

百越族群结构分析的初步思考

李辉

(复旦大学现代人类学研究中心, 上海 200433)

提 要: 百越是一个古老的族群, 至今在中国南方和东南亚各国有广泛的分布。分析这个族群的结构对于此族群的各项自然科学、社会科学研究有重要意义。历史学和语言学的研究已得出了大量的客观结论。遗传学的研究为许多其他学科未能解决的问题带来了解决途径。Y 染色体是种很好的分析材料。用数理统计的方法可以分析出可能的族群特征, 并用以研究族群结构, 分析未定群体。百越族群的遗传分析已有了良好的开端, 但是大量的古今群体还有待调查。

关键词: 百越, 群体结构, 遗传分析。

Preliminary Analyses of Genetic Structure of Daic Populations

Li Hui, Center for Anthropologic Studies at Fudan University, Shanghai 200433.

Abstract: Daic is a language family distributed in China, Vietnam, Laos, Thailand, Myanmar, India, etc, including Kam-Sui, Tai, Hlai and Kadai subfamilies. It is important to study the structure of Daic for other studies of natural sciences and social sciences of this family. History and linguistics have drawn in a lot of useful conclusions. Genetics also can resolve those questions that cannot be resolved in other field. Y chromosome is a good material to do such studies. Statistics can find genetic characters of ethnic group. The work has started well in Daic, and there are still a great number of populations to study.

Key words: Daic, Population structure, Genetic analysis.

百越族群曾经在先秦长期作为一个优势群体存在, 活跃于整个长江以南地区, 是对中国历史产生了深远影响的一个民族系统。史记对百越的分布描述为从“会稽”到“交趾”, 即从江浙到越南。也有学者认为会稽曾经在山东, 那么百越的分布更广了。对于百越内部的系统, 则史记早已指出“各有种性”, 但究竟是如何分类的古人也很难弄清。

1. 百越的概念和各学科对百越结构的研究

古籍上出现过许多被认为与百越有关的族称, 包括: 於越、甌越、闽越、扬越、南越、雒越、西瓯、滇越、干越、山越、安家、掸国、夜郎等。这些名词很值得斟酌, 因为古代汉语文献对外族名词译音的时候, 取字不会很严格, 与“越”读音相近的都有可能译成“越”, 而后人往往会望文生义, 以至以讹传讹。所以这些古民族有的可能根本不属于百越系统, 比如滇越和掸国, 云南大学的泰学专家何平用翔实的证据证明了前者是印度阿萨姆省的一个雅利安人古国, 而后者则远在叙利亚^[1]。也有些名称可能是同一个族的不同时期的称呼, 比如甌越和山越。其他一些在亲缘关系的远近上也十分值得研究。这些族群有许多留下了大量的墓葬材料^[2], 为我们用现代的人体测量和 DNA 检测研究提供了可能。



图1 古籍记载的若干越族相关称谓

现代民族中，侗台语系的各民族被认为是西部越族的直接后裔，我国境内的较大民族有壮、傣、侗、黎等。东越的各支则被认为已完全融入了汉族人群中。至于各支现代人群究竟与哪些古代越族有继承关系，还需要遗传学等各学科材料的综合比较研究来证实。民族学界一般认为壮傣与西瓯、雒越有关，仡佬与夜郎有关^[3]。侗族的来源还是个谜，其早期称谓 Klam 与“雒”较像，现在的称谓 Kam 与“干越”较像^[4]。

其他现代族群与百越是否有关，在各领域都作了深入的探讨。体质人类学的研究认为南方各族群都属于蒙古人种南亚型，很难细分^[5]。在语言学上关系词比较给这个问题的解决提供了较直接的线索。通过两语系各语言的一阶词和二阶词中关系词的数量差异比较，得出侗台语系与南岛语系有同源关系，而与汉语、南亚语系、苗瑶语系都没有同源关系^[6]。

2. 遗传学分析百越族群的结构

分子遗传学的材料发现，线粒体目前很难对族群进行区分，而 Y 染色体则相对明显的多。这可能是由于族群之间女性总是比男性的流动性大。Y 染色体在世界人群中已构成了 131 种单倍型，而且族群特异性较大^[7]。南方的许多民族已作了这方面研究，积累了大量数据。但是由于族群的交流，许多单倍型也在各族之间多多少少地散开了。但是，如果承认最早的各个民族系统的遗传结构是迥异的，那么某族势必保留一定比例的原初特征遗传成分才能归入这个系统。如果特征遗传成分少到一定程度，那么他就实质上不是这个系统的成员了。撇开有争议群体，分析各单倍型在各确定民族系统之间的分布，分清哪些单倍型是完全散在的类型，哪些是特征类型，再根据特征类型的差异来分析族群结构则是遗传结构分析的基本思路。最简单的方法是找单纯民族。就是在一个族群里寻找只有一种类型的民族，那么这个民族的单倍型很可能是此系统特征类型。我们按照宿兵的单倍型分类命名系统^[8]来看，景颇族这个典型的氏羌民族只有 H8 型^[9]，很可能就是氏羌民族的特征之一，但是可能由于汉族的扩张使各系统民族大多含有不小比例的 H8，所以这一结论也不绝对肯定。侗傣民族的结构较复杂，还没有找到单纯民族（很可能古代族群中有），这就要用数理统计来寻找特征。表 1 对侗傣民族和邻近的其他民族做主成分分析，根据各主成分对各民族进行排列，并计算出主成分与民族系统的相关性。

表 1 各民族的主成分排列（末行为主成分与民族系统的相关性）

主成分 1	主成分 2	主成分 3	主成分 4	主成分 5	主成分 6	主成分 7
BUNU	NETHAI	TUJIA	MAQIAO	CAOMIAO	BUNU	LANA
SEK	HLAI	HMONG	KAM	KINH	SHE	NETHAI
AHOM	MAONAN	LAHU	KOTA	LOLO	BLANG	TUJIA
LOLO	BOUYEI	DAYAK	MALAY	MALAY	KAM	CAOMIAO
LAKA	TORAJAN	BALI	SHAN	LAKA	LIUJIA	DAI
LANA	KOTA	JAVA	TUJIA	MUONG	LANA	HLAI
MAQIAO	MULAM	BUNU	DAYAK	LIUJIA	HLAI	BALI
SHE	LAKA	CAOMIAO	BATAK	MULAM	BOUYEI	BOUYEI
TUJIA	BATAK	NHAN	HLAI	BLANG	TUJIA	KINH
DAYAK	SUI	BATAK	ZHUANG	SHE	MAONAN	MIEN
MIEN	JAVA	SHE	NHAN	MAONAN	BATAK	BATAK
CAOMIAO	BALI	TORAJAN	LIUJIA	TUJIA	KOTA	DAYAK
MULAM	MUONG	MALAY	JAVA	NHAN	KAREN	MUONG
MALAY	DAYAK	SHAN	CAOMIAO	NETHAI	SHAN	NHAN
LIUJIA	KAM	KOTA	MULAM	SHAN	NETHAI	ZHUANG
NETHAI	HMONG	BLANG	SUI	KAM	ZHUANG	SHAN
DAI	KINH	LIUJIA	AHOM	BATAK	MUONG	LIUJIA
KOTA	BLANG	VA	KINH	LANA	KINH	LAHU
LAHU	ZHUANG	DAI	SHE	BOUYEI	SUI	KAREN
HMONG	LOLO	MIEN	KAREN	KOTA	VA	SUI
VA	BUNU	KINH	LANA	HMONG	NHAN	BLANG
NHAN	MAQIAO	MAONAN	MUONG	KAREN	CAOMIAO	KOTA
HLAI	LAHU	BOUYEI	BUNU	HLAI	MIEN	MAONAN
SHAN	TUJIA	LOLO	SEK	SEK	BALI	MALAY
KAM	LANA	KAM	TORAJAN	AHOM	LAKA	HMONG
TORAJAN	MALAY	MUONG	BLANG	BALI	MAQIAO	TORAJAN
KAREN	SHE	MULAM	NETHAI	SUI	HMONG	JAVA
JAVA	KAREN	LAKA	MIEN	LAHU	DAI	MAQIAO
ZHUANG	CAOMIAO	NETHAI	LAKA	MIEN	AHOM	VA
MAONAN	SEK	MAQIAO	LOLO	ZHUANG	TORAJAN	KAM
BATAK	AHOM	HLAI	VA	DAYAK	JAVA	SEK
BALI	SHAN	SUI	DAI	BUNU	SEK	LAKA
BOUYEI	DAI	KAREN	BALI	DAI	LAHU	AHOM
SUI	VA	AHOM	BOUYEI	TORAJAN	MULAM	SHE
KINH	NHAN	ZHUANG	MAONAN	VA	MALAY	MULAM
MUONG	LIUJIA	SEK	LAHU	JAVA	LOLO	LOLO
BLANG	MIEN	LANA	HMONG	MAQIAO	DAYAK	BUNU
-0.102	-0.435**	0.579**	-0.149	0.277	0.169	-0.070

表 1 中深蓝色为侗傣民族，浅蓝为马来民族。单星号为显著相关，双星号为极显著相关。可见主成分 2 和 3 是与民族系统相关的。

再对各主成分和各单倍型在各民族中的对应数据作相关分析，观察与民族系统相关的主成分中是哪些单倍型在起作用，这些单倍型很可能是特征型。表 2 中列出了相关分析的结果。

表 2 各单倍型与主成分的相关系数

单倍型	主成分 1	主成分 2	主成分 3
H1	-0.2806	0.1075	-0.3492*
H2	-0.1575	0.0817	0.2976
H3	-0.4078*	-0.1444	0.2370
H4	-0.2764	0.0447	0.1685
H5	0.2723	0.0547	-0.8360**
H6	0.2001	0.3863*	-0.4352**
H7	-0.3512*	-0.0361	-0.4033*
H8	-0.1563	0.7706**	0.5545**
H9	0.2815	-0.6123**	-0.0887
H10	-0.1223	-0.2164	-0.0280
H11	0.5449**	-0.8398**	0.3513*
H12	-0.2188	0.0988	0.3678*
H13	0.0748	0.2917	-0.1648
H14	-0.3211	0.0870	0.3015
H16	0.1409	-0.1682	-0.1547

表 1 中主成分 2 与百越极显著负相关，表 2 中极显著负相关的单倍型为 H9、H11，可能是百越的特征型。主成分 3 与百越极显著正相关，这与之显著正相关的为 H8、H11、H12。H8 在主成分 2 和 3 种的相关性是矛盾的，这可能是由于其他复杂的因素引起的干扰（很可能是汉人的强势影响），所以把它排除掉，则 H11、H12 可能是特征型。另外 H1、H6、H7 等则最有可能是外系统类型。再如 H10 虽然百越民族中也并不多，所以没有表现出显著信号，但在百越系统之外几乎不出现，所以也可以认为是特征型。

有了特征型的大致概念，就可以考察这些特征型在百越系统内部的分布^[10-12]，以分析百越的结构。

表 3 各百越民族群体中特定特征单倍型频率（%）

民族	样本量	H9	H10	H11	H12
黎 Hlai	11	27.3		54.5	9.1
北壮 Zhuang	29	17.9		25	10.7
布依 Bouyei	45	4.4		46.7	11.1
东北泰 Thai (N.E.)	20	5	5	45	20
兰那 Lana	20			20	20
傣泐 Dai	40			5	17.5
石 Sek	5				20
阿霍姆 Ahom	10		10		
仡佬 Mulam	40	5	25	30	
侗 Kam	40	25	10	20	
马桥 Maqiao	74	39.6	12.5	8.3	
水 Sui	50	18		44	
毛南 Maonan	32			56.3	
拉珈 Laka	10			40	

在我们研究过的百越族群中根据特征单倍型对比，发现基本可以分为 5 组。第 1 组，黎族、壮族、布依族，有较稳定的 H11、H12，H9 频率则依次降低，代表从广东到贵州的迁徙。第 2 组，东北泰（老族）—傣组，有稳定的 H12，其他种类依次减少，代表向西南部的迁徙。第 3 组，印度阿霍姆，仅有 H10，可能坎梯等族也在此组。第 4 组，侗—仡组，有稳定的 H10，H9 逐次加强，H11 渐少，代表从广东到江西和浙江的迁徙。第 5 组，水—毛组，有稳定的 H11，水族的 H9 在毛南族丢失了。这些信息在民族学、语言学等领域肯定会有深刻的意义。

在每一组的稳定特征型中，我们可以进一步分析其他遗传标记，来确定具体的迁徙方向和分化年代。这已在氏羌系统中有了先例^[9]。

3. 百越史前祖先和散在部落的研究

百越作为一个古老的族群，肯定与某些史前人群有继承关系。在百越的活动区域内，出现过河姆渡、马家浜、崧泽、良渚、大窑坑、左镇、肩头弄、马桥、吴城、湖熟、石硖、东山等大量的考古文化，这些文化有的已被确认为百越的祖先创造的，而其他一些则争议较多。而用遗传学方法把这些文化的遗骸所保留的遗传信息与现代的百越族群比较，很多问题很可能就可以迎刃而解。

在现代百越族群中，除了几个已确认的较大民族外，还有大量的小部落。这些小部落可能对百越族群的结构分析有很重要的作用。所以研究的任务还很重。有很多族群与百越的关系还没有确定，更值得深入分析。百越族群和可能与百越相关的各民族和部落列于下文和图 2。

侗水支：

1.KAM, NORTHERN 北侗；2.SUI 水；3.AI-CHAM 锦；4.MAK 莫家；5.TEN 佯僮；6.KAM, SOUTHERN 南侗；7.MAONAN 毛南；8.MULAM

佤佬：9.LAKKIA 拉珈；10.POU 标

壮傣支：

11.LINGAO 临高；12.SEK 石；13.ZHUANG, NORTHERN 北壮；14.NHANG 热依；15.BOUYEI 布依；16.TAI MÈNE 傣曼；17.ZHUANG, SOUTHERN 南壮；18.E 五色；19.MAN CAO LAN 高栏山仔；20.NUNG 侗；21.TÀY 岱；22.TS'ÜN-LAO；23.TAI DAM；24.LANA 兰那；25.PHUAN；26.SONG 宋；27.THAI 泰；28.TAI DÓN；29.TAI DAENG；30.TAY TAC；31.LAO 僚；32.NYAW 饶；33.PHU THAI 孚泰；34.THAI, NORTHEASTERN 东北泰；35.AHOM 阿霍姆；36.AITON 艾通；37.KHAMYANG 坎佯；38.PHAKAI；39.SHAN 掸；40.TAI NŪA 傣呐；41.LŪ 傣泐；42.KHAMTI 坎提；43.KHŪN 坤；44.PU KO 普格；45.PHAKTAI 帕西傣；46.TAI LONG 傣龙；47.TAI YA 傣雅；48.TAI HANG TONG；49.TAI HONGJIN 金平傣；50.TAI MAN THANH；51.YONG；52.TURUNG 茶隆；53.RIEN；54.TAY KHANG；55.TAI PAO；56.TAY JO；57.KUAN；58.YOY；59.KANG 黑傣；

黎加支：

60.HLAI 黎；61.JIAMAQ 加茂；

佧央支：

62.BUYANG 布央；63.EN 侬环；64.YERONG 夜郎；65.MULLAO 木佬；66.LACHI 拉基；67.LACHI, WHITE 白拉基；68.GELAO 仡佬；69.LAQUA 普标；70.LAHA 拉哈；71.CUN 仡隆；72.GAO 羿人

台湾原住民（争议）：

73.ATAYAL 泰雅；74.TAROKO 太鲁阁；75.AMIS 阿美；76.AMIS, NATAORAN 那陶兰；77.BASAY 玛卡道；78.BABUZA 巴布萨；79.KAVALAN 噶玛兰；80.SIRAIYA 西拉雅；81.HOANYA 和安雅；82.KETANGALAN 凯达加兰；83.KULUN 雷朗；84.PAPORA 巴布拉；85.PAIWAN 排湾；86.PYUMA 卑南；87.PAZEH 巴则海；88.SAISYAT 赛夏；89.THAO 邵；90.TAOKAS 道卡斯；91.BUNUN 布农；92.RUKAI 鲁凯；93.TSOU 曹；94.SAAROA 萨阿拉；95.KANAKANABU 卡纳卡纳布；

东南沿海不同程度汉化的越裔：

96.DANGA 疍家；97.MINYUEC 闽越；98.OUYUEC 瓯越；99.YUYUEC 於越；100.SANCHUEC 马桥。



图 2 现代百越族群的分布

图例：○ 侗水支 Kam-Sui；□ 壮傣支 Tai；△ 黎加支 Hlai；▽ 佧央支 Kadai；★ 台湾原住民 Taiwan Aborigines；◇ 东南越裔 Yue。



图3 现代百越族群可能相关的六个分支

从左到右，上行：台湾原住民（布农），黎加支（美孚），侗水支（水族）；下行：东南越裔（马桥），壮傣支（傣雅），侗央支（普标）。

- [11] 李永念,左丽,文波等. 中国布依族人的起源及迁移初探:来自Y染色体和线粒体的线索[J]. 遗传学报, 2002, 29(3).
[12] 李辉,杨宁宁,黄兴球. 广西融水北高村人的遗传学和体质人类学特征[J]. 中央民族大学学报.

在各学科共同努力下，百越族群的来龙去脉肯定会研究得越来越清晰，为中国和东南亚各国的历史文化研究提供大量可鉴的材料，也为这个族群系统中的社会变迁、群体心理学、流行病学、医学遗传学、群体遗传学等研究打下基础。

主要参考文献:

- [1] 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.
[2] 陈明芳. 中国悬棺葬[M]. 重庆: 重庆出版社, 1992.
[3] 梁敏,张均如. 侗台语族概论[M]. 北京: 中国社会科学出版社, 1996.
[4] 石林. 侗台语比较研究[M]. 天津: 天津古籍出版社, 1997.
[5] 中国人类学会. 中国八个民族体质调查报告[M]. 昆明: 云南人民出版社, 1982.
[6] 陈保亚. 语言接触与语言联盟——汉越语源关系的解释[M]. 北京: 语文出版社, 1996.
[7] Underhill PA, et al. The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations[J]. *Ann.Hum.Genet.*, 2001, 65: 43-62.
[8] 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.*, 1999, 65(6):1718-1724.
[9] Su Bing, Xiao CJ, et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas[J]. *Hum. Gent*, 2000, 107:582-590.
[10] Su B, Jin Li, Underhill Peter, et al. Polynesian origins: insights from the Y chromosome[J]. *Proc.Natl.Acad.Sci.USA*, 2000, 97:8225-8228.