



线粒体证据揭示秦陵修陵劳工的来源

袁媛¹, 徐智¹, 谭婧泽¹, 张帆¹, 许渤松¹, 李士林¹, 李春香², 周惠², 朱泓², 张君³, 段清波⁴, 金力^{1,5}, 李辉¹

1. 复旦大学 生命科学学院 生物医学研究院 现代人类学教育部重点实验室 进化生物学中心, 上海 200433; 2. 吉林大学 边疆考古研究中心 古DNA实验室, 长春 130012; 3. 中国社会科学院 考古研究所, 北京 100710; 4. 陕西省文物与考古研究所, 西安 710054; 5. 中国科学院 上海生命科学研究院-马普学会 计算生物学伙伴研究所, 上海 200031.

摘要: 越来越多的研究试图利用古 DNA 的手段揭示古代人类个体的来源。2003 年秦始皇陵秦俑坑附近出土的一批人骨被推测是建陵劳工的遗骸, 年代大约在 2200 年前。我们通过克隆测序和 PCR-RFLP 的方法, 成功地从 19 个骨样本中得到线粒体 HVR 区和编码区的 SNP 位点的序列信息。为了揭示这些劳工的来源, 我们将所得序列与中国现代人的 32 个群体(2164 个体)的线粒体序列分别在个体和群体水平上作比较, 结果表明秦陵劳工墓的劳工来源是多样的。本文详尽地介绍了古 DNA 的操作方法、序列信息的甄别和数据分析方法。

关键词: 古 DNA; 线粒体 DNA; 秦始皇陵劳工

Origins of the Workers Who Built the First Qin Emperor's Mausoleum Revealed by Mitochondrial DNA

YUAN Yuan¹, XU Zhi¹, TAN Jingze¹, ZHANG Fan¹, XU Bosong¹, LI Shilin¹, LI Chunxiang², ZHOU Hui², ZHU Hong², ZHANG Jun³, DUAN Qingbo⁴, JIN Li^{1,5}, LI Hui¹

1. MOE Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433 China; 2. Ancient DNA Laboratory, Research Center for Chinese Frontier Archaeology, Jilin University, Changchun JL130012 China; 3. Institute of Archaeology, Chinese Academy of Social Sciences, Beijing 100710 China; 4. Shaanxi Provincial Institute of Cultural Relics and Archaeology, Xi'an SA710054 China; 5. Chinese Academy of Sciences and Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031 China.

Abstract: Ancient DNA analyses are always employed to reveal the ethnic origins of the ancient human remains. In this paper, we analyzed 19 human bone remains excavated from a tomb near the Qin Dynasty's Terra-Cotta Warriors and dated some 2,200 years before present. These samples were supposed to be workers building the mausoleum for the First Emperor of China. The polymorphisms in the first hypervariable region and coding regions of mitochondrial DNA were analyzed using cloning sequencing and RFLP. The results were compared with 2,164 mtDNA profiles from 32 contemporary Chinese populations at both population and individual levels. Our results showed that mausoleum-building workers may be originated from very diverse ethnic populations.

Key words: Ancient DNA; mtDNA; Mausoleum-building workers (MBWs); the First Emperor of China

2003 年 2-3 月, 陕西省考古研究所始皇陵考古队在清理一处秦代窑址时发现了 121 具人骨架。从骨骼结构改变(可能因繁重的体力劳动造成)、几乎全为男性(除一例怀疑为女性)、屈肢的埋葬方式以及生产工具和刑具等伴随物品这四个方面, 张君等推测他们是当年修建始皇陵的劳动者。目前尚未发现任何墓志或其他证据能够说明该人群的来源。由于埋藏环境曾受到水的侵蚀, 窑址人骨保存状况并不好。多数头骨严重受压变形甚至破碎, 不能得到较好的测量数据, 因而无法从骨骼形态上分析这些人的来源。初步观察发现, 个体间的颅面部形态特征差异较大, 推测该人群来源多元化。

秦始皇(公元前 259 年-公元前 210 年)是中国历史上第一个封建集权王朝——秦王朝的开国皇帝。据史书记载, 秦始皇即位的次年即开始修陵园, 到公元前 208 年完工, 历时 39 年(另两资料分别为 36 年和 38 年)。当时的丞相李斯为陵墓的设计者, 由大将军章邯监工, 共征集了 72 万人力, 修陵人数最多时近于 80 万, 几乎相当于修建胡夫金字塔人数的 8 倍。秦始皇陵位于陕西省西安市以东 30 公里的骊山北麓(另两资料: 位于临潼以东 5 千米处的下河村; 位于西安市以东 35 公里的临潼区境内), 陵园总面积为 56.25 平方公里, 建筑材料是从湖北、四川等地运来的。陵上封土原高约 115 米, 现仍高

达 76 米，陵园有内外两重城垣，内城周长 3840 米，外城周长 6210 米。内外城廓有长约 8-10 米的城墙，今尚残留遗址。秦始皇陵是中国历史上第一个皇帝陵园，其巨大的规模、丰富的陪葬物居历代帝王陵之首；古埃及金字塔是世界上最大的地上王陵，而中国秦始皇陵是世界上最大的地下皇陵。1974 年春，在秦始皇陵坟丘东侧 1.5 公里处，当地农民打井，无意中挖出一个陶制武士头。后经国家有组织的发掘，终于发现了震惊全世界的秦始皇陵兵马俑。兵马俑坑是秦始皇陵的陪葬坑，位于陵园东侧 1500 米处。现已发掘 3 座，俑坑坐西向东，呈“品”字形排列，坑内有陶俑、陶马 8000 多件，还有 4 万多件青铜兵器，被誉为“世界第八大奇迹”。1987 年，联合国教育、科学文化组织，把秦始皇陵列入世界文化遗产保护目录，成为全人类的共同财富。

通过生物学研究分析秦俑坑附近出土的秦代劳工遗骸，无疑有助于了解秦俑和秦陵建造者的来源，进而使我们对秦代历史有更多更直接的理解。本项研究就是通过古代 DNA 研究来分析这批遗骸的族属情况。古代 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。1997 年，Kriings 等[1]从尼安德特人(*Homo neanderthalensis*)化石中成功地提取了 mtDNA 片段，并进行 PCR 扩增和测序，同全世界 1600 例现代人的 mtDNA 进行比较之后，得出尼安德特人与现代欧洲人没有直接的继承关系，只是现代人的旁系远亲而非直系祖先，支持非洲起源假说。这是古 DNA 研究用于人类学研究的里程碑事件。其中，线粒体 DNA(mtDNA)位于线粒体中，母系遗传，是闭合的环状双链 DNA 分子。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，分为 HVRI (16024-16400)、HVRII(44-340)、HVRIII (438-576)。利用 mtDNA 多态性遗传标记来追溯人类群体起源迁移事件发生的大致时间及路线，成为人类进化研究的重要手段。

一、材料与amp;方法

1. 墓葬人骨材料

我们从陕西考古所获得了所有 121 份个体的样本，形态学分析表明这些样本均来自不同个体。我们随机选取了 50 个样本进行 DNA 分析，人骨样品主要是肢骨残段。

2. 实验方法

DNA 提取: a)去污染预处理: 在股骨中段取 1cm×2cm 的骨片，打磨去除表面 1-2mm; 漂白液浸泡、无水乙醇清洗后，空气干燥; 置于 256nm 紫外灯下照射。b)样品的粉碎: 用 SPEX Sample Prep Freezer/Mills6750 粉碎处理过的骨片。c)DNA 抽提: DNA 抽提步骤按照 Rohland 和 Hofreiter[2]所描述。

PCR 扩增: a)25μl 的 PCR 体系: 1U rTaq 聚合酶(Takara), 200μM dNTP, 1.5mM Mg²⁺, 1μM 引物(表 1), 1μg BSA(Takara)以及 4μl 的 DNA 抽提液。b)PCR 程序: 94°C 预变性 15min, 45 个循环(94°C30s, 52°C30s, 72°C30s), 72°C 延伸 7min。c)PCR 产物鉴定: 5μl PCR 产物通过 2%的琼脂糖凝胶电泳, 条带单一且长度准确的 PCR 产物被用于后续操作。

PCR 产物直接测序: 所得 PCR 产物, 首先通过直接测序初步了解其 DNA 的序列情况。a)PCR 产物纯化: 1.5-2μl(10ng)的 PCR 产物与 0.1U 虾碱磷酸酶(Shrimp Alkaline Phosphatase, SAP)和 1.25U 外切酶(ExonI)混合后, 37°C 孵育 80min, 80°C 15min 变性失活酶。b)测序反应: 使用 Big Dye Terminator kit (Applied Biosystems), 程序为 96°C 10s 预变性, 25 个循环(96°C 10s, 50°C 5s, 60°C 90s), 60°C 延伸 4min; 测序引物为相应的 PCR 引物。

克隆测序: 直接测序得到的序列信息不能真实地反应古代样品中 DNA 的序列, 要获得可靠的序列信息, 需要进行克隆测序。a)PCR 产物纯化、连接: 使用 AxyPrep PCR Clean-up Kit(Axygen)纯化 PCR 产物, 所得产物 5μl 与 5μl 的 Solution1 和 0.5μl 的 pMD18-T Simple Vector(Takara)混合, 4°C 放置过夜。b)克隆: 氯化钙法制备大肠杆菌感受态细胞, 转化, 活化, 平皿培养过夜。c)PCR 鉴定克

表 1 古 DNA 分析所用引物 Tab. 1 The primers used in ancient DNA analysis

引物	序列(5'-3')	片段长度	多态位点	
测序	L16053	GGGAAGCAGATTTGGGTAC	146bp	HVRI
	H16158	GATGTGGATTGGGTTTTATG		
	L16154	TACCATAAATACTTGACCACCTG		
	H16265	GTTTGTGGTATCCTAGTGGG	156bp	HVRI
	L16263	AACTCCAAAGCCACCCC		
	H16366	TGAGGGGGGTCATCCAT	235bp	HVRI
	L16016	ATTCTCTGTTCTTTCATGGG		
	H16403	ATTGATTTACGGAGGATGG	138bp	HVRI
	L16017♪	TTCTCTGTTCTTTCATGGGGA		
	H16251♪	GGAGTTGCAGTTGATGTGTGA	408bp	HVRI
	L16201♪	CAAGCAAGTACAGCAATCAAC		
	H16409♪	AGGATGGTGGTCAAGGGA	235bp	HVRI
	RFLP	L629	CACATCACCCATAAAACAAAT	113bp
H742		ATCGTGGTGATTTAGAGGGT		
L5135		CTCCAGCACACGACCCTAC	119bp	5176AluI
H5253		GCAAAAAGCCGGTTAGCG		
L12362		CCCTAACCCCTGACTTCCCTAAT	106bp	12406HpaI
H12449		AAGTGGATGCGACAATGG		
L13188		CACTCTGTTTCGACGAGTCTG	113bp	13262AluI
H13280		GTTGGTTGATGCCGATTGTAA		
L8226		TCCCCTAAAAATCTTTGAAATA	100or91bp	9bpΔ
H8325		AAAAGGTTAATGCTAAGTTAGCTTTA		
L4798		CCTTTCACCTTCTGAGTCCCA	117bp	4833HhaI
H4914		GTGAGGGAGAGATTTGGTATATG		
L10341		ATCATCATCTAGCCCTAAGTC	99bp	10398Ddel
H10439	GTCGAAATCATTCTGTTTTGTT			
克隆测序	RV-M	GAGCGGATAACAATTTACACAGG	156bp	
	M13-47	CGCCAGGGTTTTCCCAATGACACGAC		

L 和 H 分别代表 mtDNA 的轻、重链，引物编号参照 rCRS[3](Genbank, AC_000021), ♪代表在吉林大学重复实验所使用的引物。L and H indicate the light and heavy chains of mtDNA respectively numbering according to the revised Cambridge reference sequence (rCRS) [3](Genbank, AC_000021). Symbol ♪ represents primers used for the independent replication experiment in Jilin University.

隆：PCR 体系同上，引物为 RV-M 和 M13-47(表 1)；程序为 94℃5min 预变性，35 个循环(95℃30s, 55℃30s, 72℃30s)，72℃延伸 10min；PCR 产物电泳鉴定，没有片段插入的 PCR 产物长度为 156bp，有片段插入的 PCR 产物长度约为 300bp，此为成功连接的 PCR 产物。d)克隆子测序：取 8 个成功连接的 PCR 产物 10ng 作为测序模板测序。

RFLP 分型：为进一步确定单倍群归属，根据 mtDNA 的谱系树[4]，通过限制性酶切获得编码区的 SNP 位点信息。PCR 体系同上，根据初步确立的单倍群，选择不同的引物和内切酶(表 1)。

3. 序列分析

单倍型确定：测序源文件导入到 Lasergene7.1.0(DNASTAR, <http://www.dnastar.com>)，将 rCRS 作为参照进行序列比对。通过克隆序列获得样品的单倍型，通过与已经发表的来自世界各地的 26000 多个个体的单倍型比较之后，初步确立其单倍群归属。

数据分析：研究秦陵劳工墓样品的 DNA，目的是揭示其来源。我们获取了来自中国各地的 32 个群体共 2164 个个体的 mtDNA 序列(表 2, 图 1)。根据其地理位置和民族分类以及南北人群的显著差异[5,6]，该 32 个群体被分为北方汉族(NH)、南方汉族(SH)、北方少数民族(NM)和南方少数民族(SM)。用 SPSS11.0.1 绘制基于单倍群比率的主成分分析图；通过 ARLEQUIN3.1[7]计算基因和核苷酸多样性；用 Network4.5.0.0 (www.fluxus-engineering.com)绘制基于邻接法的最大简约树。

二、可靠性分析

按照古 DNA 研究的可靠性标准[15-18]，我们对该 50 个样本采取严格的实验步骤，以避免外源污染。其中，有 22 个样本没有可扩增的 DNA，予以舍弃；将剩余的 28 个样品进行重复实验，其中 9 个不可重复，予以舍

表2 用于数据比较的32个现代中国人群
Tab. 2 Modern Chinese populations used for comparison with Qin samples

群体	编号	样本量	来源	
北方汉 (NH)	陕西	1	53 [6]	
	青海	2	44 [6]	
	内蒙古	3	45 [6]	
	西安	4	85 [8]	
	辽宁	5	51 [6]	
	甘肃	6	45 [6]	
南方汉 (SH)	浙江	7	61 [6]	
	云南	8	59 [6]	
	台湾	9	66 [9]	
	四川	10	70 [6]	
	上海	11	56 [6]	
	江西	12	23 [6]	
	江苏	13	67 [6]	
	湖南	14	16 [6]	
	广西	15	29 [6]	
	福建	16	54 [6]	
	长沙	17	82 [8]	
北方少数民族 (NM)	安徽	18	42 [6]	
	广东	19	29 [10]	
	蒙古	20	48 [11]	
	朝鲜	21	48 [11]	
	哈萨克	22	53 [12]	
	鄂温克	23	47 [11]	
	藏	24	56 [13]	
	维吾尔	25	47 [12]	
	南方少数民族 (SM)	彝	26	56 [14]
		土家	27	96 [13]
瑶		28	395 [14]	
苗		29	142 [14]	
傣		30	38 [10]	
白		31	55 [13]	
壮		32	83 [10]	

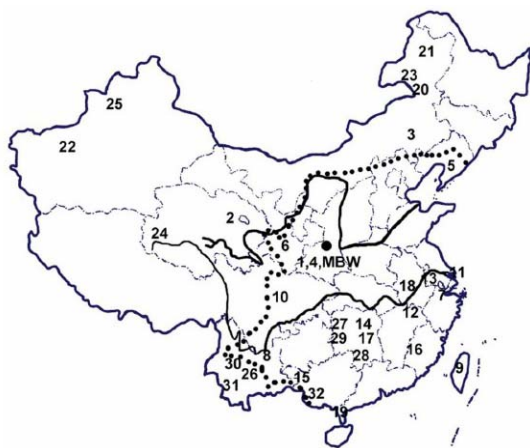


图1 秦俑劳工墓和32个现代中国人群的地理位置
虚线表示秦朝的边疆，群体的编号见表2，劳工墓位置以实心圆表示。

Fig.1 The geographical locations of Qin Mausoleum worker tomb and the 32 Chinese populations. The dot line indicates the border of Qin Dynasty; the numbers are the same as in Table 2, and the big dot indicates the location of mausoleum worker tomb.

弃。在所得的19个样本中随机挑取两个(M54和M57)在吉林大学进行独立重复实验，结果具有一致性。此外，这19个样本的DNA显

示了各种古DNA的分子特性，且与实验人员的序列不同，因此我们认为这19个样本的序列信息是可靠的。

1. 实验过程的控制

所有样品都是新近挖掘的，且尽可能避免触摸(研究表明博物馆收藏的材料有高比例的污染[19]，且污染水平多样[20])。所有的PCR前操作都在一个隔离的古DNA专用实验室中进行。此外，实验人员保持(PCR前→PCR→PCR后)的单向流动，以防止外源DNA通过粘附在衣物和皮肤上进入古DNA实验室；长时间紫外照射和次氯酸钠的使用[21]是行之有效的消除污染的方法；使用一次性手套、口罩和带滤芯的移液头等避免样品间的交叉污染。

三段PCR目的产物首尾有重叠区域，这样可以减小误扩增到核mtDNA的可能性[17,22,23](表1)。为检测可能的污染，在抽提和PCR时，都引入阴性对照，与样品的比例为1:1。

理论上，一个克隆序列代表一个DNA片段分子的插入，即序列是唯一的，若存在杂合，很可能是交叉污染，将此克隆剔除。如果某一信息位点有6个或6个以上的克隆序列显示同样的碱基，则认为该位点是可靠的(图2)。将所得序列与rCRS比对，逐一检验各个位点，最后确定该样品的单倍型。

图2中rCRS[3]为参照标准序列；“.”表示该位点与rCRS一致；a、b、c分别代表HVRI上的三段不同片段(16053-16158、16154-16265、16263-16366)。克隆测序反应了PCR产物中单分子的序列信息，可以通过不同的克隆观察到多聚酶的低保真度所造成的某一位点的假象、测序错误以及跳跃PCR[24,25]。在大多数的克隆中，位点16223、16239、16298、16327和16357上都出现了转换，因而可以认为样品M54的mtDNA HVRI的可靠单倍型为16233-16239-16298-16327。C7和C8不同于C1-C6，被认为是外源污染DNA，但和实验人员的DNA序列不同(表3)，因此并不是源于实验人员。16079、16089、16107、16109、16148、16172、16211、16227、16259、16306和16313位点上的碱基置换被认为是散在的碱基错配，是

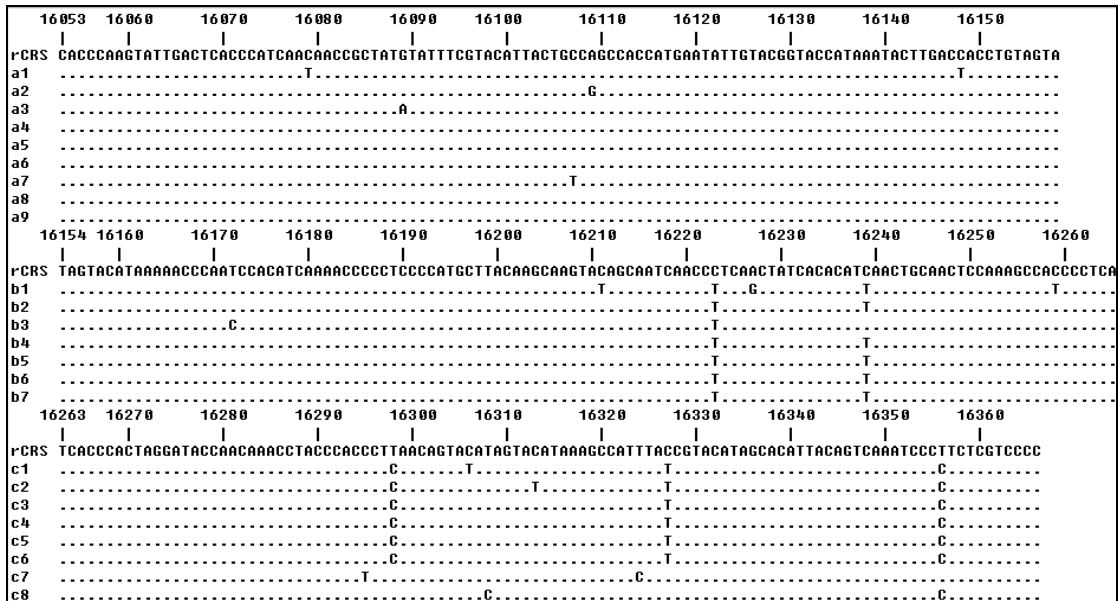


图2 秦陵劳工墓 M54 的 mtDNA 克隆序列对比 Fig.2 Alignment of cloned mtDNA sequences from sample M54.

表3 在所有 19 个劳工墓样品的克隆序列中出现的碱基错配

Tab.3 Miscoding lesions observed within clones from MBWs

Sample	Transition(Type1)		Transition(Type2)		Transversion				
	T→C	A→G	C→T	G→A	G→T	A→T	C→A	T→G	A→C
M7			4						
M17	3	1	2		2	1			
M21	1		5	1					1
M31	2	1	1	2					
M37	4		3						1
M39	4	4	1						
M41	3	3	7	1					
M43			4				1		
M47	3		3				1		
M50		1	4						
M53		2	4	1				1	
M54	2	2	7	1					
M57	1		2	2	1				
M60		3	5						
M61	1		1	1					
M71	1		4	1					
M86	4								
M91	3		2						
M98	1					1			
Total	33	17	59	10	3	2	4	1	3

表4 秦陵劳工墓的挖掘和分析人员的线粒体高变 I 区的序列特征和所属单倍群

Tab.4 HVSI motives and mtDNA haplogroups of the staff involved in the excavations and experiments for MBWs.

人员#	HVRI(16053-16366), 16000+	单倍型
F1	66-223-311	M*
F2	223-234-290-362	M*
F3	172-223-356-362	不明
F4	223-234-291-316-362	M9a
F5	189-223-311	M*
F6	111-129-140-183C-189-200T-243	B5b
A1	68-86-189-271-304-311	F1
A2	223-362	不明
J1	136-183C-189-217-218-239-248	B4b1
J1	126-174-223-311-362	不明
J1	145-182C-183C-189-217-261-360	B4a

#F1-F6, 复旦大学的研究人员; A1 和 A2, 陕西考古所的研究人员; J1-J3, 吉林大学的研究人员。 F1-F6, the staff of the lab at Fudan University; A1 and A2, the staff of Shaanxi Institute of Archaeology; J1-J3, the staff of the lab at Jilin University.

古 DNA 的一个明显特征[26]，不是污染。

2. 重复实验

虽然古 DNA 存在碱基错配，但同一个碱基错配在多次抽提中重复出现几乎是不可能的[26]，因此每个样品都经过两次独立重复抽提。重复实验从预处理样本开始，获得的序列结果与之前的序列比对，如存在较大差异，则舍弃该样品；如两次的实验结果一致，则认为该序列是可靠的；如部分信息位点一致，则再次对样品进行重复实验。

3. 不同实验室的独立重复

从经过重复实验确认的样品中随机选取 2 个样品，在吉林大学边疆考古研究中心进行独立重复实验。DNA 抽提使用 GENE CLEAN Kit 或 Ancient DNA(Q-BIOgene)；12.5μl PCR 体系包括 1U Taq 酶，200μM dNTP，2.5mM MgCl₂，50mM KCl，2μM 引物(表 1)和 2mg/mL BSA。PCR 程序为 94°C 4min，6 个循环(94°C 1min，58°C 55s，72°C 1min)，34 个循环(94°C 1min，54°C 55s，72°C 1min)，72°C 延伸 10min。PCR 产物在 ABI310 上直接测序。

4. 古代 DNA 分子特性

所有 19 个样品都显示出一定的古代 DNA 分子特性(表 3)，包括不同克隆之间观察到的嵌合(chimeric)序列[27,28](图 2)以及 408bp 的长片段的扩增失败。除了 M60 带有 16102T→C 这一罕见突变位点[29]，其他序列

都没有超出已定义的单倍群位点。对 mtDNA HVRI 区域共 314 个核苷酸(16053-16366)测序，与 rCRS 比对，共有 45 个置换(38 个转换和 7 个颠换)，没有观察到缺失和插入。

虽然本研究没有测定样本中氨基酸的外消旋性，但是该 19 个样本在实验室内和实验室间的重复实验都得到一致的结果，即可判断必定存在内源 DNA，这一替代方法被用于很多古 DNA 研究 [26,30,31]。为确定样品 DNA 的片段长度和拷贝数存在负相关性，需要测定内源 DNA 的含量，但是初步的实时定量 PCR 显示，样品中的抑制因子使荧光信号不稳定或者不存在[32]，我们进一步扩增 408bp 的片段 [15,33,34]，所有样品均无扩增结果，这证明样品中不存在长片段的 DNA(即可能的污染)。

5. 排除实验人员污染

将实验人员的 mtDNA 序列与获得的样本 DNA 序列做比较，以排除所得序列来自实验人员(表 4)。

三、序列分析

通过以上一系列筛选，我们从 50 个样品中获得 19 个可靠的古代样品的 mtDNA(表 5)，并对其进行了序列分析(<http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html>，EU700062-EU700080)。线粒体 HVRI 单倍型、RFLP 结果以及单倍群归属如表 5。

表 5 秦陵劳工墓样品的 mtDNAHVRI 单倍型、RFLP 结果以及单倍群归属

Tab. 5 mtDNA HVRI and RFLP results of MBWs and individual haplogroup assignment.

样品	单倍群	HVRI单倍型	663 HaeIII	5176 AluI ^{9bp}	12406 HpaI	13262 AluI	10400 AluI	10398 DdeI	4833 HhaI
M7	M8a	184-189-192-223-298-311-319	/	+	2	+	/	/	/
M17	B5b	140-183C-189-243-355	/	+	1	+	/	/	/
M21	F*	207-304-311	/	/	/	/	/	/	/
M31	N*	223-245	/	+	/	/	/	/	+
M37	F1b	183C-189-195G-232A-249-258C-304-311	/	/	/	/	/	/	+
M39	F1b	189-232A-249-265C-304-311	/	/	/	/	/	/	/
M41	N9a	111-129-223-257A-261	/	/	/	/	/	/	/
M43	N9a	129-223-257A-261	/	+	2	+	/	/	/
M47	N9a	223-257A-261-311	/	/	/	/	/	/	+
M50	M7a	209-223	/	/	/	/	+	+	+
M53	D*	223-299-325-362	/	/	2	+	/	/	/
M54	C	223-239-298-327-357	/	/	/	/	+	+	+
M57	F1a1	129-145-162-172-304	/	+	2	T/	/	/	/
M60	G2a	92-102-223-227-272-278-319-362	/	/	/	/	+	+	+
M61	D5	164-172-182C-183C-189-223-362	/	/	2	+	/	/	/
M71	F1a1	162-172-304	/	/	/	/	/	/	/
M86	A	223-290-311-319-362	+	+	2	+	/	/	/
M91	B5	140-183C-189-362	/	+	1	+	/	/	/
M98	B4a	182C-183C-189-217-261-360	/	+	1	+	/	/	/

1. 主成分分析

为了便于比较, 现代人数据仅使用 16053-16366 上的序列, 即 HVRI 区。其主成分分析(图 3)表明, 秦陵劳工墓 19 个样品构成的古代群体和 32 个中国现代对比人群没有群体上的相似性。

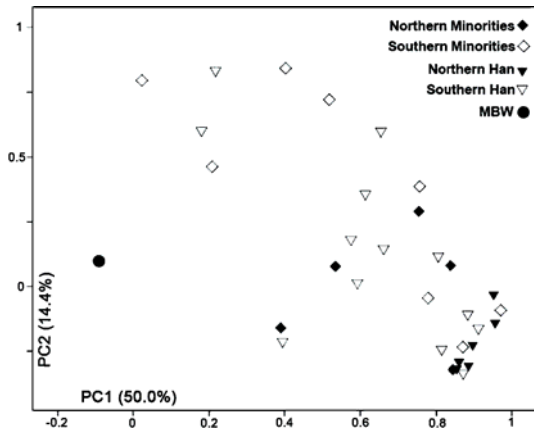


图 3 秦陵劳工墓样品和现代中国人群基于单倍群出现比率的主成分分析 32 个群体根据地理位置和民族被分成 4 组(表 2)。

Fig. 3 Principal component analysis of mtDNA haplogroup frequencies among MBWs and modern Chinese populations. The 32 modern populations were divided into four groups (see Table 2).

2. 群体多样性比较

秦陵劳工墓的 19 个样本构成的群体基因多样性和核苷酸多样性的值分别为 1.000 ± 0.017 和 0.019 ± 0.010 , 略高于其他 32 个群体(表 6), 但是差异不显著($p > 0.05$)。

3. 最大简约树

秦陵劳工墓的 19 个样品归属于 15 个单倍群, 没有显示明显的欧亚大陆西部人群特有的单倍群。由于单倍群 M7a 和 B5 在 2164 个现代人个体中很少见, 无法构树; 图 4 是其它 13 个单倍群的最大简约树。尽管 2000 年来的迁徙可能影响群体中单倍群的频率, 但聚类分析对于追溯样品的来源仍具有一定参考价值。秦陵劳工墓样品 M86(单倍群 A)与青海的一个北方汉族个体具有同样的 mtDNA HVRI 单倍型; M98(单倍群 B4a)与南方少数民族聚类在一起, 而且单倍群 B4a 在南方少数民族中的频率很高[13], 表明其来源可能是中国南方的少数民族; M17(单倍群 B5b)可能来自南方汉族(这里指南方汉族的祖先群体, 下文的类似情况不再注明),

M60(单倍群 G2a)可能来自汉族; M61(单倍群 D5)与一些南方少数民族个体聚类在一起, 而 M21(单倍群 F*)则与一些北方汉族个体聚类在一起。尽管从聚类上无法区分南方和北方, 但考虑到单倍群 F1a1 在东南亚频率较高[35], 推测 M57 和 M71 可能来自南方; 单倍群 D*在南方[13]和北方少数民族[11,12]中都高频地存在, 而在汉族中频率很低, 所以 M53 可能不是汉族。此外, 最大简约树和单倍群的频率分布无法揭示 M37(单倍群 F1b)、M39(单倍群 F1b)、M41(单倍群 N9a), M43(单倍群 N9a), M47(单倍群 N9a), M31(单倍群 N*), M54(单倍群 C)和 M7(单倍群 M8a)的来源; 由于缺少对比数据, 也无法确定 M91(单倍群 B5)和 M50(单倍群 M7a) 的来源。

四、讨论

本研究中, 我们成功提取并分析了秦陵劳工墓 19 个个体的母系遗传信息, 成功率为 38%, 印证了整个墓葬保存状况不佳。在吉林大学进行的独立重复实验能扩增出大于 200bp 的 DNA 片段, 说明这 19 个个体的保存状况相对较好。由于获得的 DNA 多样性很高(表 6), 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(表 4, 图 2)显示, Type2(C→T/G→A)的碱基错配在大多数样品中占有比较大比率, 这与之前的研究结果[36,37]一致。

通过分析, 我们认为, 秦陵劳工墓的劳工源于东亚, 且来自中国各地人群。mtDNA 的谱系树[4,38,39]显示这 19 个样品无一例外均属于欧亚大陆东部特异单倍群。这 19 个样本组成的群体的多样性大于其他 32 个现代群体, 且主成分分析也强有力地支持该古代群体和 32 个现代群体存在显著差异。同时, 个体水平上的聚类分析也显示出其来源的多样性。其中, 4 个样品(M86、M60、M21 和 M17), 占总数的 21%, 可能来自汉族; 3 个样品(M98、M61 和 M53), 占总数的 16%, 可能来自南方少数民族; 地理上, 7 个样品(M17、M60、M21、M98、M61、M57 和 M71), 占总数的 37%, 可能来自于南方; 仅有 1 个样品(M86), 占总数的 5%, 与北方少数民族

表 6 秦陵劳工墓发现单倍群在 32 个现代中国人群中的出现比率以及各个群体的基因多样性
 Tab.6 The proportion of the MBWs' haplogroups in 32 modern populations and the gene diversity of the populations

群体	样本量	mtDNA单倍群在各群体中所占比率(%)														基因多样性 (S.D.)			
		A	B4a	B5	B5b	C	D*	D5	F*	F1a1	F1b	G2a	M7a	M8a	N*		N9a	其他	
劳工墓	19	5.26	5.26	5.26	5.26	5.26	5.26	5.26	5.26	10.53	10.53	5.26	5.26	5.26	15.79		1.000(0.017)		
北方汉(NH)	1	53	5.66			1.89	20.75	1.89	3.77	5.66	1.89			1.89	5.66	1.89	49.06	0.999(0.004)	
	2	44	9.09			2.27	18.18	9.09	2.27	2.27		4.55		6.82	2.27	2.27	40.91	0.996(0.006)	
	3	45	4.44	4.44		2.22	4.44	22.22	2.22	2.22	2.22			2.22	6.67	44.44	0.995(0.006)		
	4	84	8.24	1.18		3.53	4.71	24.71	5.88	3.53			1.18		1.18	3.53	42.35	0.998(0.002)	
	5	51	3.92	3.92		1.96	1.96	21.57	3.92		3.92	1.96		7.84	1.96	3.92	41.18	0.991(0.008)	
	6	45	17.78					24.44		2.22	6.67			2.22	4.44	2.22	40	0.992(0.007)	
南方汉(SH)	7	61	4.92	6.56	1.64	1.64	6.56	6.56	1.64	1.64	3.28	1.64		3.28	1.64	6.56	50.82	0.996(0.004)	
	8	59	5.17	3.45		5.17	10.34	3.45			6.9			1.72	3.45		60.34	0.990(0.007)	
	9	66	7.58	6.06		1.52	3.03	6.06	6.06	4.55		1.52	3.03	3.03	1.52	1.52	3.03	51.52	0.995(0.004)
	10	70	4.29	4.29	1.43		1.43	17.14	1.43	1.43	2.86		4.29		1.43	2.86	57.14	0.996(0.003)	
	11	56	3.57	1.79			5.36	16.07	3.57		1.79				7.14	3.57	3.57	53.57	0.997(0.004)
	12	23		8.7	4.35	17.39		17.39	4.35		13.04	4.35	4.35					26.09	0.996(0.014)
	13	67	10.45	1.49		2.99	1.49	13.43			5.97	2.99			4.48	2.99	2.99	50.75	0.996(0.004)
	14	16	6.25	6.25				12.5									18.75	56.25	0.983(0.028)
	15	26	3.45	17.24		3.45	10.34	3.45	3.45		3.45					3.45	3.45	48.28	0.993(0.011)
	16	54		1.85		3.7	3.7	9.26	3.7	3.7	5.56	5.56			3.7	5.56	53.7	0.996(0.004)	
	17	82	6.1	1.22				23.17		1.22			6.1	1.22	1.22	4.88	4.88	50	0.995(0.003)
	18	42	11.9	7.14		4.76		9.52	7.14		2.38	4.76			2.38	4.76	45.24	0.995(0.006)	
	19	56		6.9			3.45	3.45	3.45	10.34		3.45				6.9		62.07	1.000(0.010)
	北方少数民族(NM)	20	48	8.33	2.08			6.25	31.25	6.25							6.25	39.58	0.989(0.007)
21		48	6.25			4.17		22.92	10.42				6.25			2.08	47.92	0.976(0.010)	
22		53	3.77	3.77			13.21	13.21			5.66	1.89		1.89	1.89		54.72	0.993(0.011)	
23		47	4.26			2.13	19.15	25.53	4.26								44.68	0.956(0.011)	
24		56	4.76				9.52	4.76	4.76	2.38		4.76	11.9		7.14	11.9	38.1	0.991(0.005)	
25		47	4.26				6.38	6.38	2.13			2.13	10.64		6.38		61.70	0.999(0.006)	
26		56	12.5	1.79	1.79		1.79	12.5	1.79	7.14		3.57					57.14	0.996(0.004)	
南方少数民族(SM)	27	96	3.13	5.21		3.13	9.38	12.5	3.13	8.33		3.13			1.04	4.17	46.88	0.993(0.003)	
	28	395	0.76	10.13	0.25	0.51	2.28	0.51	0.76	1.27	2.78	2.53		0.25	0.25	3.04	74.68	0.987(0.002)	
	29	142	4.93	7.04		0.7	7.04	2.82	2.11	2.11	2.82	2.11		1.41	0.7	2.11	64.08	0.983(0.005)	
	30	38	5.26	7.89			10.53	5.26	5.26	2.63							63.16	0.994(0.008)	
	31	55	5.45	1.82				14.55	1.82	3.64		3.64		1.82		3.64	63.64	0.997(0.004)	
	32	83	2.41	1.2		1.2	4.82	1.2	4.82	2.41					3.61	1.2	2.41	74.7	0.992(0.004)

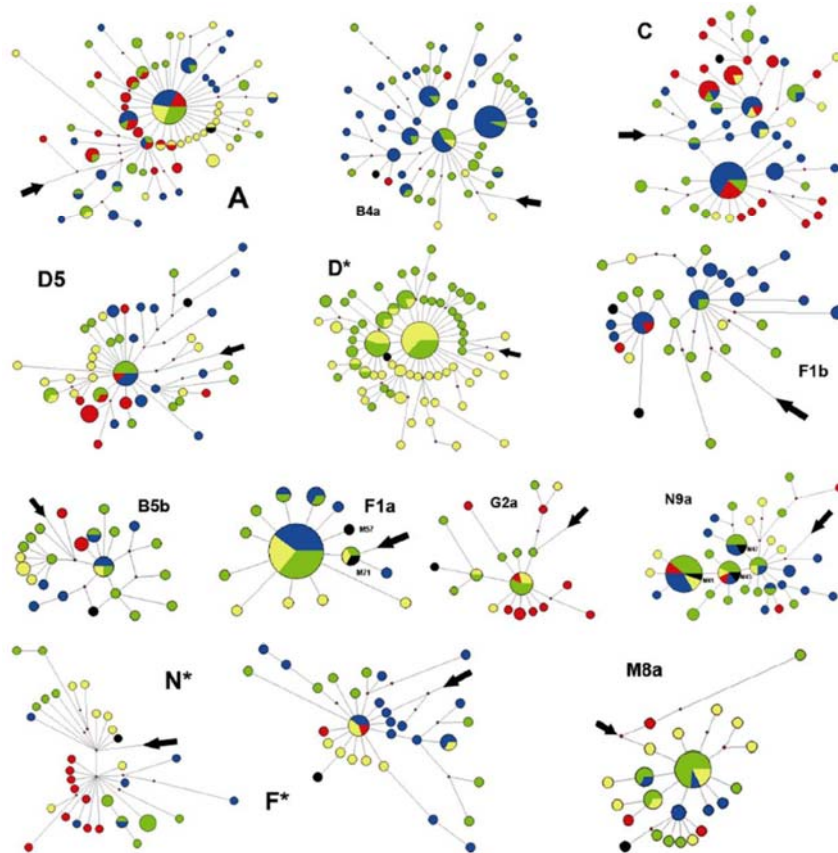


图 4 基于邻接法的 13 个单倍群的最大简约树 秦俑劳工墓样品-黑色, 北方汉-黄色, 南方汉-绿色, 北方少数民族-红色, 南方少数民族-蓝色, 外类群用黑色箭头标识。Fig.4 The maximum parsimony tree based Neighbor-joining Method. MBW-black, NH-yellow, SH-green, NM-red, SM-blue, outgroup-black arrow.

的群体相似,可能由于我们选取的参照群体的地理位置离秦朝的边界(图 1)较远,而无法与秦俑墓地样本中可能存在的更多来自北方少数民族的部分匹配。根据薛雅丽等[40]对 mtDNA 单倍群的研究,样品 M7、M53、M54、M60、M61 和 M86 的单倍群 M8a、D*、C、G2a、D5 和 A, 高频率地出现在中国北方,因此这 6 个样品可能来自于北方;而有 10 个样品(M17、M21、M31、M37、M39、M50、M71、M57、M91 和 M98), 分别属于单倍群 B5b、F*、N*、F1b、F1b、M7a、F1a1、F1a1、B5 和 B4a, 占总数的 53%,可能来自中国的南方。

由此我们认为,秦陵劳工墓的劳工有的来自汉族祖先群体,有的来自少数民族;而且在地理分布上,南方的略多于北方。这与一些考古证据不同,比如赵背户姚池头墓地发现的瓦书墓志记载了死者的来源主要集中在山东、江苏、河南、河北等地[41],但是这些死者大多是有爵位的;而从葬式、随葬物品、没有墓志和骨骼损伤等角度看,这 121 个劳工应该是处于修建秦始皇陵的劳工中最底层的,因而能更真实地反应 72 万劳工的来源。有趣的是,样品 M50(mtDNA HVRI 单倍型 16209-16223,属于 M7a 单倍群)与一个日本人[8]和几个琉球人[9,42]具有同样的单倍型。考虑到单倍群 M7a 在琉球人种的频率比中国高[35],而且其多样性在亚洲最高[42],很可能在母系关系上样品 M50 和日本人的祖先比较近。

五、结论

我们的研究显示,在母系上,秦陵劳工墓的劳工和现代中国人群具有遗传一致性,并且来源于中国各地人群,而且南方的略多于北方。尽管秦朝之后的迁徙对现代人群的遗传结构会造成一定影响,但该结果一定程度上能证明当年为修建秦始皇陵从全国征用了大量劳力。如果需要进一步明确秦陵劳工的来源,可对这批样品做更深入的研究,比如 mtDNA 全序列测序和 Y 染色体的遗传分析。

致谢: 本研究得到国家自然科学基金(30571013),上海市浦江人才计划(05pj14023),国家杰出青年基金(30625016)的资助。

参考文献

1. Krings M, Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S. (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90: 19-30.
2. Rohland N, Hofreiter M(2007)Ancient DNA extraction from bones and teeth. *Nat Protoc* 2:1756-1762.
3. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
4. 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.
5. Zhao TM, Lee TD (1989) Gm and Km allotypes in 74 Chinese populations:a hypothesis of the origin of the Chinese nation. *Hum Genet* 83:101-110.
6. Wen B, Li H, Lu D, Song X, Zhang F, He YG, Li F, Gao Y, Mao XY, Zhang L, Qian J, Tan JZ, Jin JZ, Huang W, Deka R, Su B, Chakraborty R, Jin L (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302-305.
7. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3. 0:An integrated software package for population genetics data analysis. *Evol Bioinfo* 1:47-50.
8. Oota H, Kitano T, Jin F, Yuasa I, Wang L, Ueda S, Saitou N, Stoneking M (2002) Extreme mtDNA homogeneity in continental Asian populations. *Am J Phys Anthropol* 118:146-153.
9. Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am J Hum Genet* 59:579-590.
10. Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118:63-76.
11. Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, Palanichamy MG, Zhang YP (2003) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet* 113:391-405.
12. Yao YG, Kong QP, Wang CY, Zhu CL, Zhang YP (2004) Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol Biol Evol* 21:2265-2280.
13. Wen B, Xie X, Gao S, Li H, Shi H, Song XF, Qian TZ, Xiao CJ, Jin JZ, Su B, Lu D, Chakraborty R, Jin L (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* 74:856-865.
14. Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He YG, Dong YL, Zhang YJ, Huang W, Jin JZ, Xiao CJ, Lu DR, Chakraborty R, Su B, Deka R, Jin L (2005) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol* 22:
15. Cooper A, Poinar HN (2000) Ancient DNA:do it right or not at all. *Science* 289:1139.
16. Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-679.
17. Willerslev E, Cooper A (2005) Ancient DNA. *Proc Biol Sci* 272:3-16.
18. Gilbert MT, Bandelt HJ, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends Ecol Evol*

- 20:541-544.
19. Malmström H, Storå J, Dalén L, Holmlund G, Götherström A (2005) Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Mol Biol Evol* 22:2040-2047.
 20. Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Detter JC, Pääbo S, Rubin EM (2005) Genomic sequencing of Pleistocene cave bears. *Science* 309:597-599. *Am J Hum Genet* 74:856-865.
 21. Kemp BM, Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Sci Int* 154: 53-61.
 22. Wallace DC, Stuard C, Murdock D, Schurr T, Brown MD (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. *Proc Natl Acad Sci USA* 94:14900-14905.
 23. Caramelli D, Lalueza-Fox C, Capelli C, Lari M, Sampietro ML, Gigli E, Milani L, Pilli E, Guimaraes S, Chiarelli B, Marin VTW, Casoli A, Stanyon R, Bertranpetit J, Barbujani G (2007) Genetic analysis of the skeletal remains attributed to Francesco Petrarca. *Forensic Sci Int* 173:36-40.
 24. Gilbert MT, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A (2003) Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet* 72:48-61.
 25. Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265:4718-4721.
 26. Topf AL, Gilbert MT, Dumbacher JP, Hoelzel AR (2006) Tracing the phylogeography of human populations in Britain based on 4th-11th century mtDNA genotypes. *Mol Biol Evol* 23:152-161.
 27. Hebsgaard MB, Wiuf C, Gilbert MT, Glenner H, Willerslev E (2007) Evaluating Neanderthal genetics and phylogeny. *J Mol Evol* 64:50-60.
 28. DeSalle R, Barcia M, Wray C (1993) PCR jumping in clones of 30-million-year-old DNA fragments from amber preserved termites (*Mastotermes electrodominicus*). *Experientia* 49:906-909.
 29. Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am J Hum Genet* 46:828-842.
 30. Lalueza-Fox C, Sampietro ML, Gilbert MT, Castri L, Facchini F, Pettener D, Bertranpetit J (2004) Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci* 271:941-947.
 31. Ricaut FX, Fedoseeva A, Keyser-Tracqui C, Crubezy E, Ludes B (2005) Ancient DNA analysis of human neolithic remains found in northeastern Siberia. *Am J Phys Anthropol* 126:458-462.
 32. Sutlovic D, Gamulin S, Definis-Gojanovic M, Gucic D, Andjelinovic S (2008) Interaction of humic acids with human DNA: Proposed mechanisms and kinetics. *Electrophoresis*.
 33. Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. *Nat Rev Genet* 2:353-359.
 34. Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci USA* 86:1939-1943.
 35. Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Lee ZY, Lin M (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol* 3:e247.
 36. Gilbert MT, Binladen J, Miller W, Wiuf C, Willerslev E, Poinar H, Carlson JE, Leebens-Mack JH, Schuster SC (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35:1-10.
 37. Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A (2007) Novel high-resolution characterization of ancient DNA reveals C>U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35:5717-5728.
 38. Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635-651.
 39. 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: 671-676.
 40. Xue F, Wang Y, Xu SH, Zhang F, Wen B, Wu XS, Lu M, Deka R, Qian J, Jin L (2008) A spatial analysis of genetic structure of human populations in China reveals distinct difference between maternal and paternal lineages. *Eur J Hum Genet* 16:705-717.
 41. 陕西省考古研究所临潼县文物工作队(1992) 秦陵徭役刑徒墓. 西安:陕西旅游出版社.
 42. Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo LJ, Hirose R, Fujita Y, Kurata M, Shinoda K, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res* 14:1832-1850.

译后记: 古 DNA 的研究中, 由于 DNA 严重降解并损坏, 可用于分型的样本含量极低。因此, 古 DNA 研究有成功率低、易出错、易污染的特点。为了避免出错和污染, 需要对同一样本在多个具有特殊洁净要求的实验室进行多次独立分析印证。在本研究中, 我们共对 52 个样本进行研究, 成功提取出 DNA 的有 28 个个体, 测序成功的有 19 个个体, 仅对这 19 个样本作进一步的分析。我们对所获结果进行了多次验证, 并由吉林大学的古 DNA 实验室印证后才正式发表。我们曾在内部交流时提及早期结果中有一样本疑似欧洲人种, 但在重复实验中无法验证而定为该样本分型失败。由于我们和合作者就发表结果的严格要求沟通不足, 导致了误解, 我们特此公开致歉。