



从父系遗传结构分析壮族的起源与分类

陈晶

广西中医学院第一附属医院综合实验室, 中国 南宁 530023

摘要: 壮族是中国最大的少数民族, 与东南亚广泛分布的泰老族群有着密切的关系, 在东亚人群的遗传结构研究中地位非常特殊。然而, 由于壮民族缺乏通用的本族文字, 历经数千年的发展, 壮民族内部人群的迁徙融合及交流状况很难在现有的文字资料中有所发现。仅有的材料是从语言上将壮族人群区分为壮语北部方言组和南部方言组。为了解壮族人群的内部遗传结构, 我们从父系遗传物质 Y 染色体入手, 根据东亚人群特有的 18 个 Y 染色体双等位标记, 7 个 Y-STR 标记, 对壮族各个支系共八个人群的 DNA 样本进行限制性长度片段多态酶切、荧光 PCR 确定 SNP、STR 等实验技术分析, 及主成份分析、聚类分析、分子方差分析等多种遗传学分析技术, 调查了壮族各个支系的 DNA 多样性, 分别确定其 Y-SNP、Y-STR, 揭示壮族的内部遗传结构。结果发现, 壮族的主要 Y 染色体单倍群为 O*, O2a、O1。壮族的遗传结构中有几个层次, 最早的 O*、O2a 成分普遍出现在各个支系中, 第二层是由东部来的百越核心成分 O1, 第三层是北方来的汉族成分 O3, 与壮族的历史事件和其他人类学研究基本一致。壮族支系遗传结构体现出以红水河支系为中心从东往西的梯度变化过程。传统的对壮族按方言分为南北二组的分类方法并没有在遗传学上获得相应的支持。

关键词: Y 染色体; 壮族; 内部遗传结构; 多起源

Origin and Classification of Zhuang Analyzed by Paternal Genetic Structure

CHEN Jing

Synthetic Laboratory, The First Affiliated Hospital, Chinese Medicine College of Guangxi, Nanning 530023, China

ABSTRACT: The Zhuang nationality, the largest minority in China, is quite close to the Thai and the Lao in culture, and therefore, plays an important role in the genetic structure study of the East Asian population. However, without a formal written language, it is quite difficult to trace the origins and migration history of the Zhuang with the existing Chinese records. The only classification of Zhuang populations is by language. Zhuang people can be classified into northern and southern groups based on their dialects. In order to understand the Zhuang's internal genetic structure, the DNA diversity of eight Zhuang branches were investigated after determining their Y chromosome haplotypes that represent paternal inheritance with 18 East Asia specific Y-chromosome biallelic markers, 7 Y-chromosome STR markers. RFLP, fluorescence-PCR techniques and detail genetics analyses such as principal component, hierarchical clustering and AMOVA analysis were performed. As a result, on Y-chromosome, O*, O2a and O1 were found to be the most common haplogroups in Zhuang populations. Here O* and O2a were the native component; O1, the core component of Baiyue from eastern China, was brought to Zhuang later; and finally, O3 from northern Han was introduced. This result is quite consistent with historical records and other anthropological studies. In addition, a Hongshuihe-centric, east-to-west gradient change rather than an expected south-north profile was observed in the genetic structure of the Zhuang, whereas various genetic analyses were performed, indicating that the traditionally linguistic classification of northern and southern Zhuang groups is short of genetic support.

Key words: Y chromosome; Zhuang; Internal genetic structure; Multiple origin

在中国南部, 从古至今屹立着一群人, 他们今天的主要聚居在中国广西壮族自治区的壮族人。壮族人口 1700 多万, 是中国境内最

大的少数民族, 其主要分布在广西、云南、贵州、广东部分地区, 其中大约 94% 聚居在广西。壮语属于侗台语系侗傣语族壮傣语支,

收稿日期: 2007 年 4 月 11 日 修回日期: 2007 年 5 月 2 日 联系人: 陈晶 ChenJing.GXMU@gmail.com

并在其内部按照壮语自身方言的差别以邕江为界，细分为南部方言和北部方言[1]。

据史料记载，壮族的来源可以追述到2000多年前“雒越”、“西瓯”族群。目前关于壮族来源比较一致的看法认为，壮族是百越族群中这两支的后裔[1]。在谈及与壮族密切相关名称却迥异的族群之时，我们注意到分布在广西西北部、贵州东部的布依族尽管名称上与壮族并不相同，然而究其根源，仍然可以追溯至壮族，属于学术意义上的壮民族；而云南的水族实际上是自称“水户”的布依族，也同样属于壮族群体[1]。纵观壮族长期的历史发展过程，我们可以发现其族群内部经历了复杂的迁徙演变，它与分布在东南亚的泰族、老族、掸族的起源和形成有着极其密切的关系。最后我们也发现壮族在保持其浓郁民族特色的同时，更进一步地吸收了中原地区的文化，创造了属于自己的璀璨文化，也因此壮族与汉族及其它族群存在着复杂的民族交流格局。

为了理清壮民族的起源发展人们运用了包括历史资料、民俗学等许多方法工具，现在我们在传统分析方法的基础上进一步的使用现代分子生物学理论和工具，对壮族进行分析。

一、壮族的起源

《汉书·地理志》注引臣瓚说：“自交趾至会稽七八千里，百越杂处，各有种姓，不得尽云少康之后也”[2]。现存的这些文献表明在商周时期，我国东南部和南部沿海地区，直到今天的越南北部，都是百越族群的聚居区。长期以来人们对百越族群来源的考察，多根据我国早期的历史资料，由此认为包括西瓯、骆越在内的百越族群系夏禹/少康的后裔。据史料记载，作为曾经遍布我国东南沿海乃至今天越南地区的一个族群，百越族群有着悠久的历史。而近现代的考古发掘工作更加证明了百越族群创造了包括良渚前文明在内的许许多多令人侧目的辉煌文化，百越族群同样参与了中华民族璀璨历史的构建。

至于壮族人群的历史来源说法众多，其

中最为人们所接受的是“西瓯”、“骆越”起源说。“西瓯”的记载，最早见于汉代。刘安《淮南子·人间训》[3]记述秦王朝平定岭南：“与越人战，杀西瓯君译吁宋。”，《史记·南越列传》[4]提到赵佗：“以兵威边，财物赂遗闽越、西瓯、骆，役属焉。”，《汉书·南越传》[2]也说：“蛮夷中，西有西瓯，其中半赢，南面称王。”。而“骆越”的名称据说是与其民耕“骆田”有关，其也可见于我国早期文献。《盐铁论·地广篇》[5]曾载：“荆楚罢于瓯骆”，《水注经·叶渝水》注引《交州外域记》[6]：“交趾昔未有郡县之时，土地有雒田，其田随潮水上下，民垦食其田，因名曰雒民”。其中“雒”与“骆”相通，音klok，骆越亦即垦食骆田的越人。

作为百越族群其中的一支，西瓯、骆越的命运也同中华民族的发展紧密联系在一起。自公元前220年秦始皇挥军50万南下进攻百越开始，西瓯骆越的发展一直没有离开过人们的视线。漫长的历史中，西瓯骆越的族群结构经历了复杂的变迁，其名称也随着朝代的不同而反复变动。加之历朝历代的统治者们忌讳岭南地区崇山峻岭民风彪悍，对岭南地区多采用“羁縻政策”，此举使得汉官流民不断涌入，土著居民不断迁徙[1]。于是今天壮族的复杂结构和确切起源就有了更多有待澄清释疑的地方。

为了廓清壮族发展的历史，人们一直尝试使用传统的人类学方法来研究广西壮族。以往的人类学研究方法是通过体质人类学、历史资料、民族语言、民俗等等方法来对不同的人群进行分类，也因此得到了人群的一些大体分类。然而这类方法在遇到较为复杂的民族发展变迁历史时却有些无力，因此无法对人群的精细来源和变迁有更为深入的展现。尤其在壮族这种经历了大量交流、文化在很大程度上已经汉化、历来名称变化纷繁、内部迁徙演变复杂的民族上，传统的方法更加不易进行区分。

分子人类学在充分借鉴原有人类学的研究方法的基础上，加入现代的人类基因组多态检测手段和数理统计方法，依据大范围的人群数据背景对一个民族的遗传结构进行分

析，进而能够得到人群确切的来源信息和其内部的迁徙演变信息。

分子人类学对人群的分析依据一些主要的理论基础。在人类基因组上，一些碱基的突变速度是比较稳定而缓慢的，由此产生的突变我们称之为 SNP (Single Nucleotide Polymorphism, 单核苷酸多态)。随着分子人类学的发展，研究者们发现在人类基因组一些缺乏重组的区段上的 SNP 会随着人群的历史发展、迁徙演变而有多寡的区别和不同的分布格局。因此，越古老的人群其基因组上积累的 SNP 突变也越多，SNP 产生的年代越久远；反之，如果在一个人群中检测到的 SNP 越多，SNP 出现的年代越久远，那么这个人群也越古老。另外，随着新人群的产生，新的特征性的 SNP 也往往跟着形成。根据这个理论，从上个世纪 80 年代开始，科学家们分别检测了人类线粒体 DNA、Y 染色体非重组区段上的多个 SNP，分别构建了人类母系系统发育树、父系系统发育树[7-17]，进一步分别从母系遗传角度与父系遗传角度得到了人类的“非洲起源”学说。也同时得到了各大洲不同人群的遗传背景信息，发现了不同大洲不同区域不同人群的较为特征性的 SNP 分布信息。

在东亚，研究者通过对不同人群的基因突变进行分析，发现中国南方人群积累的突变要多于北方人群[18-21]，许多相当古老的 Y 染色体单倍群（单倍群由一系列的 Y 染色体上的 SNP 定义）大量的出现在中国南方的群体中，这些单倍群包括 O、O2a。经过适当的数理计算，人们发现这些单倍群产生的年代大约是距今 20000~10000 年[7, 11-12, 22]。同时，人们也发现了不同的族群其特征的突变也不相同。例如 O1 单倍群就被发现是东部百越族群较为特征的单倍群[22]，而另一些来自北方人群的数据则表明相较于南方人群，O3 这个单倍群在北方汉族人群中有着更高的分布频率 [18-19]。依据这些已有的研究成果，我们能够有足够的比对人群进而分析今天壮族人群的整体遗传背景。

为了理清壮族的内部遗传结构，我们采集了 129 份分属于不同地区的壮族男子的 DNA 样品，并对其进行 Y-SNP 单倍群分群。

为了更为全面地了解壮族，我们还在上述数据基础上加入了已发表的布依族、水族数据，以及部分未发表数据[23-24]，使得参与分析的样本量达到 166 份，以保证每一个壮族支系包含的样本数达到统计要求，即壮族每个支系的样本数达到 10 例以上。

表 1 系壮族人群 Y 染色体 SNP 单倍群分布频率。从表中我们发现 C、D、D1a、F 这些较为古老的单倍群在壮族群体中也有一定的分布，但是频率比较低，我们考虑这些古老的单倍群更可能是作为壮族人群中一种遗传背景存在的，它们不是壮族人群的主体单倍型。单倍群 O*则需要特别的关注，这个单倍群在壮族各个支系中普遍高频，构成了壮族 Y 染色体 SNP 单倍群的主体，同时，壮族中高频的 O*单倍群分布也迥异于南方一些别的群体的 Y 染色体 SNP 单倍群分布情况。壮族人群单倍群 O*普遍高频这个结果提示我们：壮族极大的可能是别于以往一些南方群体而有着自己特征的群体。根据以往一些数据我们大体推测单倍群 O*是一个比较古老的单倍群，其产生年代至少是距今一万年以上[7, 11-12, 22]，再结合壮族该单倍群的 STR 数据我们大体了解壮族人的历史比较古老。

随后则分别是 O2a、O1 这两个 Y 染色体 SNP 单倍群，它们在壮族人群中也有较高的分布频率。而这两个 Y 染色体 SNP 单倍群在其它东亚南方群体里面也同样高频分布。这个分布格局与原有的东亚南方人群的 Y 染色体单倍型频率分布规律相一致，这说明壮族人群尽管有着自己的特点，但仍是较为典型的东亚南方群体。此外我们也发现 O3*及其下游的 Y 染色体 SNP 单倍群在壮族人群中也有一定的分布，由于这些单倍群在东亚北方部分人群中拥有比较高的分布频率，加之通过 STR 年代估计发现壮族群体中这些单倍群的发生年代均在 4000 年以内，因此可以得到提示：O3*可能并非壮族固有的遗传成分，而可能系和东亚北方人群交流获得。

从表 1 的 Y 染色体 SNP 单倍群分布情况分析，我们大体得到几个基本的结论：广西壮族人群仍旧应该也是比较典型的东亚南方人

表 1 壮族各支系的 Y 染色体 SNP 分型频率

Tab.1 Y-SNP haplogroup frequencies of branches of Zhuang

	size	C	D	D1a	F*	K*	O*	O1	O2a*	O2a1	O3	O3a	O3e*	O3e1
邕北 YB	23	8.70		4.35		4.35	21.74	17.39	8.70	21.74	8.70		4.35	4.35
右江 YJ	5			40.00			20.00		20.00					20.00
田林	22						13.64	4.55	72.73			4.55		4.55
桂边 GB	4					25.00	25.00	25.00			25.00			
红水河 HSH	39	2.56		5.13		5.13	23.08	10.26	5.13	7.69	20.51		5.13	15.38
桂北 GN	21	4.76	4.76			4.76	38.10	4.76	4.76	4.76	14.29		14.29	
邕南 YN	19				5.26	5.26	15.79	10.53	10.53	10.53	31.58	5.26	5.26	
左江 ZJ	15		6.67				33.33		40.00		20.00			
上思	15						13.33	20.00	60.00	6.67				
德靖 DJ	3						100.00							

群。在保有东亚南方人群比较特征的 Y 染色体 SNP 单倍群 O2a、O1 的同时，壮族人群也拥有自己高频分布的 O* 单倍群，这也表明壮族很可能有着自己独特的历史。最后，通过表 1 并综合 STR 的相关数据，我们也能看到壮族人群中也有一定的北方人群遗传成分。从这点而言，再对照壮族现有的历史记载，我们能发现其颇为符合之处：起源于西瓯骆越的壮族人有过自己独立的方国，参与了遍及长江南岸的百越族群辉煌的历史。随后面对汉族的崛起，也曾经为了自己的生存而一争短长，并在这种过程中艰难地融入中华民族共同进步的历程，接受了来自中原的文明还有中原的移民，最终形成自己独特的民族文化，甚至可能在当初汉民族优势压力下进行了长途的迁徙。

二、壮族支系

谈起今天壮族族名中的“壮”字还有一段小故事。壮族历来有自己的族名，在史书记载上也有着相应的汉字。解放以前是“𠬞”旁加“童”，读 *zhuàng*，解放后，认为𠬞旁是文字上歧视少数民族，原来从𠬞旁的族名如獠、猺、獠等族名都修改了，改“𠬞”旁为单人旁。其中獠改为“僮”，而“僮”字的原意是儿童听差（童工），有其原音 *tóng*，因此改音“*zhuàng*”很困难。不明就里的人往往都读成“*tóng* 族自治区”。最后为了更好的区

别，周恩来总理将族名最后确定为“壮族”，壮族才有了今天这么响亮的一个名字[1]。从“獠”到“壮”的称谓改变其背后代表的其实是一个历史事件，今天我们还能追溯其前后缘故。但是回顾历朝历代对岭南地区民族的记载，我们就会发现壮族族名历来随着时代的变迁而往往有着不同。汉代以前称为西瓯骆越，到了东汉时期逐渐演变为俚，大约到了隋唐时期称为僚，随后多俚僚连用。到了宋代史书上关于今天壮族的记载更多了一些轻蔑的记载，“蛮”、“峒獠”、“蛮獠”往往成为中原人们对壮族的称谓，直到近现代壮族的称谓又变成了前面提到过的“獠”[1]。这些称谓的变迁是否也对应着一个个典故？这些典故背面是否是一个个历史事件？如果能够这一个个历史事件串联，是否就能够更为完整地展示壮族这个族群发展的历史？很可惜的是，历来从正史也好，从稗官野史也好，想要窥得壮族内部的发展历程，都比较困难。也由于壮族本身并无自己的通用文字，其自身的历史也就少了确切的记载而只能靠代代言传口授的一些传说神话来推测了。

所幸的是运用人类学的一些工具，我们还能对今天的壮族做一些分类。其中一种分类依据就是语言。语言以其稳定的纹饰标记功能[25]，往往成为区分人群的重要工具。于是语言学家们根据不同区域壮族语言的差别，大体上以邕江为界将壮语区分成南方

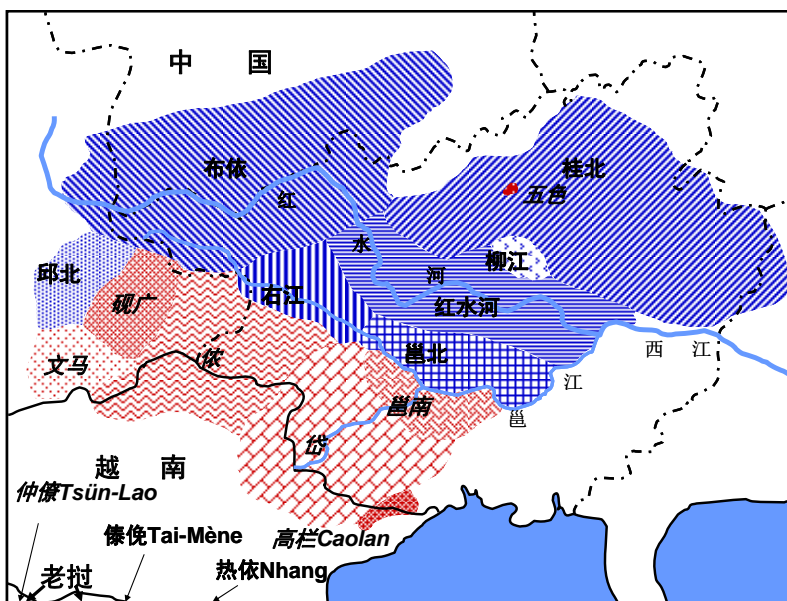


图 1 西南壮族各支系分布地图 北壮语支: Guibei 桂北, Liujiang 柳江, Hongshuihe 红水河, Yongbei 邕北, Youjiang 右江, Bouyei 桂边(布依), Qiubei 邱北, Nhang 热依, Tai-Mene 傣侬; 南壮语支: Yongnan 邕南, Tày 左江(岱), Man-Caolan 高栏, Nung 德靖(依), YanGuang 砚广, WenMa 文马, E 五色, Tsün-Lao 仲僚。

Fig.1 Distribution of Zhuang branches

言和北部方言，更进一步区分出了壮族的不同语言支系[1]。

图 1 使用不同的颜色和阴影区分了壮族不同语言支系。语言研究者们大致沿郁江(邕江)、右江走向，再到达广西平果县以后，往南偏移，后沿北回归线往西，经云南富宁、广南直至开远画一界限，并以这条线以北区分为壮语北部方言，包括了[1]：

桂北方言(河池、南丹、天峨、东兰、巴马、融水、罗城、环江、永福、融安、三江、龙胜)，柳江方言(柳江、来宾北部、宜州、柳城、忻城)，红水河方言(都安、马山、上林、忻城南部、来宾南部、武宣、象州、鹿寨、荔浦、阳朔、贵县、贺州)，邕北方言(武鸣、邕宁北部、横县、宾阳、平果)，右江方言(田东、田阳、百色)，桂边方言(田林、隆林、西林、凌云、乐业、凤山、及云南省的富宁和广南)，邱北方言(云南省邱北县和砚山县、师宗县)；

而该线以南为壮语南部方言，包括了：

邕南方言(邕宁南部、隆安、扶绥、上思、钦州)，左江方言(龙州、宁明、凭祥、崇左、大新和天等两县东部、越南北部如谅山等省)，德靖方言(德保、靖西、那坡、天等和大新两县西部、越南北部如高平等省)，砚广方言(云南省砚山、广南、麻栗坡、

马关、文山等县，以及越南北部)，文马方言(云南省文山、麻栗坡、马关和开远等，以及越南北部)。

尽管壮语南北方言有着一些明显的区别，比如北部方言一般没有 ph、th、kh 等送气塞音声母[1]。但是我们也知道，河流往往并不是阻隔人们交流的原因。纵观壮语方言的地理分布，其联系是相当紧密的。同时各个方言在通话自由程度上至少在其内部有着相当的开放性。但我们也注意到，持各个壮语方言的人其自称并不相同：桂边的人自称布依，左江的人自称岱、德靖的人自称侬等等。这些不同支系在语言上有所区别又紧密联系的情况，以及他们不同的自我称谓是否如一些人所揣测的那样代表了不同壮族人群的不同来源？有没有一些历史事件折射到壮民族本身的称谓上？更进一步这些语言支系的演变是否隐藏着壮民族本身的发展历程？

为了要廓清这些疑问，我们需要着手去理清壮族内部的遗传结构。首先我们需要通过对壮族不同语言支系的 Y 染色体 SNP 单倍群分布频率做主成分分析(principal component analysis)，去理清壮族内部不同语言支系在遗传结构上是否区分成不同的组

群。主成份分析结果显示，第一、第二主成份的累积贡献率占总方差的 82.5%，其中第一主成份贡献率为 43.3%，第二主成份贡献率为 24.3%。这说明了第一、第二主成份在壮族人群区分上占主要作用，而其他主成份的作用很弱，因此我们主要关注第一第二主成份所展示的内容。在此基础上我们发现了更重要的结果：在主成份散点图上，壮族不同的语言支系确实分成了两组。主成份分析如图 2。

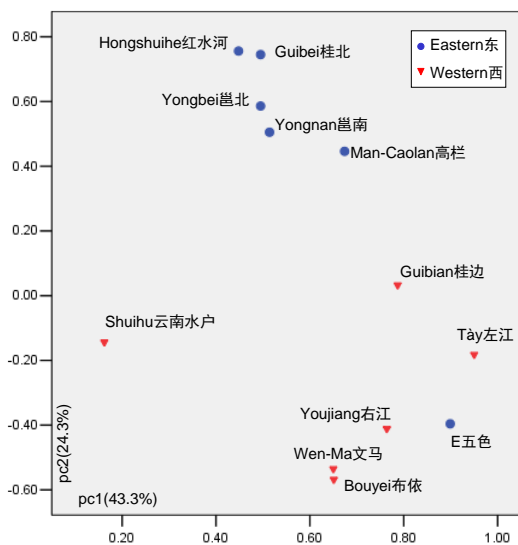


图 2 壮族 Y 染色体 SNP 分型频率的主成份分析图

Fig.2 The principle components plot of Y-SNP frequencies of Zhuang populations

从主成份散点图的横轴来看（横轴代表的是第一主成份），壮族的不同语言支系并没有被区分开；但是从散点图的纵轴来看（纵轴代表的是第二主成份），壮族不同语言支系就明显的被分成了两组。这两组可以在高栏和桂边之间划一界线，界线上方是位于主成份图上部的红水河支系、桂北支系、邕北支系、邕南支系和高栏支系，界线下方则是位于主成份图下部的桂边、左江、右江、文马、布依、五色支系。将这个结果对应到不同支系的地理分布上就不难发现，位于主成份图上部的五个语言支系大体上是分布在今天广西的东部地区，而位于主成份图下部的其他支系则分布于广西的西部。这个结果看起来和壮语方言的南北分布并不一致，这说明壮族内部的遗传结构和内部语言结构并不一

致。那么是什么因素导致了壮族内部遗传结构的区分呢？为了找出分组的因素，就需要对主要的两个主成份和 Y 染色体 SNP 的不同单倍群作相关分析（Correlation analysis），以判断是否有这样一个 Y 染色体 SNP 单倍群，这个单倍群就是将壮族不同语言支系的人群区分成两组的主要因素；如果没有这样一个单倍群，还有没有别的因素使得壮族支系分组。

相关分析是针对第一主成份、第二主成份、各个 Y 染色体 SNP 单倍群、各个壮族语言支系的经度纬度做相互之间的相关分析，得到两两因素之间的相关系数及其 p 值。做相关分析后的第一步，首先需要根据相关分析得出的结果来明确各主成分的实际意义有多大。图 3 是根据各个因素之间的相关系数大小进行排序，然后用不同的颜色表示相关系数的大小，并在正负相关系数之间以黑线标记。左侧的图形显示的就是第一主成份与不同 Y 染色体 SNP 单倍群及不同语言支系经度纬度的相关系数的排序图。我们可以在这张图上发现与第一主成份正相关的因素比较多，但是考察和 pc1 呈正相关的因素（包括单倍型、经度、纬度）与其他因素的关系时就可以发现，它们之间的相关系数并非也总是正号的，而是存在比较多的不一致性，即其间的矛盾比较多，这说明 pc1 包含的结构可能比较复杂，难以从中获得相对简单明了结论。而右侧的图形显示虽然与第二主成份成正相关关系的因素相对少一些，但是其间的矛盾比较少。综合来看，我们初步能够判断：可能第二主成份的分析结果更直观。

图 4 是各种因素相关分析结果列表。图中展示了 Y 染色体 SNP 单倍群、经度、纬度与第一主成份、第二主成份的相关系数及其 p 值。为了在繁杂的数据中显示出有意义的数值，我们使用渐变的色差来区别不同的 p 值，其中红色表示了 p 值小于 0.01，即表示两者之间极显著相关。从图中展示的结果，我们可以看到第一主成份（pc1）与单倍群 O2a 呈正相关（ $r=0.69$, $p=0.02$ ），而与单倍群 O3e 呈负相关（ $r=-0.76$, $p=0.01$ ）。而第二主成份（pc2）则是与单倍群 O* 呈正相关（ $r=0.71$, $p=0.01$ ），与单倍群 O2a 负相关

($r=-0.90$, $p<0.01$)。更为有趣的是我们还发现了单倍群 O* 与各个支系的地理经度正相关，甚至 pc2 与各个支系的地理经度也呈现正相关。这强烈的暗示我们今天壮族人群的分布有可能在地理上也有明显的东西向分布的倾向。

上述分析结果我们实际上使用的仅有一

种分析方法，是否有可能这样的方法存在统计学上的偏差呢？那么我们换一种思路来进行分析，从另一个角度来看主成份分析得到的结果是否可靠。分子方差分析 AMOVA 正是这样的一种方法。AMOVA 实际上是计算人群中间的分子方差，也就是把人群中间

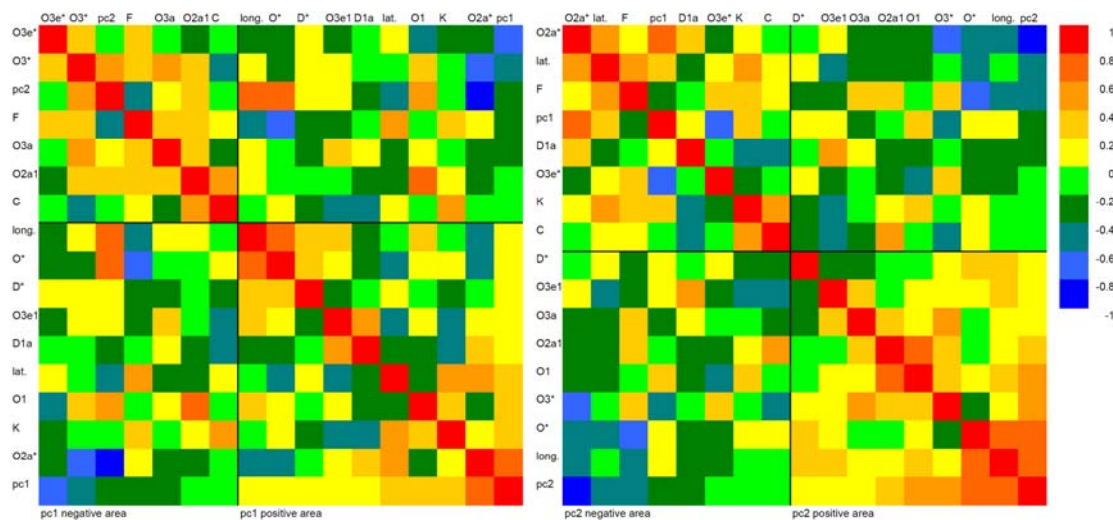


图 3 壮族各支系主成份、Y-SNP 频率、与经纬度的相关性 long.:经度; lat.:纬度; pc:主成分。

Fig.3 Correlations among principle components, Y-SNP frequencies, longitude and latitude of Zhuang branches

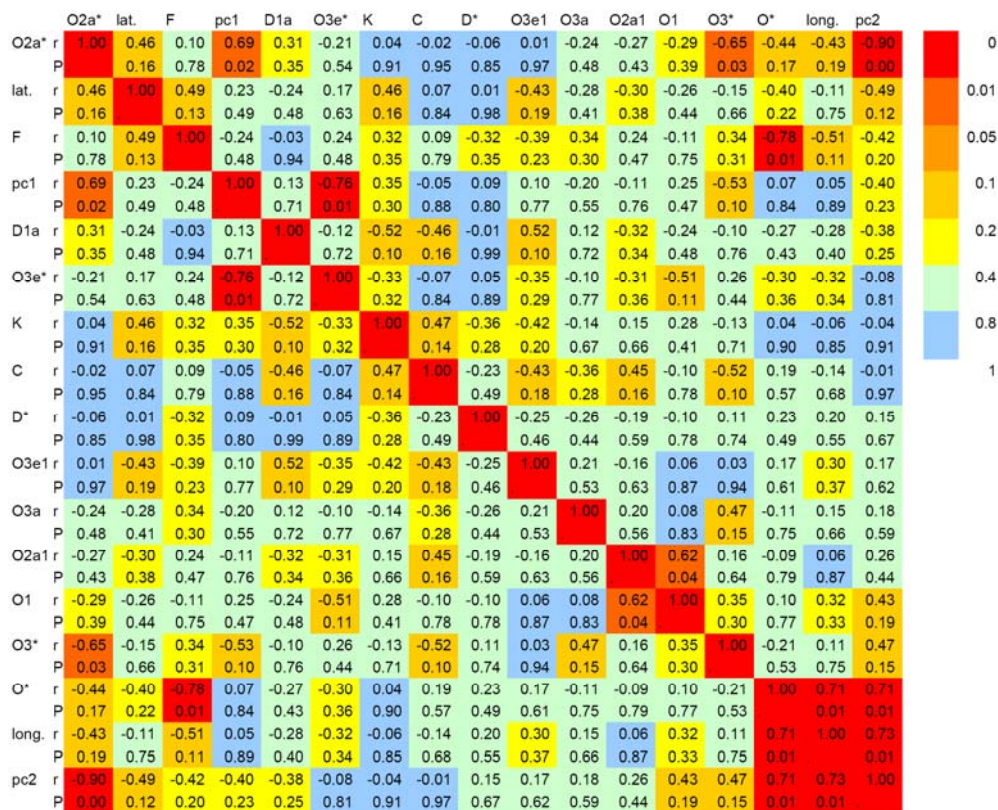

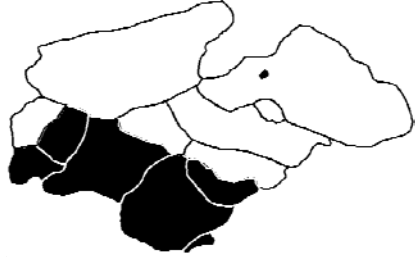


图 4 壮族各支系主成份、Y-SNP 频率、与经纬度的相关性与显著性

Fig.4 Correlations and significances among principle components, Y-SNP frequencies, longitude and latitude of Zhuang branches

表 2 壮族支系南北区分和东西区分的分子方差分析计算结果

Tab.2 Results of AMOVA of northern-southern group and eastern-western group of Zhuang branches

	东西 eastern-western	南北 northern-southern
分组 groups		
组间方差 among groups	17.851	3.461
组内方差 within groups	18.539	32.929

的差异进行层次性地分类鉴别。如果一个人群中间的遗传结构比较一致甚至是高度一致，那么人群内人与人之间在遗传上的差别会很小，他们的组内分子方差也必然比较小；如果一个人群和另一个人群在遗传结构上有所区别，那么两个人群之间的分子方差则必然比较大。根据这两点，我们就能理解：如果壮族不同支系之间存在分组，那么组与组之间的分子方差应该比较大，而组内的分子方差应该比较小。那么语言学上的南北方言的分组是否在遗传结构上也得到体现呢？主成份分析所体现的东西分组又是否在遗传结构上更为合理呢？

以上，我们按照语言南北分组方法和主成份东西分组方法使用 AMOVA 分别计算其组内组间分子方差。结果发现按照语言的南北分组其组间方差很小，远远小于其组内方差。这说明如果壮族支系是以南北分组的话，那么其组内成员在遗传结构上相似性是很小的，这种分组方法对其遗传结构而言并不特别合理。而相反的如果是按照主成份分析结果提示的东西分组，得到的结果则更为满意：尽管东西分组中，组间方差值与组内方差值相当，但是相较于南北分组，东西分组显然在体现壮族内部遗传结构上更为合理。这也就从另一个角度出发，证实了主成份分析的结果。

这种结果实际上指出一个事实：壮族内部语言支系上的区别实际上并没有体现在壮族内部的父系遗传结构上。

三、壮族支系的扩散

从非洲走出来的人最终在世界范围内的表现让我们看到了人类那非凡的开拓精神和相对于其他物种无与伦比的创造能力。就此我不得不感叹现代人不愧是这个世界上最为成功的物种，更不得不感叹在此基础上人类繁衍出来的令人动容的全部文明。因此，也许我们很难相信原来很久很久以前我们系出一根。然而，毕竟我们迥异的文化背景并不能成为人类认识自身的障碍。

就壮族而言，热爱壮族民风民俗的人们也同样热衷于用惯常的文化思维来追溯壮族今天文化的发祥地。所以今天的壮乡人们也在考证壮族起源于广西何处，考证布洛陀传说究竟暗示着壮族怎么样的史前故事，研究花山崖壁（图 5）上传颂的是什么样的神秘记忆，争辩古代称之为“牂牁江”的红水河是否壮族族名之起源，所有这些就构成了今天往往不为外人所知的壮乡人们的文化激情。壮乡并非不认识她的人所想的那样是文化的边缘地带，这里从来有着大自然赋予的旖旎风光，更有着壮族人刻苦内敛质朴而不张扬的山一般的气质。这里的壮族人从来没有辜负过丘壑起伏所蕴藏的财富，这里需要人们用心的解读。

如上一节所述，壮族语言在分布上的区别并没有反映在壮族的遗传结构上。那么，



图 5 花山崖壁画 花山位于中国广西南明县驮龙镇的左江岸边，在它临江的那面高 260 米的山崖上画满了各种呈土红色的图画。据考证，绘画年代在东汉以前，距今已有 2000 年以上的历史了。壮族先民绘制这个大型的崖壁画用意是什么，目前仍是个谜。（图片摘自中华民族艺术网中国民间艺术之旅栏目）

Fig.5 Frescoes of Huashan Precipice Huashan precipice locals at Tuolong town Ningming county of Guangxi China. The 260 meters high precipice facing the river was painted kinds of figures with red color. It was examined that the figures had been drawn before Eastern Han dynasty; thereby it is over 2000 years old. Till now, the purpose of the ancestors of Zhuang nationality painting these pictures is still unclear.

如果壮族语言并不能真正反映壮族人在广西境内的扩散迁徙情况，那么从遗传学角度而言，壮族人可能的迁徙演变情况又是怎样的呢？如果相对人类早期的迁徙，现存语言的发生以及其后的发展是相对近晚的事件，那么用分子遗传学的工具追溯壮族在广西扩散的起点或许更有助于展示壮族在广西的发展演变过程。

为此，我们进一步地展示了壮族人群的发展趋势。我们分别使用壮族各个支系的 $pc1$ 、 $pc2$ 值作为高度值，另外根据壮族各个支系的地理分布确定其经度纬度，以作等高线的方法做出主成份值在地理上的变化趋势图。

两张地形图中我们使用红色表示壮语北部方言支系，绿色则表示壮语南部方言支系。左侧是 $pc1$ 的趋势地图。在上一节内容中我们经过相关分析已经知道 $pc1$ 实际上和 Y 染

色体 SNP 单倍群 O2a 显著正相关，而与 O3e 显著负相关，考虑到 $pc1$ 实际上是壮族人群内结构的某个因素，而且 O3e 在壮族各支系中的分布频率比较低，因此与 $pc1$ 呈正相关的 O2a 才有意义。联系到现在的趋势地形图，我们就能说这张地形图实际上展示的就是单倍群 O2a 在广西内的扩散情况。观察这张地形图，我们发现 O2a 的峰值出现在两处地方：广西与越南的交界、广西北部地区。在以往的数据中我们能够了解单倍群 O2a 是一个比较古老的单倍群，很可能东亚人群在进入东南亚地区之前就已经携带它了。现在我们根据地形图可以推测携带单倍群 O2a 的人群在广西应该是由南向西北、东北两个方向扩散的。这个结果不仅印证了以往的研究分析结果，也指出单倍群 O2a、O*很可能是壮族人群中最古老的单倍群，是壮族人群遗传结构中最底层最基础的部分。

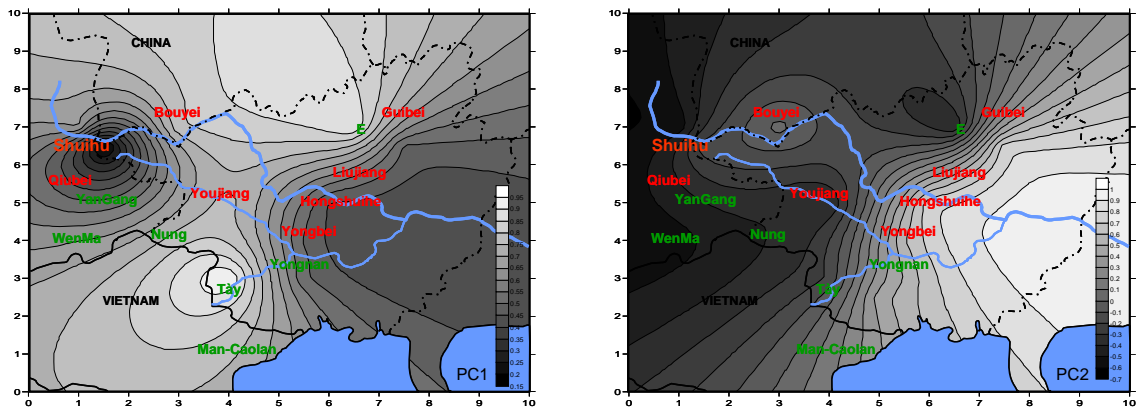


图 6 壮族 Y 染色体 SNP 主成份的地理分布图

Fig.6 Geographic map of Y-SNP principle components of Zhuang populations

相似地，右侧的地形图是 pc2 的趋势地图，由于 pc2 与 Y 染色体 SNP 单倍群 O* 显著相关，我们可以认为这张地形图展示的是携带单倍群 O* 的人在广西的发展过程。这张趋势地图的峰值出现在广西的东部，大致是壮族红水河支系分布的区域。除在桂西北与云南交界处主成分值略有升高外，整张地形图的变化均是呈现由红水河地区向西渐变扩散的趋势。在分析数据的时候我们加入了云南布依族的资料，这可能造成了桂西北与云南交界处主成分值略升高的情况，但这并不妨碍主成份值自东向西渐变下降的趋势，这就说明携带 Y 染色体单倍群 O* 的壮族人在广西境内是自东往西扩散的。再仔细观察 pc2 地形图，我们发现在广西境内的主要河流上，等高线均有所突出，这暗示我们当初壮族人扩散的时候很可能是沿着河流进行的，河流不仅不是阻隔壮族人扩散的障碍，甚至还是他们扩散的便捷通道。

最后，我们来关注另一个很重要的问题。既然我们已经了解到在父系遗传结构的角上壮族人是自东向西扩散的，那么是否存在这样的一个地点，这个地点恰恰是壮族人在广西境内扩散的起点？在第二主成份的地形图中，我们发现主成份值峰值出现在广西的东部，准确来说出现在是壮族红水河语言支系的区域。那么这个区域是否有着更有趣的故事在等待人们去发掘呢？

四、壮族扩散起点

红水河因其河水赤红而得名，在古代，红水河其实有另一个名称“牂牁江”。相传，壮族之族名即来源于红水河的古称“牂牁江”。尽管对壮族在广西境内的起源有众多看法，但是近年来的考古发现却让人不得不去关注红水河地区。

近年来红水河地区发掘了一些古人类遗址，其中就有北大岭遗址（图 7）。尽管北大岭遗址并没有太长的历史，大约 10000 多年，但是这个遗址却在时间上有着相当的连续性。从新石器时代到战国时代，再到汉代、唐宋直至明清时代，北大岭遗址均有相应时期的文物出土，风格也有延续性。而目前在广西发现的其他遗址中，尽管有比北大岭更古老的遗址，但是在时间上并没有明确的延续性。那么依据这一点，我们能否推论：在红水河地区，人类活动从新石器时代到我们能确知的历史时期是一直延续着的。这种延续性是否能给我们更多的关于壮族人的信息呢？让我们继续以分子遗传工具来探索红水河地区是否有着更为重要的信息和意义。

在上一节内容中我们通过主成份趋势地图谈到了红水河区域是第二主成份值峰值所在。我们能否依据这个结果指出红水河区域，或者说壮族人是从红水河方言区开始扩散的。如果这个推论是正确的，那么红水河区域的壮族人必然更为古老，人群基因组的多样性更高，人群 Y 染色体 SNP 单倍群分布频率也必然处在中心的位置。



图 7 北大岭遗址 广西都安瑶族自治县北大岭遗址的发掘，使广西第一次在一处遗址中发掘到自新石器时代至明清时代基本衔接的不同时期的遗迹和墓葬，文化内涵意义重大。（摘自广西壮族自治区博物馆）

Fig. 7 Beidaling site of Neolithic tool workshop The Beidaling site is located in Du'an Yao autonomous county, Guangxi. Different period vestiges and graves were unearthed from this site. It was first time that vestiges and graves were excavated with the burial time lasted uninterruptedly from Neolithic Age to Ming and Qing Dynasty at one site in Guangxi. Therefore this site has a most important significance in regional cultural studies.

前面我们使用主成份分析、相关分析、AMOVA 分析来展示壮族人基本的遗传结构，明确了壮族不同支系在遗传结构上存在的差异，那么红水河区域的壮族人是否更为古老呢？要证明这一点，我们需要做 STR 的网络结构分析。但在此之前，我们需要大致了解各个壮族支系之间的关系。

尽管已经明确壮族不同语言支系在遗传结构上存在差异，并且能区分出东西两个族群，但是我们还不能知道这些不同的语言支系之间究竟存在怎么样的关系。更简单一点说这些支系之间的“亲缘”关系是怎么样的，那些支系彼此更加靠近一些，哪一些相对远一些。这就需要描绘出不同壮语支系的树状结构图。通过聚类分析（Hierarchical

clustering analysis），我们大体能够知道在壮族发展过程中不同的壮族支系处在什么样的位置上。具体的聚类分析结果如图 8。为了做一个参照以避免偏差，我们分析的时候还需要加入外群体，即彝族、瑶族群体的数据，而这些群体已经被证明是不属于百越系统的群体。

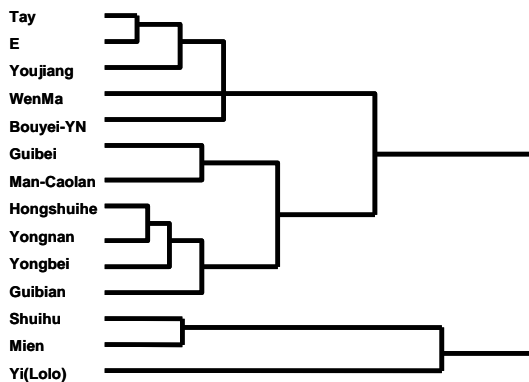


图 8 壮族各支系的 Y-SNP 聚类分析

Fig.8 Dendrogram of Y-SNP of Zhuang branches

聚类分析图实际上是一种树状结构的图形，它很形象地将不同人群的关系展示出来。从这张图中我们观察到加入的外群群体彝族、瑶族确实距离各个壮族支系比较远，符合彝族瑶族不属于百越系统群体的结论。同时我们发现仍旧是红水河、邕南、邕北、桂边这四个支系彼此靠近，左江、右江、文麻、五色、云南布依等支系彼此也比较靠近。这个图形从另一个角度支持主成份分析的结论：广西壮族自东向西的扩散过程。

另外我们也发现一个有趣的现象：主成份分析到现在的聚类分析，我们均发现广西东北部融水县永乐乡的壮族五色人始终与广西西南部的左江壮族（岱）非常接近，可能是比较特殊的迁徙事件。

既然红水河区域是第二主成分值峰值所在，同时红水河、邕南、邕北、桂边这四个支系彼此靠近，那么这些支系是否更为古老呢？我们需要进一步的解答。

尽管相对于主成份分析，聚类分析能够更为细致的展示壮族不同语言支系的关系，但是这些分析都是针对 Y 染色体上 SNP（单核苷酸多态）的分析。SNP 对人群的区分是属于一种大刻度的分析，对于人群内部发育

关系的精细分析则无能为力。具体到一个人群中，往往会存在不同的 Y-SNP 单倍群，这些单倍群在很早之前就形成了，远远早于群体的形成。所以 SNP 单倍群之间的系统关系与所处的各个群体之间的系统关系并不相关，更不能分辨群体内的相同单倍群的个体之间的关系。这就使得 SNP 无法有效的分析出群体内的历史。这时候就需要 STR (Short Repeat Tandem, 短序列重复多态) 这种多向性突变标记对群体的年龄进行估计 [10, 26]。Y-STR 与 Y-SNP 单倍群相联分析时，就成了天然的“时钟”标记。在一个固定的基因位置上，有一种特定的短序列重复了几遍，有如火车车厢。每一节火车车厢就是一小段序列，一般是两个到六个碱基。每次突变都会增加或减少一节，而这种增加或减少是随机的。所以，同一个祖先的不同个体的同一 STR 位置，可能有不同的重复数。突变的方向是双向性的，不再是 Y-SNP 那样只有向一个方向的突变。因为突变的速度是可知的，所以我们只要寻找到最大的重复数差异，再加上其它一些参数，就可以推算出他们共同祖先的生活年代。换一个角度讲，在同一个 Y-SNP 单倍群中，哪个支系的 STR 数据越丰富，这个支系发展的时间就越长，这个支系就越接近这些支系的共同祖先，也越处在中心的位置。

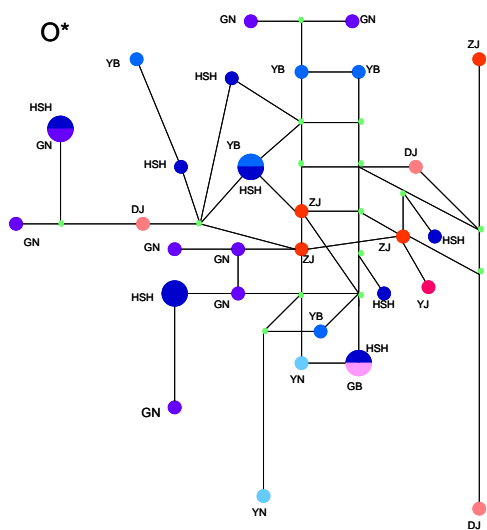


图9 单倍群 O*内部的 STR 网络结构在壮族各支系间的分布

Fig.9 STR network of haplogroup O* of Zhuang branches

明确了 STR 的作用，我们就来看看壮族不同语言支系的 STR 网络结构图 (network)。从前面的分析我们知道 Y 染色体 SNP 单倍群 O*是将壮族不同语言支系区分的主要单倍群，如果要确定红水河支系是否中心，就要看看 Y 染色体 SNP 单倍群 O*的 STR 网络结构图 (图9)。

这张网络结构图中，我们使用各种颜色代表不同的支系。图中看到红水河支系共有 8 种 STR 单倍型共 9 个个体，STR 单倍型最多，其中有三种 STR 单倍型分别与桂边、桂北、邕北的个体共享，并且红水河支系的个体与其他支系的个体的联系最多，尤其是邕北、桂北的个体，处在较为中心的位置。虽然桂北也有 8 种单倍型，但是只有其中一个个体的 STR 单倍型与红水河的个体共享，而且桂北的单倍型分布处在比较边缘的位置。从这个 STR 网络结构图，我们基本能够确定前面我们的推论：红水河支系的壮族相较于别的群体来说更为古老，红水河支系是广西壮族的源头。极有可能当初东亚人的祖先进入广西的时候，他们停驻的第一站就是今天的红水河区域，而后再以此为起点扩散到广西的各个地方。

从图中也看得到左江、右江、德靖这三个支系间联系比较紧密，而与其他支系距离比较远，这也就印证了主成份分析和聚类分析的结果：壮族不同的支系在遗传结构上应该是东西方向的分组更为合理，并且其发展趋势是自东向西的。

在这节内容中，我们通过聚类分析和 STR 网络结构图的分析，得到了壮族红水河支系是今天广西壮族人扩散的起点这个结论，那么这个结论还有没有更多的佐证呢？让我们来看看其它 Y-SNP 单倍群 STR 数据的网络结构图 (图 10)。这三张网络结构图分别是属于 Y-SNP 单倍群 O2a、O3*和 O1。O1 的网络结构图中，红水河支系的 STR 单倍型多样性也高于其他支系，同样处在相对中心的位置。但是 O2a 的图中由于红水河支系 O2a 个体比较少，并没有体现出其中心的位置，左江群体的多样性最高。在 O3*的网络图里面并没有共享的 STR 单倍型，各支系的分布也没有体现出明显的区分，可能与 O3*并非

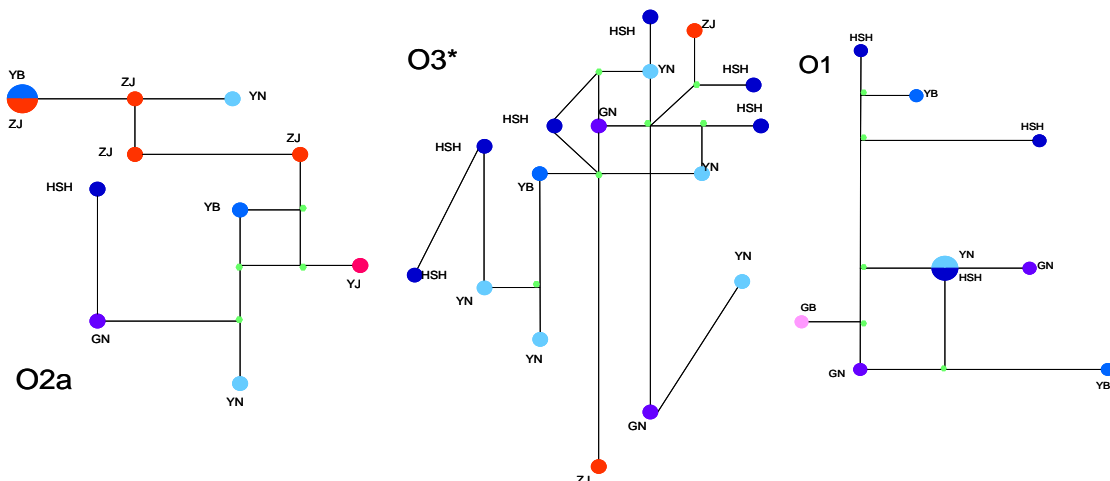


图 10 单倍群 O1、O2a、O3*内部的 STR 网络结构在壮族各支系间的分布

Fig.10 STR network of haplogroup O2a、O3*、O1of Zhuang branches

壮族的特征单倍群有关。由于属于单倍群 O1、O2a 的个体比较少，他们能够传递的信息也相对少，但是在 O1，红水河支系的中心位置还是能看出其轮廓来。更为重要的是 O3*没有出现任何的处在中心的支系个体。这也就佐证了 O3 并非壮族特征的单倍群，是属于外来的成分，从更为精细的角度证明了壮民族的起源是区别于汉民族的。

五、壮族内部遗传层次

经过上述的这些分析，我们通过两种遗传工具使用不同的统计分析方法来展示广西壮族的内部遗传结构，现在我们来整理这些分析结果，使得它更为清晰。我将这些结果概括为壮族的三个遗传层次。

第一个遗传层次是壮民族固有的民族特性。壮民族的起源问题学者们已经在民俗学上、历史学上、考古学上已经得到一个基本的论点，现在我们使用分子生物学的工具进一步肯定了这个论点的正确性。现在从广西壮族不同支系的不同 Y-SNP 单倍群频率分布上看，我们发现其与它现存百越系统人群相似，提示他们的共同起源。单倍群 O2a 在壮族各个支系的高频分布准确无误地传达了这样的信息。而单倍群 O*在壮族支系的普遍高频则提示壮族自身可能存在的特殊发展过程。这两个单倍群也就构成了壮民族遗传层次的最底层，是壮族最基本最特色的遗传结

构。所以无论是在那一直存在的考古文化中，还是在民间代代相传的神话传说中，无论是在历史书籍的明确记载中，还是在今天我们科学地分析中，壮族都是一个有着悠久历史、有着自己族群文化、有着自己独特血缘的民族，他的独特性毫无疑问。

在明确这一点之后，我们就需要去理清壮族内部支系在广西境内是怎样迁徙演变的。通过主成分分析、聚类分析、分子方差计算、趋势地图绘制等等方法，我们能够明确单倍群 O2a 很可能在东亚人进入广西之前就携带有了，其后这个单倍群在广西呈现自南向北的扩散趋势。而单倍群 O*则展示了携带这个单倍群的壮族人更为明确的扩散方式：他以红水河为中心和起点，自东向西并且很可能是沿着河流扩散的。

第二个遗传层次是来自单倍群 O1。O1 单倍群内部结构明显地体现出壮族的某些成分从东部向西部扩散的过程。结合已有的研究资料[22, 25]，我们猜测这一过程在一万年以内，这是壮族遗传结构的第二个层次。可能是西瓯混入骆越的过程。这个过程很可能与百越族群的整体命运以及百越族群的变迁衍化紧紧联系起来，成为一个宏大历史背景中的一个部分，这也将关系到壮族在整个百越族群中的地位问题。这个问题就留待以后更为详细地分析吧。

第三个遗传层次是来自 O3*及其下游单倍群的信息。在我们的分析结果中，我们注

意到 O3*及其下游单倍群在壮族人群又有着一定频率的分布,但是经过 STR 网络结构分析,这些单倍群并没有自己的中心,分布也比较零散。这就使得遗传结构与一些历史记载吻合起来。

自有文字记载的历史纪录,我们发现公元前 213 年,当时统治者秦始皇就曾从中原征调 1.2 万未婚女子至岭南配给驻守在岭南地区的军队。从那时起,岭南成为秦王朝“治狱吏不直者”的发配地。此后历代统治阶级对岭南地区多采取“羁縻政策”,使得众多来自北方的军人、流官的迁入,这导致了壮族原有的基因库经历了延续不断的扩充乃至混杂。因此属于东亚北方人群特征单倍群的 O3*,在壮族人群里面也有一定的比例。但是因为单倍群 O3*并没有区分出明显的组群,也没有明显的中心,因此单倍群 O3*及其下游单倍群应该是历代东亚北方人群进入广西和当地壮族融合留下来的痕迹,属于壮族人群遗传结构中外来的遗传成分,其历史应该在最近的数千年内,这和历史的记载也是相符合的。

壮族是否土著民族,壮族是否经历过奴隶社会,侬智高起义研究,这三个问题是长期以来壮族研究中人们关注的热点问题。近两年来随着研究的深入,学者们对第一第二个问题的看法日趋一致。现在我们使用分子生物手段来探寻壮族族源及其迁徙演变,得到了壮族土著起源的结论,更进一步的理清了壮族不同语言支系的发展演变趋势,证实了前人的研究成果,也理清了一些含混不清的问题。尽管如此,关于壮族遗传结构的研究却远未结束。壮泰族群研究历史上的“泛泰”概念是否有所根据?壮族和老族掸族又存在什么样的关系?甚至从更高的角度来看,壮族在整个东南亚民族中处于什么样的位置?这些都有待我们的努力。我们期待在更进一步的研究中,壮族的发展历史将会得到更加全面的展示。

致谢:感谢复旦大学现代人类学教育部重点实验室对本项研究的支持和帮助,尤其感谢金力教授和李辉博士的指导。

参考文献

1. 张声震(1997) 壮族通史. 北京:民族出版社.
2. 班固(汉)汉书,卷二十八下. 地理志第八下. 粤地注一.
3. 刘安(汉)淮南子,卷十八,人间训.
4. 司马迁(汉)史记,卷一百一十三,南越列传第五十三.
5. 桓宽(汉)盐铁论,卷四,地广第十六.
6. 酈道元(北魏)水注经,卷三十七,淹水叶榆河夷水油水澧水沅水浪水.
7. The Y Chromosome Consortium(2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339-348.
8. Lahn BT, Pearson NM, Jagalian K(2001) The human Y chromosome, in the light of evolution. *Nat Rev Genet*, 2(3):207-16.
9. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlffing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423(6942):825-37.
10. Seielstad M, Bekele E, Ibrahim M, Toure A, Traore M. (1999) A view of modern human origins from Y chromosome microsatellite variation. *Genome Res* 9(6): 558-67.
11. 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 K, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner P(2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358-361.
12. Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL(2001) The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations. *Annals of Human Genetics* 65 (Pt 1): 43-62.
13. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-65.
14. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325 (6099):31-36.
15. Wallace DC, Brown MD, Lott MT(1999) Nucleotide Mitochondrial DNA variation in human evolution and disease. *Gene* 238:211-30.
16. Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R(2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 19:1737-1751.
17. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503-1507.
18. Su B, Xiao J, Underhill P, Dekar R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza LL, Chakraborty R, Xiong M, Du R, Oefner P,

- Chen Z, Jin L(1999) Y chromosome Evidence for a northward migration of modern humans in East Asia during the last Ice Age. *Am J Hum Genet* 65: 1718-1724.
19. Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Dekar R, Su B, Chakraborty R, Jin L (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302-305.
20. Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69(3):615-28.
21. Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP(2003) Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73:671-6.
22. 李辉(2002)百越遗传结构的一元二分迹象.广西民族研究 70(4):26-31.
23. 董永利,杨智丽,石宏,高路,鲁靖,程宝文,李开源,咎瑞光,肖春杰(2004) 云南 18 个民族 Y 染色体双等位基因单倍型频率的主成分分析. 遗传学报, 31(10):1030-1036.
24. 李永念,左丽,文波,柯越海,黄薇,金力(2002)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. 遗传学报, 29:196-200.
25. 金力,褚嘉祐(2006)中华民族遗传多样性研究. 上海:上海科学技术出版社.
26. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int* 129(1):10-24.

编辑提要：壮族是中国人口最多的少数民族，又由于其处于东亚人群最可能的入口地区之一，在中国人群早期起源研究中有独特的地位。该文比较系统地分析了中国南方各地壮族人群的 Y-SNP，Y-STR 多态性，并对其父系遗传结构及其起源和扩散进行了探讨。结果显示壮族在父系遗传上有明显的东西分化，与语言上的南北分化完全不同。该论文提供的数据及结论对东亚人群，特别是百越系人群的遗传人类学研究有一定的借鉴意义。（文波，约翰霍普金斯大学医学系）

Editor's Summary: Zhuang is the most populated minority group in China. They mainly reside in South China, where is probably one of the most important entrances of early modern humans into East Asia. This paper analyzed Y-SNP and Y-STR variations in Zhuang populations from different regions of South China, and the results demonstrate a West-East distinction in the paternal lineage, which is different from the North-South structure as indicated by linguistic data. The author thus provides interesting clues for the genetic anthropological studies of East Asians, especially the Daic-speaking populations. (WEN Bo, Department of Medicine, Johns Hopkins University)