漢族的遺傳結構：文化傳播伴隨人口擴張

／金力、李輝、文波

摘　要

漢族是中華民族的主體。長期以來，人們一直以為漢族是一個文化的共同體，各地的漢族的血統相差較大。通過系統地對漢族群體的Y染色體和線粒體DNA多態性進行分析，我們發現漢文化向南擴散的格局符合人口擴張模式。可見漢民族在遺傳上也是有相對的一致性的，起源於黃河中上游地區。南方漢族中，男性的北方成分保留得比女性多一些。南方漢族中北方男性成分保留得最多的是福建的閩語和客家人群。

關鍵字：漢族，遺傳結構，人口擴張

壹、前言

語言和文化在人群間的擴散有兩種不同的模式：一種是人口擴張、人群遷徙模式；另一種是文化傳播模式，人群之間有文化傳播，而基因交流卻很有限。同屬印歐語系的歐洲人群的形成機制爭議頗多，爭論的焦點在於來自近東的農業文明和語言的擴散是否伴隨著大量的農業人口的遷移〔Cavalli-Sforza et al 1994,Sokal et al 1991,Chikhi et al 2002〕。有著共同的文化和語言的漢族，人口超過了11億6千萬（2000年人口統計），無疑是全世界最大的民族。因此漢文化的擴散過程廣受各領域研究者的關注。

史載漢族源於古代中國北方的華夏部落，在過去的兩千多年間，漢文化（漢語和相關的文化傳統）擴散到了中國南方，而中國南方原住民族則是說侗台、南亞和苗瑤語的人群（百越、百濮和荊蠻）〔費孝通1999,葛劍雄等，1997〕。經典遺傳標記和微衛星位元點研究顯示，漢族和其他東亞人群一樣都可以以長江為界分為兩個遺傳亞群，南方漢族和北方漢族〔趙同茂等，1989；杜若甫等 et al，1997；褚嘉佑等，1998；肖春傑等，2000〕。兩個亞群之間的方言和習俗差異也很顯著〔Xu 2003〕。這些現象看似支援文化

傳播模式，即漢族向南擴張主要是文化傳播和同化的結果。然而，兩個亞群之間有著許多共同的 Y 染色體和線粒體類型〔宿兵等，2000；姚永剛等，2002〕，歷史記載的漢族移民史〔葛劍雄等，1997〕也與漢族的文化傳播模式假說相矛盾。歷史上漢族南遷主要有3次浪潮。第一次發生於西晉時期（西元 265-316 年），遷徙人口約 90 萬（大約當時南方人口的六分之一）；第二次發生於唐代（西元 618-907 年）規模比第一次大得多；第三次發生於南宋（西元 1127-1279 年），遷徙人口近 500 萬。我們對這兩種假說進行了檢驗，證實漢文化的擴散中的確發生了大規模的人群遷徙（人口擴張模式）〔文波等，2004a〕。

貳、南北方漢族的比較

為了驗證這些假說，我們把南方漢族的遺傳組成與兩個親本群體作比較，其一是北方漢族，其二是南方原住民族，即現居於中國境內和若干鄰國的侗台、苗瑤和南亞語群體。我們分析了來自中國 28 個地區漢族群體(註1)的 Y 染色體非重組區（NRY）(註2)和線粒體 DNA（mtDNA）遺傳多態〔Cavalli-Sforza et al 2003, Wallace et al 1999, Underhill et al 2000, Jobling et al 2003〕(註3)，這些樣本覆蓋了中國絕大部分的省份（詳見圖 1）。

父系方面，南方漢族與北方漢族的Y染色體單倍群頻率分佈非常相近，尤其是具有 M122-C 突變的單倍群（O3-M122 和 O3e-M134）普遍存在於我們研究的漢族群體中（北方漢族在 37-71% 之間，平均 53.8%；南方漢族在 35-74% 之間，平均 54.2%）。南方原住民族中普遍出現的單倍群 M119-C（O1）和 M95-T（O2a）在南方漢族中的頻率（3-42%，平均 19%）高於北方漢族（1-10%，平均 5%）。而且，南方原住民族中普遍存在的單倍群 O1b-M110,O2a1-M88和O3d-M7〔Su et al 1999〕，在南方漢族中低頻存在（平均4%），而北方漢族中卻沒觀察到。如果我們假定起始於兩千多年前的漢文化擴散〔葛劍雄等，1997〕之前南方原住民族的 Y 類型頻率與現在基本一致的話，南方漢族中南方原住民族的成分應該是不多的。分子方差分析（AMOVA）(註4)進一步顯示北方漢族和南方漢族的 Y 染色體單倍群頻率分佈沒有顯著差異（Fst = 0.006, P > 0.05），說明南方漢族在父系上與北方漢族非常相似。

母系方面，北方漢族與南方漢族的線粒體單倍群分佈非常不同。東亞北部的主要單倍群（A, C, D, G, M8a, Y, Z）在北方漢族中的頻率（49-64%，平均 55%）比在南方漢族中（19-52%,平均 36%）高得多。另一方面，南方原住民族的主要單倍群（B, F, R9a, R9b, N9a）〔12,14,18〕在南方漢族中

的頻率（36-72%，平均55%）要比在北方漢族（18-42%，平均33%）高得多。線粒體類型的分佈在南北漢族之間有極顯著差異（Fst＝0.006, P＜10^{-5}）。雖然南北漢族之間線粒體和Y染色體的Fst值相近，但線粒體的南北差異Fst值占群體間總方差的56%，而Y染色體僅僅占18%。

用漢族群體的單倍群頻率資料所做的主成分（PC）分析與以上結果相一致。對NRY分析發現，幾乎所有的漢族群體都聚在圖2a的右上方。北方漢族和南方原住民族在第2主成分上分離，南方漢族的第2主成分值處於北方漢族和南方原住民族之間，但是更接近於北方漢族（北方漢族0.58 ± 0.01；南方漢族0.46 ± 0.03；南方原住民族-0.32 ± 0.05），這表明南方漢族在父系上與北方漢族相近，受到南方原住民族的影響很小。就mtDNA而言，北方漢族和南方原住民族仍然被第2主成分分開（圖2b），南方漢族也在兩者之間但稍微接近南方原住民族（北方漢族0.56 ± 0.02；南方漢族0.09 ± 0.06；南方原住民族-0.23 ± 0.04），表明南方漢族的女性基因庫比男性基因庫有更多的混合成分。

所以，總的看來，南北漢族之間的一致性還是相當高的。並沒有原先想像的那樣南方漢族大部分來源於南方原住民的同化成分，而是以北方漢族相同的成分為主。

參、南方漢族的混合結構

我們進一步用兩種不同的統計方法〔Roberts et al 1965, Bertorelle et al 1998〕來估計兩個親本（北方漢族和南方原住民）對南方漢族基因庫的相對貢獻（表1），這兩個統計量(註5)用於單位點（single-locus）分析時比其他的方法更為準確〔Wang 2003〕。兩種方法得到的混合係數估計值（M，北方漢族的貢獻比例）高度一致（Y染色體，r＝0.922, P＜0.01；線粒體，r＝0.970, P＜0.01）。就Y染色體而言，所有的南方漢族都包含很高比例的北方漢族混合比率（M_{BE}：0.82 ± 0.14，範圍0.54-1；M_{RH}：0.82 ± 0.12，範圍0.61-0.97）（M_{BE}和M_{RH}的定義分別見Bertorelle et al 1998和Roberts et al 1965），這表明南方漢族男性基因庫的主要貢獻成分來自北方漢族。相反，南方漢族的線粒體基因庫中北方漢族和南方原住民族的貢獻比例幾乎相等（M_{BE}：0.56 ± 0.24〔0.15, 0.95〕；M_{RH}：0.50 ± 0.26〔0.07, 0.91〕）。總體上北方漢族對南方漢族的遺傳貢獻父系比母系高得多（t-test, P＜0.01）；各群體分別看也是這樣：絕大部分南方漢族群體中北方漢族的貢獻在父系上大於母系（M_{BE}，11/13，M_{RH}，13/13，P＜0.01，零假設為男女的貢獻相等為二項式分佈），這表明南方漢族的群體混合過程有很強的性別偏向。南方漢族中北方漢族貢獻的比例（M）呈現出由北

向南遞減的梯度地理格局。南方漢族線粒體的 M 值與緯度正相關（$r^2 = 0.569$, $P < 0.01$），但 Y 染色體的相關性不顯著（$r^2 = 0.072$, $P > 0.05$），因為南方漢族父系的 M 值差異太小，不足以導致統計上的顯著性。

表1　南方漢族中的北方漢族混合比例

群體	Y 染色體		線粒體 DNA	
	M_{BE}（± s.e.m）	M_{RH}	M_{BE}（± s.e.m）	M_{RH}
安徽	.868 ±.119	.929	.816 ±.214	.755
福建	1	.966	.341 ±.206	.248
廣東 1	.677 ±.121	.669	.149 ±.181	.068
廣東 2	ND	ND	.298 ±.247	**.312**
廣西	.543 ±.174	**.608**	.451 ±.263	.249
湖北	.981 ±.122	**.949**	.946 ±.261	.907
湖南	.732 ±.219	.657	.565 ±.297	**.490**
江蘇	.789 ±.078	.821	.811 ±.177	**.786**
江西	.804 ±.113	.829	.374 ±.343	.424
上海	.819 ±.087	.902	**.845** ±.179	.833
四川	.750 ±.118	.713	.509 ±.166	.498
雲南 1	1	.915	.376 ±.221	.245
雲南 2	**.935** ±.088	.924	**.733** ±.192	.645
浙江	**.751** ±.084	.763	.631 ±.180	.540
平均	.819	.819	.560	.500

註：M_{BE} 和 M_{RH} 分別為參考文獻Bertorelle et al 1998和Roberts et al 1965所描述的統計量。M_{BE} 的標準誤通過1000次自展（Bootstrap）獲得。把南方原住民族和北方漢族作為南方漢族的親本群體估計北方漢族的遺傳貢獻比例，假定2000多年前開始的混合過程前後南方原住民族的等位元基因頻率基本不變，並且南北漢族之間的遺傳交流不多。實際上，從北方漢族到南方原住民族的基因流動比反向的流動大得多，所以表中的估計值在沒有適當調整前是低估的。因而漢族實際的人口擴張程度應該大於本項研究得出的數值。

南方漢族這種男女混合比例不一致的產生原因可能很複雜。最有可能的原因是，由於社會結構以父系為主，所以女性可能隨著異族通婚的過程在群體之間流動，而男性的成分則保持在群體內部。所以我們不僅看到了在漢族群體中的來自原住民成分女性比男性多，在原住民群體中也可以看到來自漢族的成分同樣女性比男性多。另外，也有可能是南遷過程中，漢族男性到達南方的比女性多一些。這種格局形成的原因還需要進一步研究。

在南方各個群體中，所含北方成分最高的是福建漢族，父系成分幾乎和北方

一致。這與福建漢族的自我認識比較一致。他們認為自己是正宗的北方漢族的後代，是較早來到南方定居的。而當地的原住民閩越族、佘族等對他們的影響並不大。所以福建的各個群體，包括閩語群體和客家人，父系結構上都有較純的北方漢族成分。臺灣的相關群體應該與福建比較相近。

肆、漢族的起源

把漢族的主成分分析得到的第一主成分數值，對應地標在地圖上，可以看到一個明顯的單中心擴散趨勢。這個擴散中心就在黃河上游的陝西一帶，以渭河平原為中心。這與傳說中的漢族發源地完全一致。

傳說中的漢族始祖，炎帝黃帝就生活在這一帶。後來他們向東發展，才進入了中原地區。實際上，黃河上游一帶，不僅僅是漢族的發源地，也是整個漢藏語系人群的發源地。根據之前的研究，藏緬語族的各個民族基本上都發源於此。在中國的 56 個民族中，包括藏、羌、普米、納西、彝、白、土家、傈僳、獨龍、怒、景頗、阿昌、拉祜、哈尼、基諾等，都屬於藏緬語族。從黃河上游發源以後，漢族向東南方向擴張，而藏緬語族向西南方向擴張，漸漸到達了中國大部分地區和周邊鄰國。所以漢族的遺傳結構與這些民族之間都有很大的共性。

伍、結語

綜上所述，我們提出了兩項證據支援漢文化擴散的人口擴張假說。首先，幾乎所有的漢族群體的Y染色體單倍群分佈都極為相似，Y染色體主成分分析也把幾乎所有的漢族群體都集合成一個緊密的聚類。再有，北方漢族對南方漢族的遺傳貢獻無論父系方面還是母系方面都是可觀的，在線粒體DNA分佈上也存在地理梯度。北方漢族對南方漢族的遺傳貢獻在父系（Y 染色體）上遠大於母系（線粒體），表明這一擴張過程中漢族男性處於主導地位；換個角度看，在漢族和南方原住民的融合過程中有相對較多的當地女性融入南方漢族中。性別偏向的混合格局也同樣存在于藏緬語人群中〔文波等，2004b〕。

據歷史記載，受北方戰亂和饑荒的影響，漢人不斷的南遷，圖 1 中畫出了 3 次大規模移民的浪潮。在兩千多年間，除了這3次大潮，各個時期幾乎都有小規模的南遷。所以，我們的遺傳研究也與歷史記載相吻合。大量的北方移民改變了中國南方的遺傳構成，而漢族人口擴張的同時也帶動了漢文化的擴散。除了大規模的人群遷徙，北方漢族、南方漢族和南方原住民族之間的基因交流造成的族群混合也在很大程度上改變了中國人群的遺傳結構。

（本文轉載自第八屆孫中山與現代中國學術研討會。作者金力為復旦大學生命科學學院院長、美國辛辛納提大學人類基因組訊息中心主任；文波為美國約翰霍普金斯大學遺傳學系博士後研究；李輝任職於復旦大學現代人類學研究中心）

註釋：

註 1 ：採集中國各地的17個漢族群體871個隨機不相關個體的血樣。用酚—氯仿法抽提基因組DNA。結合文獻報導的Y染色體和線粒體多態性資料，總共分析的樣本量是：Y染色體23個群體1289人，線粒體23個群體1119人。這些樣本涉及了中國的大部分省份。

註 2 ：通過聚合酶鏈式反應—限制性片斷長度多態性（PCR-RFLP）的方法〔Su et al 2000〕分型Y染色體上的13個雙等位元標記：YAP, M15, M130, M89, M9, M122, M134, M119, M110, M95, M88, M45, M120。根據Y染色體委員會的命名系統（YCC）〔YCC 2002〕，這些標記構成13個單倍群，在東亞人群中具有較高的信息量〔Jin et al 2000〕。

註 3 ：線粒體上，對高變1區（HVS-1）進行測序，對編碼區8個多態位點作了分型（9-bp缺失，10397 AluI, 5176 AluI, 4831 HhaI, 13259 HincII, 663 HaeIII, 12406 HpaI , 9820 HinfI），有關方法已有報導〔Wen et al 2004〕。根據東亞線粒體系統樹〔Kivisild et al 2002〕，用高變1區突變結構和編碼區多態性構建單倍群。

註 4 ：根據線粒體和Y染色體單倍群頻率，用SPSS10.0軟體（SPSS公司）作主成分分析，研究群體間關係。南北漢族的遺傳差異用ARLEQUIN軟體〔Schneider et al 2000〕做AMOVA檢驗〔Excoffier et al 1992〕。

註 5 ：南方漢族中北方漢族和南方原住民族的混合比例估計用兩種不同的統計方法〔Roberts et al 1965, Bertorelle et al 1998〕：ADMIX 2.0〔Dupanloup et al 2001〕和LEADMIX（Wang 2003）軟體。親本群體的選擇對混合比例的適當估計很重要〔Chakraborty 1986, Sans, et al 2002〕，我們通過擴大東亞的參考資料來減小偏差。分析中，10個北方漢族群體的各單倍群頻率（Y染色體和線粒體標記分別分析）的算術平均作為北方親本群體。南方原住民族的頻率平均了3個族群：侗台語群（NRY，22群體；線粒體，11群體），南亞語群（NRY，6群體；線粒體，5群體），苗瑤語群（NRY，18群體；線粒體，14群體）。通過樣本的混合比例與緯度〔Cavalli-Sforza et al 1994, Chikhi et al 2002〕的線性回歸分析揭示漢族群體的地理格局。

參考文獻：

葛劍雄，吳松弟，曹樹基（1997），《中國移民史》，（福州：福建人民出版社）。

費孝通（1999），《中華民族多元一體格局》，（北京：中央民族大學出版社）。

Bertorelle, G.& Excoffier, L（1998）. Inferring admixture proportions from molecular data. Mol. Biol. Evol. 15, 1298-1311.

Cavalli-Sforza, L. L.,Menozzi, P. & Piazza, A（1994）. The History and Geography of Human Genes （Princeton Univ. Press, Princeton.

Cavalli-Sforza, L. L. & Feldman, M.W（2003）. The application of molecular genetic approaches to the study of human evolution. Nature Genet. 33, 266-275.

Chakraborty, R（1986）. Gene admixture in human populations:Models and predictions. Yb. Phys. Anthropol.29, 1-43.

Chikhi, L. et al （2002）. Y genetic data support the Neolithic demic diffusion model. Proc. Natl Acad. Sci. USA 99, 11008-11013.

Chu, J. Y. et al（1998）. Genetic relationship of populations in China. Proc. Natl Acad. Sci. USA 95,11763-11768.

Du, R. F., Xiao, C. J. & Cavalli-Sforza, L. L（1997）. Genetic distances calculated on gene frequencies of 38 loci. Science in China Ser. C 40, 613.

Dupanloup, I. & Bertorelle, G （2001）. Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol. Biol. Evol. 18, 672-675.

Excoffier, L., Smouse, P. E. & Quattro, J. M（1992）. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479-491.

Jin, L. & Su, B（2000）. Natives or immigrants: modern human origin in East Asia. Nature Rev. Genet. 1, 126-133.

Jobling, M. A. & Tyler-Smith, C（2003）.The human Y chromosome: an evolutionary marker comes of age. Nature Rev. Genet. 4, 598-612.

Kivisild, T. et al（2002）. The emerging limbs and twigs of the East Asian mtDNA tree. Mol. Biol. Evol. 19, 1737-1751.

Roberts, D. F. & Hiorns, R.W（1965）. Methods of analysis of the genetic composition of a hybrid population. Hum. Biol. 37, 38-43.

Sans,M. et al （2002）. Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. Am. J. Phys. Anthropol. 118, 33-44.

Schneider, S., et al（2000）. Arlequin: Ver. 2.000. A software for population genetic analysis.（Genetics and Biometry Laboratory, Univ. of Geneva, Geneva）.

Sokal, R., Oden, N. L. &Wilson, C （1991）. Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature 351, 143-145.

Su, B. 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, 1718-1724.

Su, B. et al （2000）. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum. Genet.107, 582-590.

The Y Chromosome Consortium （2002）, A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. 12, 339-348.

Underhill, P. A. et al（2000）. Y chromosome sequence variation and the history of human populations. Nature Genet. 26, 358-361.

Wallace, D. C., Brown, M. D. & Lott, M. T（1999）. Nucleotide mitochondrial DNA variation in human evolution and disease. Gene 238, 211-230.

Wang, J（2003）. Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164, 747-765.

Wen, B. et al（2004a）. Genetic evidence supports demic diffusion of Han culture. NATURE, 431:302-305.

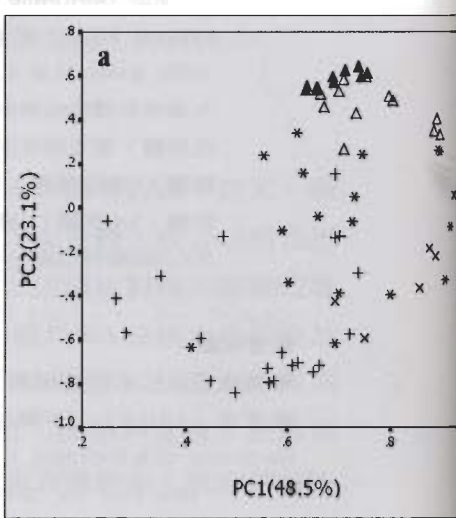
Wen, B. et al（2004b）. Analyses of genetic structure of Tibeto-Burman populations revealed a gender-biased admixture in southern Tibeto-Burmans. Am. J. Hum. Genet. 74, 856-865.

Xiao, C. J. et al（2000）. Principal component analysis of gene frequencies of Chinese populations. Sci.China C,43, 472-481.

Xu, Y. T （2003）. A brief study on the origin of Han nationality. J. Centr. Univ. Natl 30, 59-64.

Yao, Y. G. et al（2002）. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am. J. Hum.Genet. 70, 635-651.

Zhao, T. M. & Lee, T. D （1989）.Gmand Kmailotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Hum. Genet. 83, 101-110.



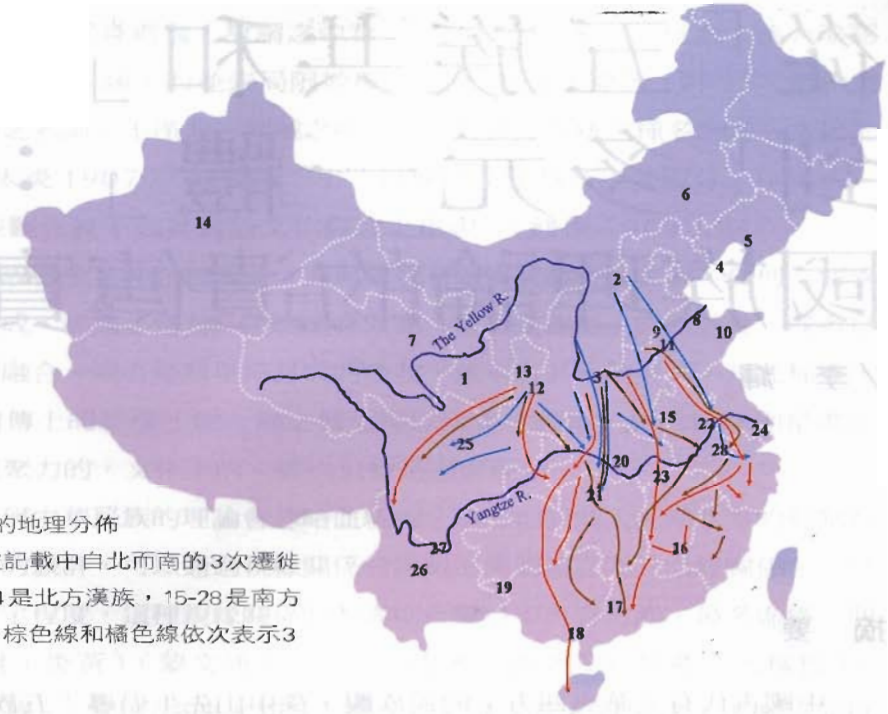


圖1　調查群體的地理分佈
圖中標出了歷史記載中自北而南的3次遷徙浪潮。群體1-14是北方漢族，15-28是南方漢族。藍色線、棕色線和橘色線依次表示3次遷徙浪潮。

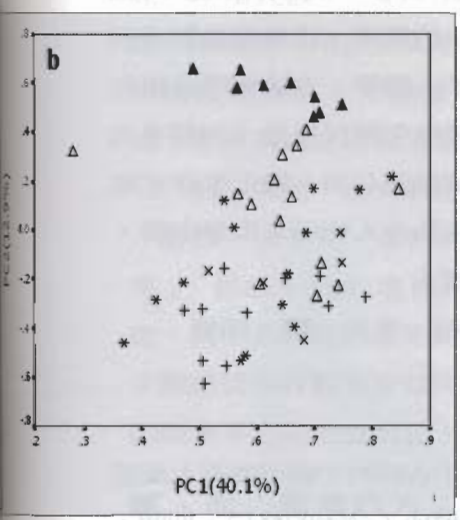


圖2　主成分散點圖
a為Y染色體單倍群散點圖，b為線粒體單倍群散點圖。
群體標記：▲北方漢族，□南方漢族，＋侗台語民族，×南亞語民族，＊苗瑤語民族。

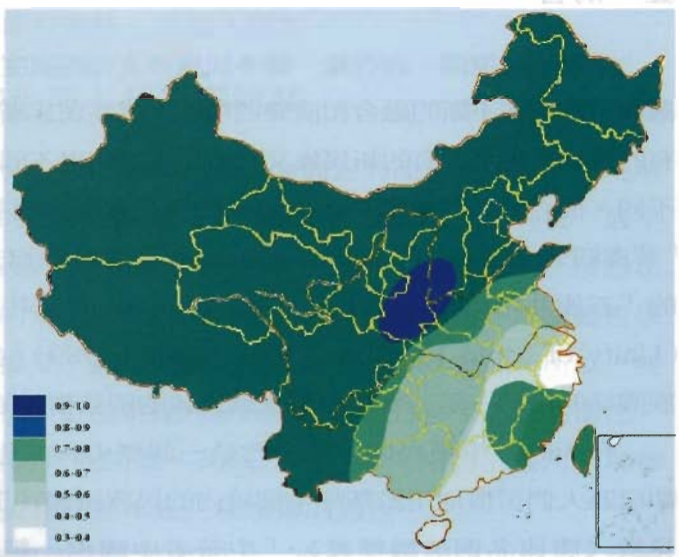


圖3　漢族Y染色體單倍群第一主成分地圖