

# 上 编

## 中国姓氏群体遗传

姓氏群体遗传学是人类群体遗传学的一个分支，是利用人类姓氏的特定传递规则，研究姓氏在人群中的分布，进而分析人群间的亲缘关系和历史上人群迁移模式等。

人类的姓在世界各地出现的时间相差很大，而且，其产生的文化背景也不同。世界上最早使用姓的国家是中国，大约在 5 000 年前的伏羲氏时代 姓被定为世袭，由父系传递，中国人的姓是世界上最古老的。在世界其他地区，从姓的产生到普遍使用姓，是从大约公元 1 000 年起的一个渐进发展的过程。在英国，当时由于名字不够用而产生使用姓的需要，这个过程延续了大约 600 年，才得以在大不列颠土地上普遍使用姓（《不列颠百科全书》第 16 卷，1999 年），因此 在欧洲大陆普遍使用姓的历史只有 400 年。

在欧洲的有些地区直到 20 世纪 姓的使用尚未普遍。比如 到 1935 年土耳其人方以法律形式规定使用姓。犹太人很晚才使用姓，而且经常出于被迫，这是由于犹太人在当时所处的国际地位，不能使用基督徒的名字，尤其在德、意、奥地区。但有一个例外 在 1 000 年前的北宋 由“丝绸之路”进入中国的犹太人定居河南开封，到明朝时已经被中国的文化所同化，全部采用了汉姓。犹太人不能忘记这一历史，是中国人给了他们祖先以姓，犹太人是西方最早有姓的民族之一。

姓的形成常常是一个民族的历史和文化偏见的反映。在西班牙，姓的形成过程同党派、家族的自豪感有关，瑞典人的姓反映了瑞典人对大自然的热爱。与我国一水相隔的日本，在公元 5 世纪晚期开始产生了姓 但仅仅是代表社会政治结构中地位和职务的世袭称号（《不列颠百科全书》第 9 卷，1999 年）。直到明治维新时期，1875 年日本政府颁布了法令，实施户籍登记，要求每一个日本人必须有姓，日本人才急匆匆地为自己找姓，多以所居地名为姓，

因此多为两字姓，这不同于中国人的复姓，日本人一下子涌现出 30 000 多个姓，在日本这才真正地普及了姓，到今天，日本的姓已达 120 000 多个 (Yasuda, 1983 ;Lasker, 1985)。

中国人，尤其在汉族社会中，宗族观念根深蒂固，直到 1949 年前，同姓同宗在汉人观念中，仍是一种很强的联系纽带。实际上西周春秋以来，在宗法制度下，中国人的亲属是以父系为中心，论亲疏。父系家族的延续被认为是至关重要的事。祭祀祖宗，不断烟火，被认为是每个家庭的头等大事，无后被认为是最大的不孝。古代的国家观念，也与家族观念相联系，所谓一家一姓的天下，实际是父系家族观念的无限扩大。在这种宗法制度和宗族观念的几千年中，中国人十分重视姓，养成了同姓聚居的习俗，很多地区流传着修谱联宗的现象，在全国形成了无数群大小不等的同姓人群。

中国人姓氏几乎都以父传子的垂直方式世代相传，女子婚后也不改姓，仅仅附上夫姓，表明已婚。所以，中国姓氏是中国几千年父系社会全过程的见证文化，以一种血缘文化的特殊形式记录了中华民族的形成和发展，在中华民族文化的同化过程和国家统一上曾经起过独特的民族凝聚力的作用。中国人姓氏是中国历史和特有的传统文化的反映，中国人的姓氏种类和分布也是一项重要的国情。研究和剖析中国人姓氏是我们的责任。

本编主要向读者介绍，自然科学家是如何认识中国人姓氏的；中国人姓氏的特征是什么？用何方法收集姓氏和确认其科学研究的价值？历代中国人姓氏的分布规律，它们揭示了什么本质？中国人的姓氏有多少？直观的姓氏遗传地形图是如何表达的？

# 第一章 绪 论

## 1.1 中国姓氏的起源

姓，代表氏族的血统，最早起源于母系社会。早期的姓出自氏族的图腾，氏族往往以天象星座、动植物作为其图腾的内容，在与周边的其他氏族的交往中以展示氏族图腾的徽记来表明其身份，区别不同的氏族和方位，这氏族徽记就是姓的最早雏形。从时间上，姓的产生远远早于文字，有了文字以后，根据图腾徽记的图形产生了有文字意义上的姓。

氏，是古代贵族标志宗族系统的称号，从夏朝中期开始，氏成为姓的支系，是私有制社会的特种产物，表示功勋和地位。先秦时女子称姓，男子称氏（郑樵《通志·氏族略》）。

中国古代姓氏传说产生于远古五帝时期。从盘古氏开天地，历三皇、五龙氏到伏羲氏定姓、氏，再经炎黄、尧舜时代，夏商周、春秋战国到秦并六国，建立了中国第一个中央集权政权，姓、氏也混而为一。中国姓氏的历史至少有五千年之久。

历来都以为中国人先有姓，后有氏。事实上，姓、氏一直在混合使用，姓和氏的关系也在转变。最早的氏为大氏族、部落、国家、氏族联盟的称呼。传说中和文献上出现的这类氏有上百个，最早的是盘古氏、天皇氏、地皇氏、人皇氏、五龙氏时代，随后为钜灵氏、黄神氏、鬼隗氏、空桑氏、次民氏等二十二氏的循蜚纪时代，接着为辰放氏、蜀山氏、混沌氏、有巢氏、燧人氏、庸成氏等十三氏的因提纪时代。在这个时期中，最重要的氏是有巢氏和燧人氏，有巢氏教民架木为巢，掘地为营，燧人氏发明了钻木取火，教人熟食，教民结绳以记事。

之后进入了伏羲氏时代，这一时期中原地区出现了共工氏、柏皇氏、朱襄

氏、昊英氏、栗陆氏、郝胥氏、昆吾氏、葛天氏、阴康氏、中皇氏、女娲氏等部落或国家。中国最早的姓就产生在伏羲氏时期，伏羲氏之后为风姓，这是中国第一个姓（范文澜《中国通史》，1978年）。这个时期的另一个重要的氏是中皇氏，中皇氏的首领为仓颉，他创造了文字，替代结绳记事。

女娲氏之后为神农氏，即炎帝，其后为姜姓，黄帝姓公孙，名轩辕，号有熊氏，有熊氏是黄帝所在的部落和国的称呼，黄帝继承炎帝之位统治了中原大地。黄帝有二十五子，得姓者十四，有姓十二：姬、西、祁、己、滕、蔽（亦作箴）、任、荀、僖、姁、依。这十二个姓事实上是有熊氏部落的分族，其他的儿子因没有实力建立起自己的族而不能有姓。这是姓为氏的支系的时代。黄帝与西陵氏女生子玄器，号青阳氏，亦名少昊，少昊氏族东迁，成为东夷族的首领，改称金天氏，其后为嬴姓、偃姓，黄帝与西陵氏女生次子昌意，昌意娶蜀山氏之女生颡顼，为高阳氏，其后为半姓、己姓、董姓、妘姓、曹姓等，黄帝之曾孙帝喾为高辛氏，其后为子姓；帝喾之子尧为陶唐氏，其后为祁姓；黄帝之裔舜为有虞氏，其后为妘姓、姚姓，颡顼之孙禹为夏后氏，其后为姁姓。所以，氏在周朝之前是部落名、国名的一种表示，与周朝的氏的含义有很大的区别。中国最古老的姓基本都出自炎黄两帝之后，据文献统计，先秦时期总共出现过30多个姓，加上考古出土的西周铜器铭文上所见的姓，总数也没有超过50个，而氏的数目却远远超过了此数。

夏朝时，氏的本意发生了变化，氏可以代表氏族中重要的分支，成为夏朝的诸侯。比如黄帝之孙颡顼帝之后的己姓，在夏朝分生出昆吾氏、苏氏、顾氏、温氏等氏族，并成为夏朝的诸侯国。夏朝时已经出现了分封制，禹为姁姓，其后分封有夏后氏、有扈氏、有男氏、斟寻氏、彤城氏、褒氏、费氏、杞氏、缙氏、辛氏、冥氏、斟氏、戈氏等诸侯国（司马迁《史记》），商朝基本上继承了夏朝的制度。

作为血缘关系的姓和氏之间发生最重要的变化是在西周初期，周武王灭商后，重新确立了姓、氏的关系：“天子建德，因生以赐姓，胙之土而命之氏。”（左丘明《左传》），姓明婚姻，世代不变，氏辨贵贱，随时更移。西周“裂土分封”，同姓、异姓诸侯国1300余，天子尊为大宗，诸侯称为小宗。西周的宗法制规定，诸侯国君的嫡长子继承国位和爵位，一姓之祖由嫡长子来祭祀和继承，虽经百代而不变更其姓，此即所谓大宗，国君嫡长子以外的儿子统称为庶子，也称别子，史称公子，他们不能继承君位，凡获得分封的另立一家，获得新氏，往往

以封邑为氏，称为小宗。别子的嫡长子继承新氏族，为新氏族的大宗，而其他儿子须再立新的支氏族 往往封于乡、亭 以乡、亭为新支氏族。没有土地可以分封的儿子 或以先祖的号、谥、爵、官、名、字、居、事、职、次等为新氏族 五代以后降为庶民，无氏可立；国亡后的国君嫡系一族往往以国名为氏，或以所继承的姓为氏。氏是特权和地位的标志，大权有大氏，小权有小氏。春秋战国时产生了无数个氏，也衰亡了无数个氏，而姓是不能改变的。所以，先秦时保留下来的姓仅仅在 30 个左右，而保留下来的氏却有近千个。我们平时提到的姓氏几乎都出自西周、春秋、战国时期的氏，与夏商以前的氏是有区别的。

直到秦始皇灭六国统一了中国，姓、氏才混合为一，汉朝以后姓和氏不再给予严格的区分（马雍，1985 年），促使姓、氏合一的根本原因是秦始皇统一天下之后 所实施的“郡县制”取代了周朝的“裂土分封制”。秦朝的郡县制既没有国、邑、乡、亭、里五等世袭封土，也没有公、侯、伯、子、男五等爵位，皇亲国戚、功勋武将也没有自己的封土。宗法分封制度的灭亡，使原先代表贵贱的“氏”与区别婚姻的“姓”之间没有本质上的区别，氏与姓都只剩下标记血缘系统符号的作用。

有关中国姓氏有史可查的既系统又可靠的记载，主要为秦汉以后的文献。正是在秦汉时期，才奠定了中国汉族姓氏世代稳定传递的基础。目前正在使用的汉族姓氏大约 3 000 余个，但大约 87% 的汉族集中在 100 个大姓上，其中 97 个大姓源自春秋战国时期形成的氏，仅有任、姜、姚三姓出自炎黄时代的古姓（袁义达，1996 年），有一点可以肯定，姓虽然起源于母系氏族时期，但中国 5 000 年姓和氏的传说和有限的文献记载表明：炎黄以来的姓和氏基本上都为父系传递下来的，这是中国汉族姓氏的遗传学的真谛。

## 1.2 同姓人群的迁移与融合

公元前 21 世纪，在黄河中游的河洛流域崛起了夏民族，这是汉民族形成的开始时期。以发源于黄河流域的炎黄部落的姜姓和姬姓为主体，融合了东夷部落的嬴姓和偃姓，形成了强大的华夏民族，以中原为中心，先后建立了夏、商、周三朝，长达 2 000 多年之久。中国的姓氏经历了严格的宗法制度时期，到秦始皇统一了中国 实施的“郡县制”取代了周朝的“分封制”，建立了户籍制度，打破了姓与氏的固定模式，姓和氏不再是贵族和特权的专用字，普天之下

的秦人 不分贵贱 都可以有姓氏 姓和氏混合为一。汉承秦制 华夏民族在这一伟大的转化中，由秦人变为汉人，逐步形成了汉民族，汉民族所拥有的姓氏就是汉姓。

西汉的建立标志了汉民族的形成，汉民族是以华夏民族为主体，融合周边的民族进化而成。汉族姓氏以先秦时期形成的姓和氏为核心，吸收周边民族的姓氏，形成了同姓不同郡望的汉族新的姓氏传递和分布规律，这不同于先秦时期的华夏姓氏。这样，随着外族的不断融入，汉族姓氏进化的结果是著名姓氏的同姓人群的规模越来越大，大姓往往拥有多个郡望，以区别不同祖先所形成的同姓群体；姓氏的种类也随着总人口的扩大而增加，或随着总人口的缩小而减少。汉族人口从西汉初的 3 000 多万，发展到明朝末的 1 亿 再发展到清朝道光年间的四亿，最终发展到今天的近 12 亿；汉族姓氏种类的数目由西汉初的 1 800 多种，增加到明末的 2 200 多种，再增加到清朝道光年间的 2 800 多种 最终增加到今天的 3 500 余种 从宋朝以来 李、王、张一直为我国人口最多的三大姓 三大姓的同姓人群的总人口在宋、元、明、当代四个时期占全国总人口的比例分别为 21%、17%、18%、22%。

有关中国姓氏有史可查的记录，主要是秦汉以后的文献。正是在秦汉时期，才奠定了中国姓氏延绵不断地稳定传递的基础。姓氏的真正含义是反映父系的血统。在这几千年里，大规模的人群迁移、民族融合和同化的进程与汉族姓氏的进化是同步的，当今汉族姓氏在全国分布特征是这几千年漫长历史进程的结果。自汉朝以来，汉民族与周边民族发生大规模融合至少有五次，汉姓也发生了五次重大的变革和发展，所经历的时代是汉朝、两晋南北朝、唐宋元朝、清朝等。

公元前 202 年汉高祖刘邦在河南荥阳即皇帝位 初建都于洛阳 不久迁陕西长安，汉朝宣告成立，史称西汉。到公元 220 年曹丕称魏帝 汉献帝让位 东汉灭亡 两汉历时 422 年。这时期的汉族所融合的周边民族的数目和人数是历史上最多的一次，是汉族第一次大融合时期。融入汉族的外民族包括北方的匈奴、鲜卑、乌桓、乌孙 东北的高句丽、扶余、濊貊、沃沮、三韩 东南的东越、闽越、山越、西瓯 南部的南越、骆越、外越、南蛮、交趾 西北的大月氏、羌、西域诸国 西南的叟、僰、冉駹、邛、蜀、巴等。

这一时期，汉朝是世界上最强大的帝国，周边的民族和国家无不向往大汉帝国，与汉人通婚和杂居，并很快被汉化，成为汉族。汉民族能够长期占据统

治地位，不间断地同化融入的其他民族，不断地加强汉民族的凝聚力，这与汉民族继承华夏族的一贯的民族协调、民族和亲的政策和推行汉化政策有关。民族间的通婚，上可追溯到4000多年前的炎黄两大部落间的世代通婚的时代，这一习俗一直延续到周朝，炎帝族的后裔姜姓与黄帝族的后裔姬姓仍世为婚姻。其他各国贵族之间的和亲通婚已经成为传统的习惯，比如姬姓鲁国与子姓宋国、嬴姓秦国与姬姓晋国的世代通婚，姬姓晋国的公族与戎族、狄族的通婚。到了春秋战国时期，华夏族的各国与周边的西戎、北狄、东夷、南蛮之间都存在这类通婚的现象和习俗。

秦始皇统一了中国，统一了文字，统一了货币，制定了统一的度量衡，废除分封制等，有利于国家统一政策的推行。汉承秦制，不但加强了民族和亲的政策，更重要的是推行了汉化政策，规定讲汉语，统一用汉字。使不同地区，特别是南北之间讲不同方言的汉族统一在汉文之下，汉字的文化认同对于汉民族的凝聚力、中华民族的凝聚力和国家的大统一起到了至关重要的作用。“没有汉字，就没有汉民族”，也就没有中国。以至于到今天，只要是中国人，不管在世界何地，讲何种中国方言，或讲何种外国语言，一看到中国字，特别是中国的姓，就会产生一种无形的亲切感和认同感。

两晋南北朝是汉族第二次大融合时期，西晋末，匈奴、羯、氐、羌、鲜卑的“五胡乱华”，在北方先后建立了十六国。南北朝时，北方几乎为鲜卑族所统治，北魏孝文帝下令全部鲜卑族所统领的各姓部落一律改称汉姓，北方地区的汉族一下子就融入了大量的外族成员。除此之外加入的民族还有北方的高丽、扶余、乌丸、柔然，西域的昭武九姓、龟兹、于阗、疏勒、车师、吐火罗、破落那，西南的天竺、叟，南方的巴、賈、獠、狸、蛮、溪、濮等族。这一时期，主要的变化是北方地区的中原汉族大量南迁江南地区，成为当地的大姓，当代中国南方著名大姓陈、林、黄等姓就是在这一时期南迁的。北方地区涌入了大量的鲜卑等族人，与当地汉人混居，形成新的华北地区的汉族和产生了一批新的汉姓。

汉族第三次大融合时期是在唐宋，唐朝推行的是民族平等政策，外族在朝中可以担任要职，外族与汉族通婚十分普遍，从而加速了外族汉化的速度。这个时期融入的外族有北方的铁勒、突厥、回纥、薛延陀、黠戛斯，西部的吐蕃、吐谷浑、羌、西域诸国，东北的高丽、百济、新罗、渤海、靺鞨、沃沮、契丹、奚、霫、鲜卑，南方的安南、占城、南蛮、乌蛮、白蛮等族。这一时期，中原百姓几次南下大迁移，聚集于赣浙苏，开发了闽粤地区，形成了著名的汉族一个分支——

客家人。中国的人口重心开始由北方向南移动，到北宋后期重心已经越过长江进入南方。从此以后，汉族姓氏在南方得到飞速的发展，形成了许多著名的大姓望族。

元人从贝加尔湖附近的蒙古东部兴起，蒙古骑兵六次南下，最终灭了南宋，成为中华大地的统治者，建立了元朝，这是汉族第一次被外族排除在统治地位之外。以蒙古贵族为主的元朝政权残酷地奴役剥削着汉人，汉族遭到了自炎黄以来从未有过的虐待和屠杀，人口急剧地减少。这是野蛮欺压文明，落后统治先进，愚昧颠倒了真理的反历史的大气候时期。

南宋宁宗时全国人口（包括北方地区）已达 1 亿，而元朝初期的全国人口仅剩 5 500 余万，汉人几乎死亡一半。后来元朝统治者逐渐恢复生产，人口逐渐地增加，到元顺帝时为元朝的人口顶峰时期，中国人（包括国土内的蒙古族等其他民族）的人口才达到 8 700 万（赵文林和谢淑君《中国人口史》，1988 年）当时汉族的张、王、李、赵、刘五大姓约占全国人口的四分之一，有人曾一再向元顺帝建议杀灭汉人五大姓，有利于元朝蒙古贵族的统治，幸亏元顺帝没有允准，汉族最终逃过了一次灭族的大劫难（巫声惠《中华姓氏大典》，2000 年）。元末天下再一次大乱，群雄起义，朱元璋统一了全国，建立了大明皇朝，在不到 100 年的时间内，中国的政权再一次回到了汉族人的手中。明初的人口再一次急剧减少到 6 000 万，汉姓的数目达到 2 000 个左右。中国姓氏在这次大的洗劫中又得到了一次大变革，吸纳了来自西亚的阿拉伯人、波斯人、突厥人、钦察人、康里人、畏吾人等新鲜血液，同时，蒙古人把西夏、高丽、契丹、女真、蒙古、吐蕃、安南、大理等外族人分散到全国各地，与汉人杂居，最后大部分外国人和外族融入到汉族之中，这是汉民族的第四次大融合。

清朝是汉族第五次大融合时期。清朝虽然再一次把汉族排除出统治地位，但很快采取民族和睦政策，发展农业，人们能够安居乐业。在其长达 268 年的统治时期，只有太平天国起义是一次大的全国性的战争，除此之外，所有的战争皆为局部的，规模有限的。在这一时期，比历史上任何一个朝代中所发生的战争的程度要小，危害也小。全国人口从清初的 8 000 多万，到清道光年间人口达到 4 亿，这是清朝人口发展的一大特点。清朝期间，北方地区的满蒙以及东北地区的其他满洲八旗、蒙古八旗纷纷改用汉姓，西部和西南地区的少数民族也在清政府的“改土归流”的政策之下，纷纷采用汉姓，各地汉化的速度加快，中国姓氏再一次获得了扩充，姓氏数目达 2 800 余种。

1911年的辛亥革命推翻了清朝，成立了推行五族共和的民族团结的中华民国政府。汉族和其他民族一起正式称为“中华民族”。1949年成立了中华人民共和国，在中国国土上的56个民族，各族之间平等友好相处，互相自由通婚。汉字是全国的通用文字，55个少数民族都有人使用汉姓。目前正在使用的汉姓有3500个左右。

中国同姓人群的迁移和融合的历史有力地证明：汉民族同化异族的能力是不能低估的，汉文化所起到的中华民族凝聚力的作用也是不可低估的。

### 1.3 姓氏传递的延续性和稳定性

中国人一般都习惯继承父亲的姓，以父系方式把姓氏传递给下一代。女子在一生内仅仅保留其父亲的姓氏，不传递给下一代。因此，绝大多数的姓氏属于一种无性别之分、以父系方式传递的“基因”，相当于位于代表人类男性染色体上的特殊遗传基因。可以这样假设，不管是X精子还是Y精子，均携有姓氏基因，均在第一代显示其姓氏的功能，惟有Y精子具有连续传递姓氏特征的性能，而X精子仅仅在第一代显示姓氏的特征而不能传递。当然，我们也注意到在中国人中一直普遍存在着改姓的现象，这包括入赘婚姻、随母姓氏、避难改姓等情况，但是在全体中国人中改姓的人数是很少的，在人群中所占的比例也很低，并不影响原有姓氏在人群中的分布。而且，不管是在什么情况下改的姓，通常从第二代起仍以父系方式传递。这种改姓现象称为姓氏的突变，突变后的姓氏仍具有正常的父系遗传的功能。中国人姓氏的悠久历史和相对稳定的传递，以及具有中国社会进化特色的改姓现象，不但增加了中国人姓氏的多样性，而且也证明了中国的任何一个大姓的组成不会是单一的，这为追踪和寻觅各个姓氏的始祖和形成的年代提供了丰富的历史依据。所以，中国人姓氏是一种含有人类父系社会全部进化信息的、赋有相对稳定传递性质的、具有中国特色的资源。

本节用自然科学的分析方法说明中国人姓氏的传递稳定性，所用的宋朝、元朝、明朝和当代的姓氏数据均来自作者的实验室。详细的样本抽样情况在以下章节中介绍。图1.3(见彩图)是宋、元、明、今四个时期的100个常见姓氏的分布曲线，中国人姓氏的分布散点存在一条符合公式 $Y = 0.185 \ln((X + 1)/X)$ (公式1.3)的分布曲线(袁义达等, 2000年b)其相关系数 $r$ 为0.99,

四条分布曲线互相吻合十分一致。说明中国人姓氏在这 1 000 年中传递的连续性和稳定性。公式中： $X$  代表大姓按人口多少排列的次序，即 1,2,3……99,100 正整数， $\ln$  表示取自然对数， $Y$  代表在总人口中排在第  $X$  位的姓氏所占人口的期望值，用百分比表示。

图 1.3(见彩图)中的  $Y$  轴上的值为各时期的姓氏在人群中的分布比率。常见姓氏在不同时期的排列次序存在一定程度上的变化，次序排列的先后不是主要的，但其在人群中的分布比率值是重要的。中国人姓氏的变化主要发生在占人口 1% 以上的 19 个最常见姓氏之中，三个时期的 19 个最常见姓氏的合计人口均占了当时全国总人口的 50% 以上。这 19 个常见姓氏是宋朝的王、李、张、赵、刘、陈、杨、吴、黄、朱、孙、郭、胡、吕、高、宋、徐、程、林等，约占了全国人口的 54%；元朝的王、张、李、刘、陈、赵、吴、杨、黄、周、徐、朱、郑、胡、高、郭、林、程、马等，约占了全国人口的 49%；明朝的王、张、李、陈、刘、杨、吴、黄、周、徐、朱、赵、胡、孙、郑、何、郭、沈、高等，约占了全国人口的 51%；当代的李、王、张、刘、陈、杨、赵、黄、周、吴、徐、孙、胡、朱、高、林、何、郭、马等，约占了全国人口的 56%。

最明显的变化是在最常见的 7 大姓氏之中，见箭头 A 所指的。元朝和明朝的 7 大姓氏的占人口的比率比宋朝和当代均低 1% 以上。反映了元明时期中国人口曾经大幅度降低过。历史上人口数据表明，由于连年战争和屠杀，元朝的北方和四川地区的人口急剧减少了，而浙江、江西、湖南和湖北一带的人口却相对有所增加。从姓氏分布上也反映了这一现象，北方地区的主要姓氏王、李、张、刘和杨的总比率，元朝和明朝时期的五大姓的人口均比宋朝和当代的都低 1 个百分点以上，尤其赵姓的比率降低了 3—4 个百分点。赵姓为宋朝的皇室之姓，在元朝自然受到更大的冲击。而东南地区的第一大姓陈姓的比率反而比宋朝高 0.7%，与当代的比率接近，成为元朝的第五大姓和明朝的第四大姓。明朝与当代之间的清朝，政局比较稳定，经济发展，玉米和红薯等高产作物广泛种植，人口增加迅速，尤其促使了北方地区人口的增加和部分“南方人口”的北归。至清末民初，全国人口已达 4 亿，南北人口比重趋于均衡。因此大姓人口迅速增加，以至恢复或超过了宋朝时期的相应姓氏的比率。

宋朝、元朝、明朝和当代四个历史时期的姓氏分布反映了一个重要的现象：中国人姓氏在历史上传递是连续和稳定的。它说明了在历史的进程中，中国人姓氏所表现的血缘文化的痕迹与生命遗传物质，尤其代表男性的 Y 染色体的进化具有基本相同的和平行的表现。

## 1.4 同姓率和同姓人群

姓氏群体遗传学的目的，是研究姓氏群体中姓氏种类组成的变化机制。因此，首先面临的问题是如何表达群体的姓氏组成，并使其形成具有既可分析的、又可与真正的遗传基因进行比较的形式。比如，提到河南地区的姓氏群体组成就是一个很笼统的概念，这样既不能与别的进行比较，又无法进行分析。在一个群体中包含了许多种的姓氏，因此，需要弄清这些不同姓氏在群体中所占的比重或称比例、比率、频率、百分比等，然后综合地考虑整个群体的姓氏组成。

研究姓氏群体遗传的首要的概念是同姓率，在一个群体内随机相遇的两个人都姓同一个姓氏的概率称为同姓率。群体内部和群体之间的同姓率，是分析群体的遗传结构、近亲系数和群体间遗传距离的重要参数。计算同姓率的公式为(Crow, et al. 1965)：

$$I_{pq} = \sum_{i=1}^k p_i q_i \quad (\text{公式 1.4})$$

公式中  $I_{pq}$  代表  $p$  群体和  $q$  群体之间的同姓率， $p_i$  和  $q_i$  分别代表  $p$  和  $q$  群体的第  $i$  个姓氏在各自群体中的所占人口的比率， $k$  表示两个群体总共拥有的姓氏数目， $\sum p_i q_i$  表示  $p_i$  和  $q_i$  乘积的全部和。中国人不同历史时期的同姓率分别为：宋朝 0.0264，元朝 0.0203，明朝 0.0213，当代 0.0276。

图 1.4(见彩图)表示四个历史时期的 100 个常见姓氏分布累计值(图中位于下方曲线)以及累计同姓率比值(图中位于上方曲线)有关四个时期的 100 个常见姓氏分布的详细情况见第二章。

图 1.4 首先反映了四个历史时期的常见姓氏分布累计值和累计同姓率比值的曲线吻合一致。其中占人口 1% 以上的 19 个最常见姓氏累计人口在不同历史时期分别为：宋朝占 53.7%，元朝占 49.5%，明朝占 50.7%，当代占 55.6%；其累计同姓率比值已分别达到宋朝 93.4%，元朝 90.3%，明朝 91.4%，当代 94.4%。而 100 个常见姓氏的累计人口分别为：宋朝占 86.2%，元朝占 84.3%，明朝占 85.1%，当代占 87.2%。其累计同姓率比值已分别快速达到宋朝 99.6%，元朝 99.3%，明朝 99.4%，当代 99.7%。说明大

约 85% 的中国人集中于 100 个常见姓氏内,100 个常见姓氏仅仅占中国总姓氏数不足 5%,而随机同姓率的比值已高达 99%。这就是说:在大街上随便找两个同姓的人在 100 次中有 99 次出现的姓氏必定属于 100 个常见姓氏,最多只有一次为其他的姓氏。所以中国常见姓氏人口的分布是研究中国人遗传结构和进化以及地域人群间的亲缘关系的主要因素。

但是,我们也注意到绝大多数姓氏的人口相对很少,但往往显示出相对隔离的地域人群的特点。因此,常见姓氏和非常见姓氏的同姓群体分布具有各自的特点和含义。中国人历来有同姓聚居和联宗修谱的习俗,而且婚姻半径较小,婚娶地域相对固定,这样易形成同姓人群的分布。中国人姓氏的分布实际上主要反映了同姓人群的分布规律。这又反映了另一个重要的现象:中国人的姓氏或同姓人群存在两种状态,大姓和小姓,或称为常见姓氏和非常见姓氏。仅占总姓氏量不足 5% 的 100 个常见姓氏已集中了 85% 以上的人口而占总姓氏量 95% 以上的 3 000 多个非常见姓氏仅代表不足 15% 的人口。

100 个常见姓氏的分布是反映各地人群的遗传组成的主要因素,它们决定着中国历史上人口迁移和地域人群间的亲缘关系的程度。而非常见姓氏人群更表现其地域特色和相对隔离的现象,保存着中华民族祖先遗传下来的各种各样的遗传资源,这是中华民族目前所拥有的最精华的民族遗产。我国是世界上人口第一的大国,拥有世界上最丰富的疾病人群资源,这包括疾病群体、家系和个体。家系就是同姓人群,特别是稀有姓氏的人群,稀有姓氏往往是单一祖先的传递。由于我国疾病谱与国外人群的疾病谱在群体遗传学上存在着差异,并占有遗传资源上的优势,特别是中国几千年姓氏制度所造成的世界上独一无二的同姓人群分布的特点,许多外国研究机构,特别是美国和日本都想进入我国这一研究领域获取宝贵的家系同姓人群和少数民族血液资源。所以,目前科技界一直在强调保护中国境内的生物多样性,保护中华民族基因组的资源,其原委和深远的战略意义就在于此。

## 1.5 姓氏遗传的南北差异

当代拥有近 12 亿人口的汉族,是世界上最大的族群。汉族在其形成的漫长历史中向四处不断的迁移与周边民族的交往外族基因不断地流入不同地区文化的互动,地理环境的影响和隔离,方言林立且差异悬殊,经济发展的

不均衡，形成了许多各具社会文化特色的汉族亚群体。

我国汉族的亚群体确实实是存在的，但从来没有一个统一的划分依据和标准，而是各学科根据各自的需要制定自己的标准来划分的。官方的划分中最著名的为语言学的划分，语言学家根据方言把汉族分为北方、湘、赣、吴、闽、粤和客家七大方言区人群，每一方言区再被细分若干亚方言区。而其他学科因没有相应的标准，往往以长江为界，把汉族简单地分为南方人和北方人。更多的情况下以省区、直辖市行政区划为单位来区分汉族亚群体，如河南人、山东人、上海人等。也有以大行政区划为单位结合人类文化学的特征来区分汉族亚群体，如华北人、华南人等。

民间存在着不正规的划分现象，主要表现以家谱或几姓联谱的形式存在，这在我国的东南和南方地区、港澳台地区和海外华侨中十分盛行。民间的家谱类的传统资料是研究中国汉族亚群体的重要材料，以美国犹他州国际家谱研究中心为代表的一批国际研究机构，长期以来一直十分重视收集中国人的家谱，一直结合各种学科的需要，研究着中国人家谱的历史、今天和将来。我国国家图书馆和上海图书馆等收藏了中国年代最古、质量最高的家谱，全国图书馆收藏的家谱种类达 2 万多种，总数大约有 20 万册，这还不包括民间收藏的和 1949 年以后重修和新修的家谱。这是一笔巨大的财富和具有持续研究价值的宝贵资源。

华夏民族在迁移和进化的过程中，北方地区主要交往和融合的是北方地区的少数民族，如匈奴、鲜卑、突厥、羌、蒙古等属北蒙古人种的民族，两广、闽、台地区主要交往和融合的是南方地区的少数民族，如南越、交趾等属南蒙古人种的民族。北方地理地形平坦，人群迁移容易，人群间的同化和语言统一要比南方地区快，南方丘陵起伏，地形复杂，气候恶劣，人群迁移困难，人群间的同化时间长，互相交往困难，造成地理上的隔离状态，语言变化很大，形成无数的“族群岛”。近 12 亿人口的汉族，分布于全国各地，虽同为汉族，南北之间方言杂乱，几乎到不能通话的地步；几千年的汉族进化历史也证明了南北地区的汉族存在差异，而且是一种群体遗传基因进化上的差异。因为文化上的差异，包括语言上的差异，政府通过政策法令，可以在不长的时间改变或缩小这类差异，这类差异不是根本性的；几千年的不同程度的隔离和遗传基因进化中所造成的不同地区汉族亚群体之间的差异才是本质上的区别。

图 1.5a(见彩图)是我们利用中国汉族全国各省的 ABO, MN, Rh, P 四个

遗传基因的 100 多万个人的数据，用主成分分析法所作地域人群血缘亲缘关系图 (Yuan Yida, et al. 1992)。这幅示意图非常清晰地说明了：南北汉族之间存在着遗传结构上的差异。南方汉族所代表的省份包括闽、台、粤、桂四省区，长江流域地区的省份为中间地区，但中间地区的汉族的遗传结构更接近于北方地区省份的汉族，而浙江省的汉族的遗传结构倾向于南方。坐标系中，横坐标代表第一主成分 (含 42% 的信息) 纵坐标代表第二主成分 (含 17% 的信息) 详细的方法请参看有关的文献。

中国汉族姓氏在历史上的传递是连续和稳定的，在历史的进程中所展现的血缘文化的痕迹与生命遗传物质，特别是代表男性的 Y 染色体的遗传具有基本相同和平行的表现。因此，用群体遗传学的分析方法可以分析汉族姓氏。我们对各省姓氏分布的数据进行生物遗传距离的分析，获得了宋朝、明朝和当代的汉族人群的三张聚类图，分别为图 1.5b、图 1.5c、图 1.5d。并发现汉族从宋朝开始已经形成了南北两大群体 (袁义达等, 1999 年)。

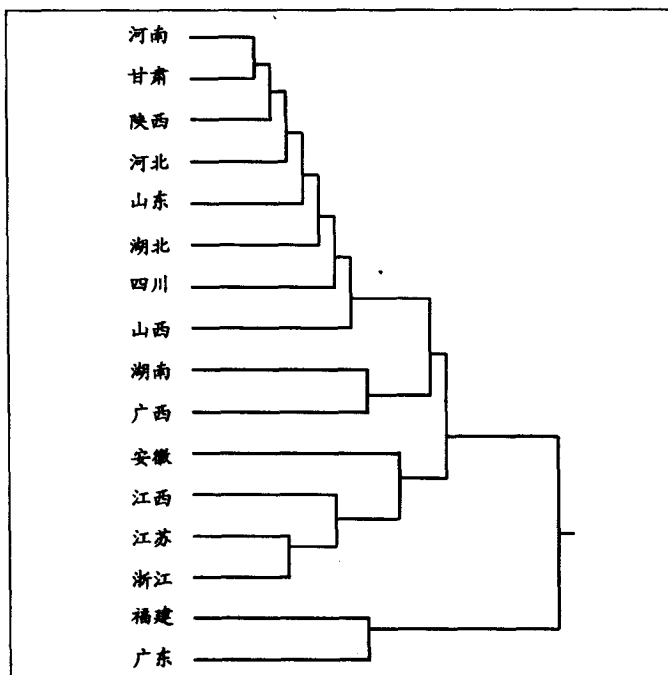


图 1.5b 宋朝姓氏分布的地域亲缘关系聚类图

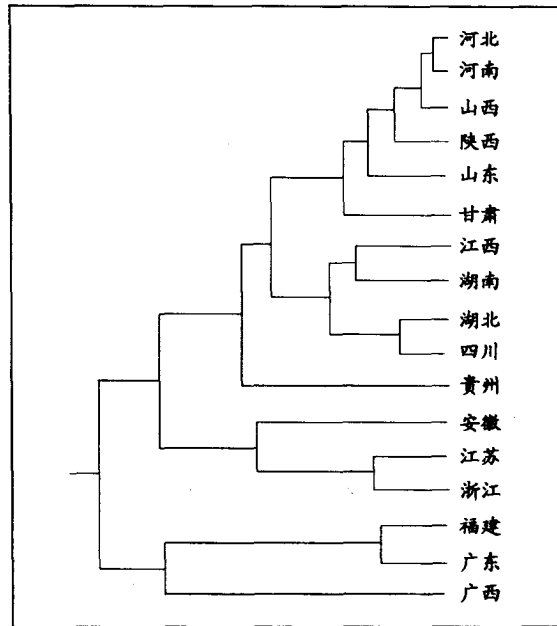


图 1.5c 明朝姓氏分布的地域亲缘关系聚类图

图 1.5b 十分明显地把宋朝 16 个省划分为两大区域：南部的广东和福建与北部的 14 个省。这两区域间的遗传距离十分大，说明了南北两大人群的异源性，其地理分界线应在中国南部福建省与江西省交界的武夷山和广东省与湖南省交界的南岭一带。北部 14 个省又明显的划分为两块，其中浙江、江苏、江西和安徽四省形成了东南地区人群，其聚类的顺序也很合理，反映了古吴越民族与中原地区南下人群混合的历史痕迹。剩下的 10 省进一步可划分成两块，其中湖南和广西二省组合在一起，在宋朝时该地区是一块很荒凉的“蛮”地，其余的 8 个省包括了全部北方省份和西南省份，其特点是，北方省区和西南省份的人群之间的姓氏遗传距离都较小，反映了该地区人群经历了长期频繁的迁移、混居和融合，形成了互相依存的北方人群。而西南省份的人群在历史上一直主要与北方省份的人群进行交流，四川地区（包括贵州和云南）的聚类位置也反映了这一历史。湖北是一个十分特殊的省份，它处于长江的中游，北面河南和陕西，西接四川，东临安徽和江西，南靠湖南，是中国历史上兵家必争之地。图 1.5b 也十分有趣地反映了湖北的地理位置和历史。总而言

之 宋朝姓氏频率分布的聚类图(图 1.5b)反映了在 1 000 年前 中国人群已经形成了南北二大人群,其基本的人群分布区域与现代的分布基本一致。

图 1.5c 非常清楚地把明朝中国人分为南北两大人群,形成福建、广东和广西三省为南方人群 其他 14 省为北方人群的局面。明朝的人群地理分布特征已与当代的分布基本一致,其南北两大异源汉族的地域分界线也应在武夷山和南岭一带,而不是在长江。与宋朝的地域人群分布相比较,广西已与湖南分离而与广东、福建聚类。长江中上游的江西、湖南、湖北和四川形成了一块更接近于北方的区域,而江苏、浙江和安徽形成长江下游区域,但仍与南方人群血缘较远 更接近北方人群。

图 1.5d 反映了当代中国汉族的南北两大人群的异源性,以及明显的分界地域在南部的武夷山和南岭一带。29 个省市区分为南北两大块,而且其遗传距离很大 在图上用虚点表示。南方汉族包括福建、台湾、广东、广西三省一区(包括海南省、香港、澳门),北方地区的东北三省、内蒙古和河北、北京和天津、以及山东和河南之间的遗传距离很近,都聚集在一起。随后,西北地区的陕西、甘肃、宁夏、山西、新疆聚在一起。安徽的姓氏分布表明其属于北方 但又接近长江流域的省份,所以安徽最后与北方的省份聚类。长江中上游的省份湖北、四川、湖南、贵州、江西、云南先后聚类在一起 形成一个亚群体 然后与北方的人群聚类 这是十分合理的。东部的浙江、江苏、上海聚类在一起 与南方的群体有一定的渊源,但几千年的进化使这一地区的人群更接近北方,所以最终与北方地区聚类。青海最后与北方聚类在一起,而不是先与陕甘宁聚类,这可能与青海的姓氏样本有关,青海的抽样点为西宁,西宁为解放后东部移民大西北的主要地区,所得到的样本可能为原先的上海和江苏地区的移民。

总而言之,宋、明、今三个时期的姓氏遗传距离分析结果表明,在中国 1 000 年前的宋朝已经形成了南北两大区域的人群,从姓氏遗传信息上明显地显示出南北两大区域人群结构的区别,表明南北两大区域人群的不同源性,其明显的地域分界线在南方的武夷山和南岭,而不是长江。中国人姓氏的历史至少已有 4 000 年以上 主要起源于中国西北和中原地区的民族 融合并同化了各个时期周边各民族的姓氏,形成了中国几千年来一直在使用的汉字姓氏。根据 1 000 年来的中国人姓氏分布和对人群迁移、混居和融合情况的分析 基本上反映了当时的民族联合和进化的历史,也说明了中原地区的人群是组成中华民族的主体,它融合了南北各个时期的民族成分,尤其是北方地区的民族

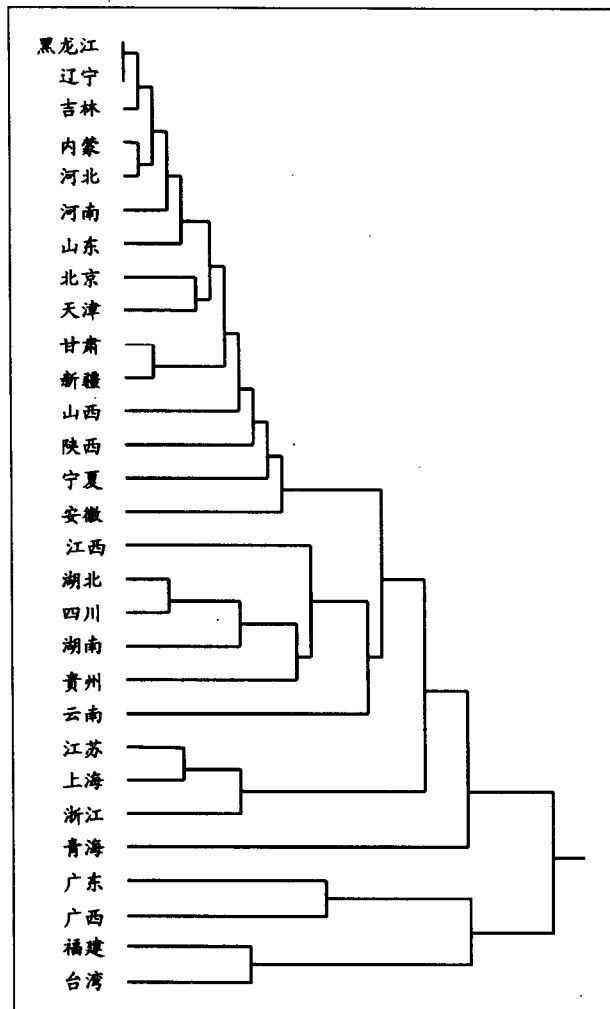


图 1.5d 当代姓氏分布的地域亲缘关系聚类图

成分，形成了互相依存不可分割的当代的中华民族的整体。当代人群分布的情况基本上与明朝宋朝人群分布具有很高相似性和渐变的过程，它们之间的区别点正是反映了这 1 000 年期间人群进一步迁移与分化的结果。