

百越遗传结构的一元二分迹象

李 辉

【提 要】最新的分子人类学材料——Y 染色体 DNA 能对民族系统进行精细的分析。百越系统被发现与中国的其他系统差异很大，而与南岛语系民族（马来系统）特别是台湾语族群体相当接近。已经研究的百越群体显示出遗传发生关系和语言文化类型的差距，这与百越的整体认同和地域分化有关。根据数据的主成分分析得到百越民族系统遗传结构的三个特点。1. 百越有单起源的遗传学迹象，可能大约三、四万年前发源于广东一带，而后慢慢扩散开来；2. 百越二分为以浙江为中心的东越和以版纳为中心的西越；3. 百越群体在发展过程中曾经由广东向东北、西北、西南三个方向迁徙。百越接触过的许多族群也涵入了部分百越的遗传类型。百越从广东经江西向浙江的扩散可能发生于很早以前，福建与浙江的越族群体是不同路线迁来的，在台湾原住民遗传结构中同样表现出这种二元性。西部傣族族群是很晚从广东迁出的。

【关键词】百越 遗传结构 主成分分析 单起源迹象。

【作 者】李辉，复旦大学现代人类学研究中心博士。上海，200433

【中图分类号】Q988 **【文献标识码】**A **【文章编号】**1004 - 454X (2002) 04 - 0026 - 006

一、遗传研究的百越范围

民族学在界定民族系统的范围时，依据的是民族系统的独有特征。长期以来，对民族进行的遗传学研究也期望找到这样的特征，但是从血型等早期的民族遗传指标中找到系统特征无异于天方夜谭。近十年来人类基因组研究的兴起才使民族遗传研究有了希望。线粒体 DNA 最先展现了全人类的系统结构，但是由于它构建的群体系统稳定性较差而遭到了许多骂名。对于精细的民族研究，现有的线粒体标记更是无能为力。随后研究热点转到了 Y 染色体 DNA 上。这是一种纯父系遗传、信息量大、稳定、特异的绝佳遗传标记，它不但构建了稳定的发生于非洲的全人类谱系树，在民族之间更是体现出不同层次的差异。

我们对亚太地区大量群体的 Y 染色体调查研究画出了图 1 的民族系统亲缘树。

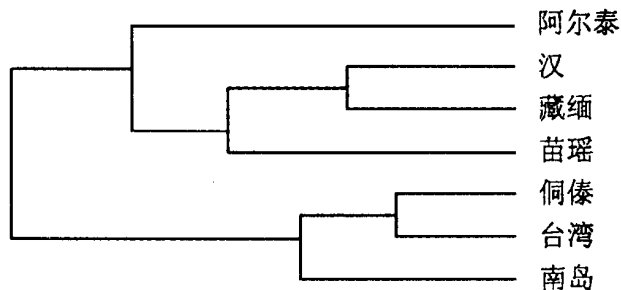


图 1 Y 染色体构建的东亚民族系统亲缘树

现代百越民族的主体——侗傣民族，显然很明显的迥异于国内其他民族系统。而有趣的是，南岛语系与之相当接近，其中的台湾语族更为接近。这可能提示了百越民族与这两个族群很近的同源关系。

当然，遗传学上界定的百越范围与语言学和民族学不一定完全相同。血统上源于百越而语言文化已然改变的群体也完全应当纳入百越研究的范围。东部的越族群体多属于此。台湾语族也可能是这一类。另一方面，语言文化上被百越同化的外源民族却不应属于百越遗传研究对象。比如湘桂黔交界的草苗是讲侗语的群体，但是遗传上明显的汉父苗母，侗族的成分很少，不应看作百越血统的群体。因此，遗传研究的百越范围至少应包括侗傣语系的僮傣语族、侗水语族、仡央语族、黎语族，还应包括东部地区的越裔群体，可能还有台湾几支原住民。从印度阿萨姆到台湾几千里地域内有大约 100 个民族群体。

二、百越的遗传结构

经过笔者及合作者近年来的收集，大部分的百越群体的遗传材料已经采集到了，许多群体的遗传数据已经实验得出，包括印度的阿霍姆，泰国的兰那、东北泰和石族，云南的傣泐、傣纳，广西的壮族、毛南族、仡佬族、仡佬族、夜郎、水族、侗族、拉伽，贵州的布依族和水族，海南的黎族和仡隆，上海古今的马桥人，浙江的於越，台湾的泰雅、布农、阿美、排湾、曹、赛夏、卑南、鲁凯、邵、巴则海。

这些群体的 Y 染色体遗传结构体现出相当大的一致性，而与其他系统中研究过的群体完全不同。他们都有大量的 M119、M110 或 M95、M88 突变，而外族极少有这些遗传标记，与百越接触少的群体则没有。

另一方面，这些群体彼此间的遗传关系则体现出与语言学和民族学分类不同的格局。特别是侗水和仡央两支，完全打破了文化的分类而重新组合。图 2 中，我们用线条表示群体间遗传关系，用虚线圈表示文化的分类。

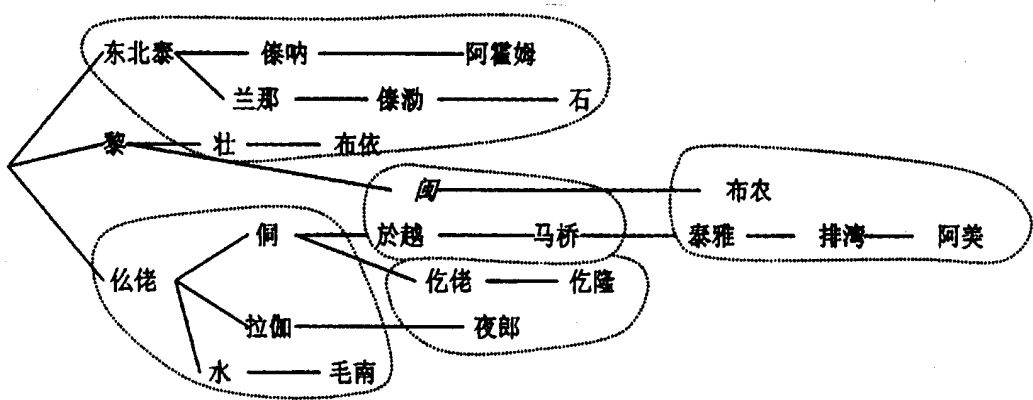


图 2 百越各群体遗传关系和文化分类

这种系统发生和文化类型之间的差异是可以理解的。整个百越民族早期是个内部认同很强的人群，所以史书上百越分布于会稽到交趾的记载也证明古人很容易把这个民族群体整体化地看待。其内部分支间的文化交流早期是没有障碍的。所以较早迁到同一地域的不同分支会在文化上一定程度的同化。这种同化有时不完全，也会保留一些发生历程的痕迹，比如夜郎和拉伽之间就有其他分支没有的词汇。

当然随着对更多群体更多材料的深入研究，这个系统结构也会调整扩充。

三、百越的地理分化

根据现有的百越民族群体 Y 染色体数据, 我们应用主成分分析的数理统计方法, 把数据中的主要趋势信息抽提出来, 得到了三个主成分 (趋势)。按各个群体的对应值把三个主成分按等高线绘制原理作成三张地图。这三张地图体现了百越遗传结构中的三个主要特点。

第一主成分占到信息总量的 47.0%, 从图 3 中单一中心、梯度平缓的分布格局明显看出, 所有的百越群体首先是有整体性的, 共性是最主要的。因此百越的血统只有一个主要来源。图中的分布中心在广东一带, 所以广东最有可能是百越民族血统最早的发源地, 而后渐渐向四周扩散。

第二主成分占到信息总量的 35.6%, 从图 4 中可以看到两个分化方向, 一个在浙江一带, 另一个在版纳一带。因此, 百越民族在一元性之后的另一主要特征是二分性, 即分化为东越和西越。但是, 特别要补充说明的是, 西部的侗族、仡佬族归入了东越类, 而福建归入西越类。这背后可能还有故事。

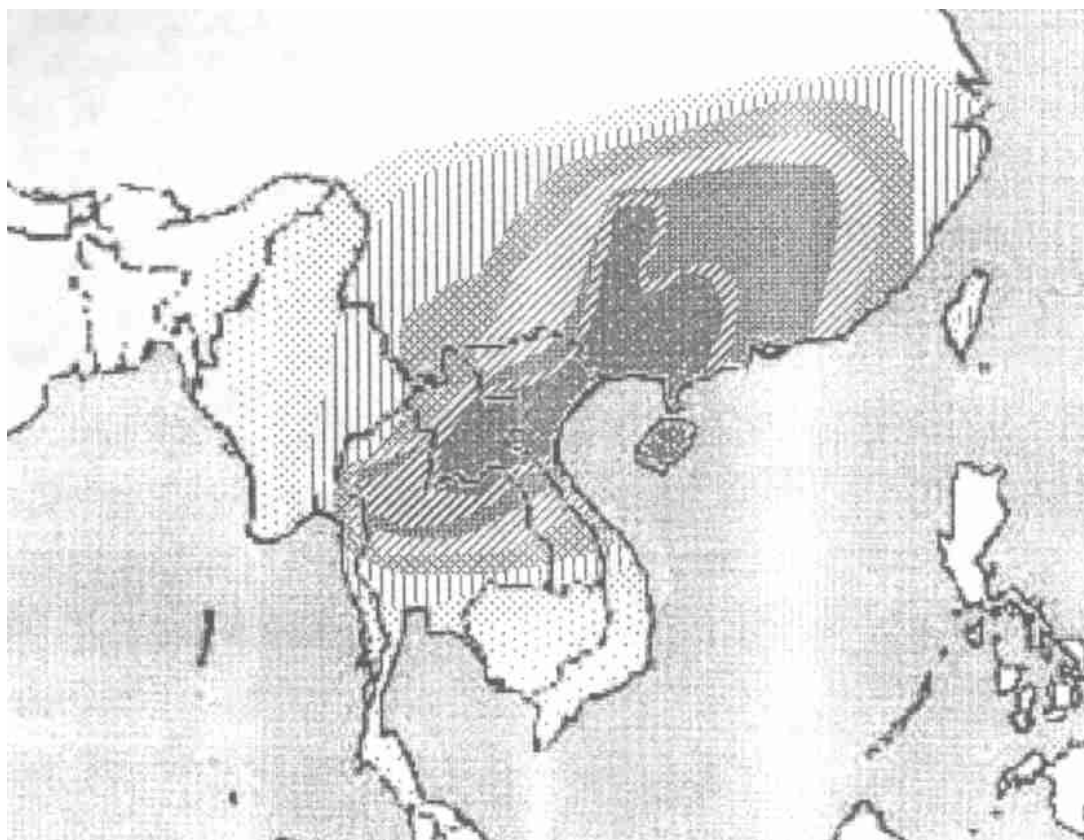


图 3 百越遗传结构第一主成分的地理分布

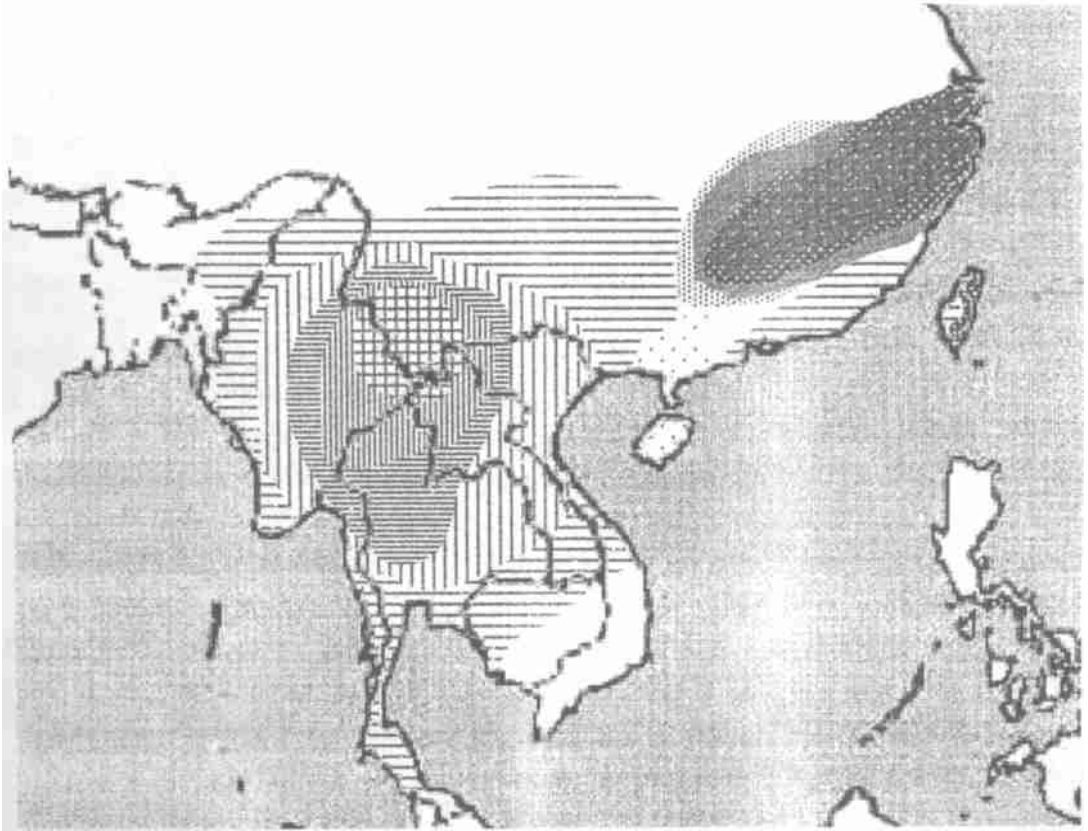


图 4 百越遗传结构第二主成分的地理分布

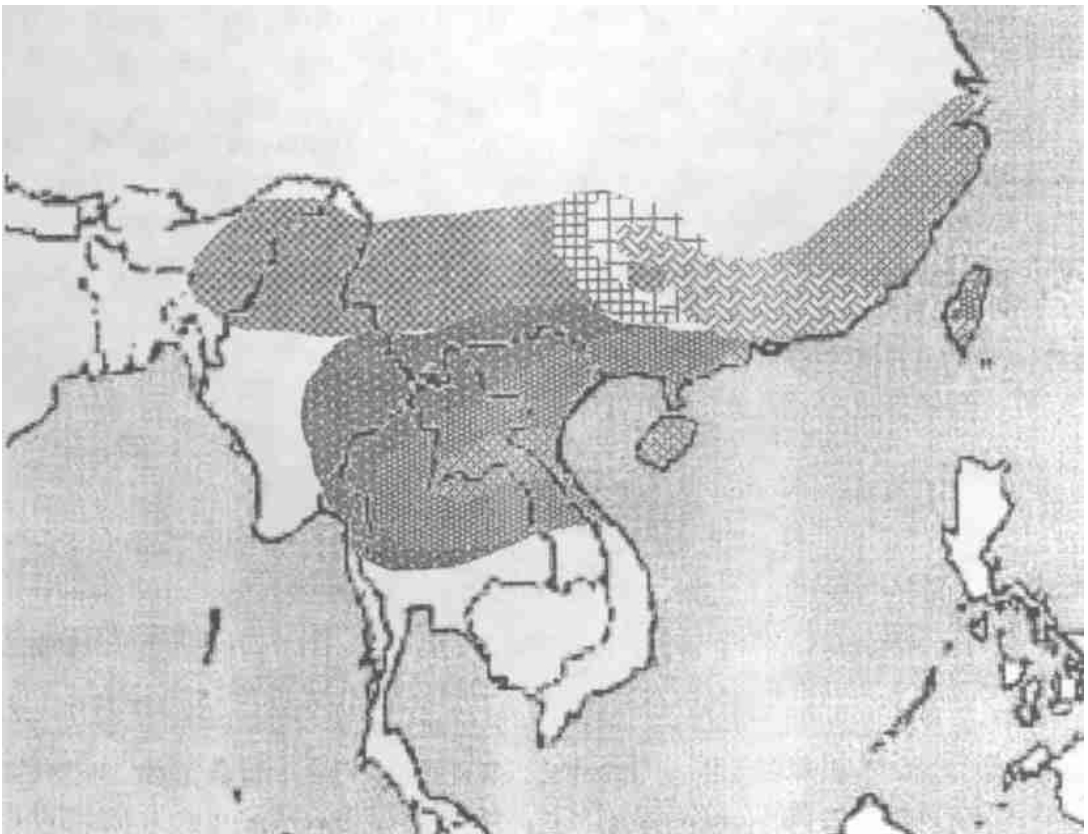


图 5 百越遗传结构第三主成分的地理分布

第三主成分占信息总量的 17.4%。图 5 中,从广东向西南、西北、东北三个方向延伸出三个不同的结构。这张图体现的是与第一主成分的扩散不同的百越民族迁徙的过程。广东等地受到汉族移民冲击之后,百越原住民开始搬迁,在别处重新定居。

所以,百越的遗传结构特点至少可以归纳为“一元二分”,至于“三角”方向的迁徙,还有待于补充论证。

四、受百越影响的族群

百越长期以来作为南方的主体民族,曾经与许多外系统民族接触交流,也发生了一定的基因交流。浙闽地区曾是百越的主要政治中心,所以这一地区现代的汉族群体中保留了相当比例的越族特征遗传标记。以 M119 突变比例为例,上海和浙江都是 26%左右,是全国汉族中最高的,其次是广东 20%、安徽 18%、辽宁和湖北 17%、江苏 16%、江西 14%、湖南 13%、河南 11%、山东 9%、四川 7%、云南 4%,实际上是以浙江为中心沿海岸线扩散。值得一提的是, M119 突变类型中的 M110 突变亚类型,更具东越特色,但是在一般汉族群体中没有发现,在上海马桥原住民中有相当比例。福建的类型不一样,有 8%的西越特有的 M88 突变,广东有 7%。

汉族吸收越族的血统的现象,符合强势民族吸收外族血统的理论。这在其他民族的遗传结构中也有表现,特别是与百越关系最密切的百濮民族(南亚语系)。这些民族中较强势的民族含有百越特征的成分,而弱势民族则没有。比如,京族是越南的主体民族,当然是相对强势的,其中含有 7%的 M119 突变;高棉族是柬埔寨的主体民族,其中含有 4%的 M119 突变、4%的 M110 突变和 12%的 M88 突变。而南亚系统中的较弱势的芒族、布朗族、佤族中则还没有发现过百越特征的成分,弱势群体向百越的基因流动可能是单向的。

苗瑶族群受百越的遗传影响现在还很不清楚。氏羌族群中也有几个地区性的强势民族吸收了百越的成分。最典型的是土家族,其中含有 20%的 M119 突变和 10%的 M88 突变,可能来自侗族、仡佬族和布依族。此外,傈僳族含有 20%的 M88 突变,拉祜族含有 8%的 M88 突变,这几乎可以肯定是来自于傣族。

特别有意思的是,在东部的阿尔泰系统的民族中也含有少量的 M119 突变,这些成分是他们原本就有的,还是百越带进来的,这有待于进一步的研究。布里亚特最多为 35%,尼乌柯(Nivkh) 6%,满族 5.6%,蒙古族 4.2%,乌尔奇(Ulchi) 3.8%,日本人 3.4%,叶尼塞鄂温克 3.2%。百越和阿尔泰系统通古斯民族的遗传和文化相似性究竟是发生于发生关系上较近还是相互交流影响,这是一个值得深入的课题。

百越与外族的交流融合是一个非常复杂的过程,既有百越的成分流向外族,又不断有外族融入百越,甚至可能有过西部的许多小民族整体性地融入百越族群的事件,这又是一个值得研究的课题。而东部的越族由于汉化程度很高,其分布和结构的研究需要大量的田野调查。像上海马桥这样的较纯的越裔群体在浙江和福建的某些地区也应该存在。对其研究的意义应该是不言而喻的。

五、对百越发生和分化的大胆猜测

对百越的遗传结构成因的解析可能是多种多样的,这里我们想提出一种猜测,以供史学家和以后的百越民族遗传学研究者借鉴推敲。

首先是百越民族的起源问题。最初从非洲来到东南亚时,百越和南岛民族可能是同一个群体。而后百越民族北迁至广东,南岛民族南迁至马来。两者分开后分别发展成两个系统。台湾语族究竟来自于南岛民族还是百越民族,这将由以后更多的遗传学研究数据来解答,但是目前的数据更偏向于后者。在两广,百越民族可能居住了很长的一段时间(几万年)而没有外迁,因此发育出了统一的体质和文化。内部几个分支的分化发生在外迁前还是外迁后还无法得知。可能百越先分成南北两

部分，南部是壮傣和黎的祖先，北部是侗水和仡央的祖先。

由于广西西部和云南居住着百濮民族，湖南北部居住着三苗族群，所以百越最早的扩散方向是空旷无人的东北和正东。这是两条不同的路线，携带了不同的遗传标记。发生的年代相当早，至少在河姆渡文化之前。向东北迁徙的越人出于北部类群，先进入江西，形成后来的干越。部分干越后来西迁，很可能就是今天的侗族（Kam），侗族的起源学说中有江西说一种。这解释了为何侗族遗传结构会归入东越类型。从江西进入浙江一带的越人形成了於越，在太湖平原和宁绍平原创造了灿烂的文化。向正东方向迁徙的越人出于南部类群，进入福建，形成了闽越各部。瓯越还没有调查，不知道属于上述哪一支。台湾原住民如果是来源于百越的话，也是分别来自于这两个分支的。明显布农与闽越相似，赛夏也有部分成分与福建相似，而其他族则多与於越相似。百越渡海到台湾的可能性是相当大的，值得深入研究。

仡央族群的迁出也是较早的事件，不知道他们是不是直接迁到贵州西北部，但至少周朝就有贵州夜郎国的记载了。他们应该是侗水族群中分化出来的。黎族进入海南岛也较早，但是全部进入海南是较晚的事件。壮傣族群的大规模西迁肯定是汉朝时南越国灭亡之后，傣族进入元南和泰国据证也只是唐朝的事情。而从云南的勐卯国西迁至印度则是公元1215年。所以西部越族发源于广东的历史是清晰的。但是西部越族的各支系的分化路线相当的错综复杂，完成对其遗传研究这项艰巨的任务对东南亚各国的意义却相当重大。

本文就目前对百越遗传结构研究得到的数据分析了其结构的“一元二分”的特点，并对这种结构特点的成因作了一些大胆的猜测，希望能抛砖引玉，在百越史研究中启发一些新的思路，也激励对百越族群的遗传研究的广泛深入的后续工作。

致谢：广西民族学院民族学人类学研究所、共青团西双版纳傣族自治州委员会、共青团文山壮族苗族自治州委员会、共青团上海市委员会、中华民族史学会、海南省迁琼先民研究会、广州军区政治部、绍兴市禹陵村委在田野调查中给予了倾力的支持和帮助。

参考文献：

- Cann RL, Stoneking M, Wilson AC. Mitochondria DNA and Human evolution [J]. Nature, 325: 3136, 1987.
- Jorde LB, Watkins WS, Bamshad MJ, et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data [J]. Am. J. Hum. Genet., 66 (3): 97988, 2000.
- Underhill PA, et al. The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations [J]. Ann. Hum. Genet., 65: 4362, 2001.
- Su Bing, Xiao CJ, et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas [J]. Hum. Genet., 107: 582-590, 2000.
- 李辉, 宋秀峰, 金力. 人类谱系的基因解读 [J]. 二十一世纪 (香港), 2002, 71 (6): 95105.
- 李辉. 百越族群遗传结构分析的初步思路 [A]. 现代人类学国际研讨会论文集 (会间交流本) [C]. 上海, 2002. 4, pp89-94.
- Su B, Xiao J, Underhill P, et al. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age [J]. Am. J. Hum. Genet., 65 (6): 17181724, 1999.
- 石林. 侗台语比较研究 [M]. 天津: 天津古籍出版社, 1997.
- 何平. 从云南到阿萨姆——傣-泰民族历史再考与重构 [M]. 昆明: 云南大学出版社, 2001.
- Buragohain Ye Hom. King Sukapha and His Journey to Assam: The Manuscript Evidence [A]. Proceedings of the 3rd International Conference on Thai Studies [C], Canberra, 1987. p18.

(责任编辑: 覃乃昌)