



Genetic affinity between the Hmong-Mien and Mon-Khmer populations

Yan Lu, Xiaoyun Cai, Hui Li

MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China

Email: qtsallow@gmail.com

Received: Sep. 18th, 2011; revised Oct. 20th, 2011; accepted Dec. 15th, 2011.

Abstract: The debate on the origin of East Asian is taking down by the molecular anthropological studies of the various populations in and around East Asia. Some previous studies revealed that most of the Y chromosome lineages of East Asian came from the South. However, the Southeast Asian populations which supposed to be the ancestral populations of East Asian lack for investigations. Therefore, the entrance of East Asian and the formation process of the unique genetic structure have never been studied. Here we present the Y chromosome data of 1652 individuals from 47 populations of Mon-Khmer and Hmong-Mien ethnic groups, which are the aborigines of Southeast Asia and extending area in East Asia. The most frequent haplogroup of the two ethnic groups is O2a*-M95, but is not ethnic specific. Haplogroup O3a4-M7 can only be found in these two ethnic groups, indicating the closest relationship between these two groups. The STR network of O3a4 displayed a layered expansion structure, with the Mon-Khmer haplotypes in the central and the Hmong-Mien haplotypes in the outer layer, the Lolo-Burman haplotypes in the margin. The layered structure was resulted from the slow and even migration of early modern human from Southeast Asia to East Asia by a “Jungle Filter” effect which includes many small bottle-neck effects. The age of O3a4 was estimated to around 27 thousand years, and the migration event was earlier than 16 thousand years ago during the last Ice Age. The unique genetic structure of East Asian might also be resulted from the “Jungle Filter” effect.

Key words: Jungle Filter effect, origin of East Asian, Y chromosome haplogroup O3a4, network analysis

现代人人类学通讯 2011年 第五卷 第 214-223 页 第二届语言进化与遗传进化国际会议

苗瑶与孟高棉人群的遗传同源

陆艳, 蔡晓云, 李辉

复旦大学现代人类学教育部重点实验室, 上海 200433

摘要: 东亚及其周边地区的现代人起源问题一直是分子人类学研究的焦点, 早前的研究显示东亚的 Y 染色体谱系大部分来源于南方。东南亚的群体被认为是东亚群体的祖先, 对这部分群体的分子人类学的调查却相对缺乏, 尤其是对现代人由东南亚进入东亚的过程及独特的遗传结构的形成的研究尚未展开。本文调查了东南亚和东亚周边地区的土著族群: 孟高棉和苗瑶语族的 47 个群体的 1652 个体的 Y 染色体, 结果显示 Y 染色体单倍群 O2a-M95 在多个群体中都存在高频现象, 而单倍群 O3a4-M7 为孟高棉和苗瑶语族所特有, 暗示了这两个族群的较近的遗传关系。基于 O3a4-M7 单倍型的 STR 网络结构图显示出分层扩散的结构, 南亚语系中的孟高棉族群所在单倍型位于网络结构的中心位置, 苗瑶族群处于孟高棉的外圈, 汉藏出现在最外围。这样的分层结构可能源于早期现代人从东南亚进入东亚过程中非常缓慢和均匀的, 由无数个并行的小的“瓶颈”效应组成的“丛林过滤”效应。单倍群 O3a4-M7 整体年龄约为 2.7 万年前, 早于 1.6 万年的末次冰川期。东亚人群的各种遗传特质可能源于这种“丛林过滤”效应。

一、前言

人类自诞生以来，从未停止过对自身的探索，现代人类的起源问题，也总能引起人们的探索热情和无休止的争论。对基因组研究的深入，为人类进化的探索注入了新的活力。对人类线粒体 DNA 和 Y 染色体非重组区上突变位点的研究推动了遗传物质在进化领域的研究，得以分别从母系和父系遗传结构角度构建系统发生树，从而阐明世界各地不同人群之间的系统演化关系。

东亚地区一直在人类进化研究中占据着非常重要的地位，其人群与东南亚和太平洋的人群关系相当紧密，与南亚也有着一定的联系，而且在当地挖掘出大量具有相当连续性的古人类化石。但是对于人群走出非洲后，是由北方还是南方进入东亚地区仍存在很大争议。走出非洲以后，需要绕过青藏高原才能到达东亚，因而不可避免会存在不同的途径，有人认为是从北边进入[1]，而更多对东亚地区人群 Y 染色体非重组区的研究则支持现代东亚人群走出非洲后，从东亚南部进入东亚地区，随后分别向南向北扩张[2-8]。

其间，也就是距今30,000年左右时，产生了对东亚地区现代人口分布影响最为深远的Y染色体O-M175突变。其下游突变形成的单倍群O3-M122、O2a-M95和O1a-M119在东亚人群中所占的频率可达到57%以上，由此不难推断出O型为东亚的主要单倍型，并且Y-SNP和Y-STR的数据也证明O单倍型下面主要的三个分支起源于南方[5,7][9-11]。所以，目前可以确定的是东亚人群的主体从南部经过东南亚进入东亚。研究东南亚和东亚的连接处的群体遗传结构可以探索东亚人群的起源发生过程。

但是由于O2a-M95、O3-M122主干单倍群起源太早，在东亚人群扩散之前就已经有了很高的多样性，所以分布广泛而且杂乱，分析的清晰度并不足以分辨群体之间的关系，因而得出的南方起源的迁徙路线略显粗略。对于较晚产生的单倍群的分析可能得到更清晰的结果。再者，先前对南方入口处人群的研究尚不够深入，对于东亚大陆入口处的两大关键土著群体——孟高棉族群(MK)

和苗瑶族群(HM)(图1)调查非常少。孟-高棉语族包含约100种语言，族群约有35,000,000人，主要生活在中南半岛地区，分布于从印度东北部到柬埔寨和越南等东南亚许多地区。而苗瑶族群则主要分布在中国南部及东南亚地区，包括32种语言，族群在中国就拥有超过12,000,000的人口(2000 Census)。相比较而言，东南亚的汉藏和侗台族群则是较晚到来的外来群体[9-12]，他们分别从东亚北部和东部迁入东南亚的历史不超过3000年。因而孟高棉和苗瑶群体在现代进入东亚过程的研究探讨中格外重要。

本文通过对这两个族群Y染色体SNP和STR位点的完整调查，以及相关群体Y染色体信息的比较，发现单倍群O2a-M95在多个群体中都存在高频现象，仅单倍群O3a4-M7为这两个族群所特有，从O3a4的STR多样性结构上看到了现代人进入东亚过程中的扩散方式信息。

二、结果

2.1 Y 染色体 SNP 单倍群分布

本项研究的样本所涉及的单倍群频率见表1。从表中我们几乎观察不到O1a2-M110的分布，仅在受侗台群体影响较深的傣人(Palyu)群体含有较少比例。P、Q的频率也极低。苗瑶群体里C、D单倍群的分布比孟高棉稍多。显而易见，最普遍存在的单倍群是O2a-M95，在苗瑶类群中最高频率可达到45.16%(O2a*与O2a1之和)，在孟高棉类群中最高频率甚至达到了87.18%，但是这个单倍群在其他群体中也是广泛出现的。而O3a4-M7在其他群体中则鲜见，却为这两个群体所特有，存在族群特色[7,13]，为我们深入研究东亚入口处的族群及其在东亚族群形成中的作用提供了契机。

2.2 群体的聚类分析

为了进一步了解这两个群体之间亲疏关系，以及他们与东亚其他人群之间的亲缘关系，我们又加入了已发表的其他六个群体的单倍型频率数据进行分析[5,9,14]。从主成分分析所绘制出的散点图中，我们可以清晰地发现，东南亚的群体(南亚和侗台)与东亚

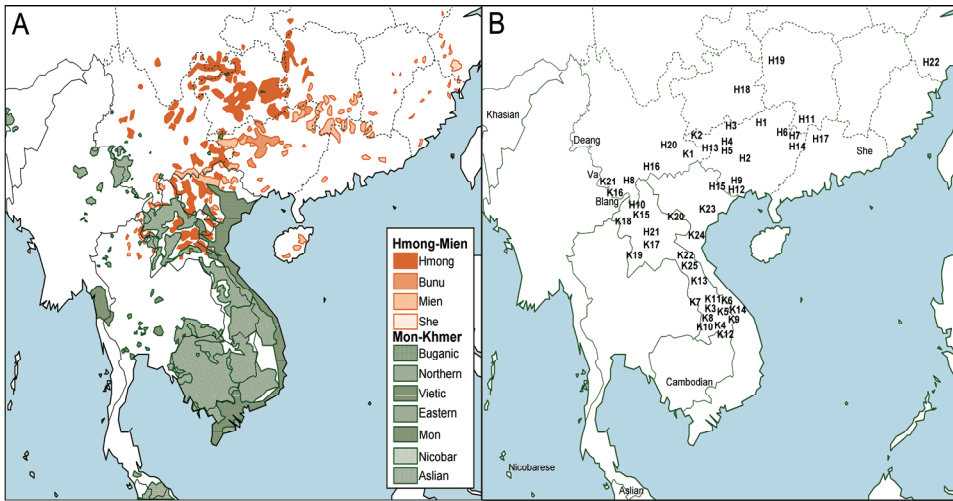


图 1 (A) 苗瑶和孟高棉族群的分布 (B) 本文涉及的采样群体的分布。B 图中群体编码与表 1 对应。

的群体分向两个极端。东亚内部的群体又分为沿海(III)和内陆(II)两大群。而苗瑶和南岛处于这两端之间。说明苗瑶群体在研究人群从东南亚进入东亚过程中的作用举足轻重。

而从同样用 SPSS13.0 绘制的聚类树分析产生的结果中，印度的 Dravidian 和 Indo-Aryan 作为外群，成为整个聚类树的根；汉族和藏缅语族人群距离最近，与我们设想的一样。南亚族群和苗瑶族群总体上非常接近，而且他们的亚群是交错的，其紧密关系可见一斑。

2.3 两大主要单倍群 M7 和 M95 的地理分布

M7 和 M95 是孟高棉语族和苗瑶语族群体两个最主要的两个突变型遗传标记，深入研究这两个单倍群是探讨这两个族群遗传结构的关键。根据这两个遗传标记的突变型频率，我们又绘制了地理分布图，以期看到两个遗传标记的频率在族群和地理上的一些规律。从表 1 和 Fig.2A 中可以清楚地看到，M7 突变型具有较高的民族性，在苗瑶中的畲族、布努中频率较高，在东高棉群体中也是相对高频。但是在侗台、汉族等其他民族系统中则几乎不存在，所以地理分布上被侗台等群体分割成零碎的块状。而 M95 突变型分布在从印度东北部到中国西南部的广大地理区域

内，在各种族群中都具有较高的频率，并没有那么强烈的民族性。这提示 M95 突变型相对于 M7 突变型出现的时间要早得多，至少在整个南亚语系人群形成之初就出现了，并很早就影响了其后形成的整个东南亚和周边的群体。所以 M95 在印度的南亚语系人群中也有发现，并且多样性最高[15]。而 M7 突变型可能只是在人群将要进入东亚的时候产生的，并且没有扩散到过多的群体中。

2.4 两大主要单倍群 M7 和 M95 的网络结构图

对于特定遗传标记的起源和扩散过程，可以通过它所在单倍群的内部多样性的分析进行探讨。我们对观察到的所有 O3a4-M7 和 O2a-M95 个体进行了 Y-STR 分析，并加入了文献发表的数据[7,11,14,16]，分别构建了单倍型网络图(图 4)。

全世界的 Y 染色体分型数据中，带有 M7 突变的 O3a4 单倍型在大量孟高棉和苗瑶人群中均有分布，而在其它人群中则极其罕见，显示出强烈的群体特异性。基于 O3a4-M7 单倍型的 STR 构建网络结构显示出极其明显的分层扩散的结构。南亚语系中的孟高棉族群所在的单倍型位于网络结构中心，苗瑶族群的单倍型大多处于孟高棉单倍型的外圈，汉藏族单倍型则出现在最外围处。这一点

表 1 苗瑶和孟高棉族群中 Y 染色体的 SNP 频率

编号	. 语系	群体	ISO639-3	样本量	单倍群频率															
					C	DE*	D1	F*	K*	O*	O3*	O3a4	O3a5	O3a5a	O1a*	O1a2	O2a*	O2a1	Q1	P*
H1	HM-N	Pahng	PHA	31					19.35	32.26	12.90	3.20	6.50	12.90		12.90				
H2	HM-N	Bunu, Southern	BWX	20	5.00				20.00	15.00	10.00			20.00	10.00	20.00				
H3	HM-N	Bunu, Northern	BWX	10	50.00				20.00		30.00									
H4	HM-N	Bunu, Woodhandle	BWX	6	50.00	16.67					16.67						16.67			
H5	HM-M	Kimmun, Blue	MJI	28			3.57		3.57		3.57	39.29	32.14			17.86				
H6	HM-M	Yao, Lowland	MJI	31	3.23	19.35	3.23			12.90			16.13			16.13	29.03			
H7	HM-M	Kimmun, Lowland	MJI	41	2.44	7.32	2.44	4.88	4.88	14.63		7.32	12.20	12.20		24.39	7.32			
H8	HM-M	Kimmun, Mountain	MJI	32	3.13				3.13	6.25		50.00	6.25			31.25				
H9	HM-M	Mien, Southern	IUM	31				9.68	32.26	6.45	6.45		6.45	6.45		19.35	12.90			
H10	HM-M	Mien, Top board	IUM	11	9.09	9.09	9.09			9.09	9.09		27.27	9.09		9.09	9.09			
H11	HM-M	Mien, Mountain Straggler	IUM	20				10.00		25.00	5.00	10.00	35.00	10.00			5.00			
H12	HM-M	Mien, Flower-head	IUM	19					21.05	15.79	5.26		31.58			21.05	5.26			
H13	HM-M	Mien, Northern	IUM	33	3.03				9.09	21.21		12.12	18.18	9.09		21.21	6.06			
H14	HM-M	Mien, Native	IUM	41		12.20				58.54	7.32	4.88	7.32			7.32	2.44			
H15	HM-M	Mien, Thin board	IUM	11			9.09		18.18	9.09			27.27	18.18		9.09	9.09			
H16	HM-M	Mien, Western	IUM	47		4.30		8.51	14.89	23.40	2.10	4.30	23.40	4.30		14.90				
H17	HM-M	Zaomin	BPN	37	5.41			2.70	5.41	16.22	32.43	2.70	13.51				21.62			
H18	HM-H	Miao, Guizhou	HMQ	49	8.16	2.04		4.08	24.49	24.49	4.08		2.04	12.24		16.33	2.04			
H19	HM-H	Miao, Hunan	MMR	100	14.00	1.00	2.00	9.00	9.00	24.00	4.00	7.00	9.00	7.00		9.00	5.00			
H20	HM-H	Miao, Yunnan	HMD	49	6.12	6.12	2.04	12.24	18.37	12.24			6.12	6.12		30.61				
H21	HM-H	Hmong, Daw	MWW	51	25.49	7.84		5.88	3.92		33.33		7.84	1.96		5.88	5.88	1.96		
H22	HM-S	She, Northern	SHX	56	3.57		1.79		3.57	30.36	25.00	16.07	7.14	8.93		3.57				
K1	MK-B	Bugan	BBH	32				3.13	31.25			6.25	9.38			50.00				
K2	MK-B	Palyu	PLY	30		3.33		3.33	10.00	30.00		6.67	10.00	10.00	3.33	23.33				
K3	MK-E	Alak	ALK	30					6.67		13.33		3.33			56.67	16.67	3.33		
K4	MK-E	Brau	BRB	32			3.13		3.13	3.13	25.00					62.50	3.13			
K5	MK-E	Inh	IRR	34				5.88	2.94		8.82		2.94			79.41				
K6	MK-E	Jeh	JEH	32					6.25		46.88					46.88				
K7	MK-E	Suy	KDT	39					5.13	5.13	2.56					56.41	30.77			
K8	MK-E	Kataang	KGD	37					5.41	21.62	16.22		16.22			10.81	27.03	2.70		
K9	MK-E	Katu	KUF	45					2.22		22.22		6.67			68.89				
K10	MK-E	Laven	LBO	50	2.00	24.00	4.00	4.00	4.00	2.00	12.00		2.00	2.00		42.00	2.00			
K11	MK-E	Ngeq	NGT	35							48.57					48.57		2.86		
K12	MK-E	Oy	OYB	50					2.00		34.00		2.00			60.00	2.00			
K13	MK-E	So	SSS	50				12.00	18.00	6.00	8.00		2.00			42.00	12.00			
K14	MK-E	Talieng	TDF	35	2.86				5.71		22.86		2.86			62.86	2.86			
K15	MK-N	Bit	BGK	28					3.57			10.71	32.14			53.57				
K16	MK-N	Blang	BLR	52	15.38				5.77	9.62	21.15		5.77	11.54		30.77				
K17	MK-N	Khmu	KJG	51	5.88				3.92	3.92	3.92	1.96		13.73		60.78	5.88			
K18	MK-N	Lamet	LBN	35	5.71						2.86	5.71				85.71				
K19	MK-N	Mal	MLF	50					4.00	4.00	2.00	2.00	14.00			66.00	8.00			
K20	MK-N	Xinhmul	PUO	29		3.45			3.45				6.90			17.24	68.97			
K21	MK-N	Ava	WBM	29	6.90				6.90	10.34	44.83		3.45	27.59						
K22	MK-V	Bo	BGL	28					14.29	3.57	7.14			3.57		64.29	3.57	3.57		
K23	MK-V	Kinh	VIE	15			6.67		6.67	33.33		6.67	6.67	6.67		33.33				
K24	MK-V	Muong	MTQ	12			8.33		8.33	25.00		8.33	8.33			41.67				
K25	MK-V	Aheu	THM	38			2.63		13.16	15.79						52.63	15.79			

a. 群体样本号和图 1B 一致。B. 语系缩写如下：HM-N(苗瑶语系，布努语支)；HM-M(苗瑶语系，瑶语支)；HM-H(苗瑶语系，苗语支)；HM-S(苗瑶语系，畲语支)；MK-B(孟高棉语族，布甘语支)；MK-E(孟高棉语族，东孟高棉语支)；MK-N(孟高棉语族，北孟高棉语支)MK-V(孟高棉语族，京芒语支)。

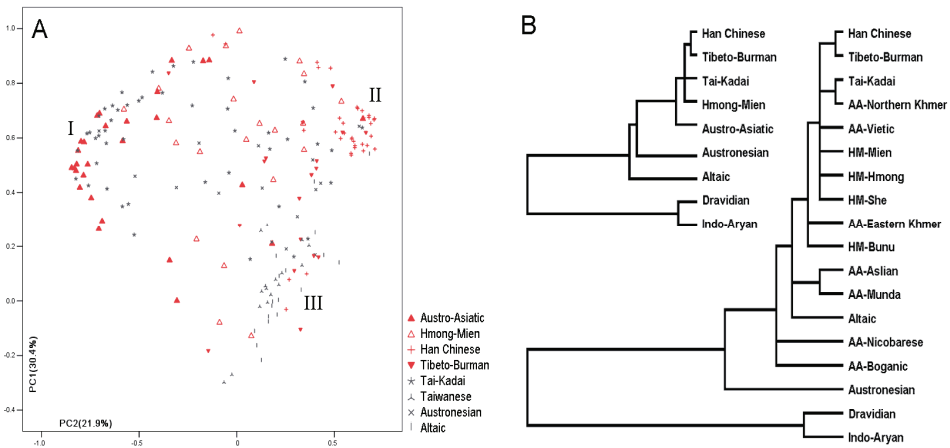


图2 基于 Y 染色体 SNP 频率数据的群体聚类分析 A 主成分分析图 B. 群体聚类分析

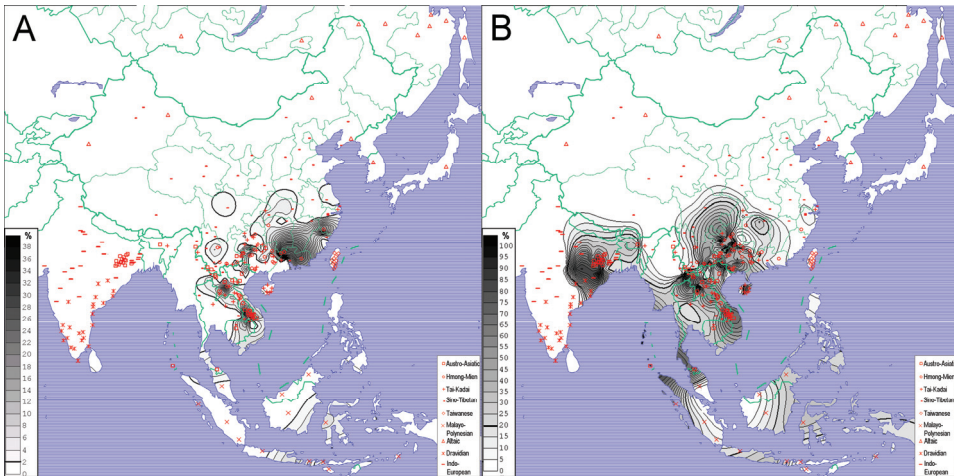


图3 单倍群 M7 (A) 和 M95 (B) 的地理分布及频率

暗示了孟高棉人群的 O3a4 更为古老。这三个层次之间的分界线比较明显，很少有不同的族群单倍型交错出现的情况。这说明整个 O3a4 单倍型在扩散过程中基本是单向地从孟高棉族群流向苗瑶族群，进而流向缅彝族群。并且这种扩散是缓慢的，以至于有足够的时间在新的群体中形成新的单倍型。同时扩散过程也是人口均匀增长的过程，所以在新的群体中多样性（单倍型种类）不断增加。在 O3a4 网络结构中，侗台族群所在单倍型的分布相对没有规律，零星的出现于中间和周边的各处，并不构成独特的分枝，说明侗台

族群中少量 O3a4 单倍型很可能是与苗瑶和孟高棉族群较晚近交流过程中获得的。

基于 O2a-M95 单倍型构建的 Y-STR 网络图中，各种族群共享了大多数的单倍型，并没有类似 O3a4 单倍型的中心扩散的结构。这可能是因为该单倍型过于古老，在族群分化之前就已经有了高度的多样性，各种单倍型分散到了后期形成的各个族群中。值得关注的是，印度的门达群体的单倍型相对处于网络结构的中心位置，这与 O2a 单倍型的早期西部起源的观点一致。

2.5 单倍群 O3a4-M7 的时间估计

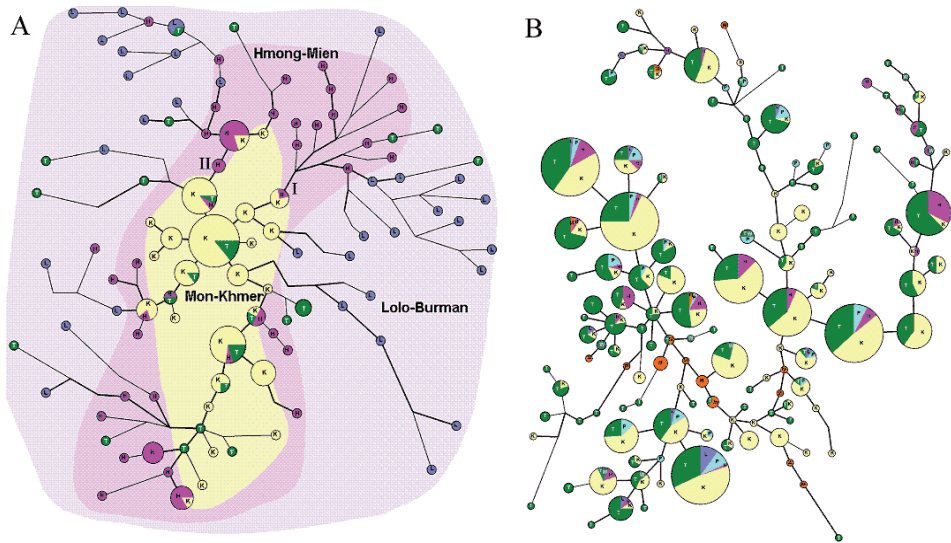


图 4 Y 染色体单倍群 O3a4-M7 (A) 和 O2a*-M95 (B) 的 Y STR 网络结构图

鹅黄色代表孟高棉群体(K), 紫红色代表苗瑶群体(H), 浅紫色代表汉藏群体(L)(在本研究中, 主要是 Lolo-Burman), 绿色代表侗台群体(T), 浅蓝色代表南岛群体(P), 橙色代表南亚门达群体(M), 青色代表台湾群体(W)

O3a4-M7 的网络结构体现出其多态性有明显的族群辨识度, 不同族群在网络结构中聚类在不同的分枝, 对这些分支估算年龄可以推算有关早期群体分化事件的大致年代。Table 2 中是根据 BATWING 计算得到的单倍群整体年龄和几个重要分支的年龄。O3a4-M7 单倍群的整体年龄约为 27 千年, 这与东亚人群最初进入中国南方的时间大致相符, 东亚地区最早发现的现代人是靠近东南亚的广西柳江人, 距今大约 30 千年[17]。在网络结构中上方有两个分支包含的苗瑶和缅彝单倍型比较集中, 在图 4A 中分别标注为分支 I 和分支 II。Table 2 中估计了这两个分支中苗瑶部分和缅彝部分的年龄。可见苗瑶部分的大致年龄在 14-16 千年, 而缅彝部分的年龄大约在 7-12 千年。这些年龄与民族学、考古学和语言学估计的苗瑶、汉藏人群的发生历史比较吻合。

根据古气候学的资料, mtDNA 的所有民族特异性 (特别是苗瑶族群) 亚单倍型都是在 Würn 冰期的最后一个冰退阶段[18], 而 Y 染色体单倍群 O3a4 中苗瑶部分的年龄, 也处

于这一时期。考古学和历史学的一些看法认为苗瑶族群至少在 1-1.7 万年前从“汉藏-苗瑶共同祖先”人群中分化并形成统一体[19]; 古人类学证据表明原汉藏祖先估计至少已经生存了 6,000 年[20,21]。因而这些数据也与古人类学证据和考古学年代基本吻合。

表 2 Y 染色体单倍群 O3a4 的时间估计 (千年)

Network branches	Age	95% CI
Total	26.555	15.145-51.9
HM-I	15.503	10.288-27.487
HM-II	13.652	7.520-30.499
ST-I	12.464	7.555-22.322
ST-II	7.102	4.110-13.104

三、讨论

3.1 苗瑶族群的孟高棉起源

本文研究的两个族群, 孟高棉语族和苗瑶语族是中南半岛和中国西南部的原住民族, 与东亚和东南亚的其他族群比较, 在遗传结构上这两个族群的相似度非常高。这在

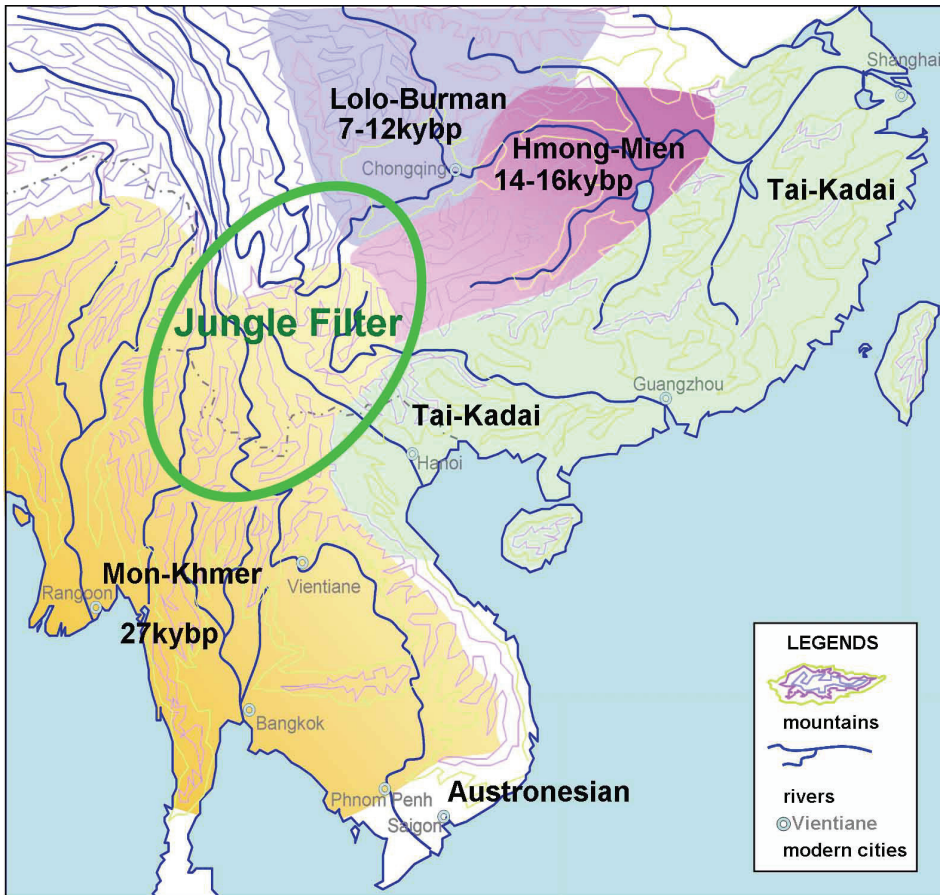


图5 东亚入口处的“丛林过滤”假说

聚类分析中有所体现，在 Y 单倍群的分布上更为明显。苗瑶和孟高棉族群的单倍群虽然都以 O2a* 为主，但它们都有区别于其他族群的特色单倍群 O3a4。O3a4 单倍群在除了孟高棉语族和苗瑶语族之外的族群中频率非常低，甚至不存在，而在孟高棉语族和苗瑶语族中却又相对较高的频率，这说明这两个族群的遗传关系必然比较亲密。O3a4 的 STR 网络结构体现出明显的分层扩散的结构，苗瑶的单倍型分布在孟高棉单倍群的外围，而且年龄在一万年以上，这说明 O3a4 在这两个族群中同时出现并不是晚近时期内族群交流的结果，而是这两个族群有发生关系。苗瑶族群是在一万多年前从孟高棉族群中分化出来的一支，是从东南亚进入东亚的现代人的一支先锋。而汉藏族群则可能是稍后在苗瑶

族群中分化出来的，所以缅彝分支的单倍型又处于苗瑶单倍群的外围。更北方的汉藏族群分支藏、羌、汉中很少见 O3a4 单倍群，可能是在族群迁徙过程中由于遗传漂变这一低频的单倍群丢失了。

在长江中游的考古学遗址中采集的 5000 年前的人类样品中就发现有 O3a4 单倍群 [13]。这一地区是古代苗瑶祖先族群的主要聚居地。这说明至少在 5000 年前，这一单倍群已经在古代苗瑶祖先族群中高频存在了，与我们的年龄估计也是吻合的。

苗瑶和孟高棉族群同源的观点也得到了语言学证据的支持。有观点认为苗瑶语族和孟高棉语族源于同一个大语系原始长江语系 Proto-Yangzian Phylum [22,23]。因此，苗瑶族群起源于南亚语系的孟高棉族群的观点在遗

传学和语言学上都有相关的证据。但是在群体文化方面，苗瑶与孟高棉族群差异非常大，这可能是因为他们的分化在一万年以上，不同的生活环境和较快的文化发展速度使得两者向不同的方向发展，以至于相互的文化面貌截然不同，只有在基因组和语言词汇保留两者关系的痕迹。

3.2 东亚入口处的单向朝北通道

Y 染色体的各种单倍群由于年龄的差异，其内部的多样性结构差异也很大。本文中分析了 O2a*-M95 和 O3a4-M7 的 STR 网络结构，发现 O2a* 的结构比较不规则，在族群之间的分布也没有规律，而 O3a4 则体现出明显的单中心扩散的结构，不同族群处于不同的扩散层次上。对南亚地区的门达族群的研究发现，O2a* 起源于这个族群，年龄将近 65 千年[15]。在网络结构中，门达族群所在的单倍群的确处于结构的中心，符合起源族群的特征。在网络结构中门达和东南亚其他族群之间的分界比较明显，也符合中心扩散的层次结构。但是东南亚和东亚族群在 O2a* 的网络结构中则完全没有结构性。现代人从南亚迁徙到东南亚大约是 5-6 万年前，而从东南亚迁徙到东亚大约是 2-3 万年前。O2a* 的产生正是在现代人从南亚迁往东南亚的过程中，所以在其内部多样性结构中可以观察到南亚向东南亚的分化。而 O2a* 的大量多样性形成在东南亚人群向东亚迁徙之前，对东亚人群的分析就没有足够的清晰度了。而 O3a4 单倍群的年龄为大约 27 千年，是最合适分析人群从东南亚向东亚迁徙过程的单倍群之一。

在东亚和东南亚之间有大量的山岭和丛林，这些是造成族群分化的地理隔离因素。比如安南山脉是分隔侗台和孟高棉的障碍物[14]。而云贵高原的丛林和峡谷是孟高棉和苗瑶之间的过渡层。根据我们的计算，群体至少在 16 千年之前开始从中南半岛穿越这片丛林向东北方向进发。这一时期是末次冰期的高峰期，高山的阻隔作用比现在更强，云贵高原北缘的几座雪山是人类无法逾越的障碍。从本文的网络结构分析中可以推断，群体在穿越这个区域是伴随着人口增长的扩散。由于受到冰川期气候的影响和复杂地形

结构的阻隔，扩散过程非常缓慢和均匀，新生的单倍型只出现在新生的群体中，而不会回流到较早的母群体中，而旧的单倍型在新生群体中却会丢失，以至于在迁徙方向越前方的群体的单倍型处在整个网络结构的越外围。这种现象就像是遗传标记被过滤过一样，新的单倍群才会被滤过，所以可以称为“丛林过滤”效应(图 5)。实际上，“丛林过滤”效应是由无数个并行的小的“瓶颈”效应组成的。东亚人群形成应该正是这种“丛林过滤”效应的结果，正是这种效应造成了东亚人群独特的体质特征和遗传结构，与东南亚的群体有了很大的不同。

研究发现的东南亚南方人群的基因多样性高于北方人群，而且南方的单倍型种类包括北方所有的种类[5]，这都无不揭示着人类进入东亚始于南方，东南亚可能是早期由非洲前来的人群进入东亚的第一站。本文的研究又发现了人类当初进入东亚过程中的“丛林过滤”效应，这不但使东亚人类的起源过程越来越清楚，也有助于东亚人群的各种遗传特质性研究。

当然，从东南亚进入东亚南部的途径可能不止穿过云贵高原丛林的孟高棉-苗瑶这一条，在其东侧可能还有沿海迁徙的侗台-南岛这一条线路[14]，在中国西北部也可能有其他人迁入[24]。通过合适的单倍群的分析，这些过程必然可以一一厘清。

四、材料和方法

4.1 群体采样

本次研究调查的群体涵盖范围较广，我们共采集了 22 个苗瑶群体和 25 个孟高棉群体总共 1652 个无关男性个体(表 1)的外周静脉血或是指尖血。采样遵循知情同意的原则，随机采集无可查亲缘关系的健康男子血样。同一群体的样本在该群体的分布区内尽量分散，以确保样本的涵盖面。这些群体几乎覆盖了东南亚和东亚的交界处。然而有些群体人口很少，因而我们的样本量也相应较小(如木柄瑶)，在研究中也是有其一定意义的。

外周静脉血样品采用常规酚-氯仿法提取 DNA，-70℃ 冻存。指尖血样品则是保

存在滤纸上。DNA 提取时,先在滤纸上切下带血样的 5mm² 小纸片放入 EP 管,用 TE 浸泡过夜,然后 10000rpm 离心 20min 弃上清。再加入 100 μ l TE 震荡混合均匀,每管加入 1U 蛋白酶 K,混匀后 55 $^{\circ}$ C 水浴 2h。之后 95 $^{\circ}$ C 变性 10min。最后每管取 2 μ l 液体进行全扩,全扩产物可用于后续的分型实验。

4.2 Genetic markers

基于东亚和东南亚人群的特殊性,我们一共进行了 20 个 Y-SNP 位点的基因分型。DE^{*}-M1 直接采用 PCR 产物电泳分型。O^{*}-M175, O3a1-M121, O3a5-M134, O3a5a-M117, D1-M15 这五个位点采用 Y 染色体荧光引物 PCR,然后在 ABI 3130xl genetic analyzer 上进行基因组扫描。C-M130, F^{*}-M89, K^{*}-M9, O3^{*}-M122, O3a4-M7, O3a2-M164, O3a3-M159, O1a^{*}-M119, O1a2-M110, O2a^{*}-M95, O2a1-M88, Q1-M120 及 P^{*}-M45 均采用 RFLP-PCR 方法。

另外,为了进一步检测样品的多样性,我们检测了 16 个 Y-STR 位点,分别是 DYS456, DYS389I, DYS389II, DYS390, DYS458, DYS19, DYS385(I 和 II), DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438 和 DYS448。采用多重降落式 PCR 反应条件,PCR 产物在 ABI 3130xl genetic analyzer 上进行基因组扫描,运用 Genemapper 软件进行基因分型数据分析处理。根据实验结果及 Y-DNA Haplogroup Tree 2007 确定各个体的 Y 染色体单倍型。

4.3 数据分析

用 SPSS13.0 软件做主成分分析[21],并进行聚类树分析,观察各群体间的亲缘远近。并使用 Golden Software Sufer7.0 绘制主成分地理分布图。

把研究对象群体中的 O3a4-M7 个体的 Y-STR 单倍型数据用 Network4.2.0.0 软件绘制网络结构图[25];并且把 O2a^{*}-M95 个体的 Y-STR 单倍型数据与相关群体中同样单倍型的 STR 数据用该软件画网络结构图,分析各族群的来源[16]。由于参考的文献报道数据[7,11,14,16]不完整,分析中只用了 DYS19, DYS389I, DYS389II, DYS390, DYS391,

DYS392, DYS393。

在 O3a4 单倍群中用七个 Y-STRs 数据,运用 BATWING 对单倍群的分化时间作出估算[26,27]。计算中假设的有效群体大小 $N_e=2000$,突变率采用 Gusmao et al 的数值 $(0.753-3.507 \times 10^{-3})$ [27]。

参考文献:

1. Nei M, Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 10(5): 927-943.
2. Turner CG, 2nd (1987) Late Pleistocene and Holocene population history of East Asia based on dental variation. *Am J Phys Anthropol* 73(3): 305-321.
3. Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA et al. (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130(1): 139-152.
4. Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ et al. (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci U S A* 95(20): 11763-11768.
5. Su B, Xiao J, Underhill P, Deka R, Zhang W et al. (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet* 65(6): 1718-1724.
6. Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70(3): 635-651.
7. Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA et al. (2005) Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet* 77(3): 408-419.
8. Zhang F, Su B, Zhang YP, Jin L (2007) Genetic studies of human diversity in East Asia. *Philos Trans R Soc Lond B Biol Sci* 362(1482): 987-995.
9. Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D et al. (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* 107(6): 582-590.
10. Wen B, Li H, Lu D, Song X, Zhang F et al. (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* 431(7006): 302-305.
11. Wen B, Xie X, Gao S, Li H, Shi H et al. (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* 74(5): 856-865.
12. Liang M, Zhang JR (2006) The relationship of the Kam-Tai languages and the original dwelling areas and the migration of their peoples. *Studies in Languages and Linguistics* 26(4): 8-26.
13. Li H, Huang Y, Mustavich LF, Zhang F, Tan JZ et al. (2007) Y chromosomes of prehistoric people along the Yangtze River. *Hum Genet* 122(3-4): 383-388.
14. Li H (2005) Genetic Structure of Austro-Tai Populations. Ph.D. dissertation of Human Biology, Fudan University.
15. Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT et al. (2007) Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol* 7:47.
16. Thanseem I, Thangaraj K, Chaubey G, Singh VK, Bhaskar LV et al. (2006) Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genet* 7: 42.
17. Wu RC (1959) Human fossils found in Liujiang Guangxi. *Acta Anthropologica Sinica* 1(3):97-104.
18. Li H, Cai X, Winograd-Cort ER, Wen B, Cheng X et al. (2007) Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol*. in press.

19. Feng H (2007) General history of Yao Nationality (Yaozu Tongshi). Beijing: Minzu Press
20. Martisoff JA (1991) Sino-Tibetan linguistics: present state and future prospects. *Annu Rev Anthropol* 20:469–504.
21. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton, NJ: Princeton University Press.
22. Starosta S (2005) Proto-East Asian and the origin and dispersal of the languages of East and Southeast Asia and the Pacific. In: Sagart L, Blench R, Sanchez-Mazas A, editors. *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London: RoutledgeCurzon. pp.182-197.
23. Sagart L, Blench R, Sanchez-Mazas A (2005) Introduction. In: Sagart L, Blench R, Sanchez-Mazas A, editors. *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London: RoutledgeCurzon. pp.1-14.
24. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I et al. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244-10249.
25. Bandelt HJ, Forster P, Rohlf A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1): 37-48.
26. Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc Ser A Stat Soc* 166(20):155-188.
27. Gusmao L, Sanchez-Diz P, Calafell F, Martin P, Alonso CA et al. (2005) Mutation rates at Y chromosome specific microsatellites. *Hum Mutat* 26(6): 520-528.