

青海大通上孙家寨古代居民 mtDNA 遗传分析

张芃胤^{1,2}, 徐智^{1,2}, 许渤松^{1,2}, 韩康信³,
周慧⁴, 金力^{1,2}, 谭婧泽^{1,2}

(1. 复旦大学现代人类学教育部重点实验室, 上海 200433; 2. 上海人类学学会, 上海 200433;
3. 中国社会科学院考古研究所, 北京 100710; 4. 吉林大学生命科学学院, 长春 130012)

摘 要 利用古 DNA 手段对考古发掘出土的人类遗骸进行遗传分析, 是揭示当地古代人群来源的重要手段。我们通过克隆测序和 PCR-RFLP 的方法, 从来自青海大通上孙家寨的约 3000-3300 年前和 2000 年前两个不同年代的牙齿样本中, 成功得到 59 个线粒体高变 I 区和编码区的 SNP 位点的序列信息。之后我们将所得序列与来自亚洲大陆的 34 个现代人群共 1833 个个体和 2 个不同年代的古代人群样本的线粒体序列分别在个体和群体水平上作比较, 结果表明这两个时期人群并不是一脉相承的。

关键词: 古 DNA; 线粒体 DNA; 卡约文化; 古羌族

中图分类号: Q987; 文献标识码: A; 文章编号: 1000-3193(2013)02-204-15

1 简介

我国青海大通上孙家寨一带地处河湟之间, 秦以前为古羌人聚居区, 两汉时期是多民族杂居区域。上孙家寨墓地发掘出土的青铜时代卡约文化(距今约 3000-3300 年)墓葬约 1113 座, 是目前已发掘卡约文化墓葬最多的一处。卡约文化是青海省古代各种文化遗址中数量最多、分布范围最广的一种土著文化, 东起甘肃青海交界处的黄河、湟水两岸, 西至青海湖周围, 北达祁连山麓, 南至阿尼玛卿山以北的广大地区均有分布, 涉及的地区有民和、乐都、平安、西宁、互助、大通、海晏、刚察、同仁、贵南、化隆、循化、贵德、尖扎、湟中、共和等 10 多个县、市, 分布范围相当广泛。在上孙家寨墓地还发掘出两汉和魏晋初(距今约 2000 年)的墓葬 182 座。从这些汉晋墓葬的文化面貌来看, 以汉文化为主体, 又明显保留有其他多种民族的文化因素。

《晋语·国语》中记载:“昔少典娶有虫乔氏, 生黄帝、炎帝。黄帝以姬水成, 炎帝以姜水成。成而异德, 故黄帝为姬, 炎帝为姜。”炎帝属古羌族部落, 部落众多。在后来的战争中, 炎帝部落大部分与黄帝部落互相融合, 成为华夏族(今汉族的先民)。炎帝部

收稿日期: 2012-07-18; 定稿日期: 2012-09-28

基金项目: 国家自然科学基金资助项目(31071102); 上海市哲学社会科学规划课题(2010BZH005)。

作者简介: 张芃胤(1984-), 男, 复旦大学现代人类学教育部重点实验室硕士研究生, 主要从事古 DNA 分子生物学研究。

通讯作者: 谭婧泽, 女, 讲师, Email: tanjingze@gmail.com

落中的另一部分则西行或南下,与当地土著居民融合,成为汉藏语系中汉族和羌族以外的其他民族的先民,如藏族、彝族、纳西族等。由此可见,古羌族是中华民族重要的组成部分。

古 DNA 是指保留在古代生物遗骸和遗迹中的遗传物质,是一种重要的遗传资源。近年来随着人类遗传学及人类基因组研究的迅速发展,促生了在 DNA 分子水平上研究人群的遗传学特征的方法,即分子人类学。其中,线粒体 DNA(mtDNA)位于线粒体中,是闭合的环状双链 DNA 分子,属于母系遗传。人类的 mtDNA 由 16569 个碱基对组成,其中非编码区里含有三个高变区(Hypervariable Region, HVR),分为 HVRI(16024-16365)、HVRII(73-340)、HVRIII(438-574)^[1]。这三个高变区的 DNA 突变是非常高频和多态的,通过检测高变区的 SNP 遗传标记,可以确定样品所属的单倍群(Haplogroup),为分析个体和群体的母系遗传特征提供重要的信息。对距今 1700 至 1900 年的青海陶家寨遗址出土的 46 个样品的遗传特征分析显示,这一人群已比较接近藏缅语族人群和现代汉族^[2]。为了进一步揭示古羌族和汉族之间的群体关系,我们分析了出土于青海大通上孙家寨两个不同时期(3000-3300 年前和 2000 年前)的古代群体的 mtDNA 遗传特征,并确定了个体的单倍群归属。群体比较分析显示,3000-3300 年前的卡约文化人群的母系遗传结构与现代西南少数民族相近,而 2000 年前人群的母系遗传构成更接近现代汉族。本研究为了解中国西北地区少数民族的迁徙和汉族的形成,提供了重要信息。

2 材料与amp;方法

2.1 材料

本文实验材料出自青海省大通县上孙家寨的大型古代墓地,由青海省文物考古研究所的多位考古学者主持发掘和悉心采集。研究随机选取其中的 78 个个体,其中卡约文化期 34 个,汉代 44 个,材料均为牙齿。

为研究青海大通上孙家寨古代人群和周边地理人群之间的异同,我们选取了来自亚洲大陆的 34 个现代人群共 1833 个个体和 2 个不同年代的古代人群作为比较(表 1),具体地理分布见图 1。

2.2 实验方法

参照 Xu Z 等的实验方法^[3]进行样品的去污染预处理和粉碎、DNA 抽提、mtDNA 片段的 PCR 扩增、克隆测序等。

2.3 序列分析及 mtDNA 单倍群归属

以 rCRS(Genbank accession number NC_012920)^[4]作为参照,使用 Lasergene 7.1.0 (DNASTAR, <http://www.dnastar.com/>),将克隆测序得到的序列进行比对。通过检索已发表的现代人 mtDNA HVR I 数据库^[5],划分每个样品的单倍群归属。选择单倍群特异的编码区 SNP,并通过微测序(Minisequencing)和限制性片段长度多态性(Restriction fragment length polymorphism, RFLP)进行分型以确定单倍群归属。Minisequencing 使用 ABI PRISM SNaPshot™(Applied Biosystems) Multiplex Kit 进行单碱基延伸反应,引物见表 2,结果用 GeneScan® Analysis Software Version 3.7(Applied Biosystems)分析,RFLP 使用的引物参照

表 1 34 个现代人群和 4 个古代人群 (加 "△" 表示) 中单倍群的分布频率
Tab.1 Haplogroup frequencies of 34 modern and 4 ancient Chinese populations used for comparison with Shangsunjiashai samples

编号	群体	数量	单倍型														
			A	B	B4	B5	C	D	D4	D5	F*	F1	F2	F3	G		
1	汉族 (青海)	44	0.09		0.02	0.02	0.05	0.18		0.14	0.02	0.02					
2	汉族 (内蒙)	45	0.04		0.06	0.02	0.09	0.22		0.02	0.02	0.09	0.02				
3	汉族 (陕西)	85	0.08		0.03	0.06	0.05	0.25		0.1	0.04	0.1					
4	汉族 (山西)	53	0.08		0.08	0.02	0.02	0.21		0.06	0.04	0.08	0.02	0.04			
5	汉族 (四川)	58	0.04		0.11	0.1	0.01	0.17		0.02	0.01	0.1					
6	汉族 (云南)	70	0.05		0.11	0.05	0.05	0.1		0.06		0.15					0.02
7	蒙古族 (新疆)	49	0.08		0.02		0.1	0.2				0.1	0.08				0.04
8	蒙古族 (青海)	15	0.13		0.07			0.2	0.13		0.13						
9	蒙古族 (内蒙)	48	0.08		0.08	0.02	0.06	0.31		0.08		0.06	0.02				
10	维吾尔族 (新疆)	89	0.04	0.01			0.08	0.06		0.05	0.01	0.06					
11	哈萨克族 (新疆)	80	0.04		0.02		0.14	0.13			0.01	0.05	0.01				0.01
12	藏族 (云南)	75	0.16		0.08	0.11	0.06	0.49		0.06			0.03				
13	藏族 (青海) I	56	0.21		0.02		0.05	0.07				0.06					0.04
14	藏族 (阿里)	46	0.2		0.02		0.04		0.11	0.04		0.04					
15	藏族 (昌都)	61	0.03	0.02	0.03	0.02	0.05	0.02	0.11	0.05		0.1	0.05				0.03
16	藏族 (拉萨)	59	0.08						0.1	0.04		0.17					0.02
17	藏族 (林芝)	53	0.18		0.02		0.05	0.01	0.12	0.05		0.14					0.02
18	藏族 (那曲)	58	0.09		0.02		0.03		0.1	0.06		0.11					0.07
19	藏族 (青海) II	44	0.18		0.1		0.07		0.18	0.05	0.02	0.02					0.02
20	藏族 (日喀则)	59	0.19				0.07		0.1			0.14	0.02				0.02
21	藏族 (山南)	56	0.2		0.03	0.01	0.08		0.18	0.07		0.04					0.01
22	藏族 (四川)	55	0.13		0.04	0.04	0.05		0.2			0.07	0.02				0.02
23	白马藏族	52	0.1		0.1		0.1		0.21	0.04	0.02	0.06	0.04				0.13
24	羌族	44	0.09		0.05				0.36	0.05	0.02	0.14					0.05
25	苗族 (湖南)	103	0.09		0.12	0.07	0.1	0.04	0.11	0.05	0.03	0.11			0.01		0.02
26	苗族 (云南)	39			0.13	0.21	0.05		0.18			0.03			0.03		0.03
27	基诺族	18	0.06		0.06	0.06				0.06		0.17					
28	拉祜族	50			0.08		0.02	0.18			0.18	0.46					
29	傈僳族	37	0.3		0.03		0.14	0.08			0.03	0.08	0.11				
30	纳西族	45	0.09		0.18	0.07	0.09	0.04				0.22	0.02				0.07
31	怒族	30	0.3		0.1			0.2					0.17				
32	普米族	35	0.14		0.03		0.23	0.17				0.03	0.06				0.06
33	土家族	35	0.11		0.09	0.09	0.09	0.26		0.06		0.09					
34	彝族	87	0.11	0.01	0.12	0.02	0.03	0.17		0.04	0.06	0.09	0.01				0.01
35	加瓦艾日克 [△]	35					0.07	0.23	0.03			0.17					
36	陶家寨 [△]	43	0.14		0.09	0.12				0.14	0.07	0.09	0.09				
37	上孙家寨 (卡) [△]	30	0.03		0.03		0.13	0.13	0.17		0.03						0.03
38	上孙家寨 (汉) [△]	29			0.07	0.03		0.17	0.07	0.24							

续表 1 34 个现代人群和 4 个古代人群 (加 "△" 表示) 中单倍群的分布频率
Tab.1 Haplogroup frequencies of 34 modern and 4 ancient Chinese populations used for comparison with Shangsunjiashai samples (Continued)

编号 群体	单倍型											M10	M11	
	G1a	G2	G3	H	HV	I	J	K	M*	M7	M8			M9
1 汉族 (青海)		0.05		0.07					0.09	0.06	0.07	0.05		
2 汉族 (内蒙)			0.02						0.04	0.11	0.02	0.04	0.02	
3 汉族 (陕西)									0.07	0.13		0.02		
4 汉族 (山西)		0.04							0.06	0.06	0.02	0.06	0.04	
5 汉族 (四川)		0.04							0.07	0.15	0.01	0.01	0.01	
6 汉族 (云南)	0.02	0.03							0.03	0.18	0.02	0.02		
7 蒙古族 (新疆)		0.12		0.04	0.02					0.02		0.02		0.02
8 蒙古族 (青海)							0.2			0.07			0.07	
9 蒙古族 (内蒙)		0.04								0.02		0.04	0.02	
10 维吾尔族 (新疆)		0.11	0.01	0.06	0.02		0.04	0.02	0.04	0.06	0.06			
11 哈萨克族 (新疆)	0.01	0.01		0.05	0.04	0.01	0.02		0.07	0.01	0.01	0.02		
12 藏族 (云南)			0.06						0.6	0.03		0.08		
13 藏族 (青海) I		0.06	0.02						0.36			0.09		
14 藏族 (阿里)		0.04	0.02									0.35	0.02	
15 藏族 (昌都)		0.03	0.05						0.02		0.02	0.26	0.02	0.02
16 藏族 (拉萨)	0.02	0.07	0.03						0.05		0.02	0.34		
17 藏族 (林芝)		0.03	0.04						0.05			0.21	0.01	0.01
18 藏族 (那曲)		0.02	0.07						0.05			0.22		
19 藏族 (青海) II		0.02	0.02						0.07	0.02		0.09	0.02	
20 藏族 (日喀则)	0.02	0.02					0.02		0.03			0.32		
21 藏族 (山南)		0.01	0.03						0.07			0.21		
22 藏族 (四川)		0.05	0.09						0.02	0.02	0.04	0.04		
23 白马藏族		0.02		0.02										0.06
24 羌族		0.02	0.07						0.02	0.04	0.02	0.05		
25 苗族 (湖南)		0.01	0.02						0.07	0.05	0.02			
26 苗族 (云南)									0.23	0.06				
27 基诺族									0.28	0.17		0.06		
28 拉祜族									0.02	0.02	0.02			
29 傈僳族									0.11	0.08				
30 纳西族									0.13	0.02				
31 怒族		0.03							0.03					
32 普米族			0.03						0.2					
33 土家族		0.03								0.09		0.03		
34 彝族		0.02							0.15	0.09		0.01		
35 加瓦艾日克 [△]				0.1				0.03	0.07				0.03	
36 陶家寨 [△]									0.05				0.02	
37 上孙家寨 (卡) [△]		0.03	0.03						0.1		0.1		0.13	
38 上孙家寨 (汉) [△]									0.07		0.14	0.07		

续表 1 34 个现代人群和 4 个古代人群 (加 "△" 表示) 中单倍群的分布频率
Tab.1 Haplogroup frequencies of 34 modern and 4 ancient Chinese populations used for comparison with Shangsunjiashai samples (Continued)

编号 群体	单倍型													
	M13	N*	N9a	R*	R9	T	T1	T2	U	U2	U4	U5	U7	V
1 汉族 (青海)		0.02	0.02											
2 汉族 (内蒙)			0.07			0.02								
3 汉族 (陕西)		0.01	0.04				0.01			0.01				
4 汉族 (山西)		0.06	0.02										0.02	
5 汉族 (四川)			0.03		0.04									
6 汉族 (云南)		0.03												
7 蒙古族 (新疆)	0.02		0.02								0.02	0.06		
8 蒙古族 (青海)														
9 蒙古族 (内蒙)			0.06		0.02		0.02							
10 维吾尔族 (新疆)		0.06		0.01		0.01		0.01		0.04	0.02		0.01	
11 哈萨克族 (新疆)		0.04	0.01			0.05		0.04				0.02		0.01
12 藏族 (云南)	0.23													
13 藏族 (青海) I		0.02	0.02											
14 藏族 (阿里)	0.04			0.02			0.04							
15 藏族 (昌都)	0.03									0.02				
16 藏族 (拉萨)													0.02	
17 藏族 (林芝)	0.05		0.01	0.01									0.01	
18 藏族 (那曲)	0.07	0.02			0.02								0.03	
19 藏族 (青海) II	0.02			0.07								0.02		
20 藏族 (日喀则)	0.02												0.03	
21 藏族 (山南)	0.03	0.01												
22 藏族 (四川)	0.13						0.04							
23 白马藏族														
24 羌族														
25 苗族 (湖南)		0.01	0.02	0.01	0.05									
26 苗族 (云南)			0.03		0.05									
27 基诺族			0.11											
28 拉祜族					0.02									
29 傈僳族														
30 纳西族	0.02				0.02									
31 怒族														
32 普米族			0.03											
33 土家族			0.03											
34 彝族		0.02												
35 加瓦艾日克 [△]	0.03	0.03		0.07		0.03			0.03	0.03	0.03			
36 陶家寨 [△]			0.12											
37 上孙家寨 (卡) [△]														
38 上孙家寨 (汉) [△]		0.03								0.03				

续表 1 34 个现代人群和 4 个古代人群（加“△”表示）中单倍群的分布频率
Tab.1 Haplogroup frequencies of 34 modern and 4 ancient Chinese populations used for comparison with Shangsunjiashai samples (Continued)

编号	群体	W	Y	Z	文献	编号	群体	W	Y	Z	文献
1	汉族（青海）			0.02	[25]	20	藏族（日喀则）			0.02	[27]
2	汉族（内蒙）		0.02	0.02	[25]	21	藏族（山南）			0.03	[27]
3	汉族（陕西）			0.01	[26]	22	藏族（四川）			0.02	[27]
4	汉族（山西）		0.02		[25]	23	白马藏族			0.12	内部数据
5	汉族（四川）		0.01	0.03	[25]	24	羌族			0.02	内部数据
6	汉族（云南）		0.02	0.03	[25]	25	苗族（湖南）	0.01	0.01		[23]
7	蒙古族（新疆）				[21]	26	苗族（云南）				[23]
8	蒙古族（青海）				[19]	27	基诺族				[18]
9	蒙古族（内蒙）			0.04	[22]	28	拉祜族				[18,24]
10	维吾尔族（新疆）	0.06			[20,21]	29	傈僳族			0.05	[19]
11	哈萨克族（新疆）			0.15	[20,21]	30	纳西族			0.02	[18]
12	藏族（云南）			0.03	[18,19]	31	怒族			0.17	[19]
13	藏族（青海）I				[18]	32	普米族			0.03	[18]
14	藏族（阿里）				[27]	33	土家族			0.06	[19]
15	藏族（昌都）			0.03	[27]	34	彝族				[18,19]
16	藏族（拉萨）			0.05	[27]	35	加瓦艾日克 [△]				[28]
17	藏族（林芝）				[27]	36	陶家寨 [△]			0.07	[3]
18	藏族（那曲）			0.02	[27]	37	上孙家寨（卡） [△]			0.03	
19	藏族（青海）II				[27]	38	上孙家寨（汉） [△]	0.03	0.03		

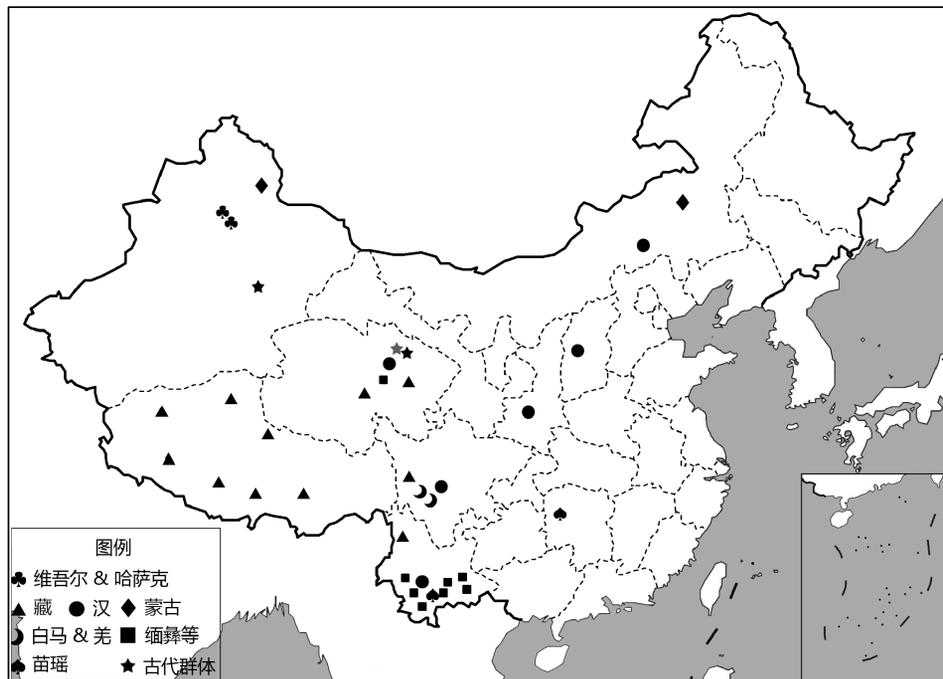


图 1 34 个现代人群及 4 个古代人群地理分布示意图（浅灰色★为上孙家寨）

Fig.1 The geographical locations of 34 modern and 4 ancient Chinese populations

表 2 SNaPshot 对 mtDNA 编码区 SNP 位点分型所用引物
Tab.2 Primers for mtDNA coding region SNPs assay with PCR-RFLP

SNP	单倍型	扩增引物	延伸引物
663	A	FOR TTAGACGGGCTCACATCA	TAATCTACTAAGAGCTAATAGAAAGGC
		REV GGATGCTTGCATGTGTAATCT	
3010	D4	FOR ATAACAGCGCAATCCTATTCTAGA	TTAATAGCGGCTGCACCAT
		REV ACTTTAATCGTTGAACAAACGAA	
4216	TJ	FOR TCCTATGAAAAAACTTCTACC	T ₁₀ CTCACCCCTAGCATTACTTATATGA
		REV GTTTGAGGGGGAATGCTG	
4715	M8	FOR GCATCCATAATCCTTCTAATAGCTA	T ₂₀ GGTAGTATTGGTTATGGTTCATTG
		REV ATGATTATTAATGATGAGTATTGATTGG	
4833	G	FOR AACTAGGAATAGCCCCCTTTC	T ₂₉ AGGCCGGATGTCAGAGGGG
		REV GCTAGTTTTGTGCATGTGAGAAGA	
5178	D	FOR ACCGCATTCTACTACTCAACTT	T ₅₁ ATCTCGCACCTGAAACAAG
		REV GCCTCTAGGAGAGGAG	
5301	D5	FOR GCTTTTTGCCCAAATGGG	T ₃₁ GGTGGCTATGATGGTGGGGA
		REV AAGTAGAGGTTAAGGAGGGTGATG	
5417	N9	FOR TAATCTACTCCACCTCAATCACACT	T ₃₄ GGGGTGGGTTTTGTATGTTCAAA
		REV TAAGGGCGATGAGTGTGG	
7028	H	FOR ACTGGCATTGTATTAGCAAATC	T ₃₃ CGACACGTACTACGTTGTAGC
		REV ATGATGGCAAATACAGCTCC	
9824	M7	FOR ATTTCCGACGGCATCTAC	T ₄₈ CCACAGGCTTCCACGGACT
		REV AAATATTAGTTGGCGGATGA	
10400	M	FOR CTCTTATTAATCATCCTAGCCC	TTTTCGTTTTGTTTAACTATATACCAATTC
		REV ATAATTTAATGAGTCGAAATCATTCCG	
12308	U	FOR ATGTCTAACAACATGGCTTTCTC	T ₄₇ GGAGTTGCACCAAAATT
		REV TGGTTATAGTAGTGTGCATGGTTATT	
12705	R	FOR ATAGAATTCTCACTGTGATATATAAACTCAG	T ₃₅ ATCAGTTCTCAAATATCTACTCAT
		REV AACAGTTGGAATAGGTTGTTAGC	
13263	C	FOR GCAGCAGTCTGCGCCCTTA	T ₃₅ AGCCTTCTCCACTTCAAGTCA
		REV CAGGAATGCTAGGTGTGGTT	

Xu Z 的方法^[3]。

2.4 重复实验

随机选取 10 个样品，在吉林大学边疆考古研究中心古 DNA 实验室进行独立的重复实验^[3]，以尽可能减小污染的可能性。

2.5 可靠性分析

按照古 DNA 的可靠性标准^[6,7]，进行严格的实验程序和分析。所有样品都被保存在恒温箱内，且尽可能少接触。所有的 PCR 前操作都在一个与 PCR 循环和 PCR 后操作隔离的专门的古 DNA 实验室中进行。此外，人员严格保持（PCR 前→PCR→PCR 后）的单向流动，以防 DNA 通过粘附在衣物和皮肤上进入古 DNA 实验室内^[8]。实验过程中使用一次性手套、口罩、带滤芯的 tip 等，以避免交叉污染。为检测可能的污染，在抽提和 PCR 时，使用阴性对照。实验人员的 mtDNA 序列被用以和获得的 DNA 序列作对照，以排除实验人员污染的可能性。只有重复实验获得一致结果，所有对照都为阴性结果的古代 DNA 才被包含在后续的分析中。使用覆盖三段重叠片段的引物以减小扩增 numtDNA 的可能性^[9,10,11]。古 DNA 的碱基错配在不同的抽提中是不可重复的^[12]，多聚酶的低保真度

易造成的某一位点的测序错误, 以及由于作为 PCR 模版的古 DNA 普遍较短而易导致跳跃 PCR^[13, 14], 因此, 我们对每个样品的 mtDNA HVR I 序列均采用克隆测序。单倍群归属通过 HVR I 序列和单倍群特异编码区 SNP 的匹配来确定^[16]。虽然没有对氨基酸外消旋进行测定, 但是对于实验室内和实验室间重复实验都得到一致结果的样品中, 必定存在内源 DNA, 这一替代方法在很多研究中都曾被使用^[14, 15, 17]。

2.6 主成分分析图绘制

基于单倍群分布比率的主成分分析图 (Principal Component Analysis, PCA) 由 SPSS for Windows Release 11.0.1 (15th Nov 2001) 绘制。

3 结果

我们共对青海大通上孙家寨墓地出土的 78 个样品进行了分析, 其中 59 个样品在实验室内获得了可重复结果 (表 3), 并且与实验人员的 mtDNA 谱系 (表 4) 不存在一致性。其余 19 个样品中的 17 个样品的结果不可重复或存在较多杂合疑为污染, 2 个样品的 DNA 不可扩增。59 个成功样品中, 10 个样品在吉林大学边疆考古研究所同样得到一致的可重复结果。59 个样品均经过克隆测序, 测序结果 JX871470-JX871835 见 <http://www.ncbi.nlm.nih.gov/genbank/>。从不同克隆子测序观察到的两种类型的碱基错配情况统计来看 (表 5), Type2 (983 个) 多于 Type1 (240 个), 表现出古代 DNA 分子的重要特性^[29, 30]。因此, 我们认为这 59 个样品的 mtDNA 谱系是可靠的。

基于比较群体的主要单倍群频率的分布在表三中显示, PCA (图 2) 结果显示, 从 PC1 上看, 可以大致分为三个群体: 左侧的汉藏, 中间的蒙古 (新疆), 右侧的维吾尔、哈萨克和且末加瓦艾日克。从 PC2 上看, 汉和藏分开, 其中青海大通上孙家寨汉代样本与现代汉族聚在一起, 卡约时期样本与西南少数民族较为接近。值得注意的是, 卡约时期样本与来自我国云南省宁蒗的现代普米族最为接近。陶家寨样本与云南藏族、纳西族最为接近。

从卡约文化和汉代两个时期的样本中各选取一个有代表性的单倍群绘制基于邻接法的最大简约树 (图 3)。

4 讨论

距今约 3300 至 2000 年前的青海大通上孙家寨墓地, 由于其所处的时间和地理位置的特殊性, 结合现代人群的数据进行分析, 能够比较清晰地揭示当地古代居民的迁徙路线。

本研究中, 我们成功提取并分析了 59 个个体的母系遗传信息, 成功率为 75.6%, 说明整个墓葬保存状况良好。由于获得的 DNA 多样性很高 (表 3), 研究人员污染和样品之间交叉污染的可能性很小。同时, 较长片段 (408bp) 扩增失败也间接证明序列信息的可靠性。克隆产物显示 (表 5), Type2(C→T/G→A) 的碱基错配在大多数样品中占有比较

表 3 青海大通上孙家寨墓地 59 个样品的 mtDNA HVR 序列特征、编码区 SNP 和单倍群归属

Tab.3 mtDNA HVRI and RFLP results of Shangsunjiazhai and individual haplogroup assignments

年代	编号	性别	年龄	单倍型	高变I区 (16053-16366)	SNPs
T	479	F	25-30	B4a	189 217 261 299	9bp 12705C
T	484	M	22-24	G3	215 223 274	10400T 4833G
T	495	-	-	C	93 223 298 327	13263G
K	497	-	-	D	92 223 294 362	5178A(-Alu I) 4833A (-Hha I)
T	503	-	-	M8	185 223 260 298	4715G 13263A(-Alu I) 9090T
T	526	M	35-40	D4	92 223 294 362	5178A 3010A
T	529	-	-	M8	223 298 327	4715G 13263A(-Alu I) 9090T
K	538	-	-	M10	93 129 193 223 311 320 357	10646A
K	544	-	-	M10	93 129 193 223 311 320 357	10646A
T	547 [▲]	F	35-40	M8	223 298 327	4715G 13263A(-Alu I) 9090T
T	564	F	50-55	D4	92 223 294 362	3010A
T	576	F	~40	G	223 229 274 362	4833G
K	604 [▲]	M	~45	M10	129 193 223 311 320 357	10646A
K	723	F	22-24	M12	223 278 362	4833A (-Hha I) 14569A
T	758	F	~50	C	93 129 223 298 327	4715G 13263G(+Alu I)
K	836 [▲]	M	22-24	C	93 129 223 298 327	4715G 13263G(+Alu I)
K	909	F	~25	D4	223 260 362	3010A
T	927	M	~30	C	129 192 223 298 327	13263G(+Alu I)
K	937	M	45	M10	93 129 193 223 311 320 357	10646A
T	959	-	-	G2a	223 227 278 362	4833G
T	965	M	40-45	A	102 223 290 311 319 362	663G
T	989II [▲]	F?	~30	M [*]	223 325 362	5178C
T	1012	M	20-24	D4	86 223 274 362	10400T 5178A 3010A
K	1026	-	-	D	223 362	5178A(-Alu I)
T	1031A	F	20-25	Z	185 223 260 298 327	4715G 13263A (-Alu I) 9090C
T	1043	F	>35	F	304	12705C 10310A
T	1048 [▲]	M	14-15	D	189 203 223 362	10400T 5178A
T	1073	F	~25	D4b2	193 223 362	5178A 3010A 9824A
T	1079	M	~35	M [*]	111 129 223 311	10400T(+Alu I)
T	1092	M	25-35	D	223 266 362	5178A
H	3	M	60+	D5	223 311 316 362	5178A
H	4III	-	-	M8	223 298 311 319	4715(-BspEI)
H	11	M	-	D5	189 223 362	5178A
H	19II	M	-	B4	150 189 217 234	9bp
H	37	M	-	D5a	189 223 266 362	5178A
H	37I	F	-	D5a	172 189 223 266 362	5178A
H	41I	-	-	D5a	92 164 223 266 362	5178A(-Alu I)
H	42II	M	-	D5a	164 172 189 223 266 362	5178A
H	44	F	-	D5	189 223 362	5178A(-Alu I)
H	53I	M	-	M	223 362	10400T 5178C(+Alu I)

续表 3 青海大通上孙家寨墓地 59 个样品的 mtDNA HVR 序列特征、编码区 SNP 和单倍群归属
Tab.3 mtDNA HVRI and RFLP results of Shangsunjiashai and individual haplogroup assignments (Continued)

年代	编号	性别	年龄	单倍型	高变I区 (16053-16366)	SNPs
H	58 [▲]	-	-	B4	189 217 223	9bp
H	59	-	-	U2e	93 129C 145 183C 189 362	12705C 12308G
H	82 [▲]	-	-	Y1	126 231 266 319	14178C
H	92I	-	-	N	93 223 304	10400C(-Alu I)
H	92II	-	-	M8	223 294 298 327	13263A(-Alu I) 4715(-BspEI) 9090T
H	98 [▲]	-	-	D4b2	173 223 362	5178A 3010T 9824A
H	99	M	-	M8a	148 184 189 192 223 298 311 319	13263A(-Alu I) 4715(-BspEI) 9090T
H	106I	M	-	M9a	223 234 291 316 362	4491A 9242G
H	108	M	-	D	223 266 362	5178A
H	109II	F	-	B5a	129 140 266G	9bp
H	119I	M	25-30	Z	185 223 260 293 298	9090C 4715G
H	125	M	-	D	112 223 274 362	5178A(-Alu I)
H	128II [▲]	F	-	D4b2	223 362	5178A 9824A
H	132I [▲]	M	-	D	93 223 362	5178A
H	137	-	-	M9	223 234 286 316 362	10400T 4491A 9242A
H	138	M	-	D	93 176 223 362	10400T 5178A
H	150	M	-	M8a	184 189 223 298 319	4715(-BspEI)
H	156	F	-	M	223 362	5178C(+Alu I) 10400T(+Alu I)
H	159	M	-	D5a	164 172 189 223 266 362	5178A(-Alu I)

[▲]代表在吉林大学进行了独立重复实验。

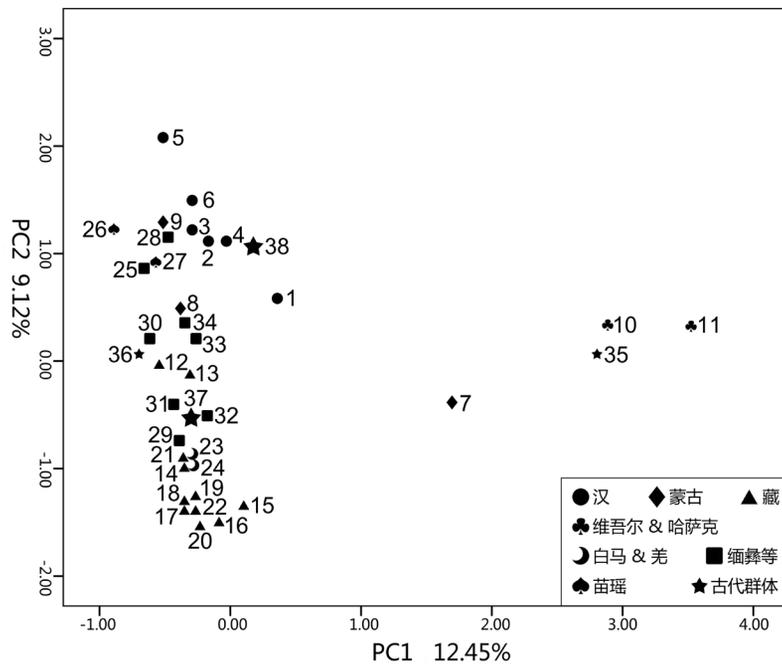


图 2 基于单倍群频率分布的 PCA 图 (37 与 38 为上孙家寨)
Fig.2 Principal component analysis of mtDNA haplogroup frequencies

大比率，这与之前的研究结果^[29,30]相一致。

从表 1 中我们可以看到，上孙家寨卡约时期群体与汉代群体相比，卡约时期群体中单倍群 C、D4、M* 和 M10 频率较高。在我国现代人群中单倍群 C 频率较高的为普米族，单倍群 D4 为藏族、羌族和苗族，单倍群 M* 为苗族和基诺族。这两个时期群体中的单倍群 M8 和 M10 在我国现代人群中均比较少见。在图 2 中也可以看出青海大通上孙家寨卡约时期的样本与西南少数民族聚在一起，尤其是与来自云南宁蒗

普米族最为接近。陶家寨群体与藏缅群体单倍群结构类似^[2]，而本研究中它们与云南藏族、纳西族最为接近（图 2），一定程度上印证了之前的研究结果。从个体水平上分析，图 3 中单倍群 G3 卡约时期的 T484 个体，与 2 个云南藏族和 1 个云南普米族非常接近，与现代汉族和维吾尔族、哈萨克族的距离最远；而单倍群 M9a 汉代的 H106I 个体，与现代汉族最近，与藏族等最远。这一结果也在一定程度上反映了两个不同时期人群的差异。

卡约文化时期相当于中原地区的西周，这一研究结果揭示了当时在青海河湟地区居住的可能是古代羌人，而他们正是现代西南地区部分少数民族的共同祖先之一。汉代群体则与现代汉族较为接近，从一定程度上说明汉代居住在青海大通上孙家寨的居民已经变成了当地现代汉人的祖先，之前卡约文化时期的古代羌人可能已经向西南方向迁徙，与历史上秦朝大迁徙和汉代汉人进入湟水地区的史料记载相吻合。

历史上羌人分布的区域很广，但无论从文献记载还是从考古遗存看，羌人的中心都在今青海省境内，向东沿及甘肃等省的部分地区。因此，我们可以推测，公元前 3000 年左右活动于河湟地区的古代居民为羌族。卡约文化的分布范围也表明，它是古代羌族的文

表 4 青海大通上孙家寨墓地样品分析人员的 mtDNA HVR I 的序列特征和所属单倍群（1-6 为复旦大学的研究人员，7-9 为吉林大学的研究人员）

Tab.4 HVRI motifs and mtDNA haplogroups of the staff involved in the excavations and experiments

编号	HVR I (-16000)	单倍型
1	66 223 311	M*
2	223 234 290 362	M*
3	172 223 356 362	Not determined
4	223 234 291 316 362	M9a
5	189 223 311	M*
6	111 129 140 183C 189 200T 243	B5b
7	136 183C 189 217 218 239 248	B4b1
8	126 174 223 311 362	Not determined
9	145 182C 183C 189 217 261 360	B4a

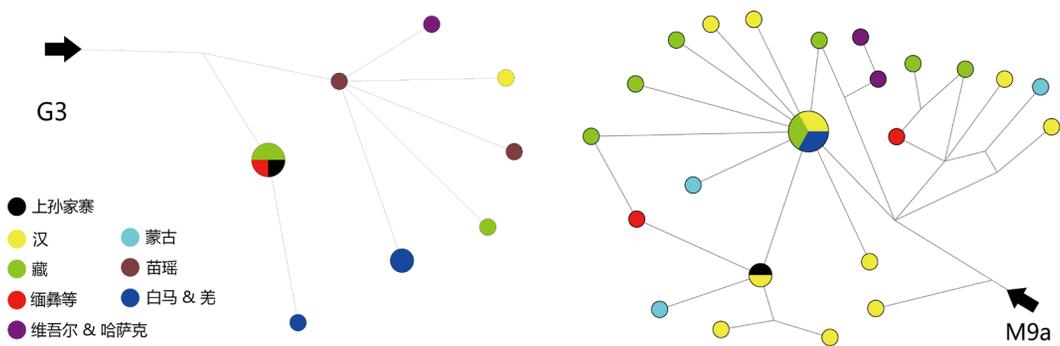


图 3 基于邻接法的 2 个单倍群的最大简约树（箭头表示根）
Fig.3 The maximum parsimony tree based on the Neighbor-joining method

表 5 59 个青海大通上孙家寨墓地样品的克隆序列中反映的碱基错配

Tab.5 Miscoding lesions observed within clones from Shangsunjiashai

年代	编号	Transition				Transversion						
		Type 2		Type 1		C→A	G→C	A→C	A→T	T→A	T→G	
		C→T	G→A	T→C	A→G							
T	479	36		1	1							
T	484	24	2									
T	495	36	2	1	1							
K	497	32	4									
T	503	4										
T	526	6	4									
T	529	4		2	1	1						
K	538	4	3		1							
K	544	7	1									
T	547	18		8				1				
T	564	20	1	1	1							
T	576	20	1		3			1				
K	604	24		2	1							
K	723	18	10	1								
T	758	5		3	2							
K	836	13	4		1							
K	909	28	2		4							
T	927	16	5	6	1							
K	937	7	1	1	2							
T	959	10		5								
T	965	2										
T	989II	5	2	7	2							
T	1012	3		1	1							
K	1026	20	7									
T	1031A	11	1	6	2							2
T	1043	10	3		2							
T	1048	1	2		1							1
T	1073	9	4									
T	1079	2	1	4	2							
T	1092	6	1	6	5			1				6
H	3	18		7	5							
H	4III	17	5	1	6							
H	11	19	2	6	2							1
H	19II	26			1							6
H	37	8	1	4	3							
H	37I	11	5	4	4							
H	41I	4	1									
H	42II	10	2	4	2							
H	44			1	1							
H	53I	8	1	16	2							
H	58	18	2	2								
H	59	23	1	1	2							3
H	82	3	1	1				1				
H	92I	7		3	1							

续表 5 59 个青海大通上孙家寨墓地样品的克隆序列中反映的碱基错配
Tab.5 Mismatching lesions observed within clones from Shangsunjiashai (Continued)

年代	编号	Transition				Transversion					
		Type 2		Type 1		A→T	T→A	T→G	C→A	G→C	A→C
		C→T	G→A	T→C	A→G						
H	92II	2		3							
H	98	36		3	2				1		1
H	99	12	6		4						
H	106I	1	2	11	2						
H	108	7	4	3							
H	109II	25	3	6							
H	119I	5	3	3	1						
H	125	9	4	1	6				1		
H	128II	18									
H	132I	16		6	1						1
H	137	27	1	4	1						
H	138	48			1						
H	150	37	2		4						
H	156	41		6	2						
H	159	18	1		2						6
	总计	875	108	151	89	1	0	0	6	0	27

化遗存。在我国古代文化的发展过程中，中原地区的文化曾经对周围产生很大的影响，同时周围地区的各种文化也给中原地区以重要影响，卡约文化也是如此。在青铜时代，古代羌族曾多次参加过重要的政治、军事行动，在以后的历史时期里也发挥过重大影响。商周时代，大部分分布于山西、河南、陕西的羌人已融入华夏族，而一部分羌人则退回到陇东。春秋战国时，尤其是秦穆公以后，西北羌人迫于秦的愈加强大的军事压力，在原来向外流动的路线基础上，开始了更大规模、更远距离的迁徙。西汉初年，进入中原的羌人已与汉族融合；未进入中原的羌人则分布于陇西、甘南、河湟一带。汉武帝北却匈奴，西逐诸羌，关内汉人进入河湟地区，而当地古代羌族居民有的向西发展，后来成为藏族先民的一部分；有的则可能长途跋涉到了新疆天山南麓，成为后来文献上所记载的“媯羌”的组成部分；也有少数可能北迁至今内蒙古西部额济纳旗一带；此外还有大量羌人继续向西南移动，成为西南地区彝、白、哈尼、纳西、傈僳、拉祜、普米等少数民族的先民^[31]。由此可见，古代羌族与国内现存很多民族都有十分密切的关系，对中华民族的形成有着巨大的贡献。

由于 mtDNA 高变区的高突变率，本研究中所讨论的大多数单倍型不是稀有的，因此可以认为相同的高变区序列可能是状态相同 (identity by state, IBS) 而非血缘相同 (identity by descent, IBD)^[32]，即拥有相同高变区序列的个体有可能并非来自同一个祖先。为了尽可能结合多个位点信息，来降低 IBS 的可能性，我们分析了每个个体的高变区序列和编码区的主要酶切位点，这个方法肯定无法完全排除 IBS 的可能性。为了了解 3000 年前青海人群和现代人群之间的关系，更多的分子人类学研究有待进一步开展，诸如线粒体基因组全测序、Y 染色体和常染色体序列分析等。

5 结 论

综上所述, 从青海大通上孙家寨墓地中获得的 59 个个体的遗传学分析结果显示, 3000 年前卡约文化时期的上孙家寨先民与现代西南地区藏缅语族人群较为接近, 与汉代的居民并不是同一人群, 汉代的居民与现代汉人更加接近。居住在青海河湟一带的古代羌人随着历史上的数次人群迁徙, 发展成为现代西南少数民族的一部分。

参考文献

- [1] Warner JB, Bruin EJ, Hannig H, et al. Use of sequence variation in three highly variable regions of the mitochondrial DNA for the discrimination of allogeneic platelets[J]. *Transfusion*, 2006, 46(8): 1461-1462
- [2] Zhao YB, Li HJ, Li SN, et al. Ancient DNA evidence supports the contribution of Di-Qiang people to the Han Chinese gene pool[J]. *Am J Phys Anthropol*, 2010, 144: 258-268
- [3] Xu Z, Zhang F, Xu B, et al. Mitochondrial DNA evidence for a diversified origin of workers building mausoleum for First Emperor of China[J]. *PLoS ONE*, 2008, 3(10): e3275
- [4] Andrews RM, Kubacka I, Chinnery PF, et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*, 1999, 23: 147
- [5] MITOMAP. A Human Mitochondrial Genome Database[DB/OL]. <http://www.mitomap.org>, 2011
- [6] Cooper A, Poinar HN. Ancient DNA: Do it right or not at all[J]. *Science*, 2000, 289: 1139
- [7] Pääbo S, Poinar H, Serre D, et al. Genetic analyses from ancient DNA[J]. *Annu Rev Genet*, 2004, 38: 645-679
- [8] MacHugh DE, Edwards CJ, et al. The extraction and analyses of ancient DNA from bone and teeth: A survey of current methodologies[J]. *Anc Biomol*, 2000, 3: 81-102
- [9] Willerslev E, Cooper A. Ancient DNA[J]. *Proc Biol Sci*, 2005, 272: 3-16
- [10] Wallace DC, Stugard C, Murdock D, et al. Ancient mtDNA sequences in the human nuclear genome: A potential source of errors in identifying pathogenic mutations[J]. *Proc Natl Acad Sci USA*, 1997, 94: 14900-14905
- [11] Caramelli D, Lalueza-Fox C, Capelli C, et al. Genetic analysis of the skeletal remains attributed to Francesco Petrarca[J]. *Forensic Sci Int*, 2007, 173: 36-40
- [12] Töpf AL, Gilbert MT, Dumbacher JP, et al. Tracing the phylogeography of human populations in Britain based on 4th-11th century mtDNA genotypes[J]. *Mol Biol Evol*, 2006, 23: 152-161
- [13] Gilbert MT, Hansen AJ, Willerslev E, et al. Characterization of genetic miscoding lesions caused by postmortem damage[J]. *Am J Hum Genet*, 2003, 72: 48-61
- [14] Pääbo S, Irwin DM, Wilson AC. DNA damage promotes jumping between templates during enzymatic amplification[J]. *J Biol Chem*, 1990, 265: 4718-4721
- [15] Yao YG, Kong QP, Man XY, et al. Reconstructing the evolutionary history of China: A caveat about inferences drawn from ancient DNA[J]. *Mol Biol Evol*, 2003, 20: 214-219
- [16] Lalueza-Fox C, Sampietro ML, Gilbert MT, et al. Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians[J]. *Proc Biol Sci*, 2004, 271: 941-947
- [17] Ricaut FX, Fedoseeva A, Keyser-Tracqui C, et al. Ancient DNA analysis of human neolithic remains found in northeastern Siberia[J]. *Am J Phys Anthropol*, 2005, 126: 458-462
- [18] Wen B, Xie XH, Gao S, et al. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans[J]. *American Journal of Human Genetics*, 2004, 74: 856-865
- [19] Yao YG, Nie L, Harpending H, et al. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity[J]. *American Journal of Physical Anthropology*, 2002, 118: 63-76
- [20] Yao YG, Lu XM, Luo HR, et al. Gene admixture in the Silk Road region of China: Evidence from mtDNA and melanocortin 1 receptor polymorphism[J]. *Genes & Genetic Systems*, 2000, 75: 173-178

- [21] Yao YG, Kong QP, Wang CY, et al. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China[J]. *Molecular Biology and Evolution*, 2004, 21: 2265-2280
- [22] Kong QP, Yao YG, Liu M, et al. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China[J]. *Human Genetics*, 2003, 113: 391-405
- [23] Wen B, Li H, Gao S, et al. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages[J]. *Molecular Biology and Evolution*, 2005, 22: 725-734
- [24] Qian YP, Chu ZT, Dai Q, et al. Mitochondrial DNA polymorphisms in Yunnan nationalities in China[J]. *Journal of Human Genetics*, 2001, 46: 211-220
- [25] Wen B, Li H, Lu DR, et al. Genetic evidence supports demic diffusion of Han culture[J]. *Nature*, 2004, 431: 302-305
- [26] Oota H, Kitano T, Jin F, et al. Extreme mtDNA homogeneity in continental Asian populations[J]. *American Journal of Physical Anthropology*, 2002, 118: 146-153
- [27] 杨亚军. 北方藏缅群体的精细遗传结构分析和基因交流 [D]. 上海: 复旦大学, 2008
- [28] Zhang F, Xu Z, Tan JZ, et al. Prehistorical East-West admixture of maternal lineages in a 2,500-year-old population in Xinjiang[J]. *American Journal of Physical Anthropology*, 2010, 142: 314-320
- [29] Gilbert MT, Binladen J, Miller W, et al. Recharacterization of ancient DNA miscoding lesions: Insights in the era of sequencing-by-synthesis[J]. *Nucleic Acids Res*, 2007, 35: 1-10
- [30] Brotherton P, Endicott P, Sanchez JJ, et al. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions[J]. *Nucleic Acids Res*, 2007, 35(17): 5717-5728
- [31] 何光岳. 氏羌源流史 [M]. 江西教育出版社, 2000
- [32] Strachan T, Read AP. *Human Molecular Genetics (2nd edition)*[M]. New York: Wiley-Liss, Chapter 12, Genetic mapping of complex characters, 1999

Genetic Analysis of People Who Lived in Shangsunjiazhai, Datong Qinghai as Revealed by Mitochondrial DNA

ZHANG Peng-yin^{1,2}, XU Zhi^{1,2}, XU Bo-song^{1,2}, HAN Kang-xin³,
ZHOU Hui⁴, JIN Li^{1,2}, TAN Jing-ze^{1,2}

(1. MOE Key Laboratory of Contemporary Anthropology, Fudan University, Shanghai 200433;

2. Shanghai Society of Anthropology, Shanghai 200433; 3. The Institute of Archaeology, Chinese Academy of Social Science, Beijing 100710; 4. College of Life Science, Jilin University, Changchun 130012)

Abstract: Ancient DNA analysis has been used to reveal the ethnic origins of ancient human remains. In this paper, we analyzed 59 human teeth excavated from two sites near Shangsunjiazhai, Datong in Qinghai, dated between 3000/3300 and 2000 years before present. The polymorphisms in the first hypervariable region and coding regions of mitochondrial DNA were analyzed using cloning sequencing and RFLP. Results were compared with 1833 mtDNA profiles from 34 contemporary populations and two ancient northwestern Chinese populations at both population and individual levels. Our results showed that people who lived in these two different periods were not the same ethnic population.

Keywords: Ancient DNA; mtDNA; Kayue Culture; Ancient Qiang