

Y 染色体单倍型在中国汉族人群中的多态性分布与中国人群的起源及迁移*

柯越海^① 宿兵^{①②③} 肖君华^① 陈华^① 黄薇^④ 陈竺^④
褚嘉佑^⑤ 谈家桢^① 金力^{①③} 卢大儒^{①**}

(^①复旦大学遗传学研究所摩尔根-谈家桢国际生命科学研究中心, 上海 200433; ^②中国科学院昆明动物研究所, 昆明 650031; ^③德克萨斯大学休斯顿分校人类遗传中心, 休斯顿 77225, 美国; ^④中国国家人类基因组南方研究中心, 上海 201203; ^⑤中国医学科学院昆明医学生物研究所, 昆明 650031)

摘要 观察了由 19 个单核苷酸多态位点(SNP)组成的 Y 染色体单倍型在全国 22 个省市汉族人群中的分布. 结果表明, 中国南北人群的 Y 染色体单倍型组成有较大差异, 南方人群的多态性明显高于北方人群, 而后者中的单倍型仅包含前者的一部分, 其中单倍型 H7, H10, H11 和 H12 仅出现在南方群体. 这一观察结果与中国南北少数民族人群间差异相符, 提示现代人类自南方进入中国, 随后由南向北逐渐迁移. 同时对携带南北人群共同的单倍型个体在 3 个 Y 染色体微卫星标记位点进行了基因组分型, 据此估算了现代人类进入中国的时间大致在 18 000~60 000 a 前.

关键词 Y 染色体单倍型 SNP 现代人起源与迁移 中国人

亚洲与其他各大洲直接或间接相连, 是研究早期现代人类起源与迁移的一个关键地区. 近年来, 越来越多的遗传学证据支持现代人类起源于非洲, 即所谓的“走出非洲”假说^[1, 2]. 尽管我们对现代人从非洲向欧洲、美洲和大洋洲的迁移已有许多研究, 但对早期现代人类如何到达东亚(包括中国)地区以及而后怎样遍布东亚大陆所知甚少^[3]. 另外, 对于现代人的起源一直存在着多地区起源的不同观点, 认为包括中国在内的亚洲、非洲和欧洲的现代人是由当地的智人演化而来的, 因此认为现代中国人是独立起源的, 这一假说的主要支持证据来源于对中国晚期智人骨骼的体质形态特征研究和在东亚发现的人类化石(包括从直立人到现代人的化石)在时间及形态变化上均呈较好的连续性的考古发现^[4~8], 尽管这一连续性的可靠性尚有争议^[9, 10], 但仍被认为是支持东亚现代人独立起源的有力证据. 近来褚嘉佑等人利用 30 个常染色体微卫星位点研究了中国人群的遗传结构, 证实了中国南北人群间的遗传差异^[11, 12], 其结果与人类非洲起源说一致, 对东亚人群的独立起源说提出了质疑^[13].

众所周知, 由于微卫星位点突变率较高, 其作为遗传标记研究人类进化和人群迁移有一定的局限性, 尤其对于较为久远的进化事件往往缺乏分辨力, 因而褚嘉佑等人的研究难免缺

1999-11-29 收稿, 2000-05-15 收修改稿

* 国家自然科学基金重大资助项目(批准号: 39993420)

** 联系人(E-mail: drlu@fudan.edu.cn)

乏令人信服的证据支持非洲起源假说, 近来许多研究者意识到 Y 染色体上的遗传标记是研究早期人类起源和迁移的理想工具^[14~18]. 作为父系遗传的 Y 染色体, 它的绝大部分为非重组区, 故单倍型保持完整, 不易受重组和回复突变影响, 突变率低, 比微卫星标记更能稳定遗传, 是进化事件的忠实记录者. 再者, 由于 Y 染色体为单倍体, 其有效群体大小远小于常染色体位点, 较易产生人群特异的单倍型. 通过对 Y 染色体 19 个单核苷酸多态性(single nucleotide polymorphism, SNP)组成的单倍型在中国汉族人群中的分布, 研究早期现代人类在东亚地区的迁移路线, 同时通过对特定 Y 染色体单倍型的微卫星位点多态性的分析, 推算早期现代人进入东亚大陆的时间.

1 材料与方法

1.1 样品采集和基因组 DNA 制备

共采集了来自中国不同地区的 362 份男性个体血样, 这些人群分别来自中国 22 个省、市、自治区. 每个人的省籍根据其祖父母及外祖父母的出生地与籍贯一致而定, 并将其归为 16 个地区. 北方人数较少者归入其他类. 外周血样按文献^[19]常规操作酚氯仿提取基因组 DNA, 定量标化 DNA, -70°C 保存备用.

1.2 Y-SNP 和微卫星(STR)遗传标记的基因分型

Y 染色体的单倍型由 19 个 SNP 组成, 其中 7 个 SNP 文献中已有报道^[17, 18, 20~22], 而其他 12 个 SNP 则是在我们最近的研究中报道的^[23]. SNP 位点的基因分型采用等位基因特异 PCR 方法 (allele specific PCR), 对于每一个 Y 染色体的 SNP 位点, 设计两个特异等位基因引物分别在其上游或下游单侧引物的 3' 端含有 SNP 突变的位点, 来识别单碱基差异的等位基因. PCR 反应基本条件: 10 μL 体系, 含 20 ng/ μL DNA 模板 1 μL , 两个引物各 2 pmol, 2.5 mmol/L $4 \times \text{dNTP}$, 10 $\times$ PCR buffer, 0.5 U Taq DNA 聚合酶. 本研究 16 个 Y 染色体 SNP 的引物序列及 PCR 扩增条件见表 1. M15Taq DNA 聚合酶采用普通 Taq 酶, 其余均为 Bioline 公司的 Biolase Platinum DNA 聚合酶. PCR 产物取 2 μL , 4% 琼脂糖 DNA 凝胶电泳观察结果.

Y 染色体微卫星位点 DYS389, DYS390 和 DYS391 的基因分型, 3 对荧光标记引物购自 Perkin-Elmer 公司. PCR 反应体系 5 μL , 含 10 $\times$ PCR buffer 0.5 μL , 3.0 mmol/L Mg^{2+} 0.6 μL , 4×0.2 mmol/L dNTPs 0.1 μL , 0.2 U TaqDNA 聚合酶 0.1 μL , 3 对引物各 0.05 μL , 20 ng/ μL DNA 模板 1 μL , 其余用去离子水补足 5 μL , 采用多重降落式 PCR 反应条件, PCR 产物在 ABI373 全自动 DNA 测序分析仪(Perkin Elmer)上进行基因组扫描, 运用 GenotyperTM 软件进行基因分型数据分析处理, 具体基因分型方法见以前的报道^[18].

1.3 人群迁移时间推算的统计方法

根据一步突变模型, 推导出以下公式作为推算时间的依据:

$$t = -N_e \ln(1 - V / N_e i),$$

其中, t 为时间(以代数计); i 为突变率; N_e 为有效群体大小; V 为微卫星位点重复单元数量的方差.

表 1 16 个 Y 染色体 SNP 的引物序列及 PCR 扩增条件

位点		引物序列	Mg ²⁺ 浓度/mmol · L ⁻¹	退火温度/℃
M122	上游引物(1)	AAT TGA GAT ACT AAT TCA C	6	50
	上游引物(2)	AAT TGA GAT ACT AAT TCA T		
	下游引物	AAA ACT TTA TCA TAT TGA G		
M120	上游引物	GAG CTT GGA CTT TAG GAC GG	4	60
	下游引物(1)	TTT CCC TTA AAA ACA GCA TGG		
	下游引物(2)	TTT CCC TTA AAA ACA GCA TGA		
M119	上游引物	ACT CAC CCT AAG GAA GTC ACG A	4	60
	下游引物(1)	CCA ATT CAG CAT ACA GGC G		
	下游引物(2)	CCA ATT CAG CAT ACA GGC T		
M7	上游引物	TGG ATT AGT CAC TTT AGA CCT AC	4	60
	下游引物(1)	GTA GTT GAG TTA CTG TTC TTC		
	下游引物(2)	GTA GTT GAG TTA CTG TTC TTG		
M50	上游引物(1)	GCC TAC CCA AAC CAC ACC T	4	60
	上游引物(2)	GCC TAC CCA AAC CAC ACC C		
	下游引物	GGA AGC CTG TGT CTA CTC TGC		
M110	上游引物	AAA AGG CTG TGG TGC TGA TC	4	60
	下游引物(1)	AAT GCA ACA GTT TAC AAG AAC ATG		
	下游引物(2)	AAT GCA ACA GTT TAC AAG AAC ATA		
M111	上游引物(1)	AGA ACA AGT TCT GTT TTT CAC ATT G	4	54
	上游引物(2)	AGA ACA AGT TCT GTT TTT CAC AG		
	下游引物	AAG AGA TGA AGA TAC CTT ATA TGC CC		
M89	上游引物	AGA AGC AGA TTG ATG TCC C	4	60
	下游引物(1)	TCA GGC AAA GTG AGA GAT G		
	下游引物(2)	TCA GGC AAA GTG AGA GAT A		
M88	上游引物(1)	CTT ATT CCT GCT TCT TCT GCG	4	58
	上游引物(2)	CTT ATT CCT GCT TCT TCT GCA		
	下游引物	AGG TGT GAC CAC AGA GAC TCA G		
M103	上游引物	CTC CAA GGA CAC AGA ACA GG	4	58
	下游引物(1)	CAA GAC CAG AGA AGG TGG G		
	下游引物(2)	CAA GAC CAG AGA AGG TGG A		
M45	上游引物(1)	GGC AGT GAA AAA TTA TAG ATA G	4	58
	上游引物(2)	GGC AGT GAA AAA TTA TAG ATA A		
	下游引物	ACC TTC CAC AGA CCA CAG		
M9	上游引物(1)	GCC TAA GAT GGT TGA ATC	4	54
	上游引物(2)	GCC TAA GAT GGT TGA ATG		
	下游引物	CTC AAG CGT AAA TGT ACT GT		
M95	上游引物(1)	GAT AAG GAA AGA CTA CCA TAT TAG TGC	4	62
	上游引物(2)	GAT AAG GAA AGA CTA CCA TAT TAG TGT		
	下游引物	GGG TGG GTG TGT TTG AAG G		
M17	上游引物(1)	TTG CTG GTT GTT ACG GGG	4	52
	上游引物(2)	GTTG CTG GTT GTT ACG GGT		
	下游引物	GCT ATT CTT GTT TCT CCA GGC		
M134	上游引物	AGA ATC ATC AAA CCC AGA AGG	4	52
	下游引物(1)	TAC TTT TGA TCC CCA CCA ATC		
	下游引物(2)	ATAC TTT TGA TCC CCA CCA ATA		
M15 ^{a)}	上游引物	GAG CTT GGA CTT TAG GAC GG	2	56
	下游引物	GTC TGG GAA GAG TAG AGA AAA G		

a) 可使用普通 DNA Taq 扩增

2 结果

从 19 个 Y 染色体 SNP 的基因分型中得到 17 个单倍型 (图 1). 表 2 列出了 17 个 Y 染色体单倍型在中国汉族及世界各大洲人群中的分布. 其中, 非洲人为 Lissongo, Biaka Pygmy 和 Mbuti Pygmy; 美洲人为 Karitiana, Guavm 和 Trapa; 欧洲人为意大利人及北欧人; 大洋洲人为 Nasioi Melanesian 及新几内亚 Highlander.

3 讨论

H1 和 H2 两种单倍型相对古老一些, 在非洲和非洲以外的群体中均有发现, 这意味着这样的突变在早期现代人类走出非洲之前就已经发生了. 在 M9 位点上 C→G 的突变, 所有非洲人都是 C, 而 G 的频率在非洲以外的人群中占 84.3%, 这显示大多数非洲以外的群体具有较近的共同祖先^[17]. H5 单倍型是其他各种区域性特征分布的单倍型(H6~17)的祖先单倍型, 因此, 可能是在走出非洲后发生的. H15 是美洲印第安人特有的单倍型, H17 是大洋洲人特有的单倍型^[20]. H14 在欧洲人群中具有较高的频率, H6~13 则为东亚人群特有的单倍型, 这一明显的群体特异性为研究人群迁移和人群间基因交流提供了理想的工具. H6~13 单倍型仅出现于东亚地区, 在东亚以外的群体中尚未发现, 尤其是 H6~8 均呈 M122 位点的突变(M122C), 这些单倍型在东亚人群中占大多数, 尤其是在汉族中占 54.1%. 因此, M122C 在东亚人群中的特征分布提示东亚人群来自同一祖先. 然而在比较中国以长江为界的南北人群的差异时, 发现南方人群中的遗传多态性明显高于北方人群. 这一差别表现如下: (i) 南方人群中的亚洲特异单倍型的数目多于北方人群, 其中 H7, H11 和 H12 仅出现在南方汉族, 在北方汉族中没有, 而北方汉族之单倍型在南方汉族中均有出现. 这些在南北汉族群体中的差异在南北非汉族人群比较中更为显著^[23], 这可能是由于汉族的南北基因交流远大于非汉族少数民族. (ii) 南方人群的基因多样性高于北方人群, 各人群的遗传多样性按北方非汉族人群、北方汉族人群、南方汉族人群、南方非汉族人群(东南亚人群)排列分别为 62.4%, 71.5%, 76.1%和 81.3%, 逐渐上升, 揭示人类进入东亚始于南方. 东南亚群体中的单倍型几乎涵盖全部中国和亚洲的特异单倍型, 因此东南亚可能是早期由非洲迁来的人群进入东亚的第一站, 从那儿开始我们的祖先从东南亚

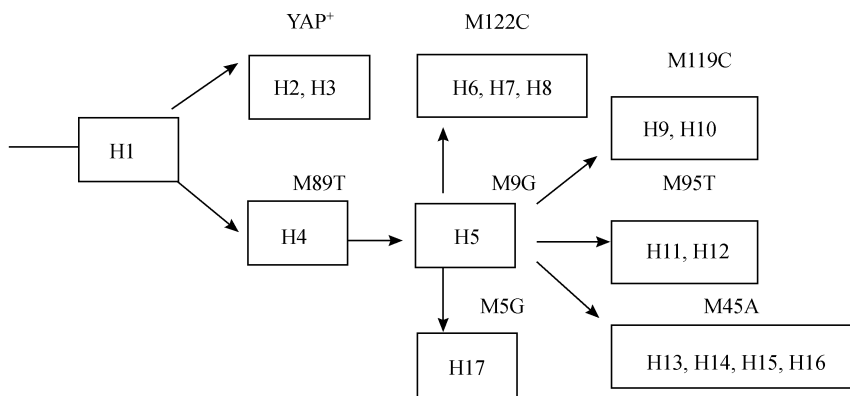


图 1 17 个 Y 染色体单倍型的系统进化关系图

H1 为祖先类型, 其他单倍型均由 H1 衍生而来. 方框上为各单倍型所特有的多态位点类型. 各单倍型在所有 19 个位点上的等位基因类型见文献^[21]

表 2 Y 染色体单倍型在 16 个汉族群体以及世界其他人群中的频率分布

地 区	群体	样本	H1	H2	H3	H4	H5	H6 ^{a)}	H7 ^{a)}	H8 ^{a)}	H9 ^{a)}	H10 ^{a)}	H11 ^{a)}	H12 ^{a)}	H13 ^{a)}	H14	H15	H16	H17
东北亚	布利亚特人	4	75.0			25.0													
	鄂温克族	8	50.0				12.5	12.5		25.0									
	满族	18	16.7			11.1	22.2	27.8		16.7	5.6								
	蒙古族	24	58.3	4.2		8.3	12.5	4.2		4.2	4.2							4.2	
北方汉族	藏族	8		12.5	25.0	12.5				50.0									
	辽宁	6	16.7				50.0				16.7				16.7				
	河北	6					16.7	50.0		33.3									
	山东	32	9.4			3.1	18.8	28.1		28.1	9.4				3.1				
	河南	28	7.1			3.6	25.0	32.1		14.3	10.7				7.1				
	其他	10	10.0				10.0	30.0		40.0	10.0								
	安徽	22	13.6				18.2	27.3		18.2	18.2				4.6				
	浙江	50	12.0				6.0	24.0		26.0	26.0		6.0						
	江苏	55	7.3	1.8		1.8	18.2	23.6	3.6	21.8	16.4		3.6		1.8				
	上海	30	6.7				16.7	26.7	3.3	16.7	26.7				3.3				
南方汉族	湖北	18	5.6				11.1	27.8	5.6	33.3	16.7								
	四川	14	7.1				28.6	7.1	7.1	35.7	7.1		7.1						
	江西	21	4.8			9.5	19.1	23.8		23.8	14.3		4.8						
	湖南	15					13.3	33.3		26.7	13.3		13.3						
	福建	13	7.7				7.7	38.5		38.5				7.7					
	云南	27	11.1			3.7	3.7	18.5		55.6	3.7		3.7						
	广东	15						40.0		26.7	20.0			6.7	6.7				
	壮族	28	3.6		3.6	7.1	3.6	3.6		25.0	17.9		25.0	10.7					
	柬埔寨	26	3.8		3.8	11.5	11.5	3.8		15.4	3.8	3.8	23.1	11.5		3.8		3.8	
	泰国人	20				5.0	5.0	5.0	5.0		5.0	5.0	45.0	20.0		5.0			
东南亚	马来亚人	13				7.7	7.7	30.8		15.4	7.7	23.1	7.7						
	Batak	18	5.6			5.6	11.1	11.1	16.7		22.2		27.8						
	爪哇人	11	9.1			9.1	27.3	9.1			18.2	9.1	18.2						
	非洲人	24	20.8	79.2															
世界其他各洲	美洲人	26														3.8	96.2		
	欧洲人	39	10.3			12.8	25.6									51.3			
	澳大利亚人	93	15.1			2.2	41.9												40.9

a)为东亚人特有的单倍型

进入中国的南方,而后越过长江进入北方地区.这一发现与线粒体 DNA 单倍型分布相符^[24].

为了估计早期现代人类进入中国的年代,我们对含 3 个具有汉族特异性的单倍型(H6~8)的个体在 3 个微卫星标记位点 DYS389, DYS390 和 DYS391 进行基因型分型.为了尽量减少由于估算群体结构的误差带来的影响,共筛选并分析了 160 个携带有 M122C 汉族特异性单倍型的个体.根据一步突变模型推导出计算瓶颈效应年代公式 $t = -N_e \ln(1 - V/N_e i)$, 该公式适用于经历了一次较强的瓶颈效应后产生快速增长的群体,对于公式中有效群体大小的估算,目前广泛接受的现代人的有效群体大小为 5 000~10 000, 因为考虑到亚洲群体比非洲群体的遗传多样性要小,我们假设亚洲群体的有效群体大小为 750~2 000, 突变率为 0.18%^[25, 26], 20 年为一代,据此得出 M122C 突变发生的年代大致是在 18 000~60 000 年前(此年代由 DYS390 推算而得,是 3 个微卫星位点中最大的).一般而言,单纯从遗传学角度对古代人类的群体迁移和基因突变年代的精确估算困难非常大,因为在上述公式中有效群体大小和突变率的估计带来的偏差直接影响结果,但我们作出这样的年代估算与体质形态学和考古学在齿型、石器和化石的发现相符^[27~30].考古研究表明,涵盖中国北方地区的北亚人种特有的 *Sinodont* 齿型发生在 18 000~25 000 年前,而这种齿型是由东南亚地区的类 *Sinodont* 齿型演变形成的,这与我们公式推算出的 M122C 发生年代下限相符.此外,在西伯利亚、贝加尔湖和阿尔泰地区近来的考古研究表明,这一地区的人类石器文明出现在 25 000~45 000 年前,这样当我们接受了包括西伯利亚和大洋洲在内的整个东亚群体是由非洲起源并经东南亚迁移而形成的事实,那么现代人类进入东亚地区应该在北亚出现石器文明之前,这样认为 M122C 突变估算年代的下限 60 000 年前更接近于事实了,因此我们认为 18 000~60 000 年是早期现代人迁入东亚所造成的瓶颈效应的年代.

据此,我们认为随着冰川期逐渐消亡,非洲起源的现代人约在 60 000 年前从南方进入东亚,在以后的数万年中逐渐向北迁移,遍及中国大陆,北及西伯利亚.大约在 8 500 年前,经历了漫长的蒙昧时期后,以仰韶文化为代表的最早的中华文明开始在黄河中上游地区萌芽.

参 考 文 献

- 1 Cann R L, Stoneking M, Wilson A C. Mitochondria DNA and human evolution. *Nature*, 1987, 325: 31~36
- 2 Vigilant L, Stoneking M, Harpending H, et al. African populations and the evolution of human mitochondrial DNA. *Science*, 1997, 253: 1 503~1 507
- 3 Cavalli-Sforza L L, Piazza M P. *The History and Geography of Human Genes*. Princeton: Princeton University Press, 1994
- 4 Brooks A S, Wood B. Paleoanthropology. The Chinese side of the story. *Nature*, 1990, 344: 288~289
- 5 Li T, Etler D A. New middle Pleistocene hominid crania from Yunxian in China. *Nature*, 1992, 357: 404~407
- 6 Wu X Z, Poirier F E. *Human evolution in China*. Oxford: Oxford University Press, 1995
- 7 Etler D A. The fossil evidence for human evolution in Asia. *Annu Rev Anthropol*, 1996, 25: 275~301
- 8 Wolpoff M H. Interpretations of multiregional evolution. *Science*, 1996, 274: 704~707
- 9 Stringer C B, Andrew P. Genetic and fossil evidence for the origin of modern humans. *Science*, 1988, 239: 1 263~1 268
- 10 Wilson A C, Cann R L. The recent African genesis of humans. *Scientific American*, 1992, 4: 68~75
- 11 Weng Z, Yuan Y, Du R. Analysis of the genetic structure of human populations in China. *Acta Anthropol Sin*, 1989, 8: 261~268
- 12 Zhao T, Zhang G, Zhu Y, et al. The distribution of immunoglobulin Gm allotypes in forty Chinese populations. *Acta Anthropol Sin*, 1986, 6: 1~8

- 13 Chu J Y, Huang W, Kuang S Q, et al. Genetic relationship of populations in China. *Proc Natl Acad Sci USA*, 1998, 95: 11 763~11 768
- 14 Jobling M A, Tyler-Smith C. Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics*, 1995, 11: 449~455
- 15 Oefner P J, Underhill P A. Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC). *Am J Hum Genet*, 1995, 57: A266
- 16 Oefner P J, Underhill P A. *Current Protocols in Human Genetics*. New York: Wiley & Sons, 1998. 1~12.
- 17 Underhill P A, Jin L, Zemans R, et al. A pre-Columbian human Y chromosome-specific transition and its implications for human evolution. *Proc Natl Acad Sci USA*, 1996, 93: 196~200
- 18 Underhill P A, Jin L, Lin A A, et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research*, 1997, 7: 996~1 005
- 19 Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning, A laboratory Mannual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989
- 20 Vollrath D S, Foote S, Hilton A, et al. The human Y chromosome: A 43-interval map based on naturally occurring deletions. *Science*, 1992, 258: 52~59
- 21 Underhill P A, Jin L, Lin A A, et al. An African origin of human Y chromosomes. Program and Abstract, 1997 Annual Meeting of The American Society of Human Genetics. *Am J Hum Genet*, 1997, 61(supplement): A18
- 22 Kayser M, Caglia A, Corach D, et al. Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med*, 1997, 110: 125~133
- 23 Su B, Xiao J H, Underhill P, et al. Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age. *Am J Hum Genet*, 1999, 65(6): 1 718~1 724
- 24 Ballinger S W, Schurr T G, Torroni A, et al. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics*, 1992, 130: 139~152
- 25 Heyer E, Puymirat J, Dieltjes P, et al. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human Molecular Genetics*, 1997, 6: 799~803
- 26 Bianchi N O, Catanesi C I, Bailliet G, et al. Characterization of ancestral and derived Y-chromosomal haplotypes of New World native populations. *Am J Hum Genet*, 1998, 63: 1 862~1 871
- 27 Turner II C G. Shifting continuity: modern human origin. In: Brenner S, Hanihara K, eds. *The Origin and Past of Modern Humans as Viewed from DNA*. New York: World Scientific, 1993. 216~243
- 28 Vasil'ev S A. The upper Paleolithic of northern Asia. *Current Anthropology*, 1993, 34: 82~92
- 29 Brown P. Recent human evolution in East Asia and Australia. *Philos Trans R Soc*, 1992, 337(1 280): 235~242
- 30 Swisher C C 3rd, Rink W J, Anton S C, et al. Latest *Homo erectus* of Java: potential contemporaneity with *Homo sapiens* in Southeast Asia. *Science*, 1996, 274: 1 870~1 874