree has lasted for a thousand years with full records, which provides a good case for cross-disciplinary studies. In this study, eleven clans claiming to be descendants of Sima Guang were genotyped for 17 Y chromosomal short tandem repeats (STRs). The results indicated that five clans of them showed very close (the maximum number of mutation steps was four) relationships on their STR haplotypes. Further tests of single nucleotide polymorphisms (SNPs) showed that they all belonged to a downstream haplogroup O1a1a1a1a1a-F492+,F656-. So, we

收稿日期: 2017-07-24; 定稿日期: 2017-09-28

项目基金: 国家自然科学基金委 (31671297), 上海市曙光计划 (14SG05), 国家科技攻关计划 (MOST2016YFC0900300), 国家体育总局专项 (2015B086)

作者简介: 杜盼新 (1995-), 男, 硕士研究生, 主要从事历史人类学研究。E-mail: dupanxin@163.com

通讯作者: 李辉, E-mail: lihui.fudan@gmail.com

Citation: Du PX, Wen SQ, Wang LX, et al. Reconstructing the genetic stemma of Sima Guang family[J]. Acta Anthropologica Sinica, 2019, 38(2): 304-316

can infer Sima Guang's paternal lineage was most likely to be that haplogroup. Furthermore, a genealogical tree based on stemmata from these clans was constructed. Then, in order to verify the authenticity of the tree, we calculated the most recent common ancestor (TMRCA) of these clans hierarchically using BATWING method. And we found the time nodes recorded in the tree were in accordance with the TMRCA. This study offered a successful case of using paternal lineage to collate and reconstruct family pedigree. In addition, this investigation will facilitate studies on Sima Qian and the imperial house of the Western Jin Dynasty

Key words: Sima Guang; Genetic genealogy; Y chromosome; Paternal genetic lineage; SNaPshot

1 引言

Y 染色体在世代传递中严格遵守父系遗传法则, 具有同一父系祖先的男性携带着相同或相似的 Y 染色体。由于修著家谱时也严格按照“父子相承”的规定, Y 染色体谱系与文字家谱在理论上具有较好的平行对应关系。基于上述原理, 我们可以将 Y 染色体遗传标记体系应用到传统的谱牒学中, 用于理清历史名人家族内部的亲缘关系及其遗传类型归属的问题。其中, 谱牒学是一门专门研究家谱、族谱的学科。1997 年, Skorecki K 等^[1]和 Foster EA 等^[2]首次通过 Y 染色体上的 STR 和 SNP 遗传标记分别对犹太祭司和美国第三任总统托马斯·杰斐逊的遗传类型进行了推断; 2000 年以后, 一系列研究^[3-5]探讨了成吉思汗的父系遗传类型; 2009 年, Bogdanowicz W 等^[6]对疑是天文学家尼古拉·哥白尼的遗骸进行身份鉴定。上述研究都是利用 Y 染色体上几个有限的 STR 和 SNP 位点, 在增强个体识别能力、获取更多遗传信息和排除外源污染上存在诸多不足。近几年, 随着高通量测序技术的发展, 获取全基因组序列成为了可能, 以上问题也得到了有效解决。例如, 2014 年, Olalde I 等^[7]对路易十六的血手帕进行采样和全基因组测试, 他们发现, 该样本的祖先来源和表型信息与历史记载不符, 从而推测这个血手帕并非来源于本人。之后, King TE 等^[8]也通过二代测序的方法对英格兰国王理查三世的遗骸进行 Y 染色体、线粒体全序和部分表型基因的检测, 并结合详细的家谱挑选理查三世的现存父系和母系后人进行了测试, 他们发现, 理查三世虽然在父系上不能和其现代后裔匹配, 但在母系上可以较好匹配, 最后结合表型、考古学、骨骼学和碳同位素年代测定等信息推断测试遗骸为理查三世本人。国内历史名人的遗传调查始于 2011 年, 王传超等^[9-12]通过一系列研究推测曹操的父系遗传类型为 O2*-F1462 型; 2015 年后, 严实等^[13-14]确定清朝皇室爱新觉罗家族的遗传类型为 C3b1a3a2a-F14735。然而, 在充分利用家谱资料和遗传谱系, 进行相互验证, 最终构建基因家谱方面的工作尚待展开。

司马光(1019—1086), 是北宋时著名的政治家、史学家和文学家, 曾被奉为儒家三圣, 受人景仰^[15]。司马光家族累世聚居于今山西夏县, 祖辈诗礼传家, 家族谱系完整^[16]。现存一部记载较为详细的《梁溪司马氏宗谱》^[17-18]和一本简易的手抄族谱。根据《梁溪司马氏宗谱》的记载, 北宋靖康年间(公元 1127 年), 司马光曾孙司马宗召和司马极为避战

乱由山西夏县分别迁居至今江苏无锡和浙江绍兴。查阅手抄族谱和《涑水司马氏源流集略八卷》记载：明万历年间，司马光十六世孙司马晰由浙江绍兴重新迁回故地山西夏县，因此，无锡、绍兴和夏县三地有家谱明确记载有司马光后裔聚居的地方。

我们系统地整理了司马家族的谱牒资料，对现存司马光后裔进行了Y染色体分型测试，获得了司马光及其后代可能的遗传类型，重构了其家族内部的谱系图。与国内之前的研究相比，本次研究具有两大优势：1) 现存两部家谱在家族谱系的记载上十分详细和完备，从五代十国延续至今，其间没有中断和佚失；2) 本次调查的几个家族之间互无联络，可以从父系遗传类型和家谱记载的一致性上，验证家谱的真实性。因此，本研究是分子人类学、历史学和谱牒学的一次跨学科的有益的尝试。

2 材料和方法

2.1 实验材料

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，在山西夏县、河南洛阳、江苏无锡和浙江绍兴四地采集了 11 个司马以及改姓马的家族（后文统称为司马家族）的 50 名男性的口腔唾液样本。在这些声称为司马光后裔的家族中，有 5 个家族的家谱明确记载他们为司马光后裔，并且他们聚居于山西夏县和江苏无锡。另外，河南洛阳和浙江绍兴两地的家族只是声称自己为司马光后裔。样本采集地在全国的分布见图 1，每个样本的详细分布情况见表 2。谱牒资料也在家族成员签署知情同意后影印保存。



图 1 样本采集地图

Fig.1 Geographic location of the sampling sites in this study

注：图中家族迁徙方向及时间主要参考《梁溪司马氏宗谱》和夏县手抄本家谱

2.2 方法

2.2.1 DNA 提取和 Y 染色体 STR 扩增

采集的唾液样本利用唾液 DNA 抽提试剂盒 (英芮诚, 上海) 和 ETP-300 型全自动核酸提取仪 (英芮诚, 上海) 提取基因组 DNA。随后, 对于提取的 DNA, 采用 AmpFLSTR® Yfiler® PCR Amplification Kit (Applied Biosystem, US) 进行扩增^[19], 扩增产物于 ABI3730XL 测序仪上样, 进行基因分型, 运行 GeneMapperID v3.2 分析实验结果。

2.2.2 Y-SNP 类型的检测

基于分层测试的策略, 我们对所有的样本进行了 100 个 SNP 位点的 SNaPshot 测试^[20], 确定了每个样本的单倍群。对于需要进一步分型的个体, 我们采用先推测后验证的方案: 首先, 通过样本的 17 STR 单倍型结合本实验室基于贝叶斯等位基因频率法的数据库^[11-12, 21-22]推测其可能所属的单倍群, 然后, 参考国际遗传谱系学会 (<http://isogg.org>) 的最新谱系树, 挑选相应的 SNP 位点 (位点信息及引物设计见表 1), 进行相应地分型测试, 验证所推测 Y-SNP 类型的准确性, 最终确定单倍群归属。

2.2.3 家族谱系图的构建

为了构建家族谱系图, 我们查阅《梁溪司马氏宗谱》、简易手抄家谱和相关历史资料, 提取重要信息, 将各支系整理合并成一张家族谱系图, 最后对谱系图的可靠性进行检验。具体操作方法和规则如下:

家谱中的重要信息包括: 人物的背景信息、家族迁徙事件的具体年代以及家族各成员间基本关系。重要人物包括: 历史著名人物如司马光和司马康; 与家族迁徙事件相关的人物如无锡始迁祖司马宗召、绍兴始迁祖司马伋和司马晰 (于明万历年间迁回夏县) 等; 与家族内部过继事件相关的人物, 如司马枢将其子司马宗召过继给司马植; 处于各支系“交汇点”处的人物, 我们将这些人物称为“节点人物”, 这些人物分别是各支系成员的最近共同祖先, 分别为无锡支的司马文焘、夏县支的司马衡以及三个支系 (无锡支、夏县支和绍兴支) 的司马池。

家族迁徙事件主要有司马宗召和司马伋于北宋靖康年间 (公元 1126-1127 年) 分别由山西夏县南迁至无锡和绍兴, 以及司马晰于明万历年间 (公元 1573-1620 年) 重新由绍兴

表 1 位点 F492 和 F656 的基本信息

Tab.1 The information of SNP markers for F492 and F656

SNP	物理位置(hg 19)	突变	PCR引物	延伸引物
F492	18615814	G->C	FOR TCTCCCAAGTTTGCATCCCA REV CACGGAATTCAGCTTCTGATC	AAGAAAATGACCTTTGTT
F656	23757974	C->T	FOR GGGCAGGATGAACGGAGGCA REV GGAGGACGCGGAAACTGCT	TTTTGTCTGTGGAAGATGGCC (反向)

注: SNaPshot 分型体系的防污染策略: 为了确保实验的准确性和防止污染, 我们设置了空白对照、阴性对照和阳性对照。其中的阴性对照为 Control DNA 007 (取自 AmpFLSTR® Yfiler® PCR Amplification Kit), 已知其单倍群为 R1b, 分型结果为 F492-F656。另外, 取同一试剂盒中的 Control DNA 9947A (女性) 为另一阴性对照。阳性对照为 D33 (福建蛋民样本), 属于 O1-F492+, F656-

迁回夏县。在上述重要家族人物和历史事件的记载中,有关四世祖司马光至七世祖司马伋以及二十世祖司马晰的内容,在两本家谱中存在很多一致性,这些一致性也是我们绘制家族谱系图的一个重要依据。

我们以家族世代为时间轴,按照重要人物出现的先后顺序(以司马光曾祖父司马政为一世),将绍兴、无锡和夏县的家谱合并成一张完整的家族谱系图。对于本次研究涉及到的各家族支系,在合并时,除重要人物保留外,其余家族人物均予以省略,并用数字代替省略的世代数。同时,对于本研究未涉及到的家族支系,我们在合并时进行了省略。

最后,构建的谱系图还需要用 BATWING 的方法对“节点人物”所处的时间和家族迁徙事件的具体年代进行检验。

2.2.4 共祖时间的估计

在共祖时间估计中,常用的 STR 突变速率是进化突变速率(EMR^[23])和三种家系突变速率(OMRB^[24]、ImMR^[24]和 OMRS^[25]);常用的计算方法是 BATWING^[20,26](Bayesian analysis of trees with internal node generation)和 ASD^[27-30](average squared distance)两种。我们前期^[31]在对家系共祖时间的计算中发现,采用家系突变速率和 BATWING 相结合的方法得到的结果最为准确,因此,本研究将采用 BATWING 法,在具体计算中使用 15 个 Y-STR 位点(去除 DYS385a、DYS385b)和初始人群数目恒定的指数增长模型^[20]。

为了检验谱系图的可靠性,我们通过 BATWING 的方法,对谱系图逐层进行共祖时间估计,并将其与相对应的“节点人物”生活的年代进行比较。首先,我们对无锡和夏县两个支系单独进行共祖时间估计,并将估计的时间分别与司马文焯和司马衡生活的年代进行比较。之后,将无锡、绍兴和夏县三个支系两两一组进行共祖时间估计,得到的时间分别与对应的“节点人物”生活的年代进行比较。最后,将三个支系一起进行共祖时间估计,并与司马池生活的年代进行比较。

2.2.5 遗传关系网络图的构建

为了直观地展示疑是司马光后裔的样本之间以及他们与全国其他同属单倍群 O1a1a-P203 的样本之间的遗传关系,我们使用 NETWORK v. 4.5.1.6 (Fluxus-engineering.com) 的 Reduced media-joining^[32]法,对司马光后裔数据、文献数据^[20,33-35]和实验室未公开数据共计 96 个样本的 15 个 Y-STR 位点(去除 DYS385a、DYS385b)构建遗传关系网络图。

3 结 果

3.1 Y 染色体分型

11 个司马家族共 50 人的 STR 和 SNP 的分型结果见表 2。根据表 2,我们可以发现,绍兴、夏县、无锡 2、无锡 3 和无锡 4 这 5 个家族(除绍兴家族外,其余 4 个家族均有家谱)中有 11 个样本的 STR 单倍型彼此之间十分接近,且同属 O1a1a-P203 单倍群。为了进一步确定其下游单倍群归属,我们先对其所属的下游 SNP 类型推测的结果为 O1a1a1a1a-F492,

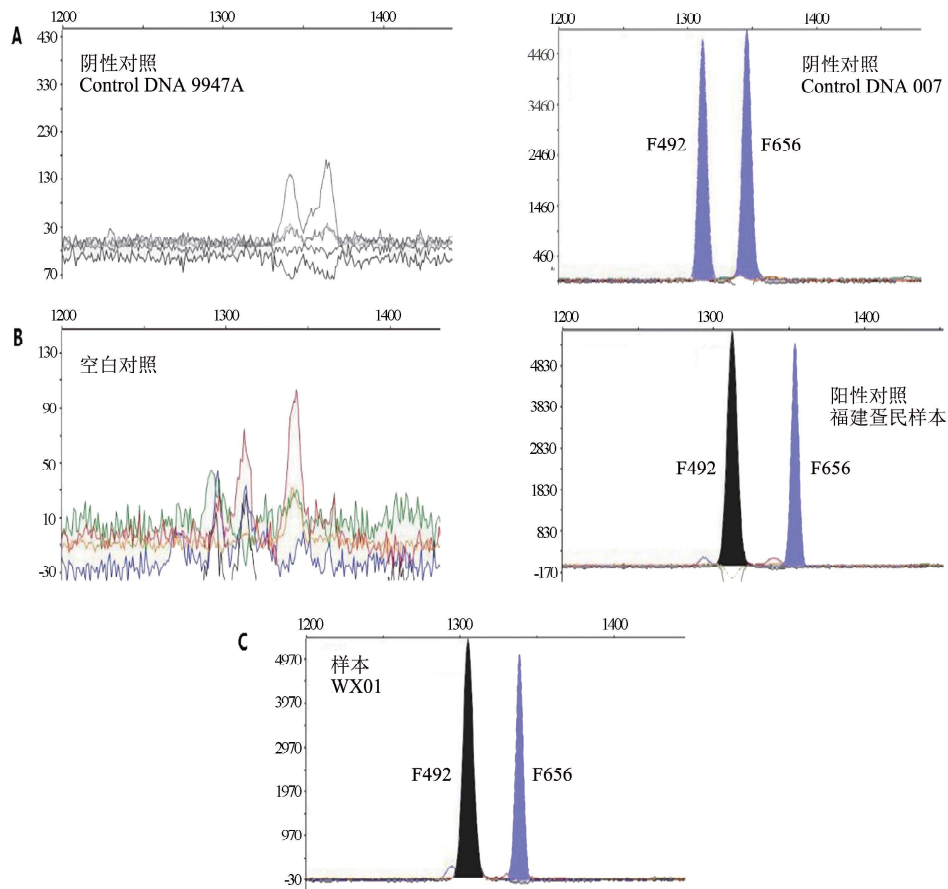


图 2 样本和不同对照组的 SNaPshot 分型结果

Fig.2 The results from a SNaPshot assay of objective clans and control groups

(注: 因测试的 11 个样本具有相同的分型测试结果, 故仅将 WX01 样本的测试结果列出作为代表)

后用 SNaPshot 测试进行验证。测试引物序列和部分分型测序峰图见表 1 和图 2。

通过对比各个家族中的共享单倍群 (表 2), 除了本次研究的目标单倍群 O1a1a-P201, 无锡 1 家族中的 WX04-08 和无锡 5 家族的 WX09 共 6 人同属单倍群 O2a2b1a1-M117, 无锡 4 家族中的 WX12 和无锡 6 家族的 WX16-17 共 3 人同属单倍群 O2a2b1*-M134+,M117-。这两个单倍群内部成员的 STR 单倍型彼此之间差异很大 (最大差异超过六步), 我们用 BATWING 法分别对这两个单倍群内的样本进行最近共祖时间估计, 属于单倍群 O2a2b1a1-M117 和 O2a2b1*-M134+,M117- 家族的最近共祖时间分别为距今 1974-2403 年和距今 2275-2677 年, 与家谱记载的时间明显不符。因此, 我们不再对这些单倍群做进一步的研究。

根据图 2 的 SNaPshot 分型结果, 阴性对照组的女性 9947A 只有很低的干扰峰 (图 2A), 男性对照 007 为 F492-,F656- 样本 (图 2A), 阳性对照 D33 (图 2B) 的分型结果为 F492+,F656-。另外, 空白对照也未发现污染 (图 2B), 说明 SNaPshot 分型体系的结果十分准确。最后, 经过三次重复试验, 被测试的 11 个样本的遗传类型确定为 O1a1a1a1a1a-

表 2 17Y-STR 和 SNP 类型的分型结果
Tab.2 The genotyping results of 17Y-STR and SNP

家族	支系	样本编号	G ₁ DYS19	B ₁ DYS389I	B ₁ DYS389II	B ₁ DYS390	Y ₁ DYS391	Y ₁ DYS392	Y ₁ DYS393	R ₁ DYS437	R ₁ DYS438	Y ₁ DYS439	R ₁ DYS448	B ₁ DYS456	G ₁ DYS458	Y ₁ DYS635	R ₁ Y ₁ GATA ₁ H4	G ₁ DYS385a	G ₁ DYS385b	SNP单倍群	
洛阳1	SMG	SMG01-3	15	14	16	24	9	13	14	14	10	11	18	15	14	21	13	13	18	O1b-M268+, P31+, PK4-, M176-	
		SMG04-8	15	14	17	24	9	13	14	14	10	11	18	15	14	21	13	13	18	O1b-M268+, P31+, PK4-, M176-	
		SMG09-11	15	14	17	24	9	13	14	14	10	11	18	15	14	21	13	13	19	O1b-M268+, P31+, PK4-, M176-	
洛阳2	WGD	WGD01-2	17	13	16	24	10	11	13	14	10	9	20	15	16	20	12	12	19	C2-M217	
		WGD03-7	17	13	16	24	10	11	13	14	10	9	20	15	16	20	12	12	20	C2-M217	
		WGD08-9	17	13	16	24	10	11	13	14	10	9	20	15	16	20	12	13	20	C2-M217	
		WGD10	17	13	16	24	10	11	14	14	10	9	20	15	16	20	12	12	20	C2-M217	
		WGD11	17	13	16	24	10	10	13	14	10	9	20	15	16	20	12	12	20	C2-M217	
无锡1	WXA	WX03	17	12	17	24	10	13	12	15	10	12	21	15	17	21	12	14	18	O2a2*-P201	
		WXB	WX04	14	12	16	24	10	14	12	15	11	12	20	15	19	20	11	19	O2a2b1a1-M117	
		WX05-7	14	12	16	24	10	14	12	15	11	12	20	15	18	20	11	19	19	O2a2b1a1-M117	
		WX08	14	12	16	24	10	14	12	15	11	12	20	15	18	20	11	19	20	O2a2b1a1-M117	
绍兴	SX	SX	15	12	17	23	11	14	13	14	10	11	18	17	15	19	12	13	14	O1a1a-P203	
夏县	XX	XX01-5	15	12	17	23	10	14	13	14	10	11	18	17	15	19	12	14	14	O1a1a-P203	
		XX06	15	12	17	23	10	14	13	14	10	11	18	17	15	19	12	13	14	O1a1a-P203	
无锡2	WXC	WX01	15	12	17	23	11	14	13	14	10	12	18	17	16	19	12	10	13	O1a1a-P203	
无锡3		WX02	15	12	17	23	11	14	13	14	10	12	18	16	16	19	12	10	13	O1a1a-P203	
无锡4		WX11,13	15	12	17	23	11	14	13	14	10	12	18	17	15	19	11	10	13	O1a1a-P203	
		WXE	WX10	16	12	16	23	10	14	13	14	10	11	18	16	15	19	11	13	13	O1a1a1a-F140+, F492-
		WXF	WX12	15	12	16	23	10	12	13	14	10	12	19	16	17	19	13	12	18	O2a2b1-M134+, M117-
		WXG	WX14	14	14	17	23	10	14	13	14	11	11	19	16	17	21	13	10	13	N-M231
		WXH	WX15	15	13	16	25	11	13	14	14	10	11	18	15	18	22	10	13	18	O1b1a1a-M95
无锡5	WXD	WX09	15	14	17	24	10	14	12	15	11	12	20	16	18	21	11	13	17	O2a2b1a1-M117	
无锡6	WXI	WX16-17	14	12	16	23	10	12	12	15	10	13	19	15	17	19	12	12	16	O2a2b1-M134+, M117-	
无锡7	WXJ	WX18-21	15	12	17	25	9	13	12	15	10	11	21	15	18	22	12	10	22	O2a2b-P164+, M134-	

注：1、斜体标注的支系最可能为司马光后裔；2、SMG：洛阳市的一个样本采集地的简称；WGD：洛阳市的另一个样本采集地的简称；WX：无锡的简称，为了对无锡的不同支系进行区分，在“WX”后面加上大写英文字母A-J，故出现编号“WXA”、“WXB”等；SX：绍兴的简称；XX：夏县的简称

F492+,F656- (图 2C)。根据所有样本的单倍群类型所构建的最简谱系树见图 3。

通过比较表 2 的分型结果，我们可以发现，同一家族中可能包含不止一种父系遗传类型。为了便于区分，在同一家族中，按照单倍群的差异，又划分为若干不同支系。因此，按照上述方法，11 个家族可划分为 14 个支系。其中，“无锡 2”家族的 WX01、“无锡 3”家族的 WX02 和“无锡 4”家族的 WX11 和 WX12，他们之间的 STR 差异在 3 步以内，同属 O1a1a1a1a1a-F492 单倍群，并且家谱记载他们的共同祖先为司马文煊，故将他们归为同一支系，命名为 WXC。通过表 2 可以看出，夏县 XX 和绍兴 SX 两个支系的

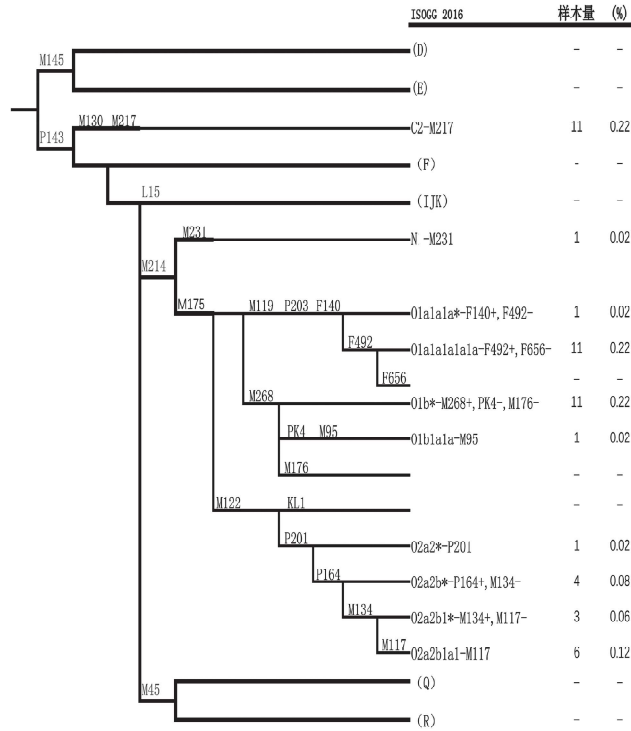


图 3 基于单倍群类型所构建的进化树

Fig.3 Phylogenetic tree of samples in this study

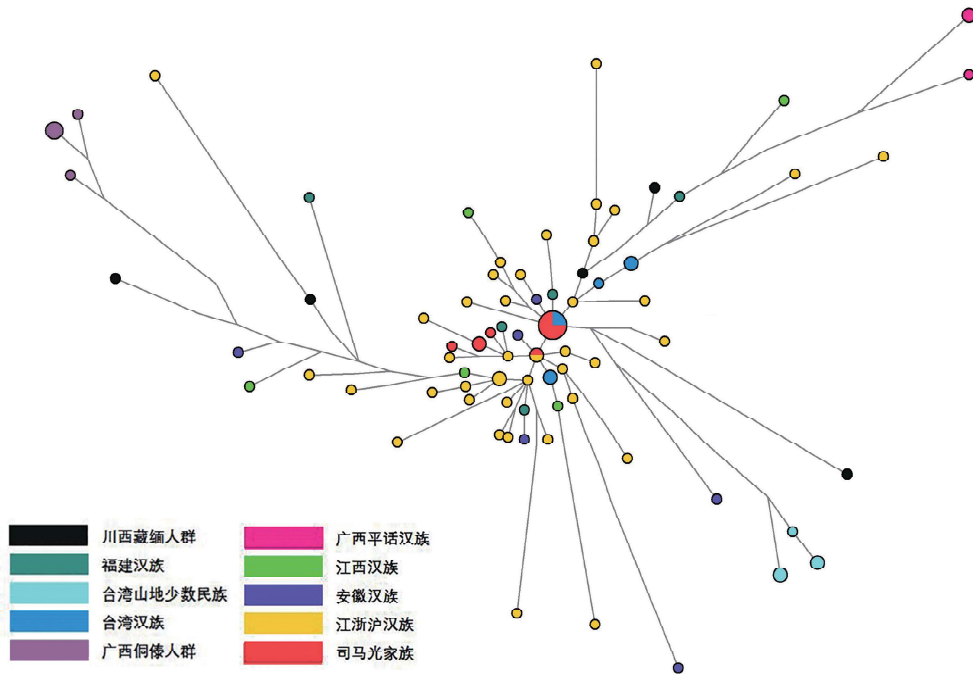


图 4 单倍群 O1a1a-P203 的遗传关系网络图

Fig.4 The network of haplogroup O1a1a-P203

注: 图中每一个圆点代表一种 STR 单倍型, 圆点的面积与人数成正比; 连线长短与 STR 步数的差异成正比; 不同颜色代表不同群体

STR 单倍型彼此之间的突变在两步以内，当不考虑突变率较高的 DYS385a、DYS385b 两个位点时，二者与无锡的 WXC 支系的突变步数的最大差异也在四步以内，并且他们同属 O1a1a1a1a-F492 单倍群。为了更加直观地展示上述三个支系间的关系，我们绘制了遗传关系网络图（见图 4）。从网络图可以看出，三个目标支系的所有成员聚集在一起，他们之间的最大突变步数在四步以内，说明这三个支系在父系上具有密切的遗传关系。同时，相比其他群体，他们与江浙汉族、安徽汉族和江西汉族等群体的遗传关系更为密切。

考虑到这三个支系在 STR 和 SNP 分型结果的一致性，我们认为其单倍群为司马光单倍群的可能性较大。

3.2 谱系图构建和检验

根据谱系图构建规则，我们将三个支系的谱牒资料进行整理合并，最后绘制成一张完整的家族谱系图（图 5）。我们通过 BATWING 方法计算各支系的共祖时间（表 3），并对家族谱系图进行逐层检验。

根据谱系图，三世祖司马池（距今 976-1041 年）为三个支系（无锡、绍兴和夏县）的最近共同祖先，同时也是绍兴和无锡支系（SX-WXC）以及无锡和夏县支系（WXC-XX）的最近共同祖先。通过 BATWING 方法估计以上三个支系的共祖时间分别为距今 911.6-1160.7 年、898-1145.7 年和 948.9-1213.3 年，都基本与司马池生活的年代相一致；二十世祖司马晰于明万历年间（距今 397-444 年）从绍兴迁至夏县，这一历史事件可大致通过 SX 和 XX 两个支系的共祖时间（距今 247.3-328.3）来估计；十四世祖司马文焘，此人为无锡 WXC 支系的最近共同祖先，推测其生活年代为距今 701-766 年（从司马池算起，以 25 年为一代）。通过 BATWING 方法计算得到的时间为距今 774.5-1007.8 年，二者十分接近。另外，由于夏县支系中所有成员的最近共祖时间仅为 150-200 年（从司马衡算起，以 25 年为一代），导致所有成员均共享同一种 STR 单倍型（去除 DYS385a、DYS385b），因而在如此短的时间内进行共祖时间估计就没有了实际意义。

综上所述，谱系图记载的历史时间与计算所得共祖时间较为一致，说明根据家谱资料构建的谱系图比较可靠。

表 3 目标支系共祖时间的估计
Tab.3 The TMRCA of objective clans

Clan	EMR		lmMR		OMRB		OMRS	
	TMRCA	95% CI	TMRCA	95% CI	TMRCA	95% CI	TMRCA	95% CI
SX-WXC-XX	4371.3	569.2-31873.6	1160.7	159.5-7616.8	1147.6	157.3-7545.2	911.6	121.9-6086.5
SX-WXC	4405.4	353.1-48602.5	1145.7	87.6-12491.4	1120.5	85.9-12328.3	898.0	66.1-10179.8
SX-XX	1204.4	50.0-16042.0	328.3	12.6-4604.3	312.7	12.1-4392.6	247.3	9.3-3589.0
WXC-XX	4446.4	553.3-32870.1	1213.3	156.0-8263.8	1182.5	152.2-8083.7	948.9	119.2-6623.7
WXC	3833.7	210.6-53229.6	1007.8	48.2-15604.5	978.1	45.8-15397.8	774.5	34.8-12630.9
XX	--	--	--	--	--	---	--	--

注：TMRCA (Time to Most Recent Common Ancestry) 为最近共祖时间；CI(Confidence Interval) 为置信区间；EMR, lmMR, OMRB, OMRS 为四种 STR 突变速率

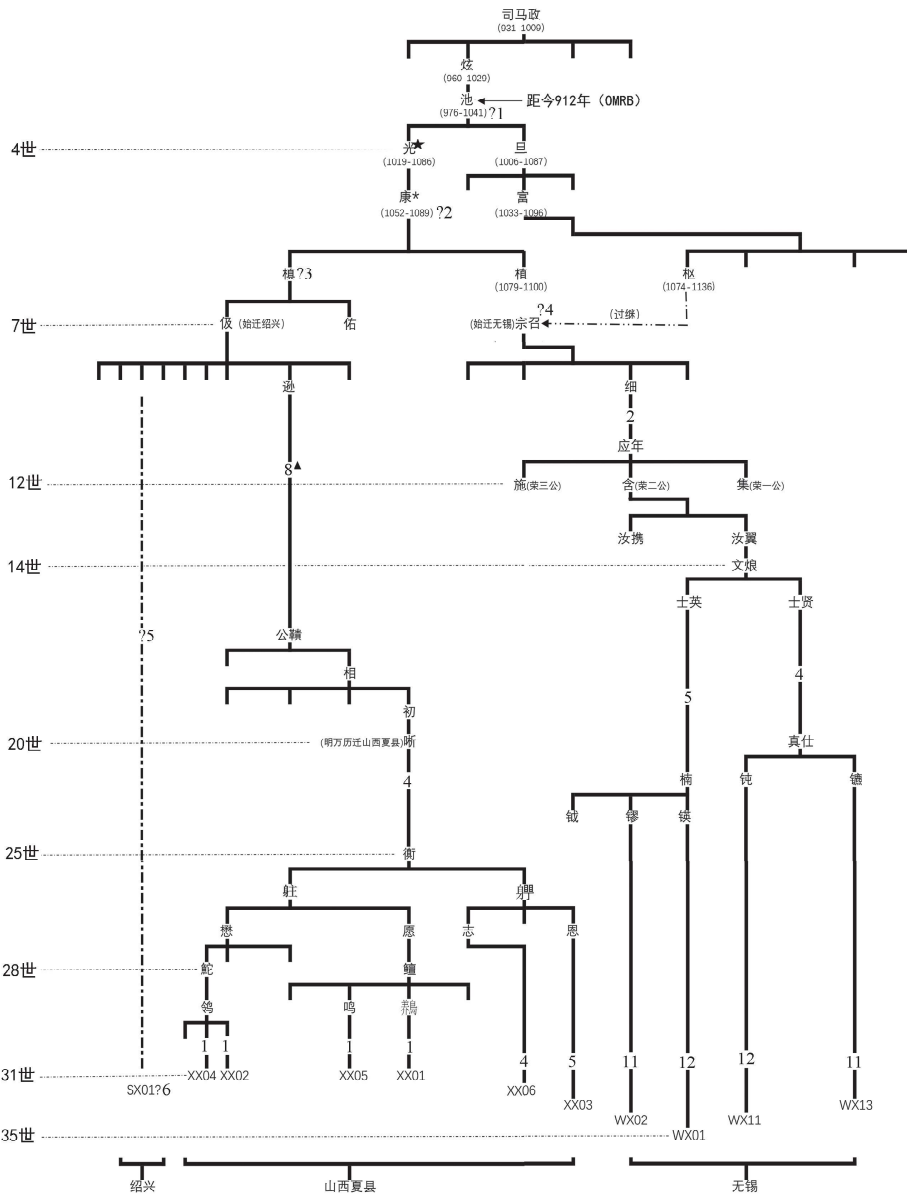


图 5 司马光家族谱系图

Fig.5 The genealogical tree of Sima Guang family

- 1) 此处“8[▲]”表示从司马逊到司马公贽中间隔有 8 代人, 谱系图中类似地方也用数字表示间隔代数。
- 2) 此处“*”：《宋史》等资料记载司马康为司马光之子, 然而其他私家著作 (如苏轼《西楼帖·与堂兄三首》和邵雍《闻见录》) 提到康是由光兄司马旦之子过继 (详参颜中其《司马康为司马光兄亲子》, 《古籍整理研究学刊》1988 年第 3 期), 《梁溪司马氏宗谱》沿袭《宋史》观点。实际上, 由于 Y 染色体父系遗传的特点, 司马康究竟是司马光或者是司马旦的子嗣不影响本文的结论。
- 3) 此处“?1”：根据家谱记载, 司马池生卒时间为公元 976-1041, 与《司马光年谱》记载的公元 979-1041 有差别, 特此说明。
- 4) 此处“?2”：根据家谱记载, 司马康生卒时间为公元 1052-1089 年, 与《续资治通鉴长编》记载的公元 1050-1090 年有差别, 特此说明。
- 5) 此处“?3”：根据家谱, 司马慎为司马康亲子, 但《建炎以来系年要录》载司马光孙司马植死后, 以“再从孙慎”继嗣, 即司马慎可能并非司马植亲子。需要指出的是, 《要录》中的“司马慎”与家谱中的“司马慎”应是同一人。
- 6) 此处“?4”：家谱记载司马宗召生卒时间为 1156-1214, 但后又说“宗召于靖康 (1127 年) 中以遗荫佐无锡”, 并且与其父司马枢的生卒时间 (1074-1136) 不符。同时, 《建炎以来系年要录》记载, 绍兴六年 (1136 年) “吉州万安县丞司马宗召添差两浙路转运司干办公事”。因此, 我们推断宗召生于“绍兴二十六年 (公元 1156 年)”一句有误。
- 7) 此处“?5”和“?6”：绍兴的这一支因没有家谱, 已不清楚是司马光的几世孙。

4 讨论

4.1 司马氏的父系遗传组成

本研究对司马家族中 14 个支系进行了 Y 染色体的分型测试, 发现有三个支系的 STR 单倍型彼此之间比较接近(最大突变步数在四步以内), 且同属于 O1a1a1a1a1a-F492+, F656- 单倍群。然而, 其余各支系之间的 STR 单倍型差别很大, 并且他们所属的单倍群类型也不相同(表 2)。

根据图 3 的结果, 可以进一步发现各单倍群在本次调查的总人群中所占比例分别为 O2-M122 (28%)、O1a-M119 (24%)、O1b-M268 (24%)、C2-M217 (22%) 和 N1-M231 (2%)。这一结果表明司马氏在父系遗传组成上具有多样性。国内外关于姓氏的研究也有类似结果, 如侯伟光等^[19]对曲阜地区孔姓人群进行 Y 染色体遗传多态性研究时, 发现存在 C3、Q1a1 和 O3 等多种单倍群; 为了确定著名航海家克里斯托弗·哥伦布的父系遗传类型进而推断其出生地, Martínezgonzález LJ^[36]等分别对西班牙加泰罗尼亚地区和意大利北部地区的以哥伦布为姓的人群进行研究, 同样发现有 R1b、J1 和 J2a1 等多种单倍群。上述情况, 可能是由于在姓氏传递过程中发生了“非父事件”^[37-39]或是由姓氏“非单一起源”^[40-41]的特点决定。

4.2 家族遗传谱系的可靠性

结合谱牒资料和 Y 染色体分型结果, 本次研究重构了司马光家族基因谱系。对于其可靠性, 我们将从以下几个方面进行探讨: 首先, 在对各地独立记载的谱牒资料进行整理时, 发现它们在重要家族人物和历史事件的记载上存在一致性, 因此, 根据这些一致性, 将山西夏县、江苏无锡和浙江绍兴三地的司马光家族谱系进行整合, 绘制成一张详细的家族谱系图; 之后, 依照谱系图分别对三地的支系成员进行 Y 染色体分型, 发现他们同属下游单倍群 O1a1a1a1a1a-F492+, F656-, 据此, 我们可以推断司马光父系遗传类型极有可能属于该单倍群; 同时, 采用 BATWING 方法对所有支系逐层进行最近共祖时间估计, 得出的时间与谱系图中相应的时间较为吻合。因此, 基于家谱文字资料和 Y 染色体遗传谱系构建基因家谱还是可行的。

另外, 对于家族中的一些事件, 通过 Y 染色体基因分型和共祖时间估计也难以检验其真实性, 如司马康是司马光亲子, 还是由司马旦过继给司马光, 司马槿是否为司马康亲子; 绍兴这一支系与夏县支系分开的准确时间, 以及其在谱系图中所处的确切世代。要解决上述问题, 我们可能需要挑选代表性个体进行 Y 染色体全序测试。

4.3 司马光遗传类型的归属

根据上述对家族基因谱系可靠性的论证, 可以推测司马光的遗传类型很有可能为 O1a1a1a1a1a-F492+, F656- 型, 但要最终确定其遗传类型, 还需要: 1) 对司马光或其父系亲属的遗骸进行古 DNA 测试, 获取直接证据; 2) 寻找其他具有可靠家谱的司马光后裔, 进一步扩大样本量, 对其进行 Y 染色体分型, 进而确定司马光的遗传类型以及其家族内

部的谱系关系, 完善本次研究所构建的基因家谱。

在本研究的基础上, 我们可以进一步对司马氏的源流进行探索, 如司马光、西汉史学家司马迁和西晋司马王室三个家族之间潜在的遗传关系。根据现存历史资料^[42-46]的记载, 以上三个家族很有可能源于同一父系祖先。但由于史料残缺、年代久远, 我们无法判断这些记载的可信度, 因此, 仅仅通过历史学研究很难论证他们三个家族之间的关系。借助分子人类学的跨学科研究或许是我们揭开这些历史谜题的有益的尝试。

致谢: 感谢上海复信生物科技有限公司在样本采集方面的帮助; 感谢无锡祠堂文化研究会马森根、夏县司马温公祠负责人王在京、洛阳市洛龙区档案史志局马正标、江南晚报张月、绍兴晚报顾鹏程和黄武妹等联络人, 以及司马扩军、马兴生、马炳德和司马国昌等家族召集人的无私帮助。

参考文献

- [1] Skorecki K, Selig S, Blazer S, et al. Y chromosomes of Jewish priests[J]. *Nature*, 1997, 385(6611): 32-32
- [2] Foster EA, Jobling MA, Taylor PG, et al. Jefferson fathered slave's last child[J]. *Nature*, 1998, 396(6706): 27-8
- [3] Zerjal T, Xue Y, Bertorelle G, et al. The genetic legacy of the Mongols[J]. *Am J Hum Genet*. 2003, 72(3): 717-21
- [4] Abilev S, Malyarchuk B, Derenko M, et al. The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan[J]. *Hum Biol*. 2012, 84(1): 79-89
- [5] Batbayar K, Sabitov ZM. The Genetic Origin of the Turko-Mongols and Review of The Genetic Legacy of the Mongols. Part I: The Y-chromosomal Lineages of Chinggis Khan[J]. 2012
- [6] Bogdanowicz W. Genetic identification of putative remains of the famous astronomer Nicolaus Copernicus.[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(30): 12279-12282
- [7] Olalde I, Sánchezquinto F, Datta D, et al. Genomic analysis of the blood attributed to Louis XVI (1754–1793), king of France[J]. *Scientific Reports*, 2014, 4(7497): 4666
- [8] King TE, Fortes GG, Balaesque P, et al. Identification of the remains of King Richard III[J]. *Nature Communications*, 2014, 5(1828): 5631
- [9] 王传超, 严实, 侯铮, 等. Y 染色体揭开曹操身世之谜 [J]. *现代人类学通讯*, 2011, 5: 107-111
- [10] Wang CC, Yan S, Hou Z, et al. Present Y chromosomes reveal the ancestry of Emperor CAO Cao of 1800 years ago [J]. *Journal of Human Genetics*, 2012, 57(3): 216-218
- [11] Wang CC, Yan S, Yao C, et al. Ancient DNA of Emperor CAO Cao's granduncle matches those of his present descendants: a commentary on present Y chromosomes reveal the ancestry of Emperor CAO Cao of 1800 years ago [J]. *Journal of Human Genetics*, 2013, 58(4): 238-239
- [12] 文少卿, 王传超, 敖雪, 等. 古 DNA 证据支持曹操的父系遗传类型属于单倍群 O2[J]. *人类学学报*, 2016(4): 617-625
- [13] Yan S, Tachibana H, Wei LH, et al. Y chromosome of Aisin Gioro, the imperial house of the Qing dynasty.[J]. *Journal of Human Genetics*, 2015, 60(6): 295-298
- [14] Wei LH, Yan S, Yu G, et al. Genetic trail for the early migrations of Aisin Gioro, the imperial house of the Qing dynasty[J]. *Journal of Human Genetics*, 2016
- [15] 李昌宪. 司马光评传 [M]. 南京: 南京大学出版社, 2002
- [16] 四库全书存目丛书编纂委员会编. 四库全书存目丛书 史部 第 84 册. 济南: 齐鲁书社, 1996: 72-190
- [17] 司马栋. 梁溪司马氏宗谱: 二十卷 [M]. 1916
- [18] 马森根. 梁溪司马氏宗谱尤家坦横街支 [M]. 2013
- [19] 侯伟光, 王传超, 蒋世洪, 等. 曲阜地区孔姓人群 17 个 Y-STR 基因座遗传多态性分析 [J]. *人类学学报*, 2016, 35(1): 125-131
- [20] Wang CC, Wang LX, Shrestha R, et al. Genetic Structure of Qiangic Populations Residing in the Western Sichuan Corridor[J]. *Plos*

- One, 2014, 9(8): e103772-103772
- [21] Wang CC, Wang LX, Shrestha R, et al. Convergence of Y Chromosome STR Haplotypes from Different SNP Haplogroups Compromises Accuracy of Haplogroup Prediction[J]. *Journal of Genetics and Genomics*, 2015, 42(7): 403-407
- [22] Wen SQ, Tong XZ, Wang CZ, et al. Y-chromosomes from skeletal remains of Chinese Expeditionary Force offer a clue to their paternal relatives[J]. *Science Bulletin*, 2016, 61(6): 425-427
- [23] Zhivotovsky LA, Underhill PA, Cinnioğlu C, et al. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time.[J]. *American Journal of Human Genetics*, 2010, 74(1): 50-61
- [24] Burgarella C. Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data.[J]. *European Journal of Human Genetics*, 2011, 19(1): 70-5
- [25] Shi W, Ayub Q, Vermeulen M, et al. A Worldwide Survey of Human Male Demographic History Based on Y-SNP and Y-STR Data from the HGDP-CEPH Populations[J]. *Molecular Biology and Evolution*, 2010, 27(2): 385-93
- [26] Wilson IJ, Balding DJ. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities[J]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2003, 166(2): 155-188
- [27] Zhivotovsky LA. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow.[J]. *Molecular Biology & Evolution*, 2001, 18(5): 700.
- [28] Behar DM, Thomas MG, Skorecki K, et al. Multiple Origins of Ashkenazi Levites: Y Chromosome Evidence for Both Near Eastern and European Ancestries[J]. *American Journal of Human Genetics*, 2003, 73(4): 768
- [29] Ramakrishnan U, Mountain JL. Precision and Accuracy of Divergence Time Estimates from STR and SNPSTR Variation[J]. *Molecular Biology & Evolution*, 2004, 21(10): 1960
- [30] Sengupta, Sanghamitra, Zhivotovsky, et al. Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists[J]. *American Journal of Human Genetics*, 2006, 78(2): 202-21
- [31] Wang CC, Li H. Evaluating the Y chromosomal STR dating in deep-rooting pedigrees[J]. *Investigative Genetics*, 2015, 6(1): 1-3
- [32] Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies.[J]. *Molecular Biology and Evolution*, 1999, 16(1): 37-48
- [33] Trejaut JA, Poloni ES, Yen JC, et al. Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia[J]. *BMC Genetics*, 2014, 15(1): 77
- [34] Deng QY, Wang CC, Wang XQ, et al. Genetic affinity between the Kam-Sui speaking Chadong and Mulam people[J]. *Journal of Systematics and Evolution*, 2013, 51(3): 263-270
- [35] Yan LU, Pan SL, Qin SM, et al. Genetic evidence for the multiple origins of Pinghua Chinese[J]. *Journal of Systematics and Evolution*, 2013, 51(3): 271-279
- [36] Martínez-González LJ, Martínez-Espín E, Álvarez JC, et al. Surname and Y chromosome in Southern Europe: A case study with Colom/Colombo[J]. *European Journal of Human Genetics*, 2012, 20(2):211-6
- [37] Sykes B, Irven C. Surnames and the Y Chromosome[J]. *American Journal of Human Genetics*, 2000, 66(4): 1417-1419
- [38] King TE, Ballereau SJ, Schürer KE, et al. Genetic Signatures of Coancestry within Surnames[J]. *Current Biology*, 2006, 16(4): 384-388
- [39] King TE, Jobling MA. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution[J]. *Trends in Genetics*, 2009, 25(8): 351-360
- [40] 李占军. 中华姓氏起源 [M]. 世界社会文献出版社, 2007: 404-405
- [41] 廖康强. 中华姓氏起源通史 [M]. 北京: 中国商业出版社, 2014: 431-433
- [42] (宋)司马光著, 李之亮笺注. 司马温公集编年笺注 (第五册) [M]. 成都: 巴蜀书社, 2009: 496-498
- [43] (宋)庞籍. 天章阁待制司马府君碑铭 [A]. 见: 司马光年谱 [M]. 北京: 中华书局, 1990: 483
- [44] (宋)苏轼. 苏东坡全集 [M]. 北京: 中国书店, 1986: 420
- [45] (唐)房玄龄等. 晋书 [M]. 北京: 中华书局, 1974: 1
- [46] (西汉)司马迁. 史记 [M]. 北京: 中华书局, 1974: 3285-3286