

DOI: 10.16359/j.cnki.cn11-1963/q.2017.0067

应用 Y 染色体 SNP 对新疆三个隔离人群 遗传多样性的研究

刘树虎, 伊力哈木·乃扎木, 热比亚木·巴克,
布帕提玛穆·阿布都克热穆, 多力坤·买买提玉素甫

新疆大学生命科学与技术学院, 乌鲁木齐, 830046

摘要: 克里雅人、罗布人、刀郎人是生活在我国西部边疆沙漠腹地、人口稀少的隔离人群。基于对这三个隔离人群 179 人 Y 染色体全序列的测序和分型, 得到每个个体 Y 染色体所有突变的 SNP 位点和隶属的单倍群, 并对各单倍群类型和频率进行了分析。以探知三个隔离人群的 Y 染色体遗传结构和遗传多样性。通过研究结果表明: 克里雅人群检出 12 个单倍群, 高频单倍群有 J2a1b1(25.64%), R1a1a1b2a(20.51%), R2a(17.95%), R1a1a1b2a2(15.38%); 罗布人群检出 16 个单倍群, 高频单倍群有 J2a1(43.75%), J2a2(14.06%), R2(9.38%), L1c(7.81%); 刀郎人群检出 40 个单倍群, 高频单倍群有 R1b1a1a1(9.21%), R1a1a1b2a1a(7.89%), R1a1a1b2a2b(6.58%), C3c1(6.58%)。三个隔离人群与维吾尔族、蒙古族、撒拉族亲缘关系较近; 在单倍群类型和频率上与维吾尔族最接近且无显著性差异 ($f=0.833$, $p=0.367$)。此外, 三个隔离人群单倍群类型和频率显示明显的亚欧混合现象, 经过长期基因融合使其具有中亚人群的典型特征, 适用于法医遗传学。

关键词: 隔离人群; Y 染色体; 单核苷酸多态性; 单倍群; 遗传多样性

中图法分类号: Q987; 文献标识码: A; 文章编号: 1000-3193(2018)01-0146-11

A study of genetic diversity of three isolated populations in Xinjiang using Y-SNP

LIU Shuhu, NIZAM Yilihamu, RABIYAMU Bake, ABDUKERAM Bupatima, DOLKUN Matyusup

College of the Life Sciences and Technology, Xinjiang University, Urumqi 830046

Abstract: The Keriyan, Lopnur and Dolan peoples are isolated populations with sparse numbers living in the western border desert of our country. By sequencing and typing the complete

收稿日期: 2017-05-02; 定稿日期: 2017-09-29

基金项目: 国家自然科学基金项目资助(31460285); 中科院上海生命科学研究院计算生物学研究所重点实验室开放课题(2014KLCB02)

作者简介: 刘树虎, 男, 山东莒南人, 硕士研究生, 主要从事人类分子遗传学研究。E-mail: 763656707@qq.com

通讯作者: 多力坤·买买提玉素甫, 男, 维吾尔族, 教授, 硕士生导师, 主要从事人类与医学分子遗传学研究。

Citation: Liu SH, N, Yilihamu, R Bake, et al. A study of genetic diversity of three isolated populations in Xinjiang using Y-SNP[J].

Acta Anthropologica Sinica, 2018, 37(1): 146-156

Y-chromosome of 179 individuals in these three isolated populations, all mutations and SNPs in the Y-chromosome and their corresponding haplotypes were obtained. Types and frequencies of each haplotype were analyzed to investigate genetic diversity and genetic structure in the three isolated populations. The results showed that 12 haplogroups were detected in the Keriyan with high frequencies of the J2a1b1 (25.64%), R1a1a1b2a (20.51%), R2a (17.95%) and R1a1a1b2a2 (15.38%) groups. Sixteen haplogroups were noted in the Lopnur with the following frequencies: J2a1 (43.75%), J2a2 (14.06%), R2 (9.38%) and L1c (7.81%). Forty haplogroups were found in the Dolan, noting the following frequencies: R1b1a1a1 (9.21%), R1a1a1b2a1a (7.89%), R1a1a1b2a2b (6.58%) and C3c1 (6.58%). These data show that these three isolated populations have a closer genetic relationship with the Uygur, Mongolian and Sala peoples. In particular, there are no significant differences in haplotype and frequency between the three isolated populations and Uygur ($F=0.833$, $p=0.367$). In addition, the genetic haplotypes and frequencies in the three isolated populations showed marked Eurasian mixing illustrating typical characteristics of Central Asian populations.

Key words: Isolated population; Y-chromosome; SNP; Haplogroup; Genetic diversity

1 前言

Y 染色体只能由父亲传给儿子, 而 Y 染色体非重组区域 (Y-NRY) 在减数分裂过程中不与 X 染色体配对, 这使其只通过碱基突变来积累变异信息^[1]。因此通过测定 Y-NRY 的突变标记, 可追踪父系遗传的进程。加上 Y 染色体单核苷酸多态性 (Y-SNP) 具有很强的地域特异性, 测定不同地域人群的 Y 染色体单倍群类型, 可推断人群进化、迁徙及相关历史活动, 能有效地评估群体的遗传结构^[2,3]。SNP 即单核苷酸多态性, 指在基因组水平上由单个核苷酸的变异所引起的 DNA 序列多态性, 其多态性丰富。2002 年 YCC^[4] (国际 Y 染色体协会) 对世界不同地域人群 74 名男性个体的 245 个 Y-SNP 位点检测分析构建了包含 153 个单倍群的进化树。2008 年 Karafet 等^[5] 用 599 个 Y-SNP 位点构建了含 311 个单倍群进一步拓扑了 Y 染色体进化树。Su 和 Jin 等^[6,7] 运用 19 个位点构建了 17 种单倍型群, 7 种单倍型只存在东亚人群中, 通过研究发现南方群体的单倍群多样性比北方群体高, 认定东亚人群起源于南方, 并构建了父系迁徙线路。因 Y-SNP 具有父系遗传、地域特异性强、多态性丰富等特点广泛应用到人类的起源、演化、法医遗传学、疾病等研究中^[8-11]。

克里雅人处于塔克拉玛干沙漠腹地, 高温干旱、沙尘暴频繁^[12-14], 生活环境非常恶劣。由于地理位置偏远、交通不便, 只与邻近地区的来往, 克里雅人至今还过着半定居、半游牧、与世隔离的生活。关于其起源如今还不清晰: 有人支持是西藏藏族的后裔; 有人认为是当地沙漠土著居民; 还有人猜测克里雅人是 2000 年神秘消失的古楼兰人的一支。

罗布人地处塔克拉玛干沙漠北延的罗布泊地区, 近年由于环境的恶化、罗布泊干涸导致动植物的消亡, 罗布泊地区变为“死亡之地”, 生活条件更加恶劣^[15, 16]。罗布人起源有不同的说法: 普尔热瓦尔斯基根据罗布人的外貌特征判断其是蒙古人和雅利安人的混

血；以王野苹为代表的一些学者认为是塔里木盆地古老的土著人；有些学者还认为罗布人是在公元前后出现于今天塔里木盆地中南部的楼兰国的居民之一。

刀郎人地处塔克拉玛干沙漠西北边缘交通状况不佳，与外界交流很少，生活环境恶劣^[17, 18]。其起源主要依据历史记载、文化生活习俗、村镇废墟及古墓研究，有 4 种说法：“土著说”、“回纥说”、“蒙古说”、“混成说”，持“混成说”的人最多。

本研究目的为：收集、保存特殊人群的基因资源；通过 Y 染色体全序列测序、分型全面的探究三个隔离人群 Y 染色体遗传结构和遗传多样性；为三个隔离人群的起源在分子遗传学角度提供凭证，进而丰富、完善中华民族遗传多样性研究。

2 材料与方法

2.1 血液样本采集

研究所用的 179 份血液样本来源如图 1。克里雅人血样采自于田县达里雅布依乡（39 例）；罗布人血样采自尉犁县喀尔曲尕乡（64 例）；刀郎人血样采自阿瓦提县乌鲁却勒镇（76 例）。取样在当地医护人员的帮助下，随机选取身体健康、个体之间无相关、三代以上均生活在本地、无异族通婚史的个体，并与受检者签署知情同意书。每人抽取外周血 5 mL，EDTA 抗凝，超低温冰箱保存。

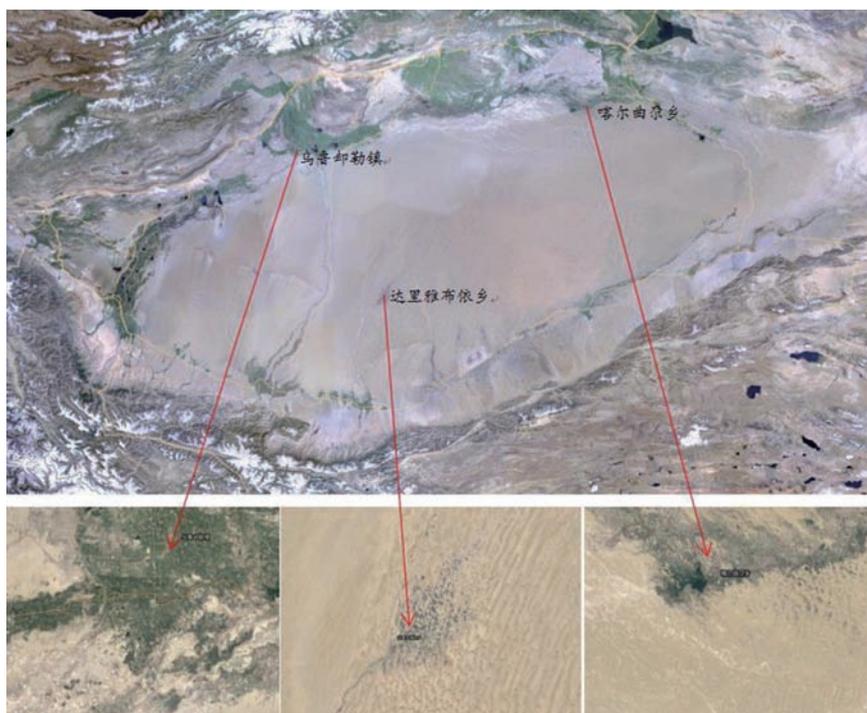


图 1 血液样本来源图

Fig.1 The populations distribution map

2.2 基因组 DNA 提取

Genview 公司基因组 DNA 抽提试剂盒（离心柱法）：严格按照试剂盒操作步骤提取基因组 DNA；主要仪器：台式高速离心机：TG24-WS, 郑州宏郎仪器设备有限公司；移液枪：Nichipet V, Nichiryo 公司；漩涡混合器：安胜 V6, 广东安胜仪器有限公司；电热三用水箱：HHW21.420ALL, 天津泰斯特仪器有限公司。

2.3 Y 染色体全序列测序及单倍群分型

测序步骤（测序公司提供）：NanoDrop 测定 gDNA 浓度 → 琼脂糖凝胶电泳检查 gDNA 的完整性 → 使用 Covaris 将 gDNA 打断 → 用 sample purification beads 进行纯化 → 修补产物末端成平末端用磁珠进行片段筛选 → 在 3' 末端加 A → 片段的两端分别与通用的接头相连接（其中一个接头带有特异性的 barcode 序列，用以区分样本的来源） → 纯化连接产物去除连接不完整的产物以及接头自连产物 → 运用与接头序列互补的通用引物对之进行 8 个循环的 PCR 扩增 → 纯化 PCR 扩增产物去除引物二聚体 → 琼脂糖凝胶电泳质检及 Qubit 浓度测定和 2100 Bioanalyzer 片段长度测定 → gDNA 文库构建完成 → 调整待测 gDNA 文库浓度至 3 nM → 在 cBot 上完成 cluster generation → 将 flowcell 转移到 HiSeq X 测序系统上，按照 Illumina 的标准流程进行测序得到所有个体的完整的碱基序列。单倍群分型：个体突变的 SNP 位点依据 YCC 网站（<https://isogg.org/tree/OLDISOGG-YDNA-SNP-Index.html>）查询比对确定个体隶属的单倍群。

2.4 数据分析

通过直接计数法统计各人群所含单倍群的频数及频率；运用 Mega7.0 软件依据突变的 SNP 碱基序列绘制人群 NJ 系统进化树；主成分分析使用 SPSS19.0 软件。我们还收集了已报道的我国不同人群的 Y 染色体单倍群频率数据进行比较研究。

3 结果

3.1 Y-SNP 单倍群的频率分布

三个隔离人群 Y 染色体单倍群类型及频率见表 1。克里雅人检测出 6 大主单倍群共 12 个单倍群类型，其中单倍群 J2a1b1 (25.64%), R1a1a1b2a (20.51%), R2a (17.95%), R1a1a1b2a2 (15.38%) 频率较高。罗布人检测出 8 大主单倍群 16 个单倍群类型，其中单倍群 J2a1 (43.75%), J2a2 (14.06%), R2 (9.38%), L1a2 (7.81%) 频率较高。刀郎人检测出 10 大主单倍群共 40 个单倍群类型，各单倍群频率都不太高但是类型分布广泛，其中主单倍群 R, C, J, Q 频率较高，特别是 R 单倍群及下游亚单倍群共检出 14 个单倍群类型频率高达 47.39%。单倍群 O 在参考的各地汉族、回族、哈萨克族、锡伯族、藏族和俄罗斯族中有高频分布，三个人群中都有检测到但频率有一定差异（分别为 2.56%, 7.81%, 5.27%），这与此单倍群广泛高频分布在东亚汉族占 75% 以上，在亚洲西部和欧洲不存在相吻合^[23]。

3.2 主成分分析

依据表 2 各参考人群 Y 染色体单倍群数据运用 SPSS19.0 作的主成分分析如图 2 和图 3, 其中 3 个主成分合理解释了 90.26% 的总方差。克里雅人、罗布人、刀郎人、和田维吾尔族、吐鲁番维吾尔族、塔吉克族、塔塔尔族、乌孜别克族、柯尔克孜族、撒拉族、蒙古族

表 1 Y 染色体单倍群频率分布

Tab.1 Y-chromosome haplogroup frequency distribution

单倍型	频率(%)	单倍型	频率(%)	单倍型	频率(%)	单倍型	频率(%)
C3c1	2.56	C3c1	3.13	C3b	3.95	Q1b1	2.63
G2a2b2a1a2a	2.56	H1	1.56	C3f	1.32	Q1a2	1.32
H	2.56	J2a1	43.75	C3c1	6.58	Q1a1a1	1.32
H1	2.56	J2a2	14.06	D1a	1.32	Q1a2a1c	1.32
J1a2b3	2.56	J1a2b	1.56	D3a	2.63	Q1a2a1c1	1.32
J2a1b1	25.64	L1c	7.81	G2a	2.63	R1b1	2.63
O3a2c1a	2.56	N1c1a1a	1.56	G2a1a1a	1.32	R1b1b	1.32
R1b1a1	2.56	O2a1a	1.56	G2a2b1b	1.32	R1b1a1a	1.32
R1a1a1b2a	20.51	O3a2c	4.69	G2a2b2a	1.32	R1b1a1a1	9.21
R1a1a1b2a2	15.38	O3a2c1a	1.56	H2	1.32	R1a1a1b2	1.32
R1a1a1b2a2a	2.56	Q1a1a1	1.56	J2a1	3.95	R1a1a1b2a	1.32
R2a	17.95	R1b1a1	1.56	J2a2	3.95	R1a1a1b2a1	1.32
		R1a1a1b2	1.56	J1a2b2	1.32	R1a1a1b2a2	5.26
		R1a1a1b2a2	3.13	N1b	1.32	R1b1a1a2a2	1.32
		R1a1a1b2a1a	1.56	N1c2	1.32	R1a1a1b2a1a	7.89
		R2	9.38	N1c2b	1.32	R1a1a1b2a2a	5.26
				N1c1a1	1.32	R1a1a1b2a2b	6.58
				O3a2c	2.63	R1a1a1b1a2b3	1.32
				O3a2c2	1.32	R2	1.32
				O3a2c1a	1.32	T1a1a	1.32

注：一二列为克里雅人单倍型和频率；三四列为罗布人单倍型和频率；后四列为刀郎人单倍型和频率。

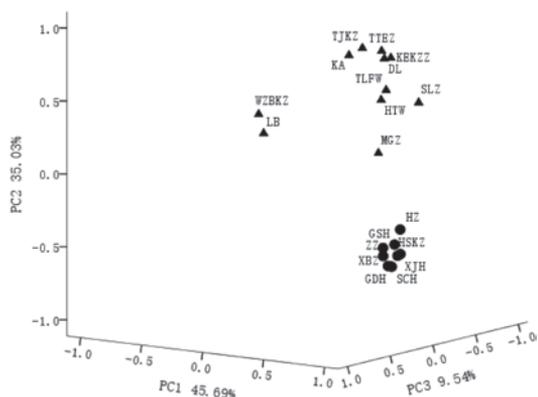


图 2 Y 染色体主成分分析三维图

Fig.2 The principal component analysis of Y chromosome

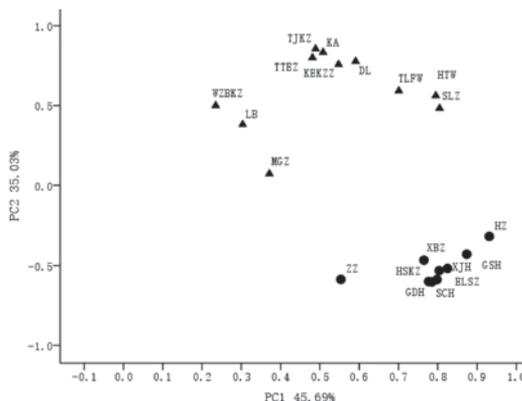


图 3 Y 染色体主成分分析二维图

Fig.3 The principal component analysis of Y chromosome

表 2 20 个人群 Y 染色体单倍群频率分布
 Tab.2 Y-chromosome haplogroup frequency distribution of 20 populations

主单倍群→ 人群(数量)↓	C	D	F	G	H	J	K	L	N	O	P	Q	R	T
KA(39)	2.56			2.56	5.12	28.20				2.56			58.96	
LB(64)	1.56					59.37		7.81	1.56	7.81		1.56	17.19	
DL(76)	11.85	3.95		6.59	1.32	9.22			5.28	5.27		7.91	47.39	1.32
HTW(478) ^[20]	8.37	2.72	0.84	5.65	3.56	15.69		3.56	5.02	14.23		7.74	28.24	
TLFW(143) ^[20]	16.78	2.80		0.70		12.59		0.70	4.90	10.49		15.38	32.87	
WZBKZ(23) ^[21]	4.35			8.69		34.78			8.69				13.04	
TJKZ(31) ^[21]	3.23			6.45		16.13					3.23		61.29	
KEKZZ(45) ^[21]	8.89								4.44	4.44	2.22		71.11	
SLZ(135) ^[20]	5.93	4.44		2.22	5.19	2.96			1.48	22.22	5.19	2.96	44.44	
HZ(175) ^[20]	9.71	4.00		1.71	1.71	11.43			4.57	46.29		4.57	13.71	
HSKZ(87) ^[20]	32.18					1.15			2.30	59.77			3.45	
ZZ(289) ^[22]	6.90	25.80	6.60				11.40			33.50				
XJH(32) ^[19]	6.25	3.13								56.24	3.13		6.25	
GSH(30) ^[19]	20.00					10.00			6.67	40.00			6.67	
GDH(35) ^[19]	8.57								2.86	65.71				
SCH(34) ^[19]	11.76								2.94	55.88				
TTEZ(33) ^[21]													60.60	
ELSZ(19) ^[21]	10.53	5.26				5.26				47.37				
XBZ(32) ^[21]	12.50					3.13			15.63	28.12			3.13	
MGZ(129) ^[20]	27.13	34.11		0.78	4.65	1.55			2.33	7.75	0.78	5.43	15.50	

注: ① KA: 克里雅人, LB: 罗布人, DL: 刀郎人, HTW: 和田维吾尔族, HTWZ: 和田维吾尔族, TLFW: 吐鲁番维吾尔族, HZ: 回族, HSKZ: 哈萨克族, WZBKZ: 乌孜别克族, TJKZ: 塔吉克族, KEKZZ: 柯尔克孜族, TTEZ: 塔塔尔族, ELSZ: 俄罗斯族, XBZ: 锡伯族, MGZ: 蒙古族, SLZ: 撒拉族, XJH: 新疆汉, GSH: 甘肃汉, GDH: 广东汉, SCH: 四川汉。② 参考人群数据来源文献 19-22, 具体已标注表内人群名后, 文中所有人群简写都参见此表。③ 因各参考人群单倍群分类到各个亚群分支的程度不同, 所以在汇总人群单倍群频率数据时将同一单倍群大类下的亚群分支进行合并, 例如: 对于单倍群 G, 一些人群分到 G1a, G2a 这个层次, 另一些人群分到 G1, G2, G3, 而有的人群只能确定到 G 这个层次, 故而将各个亚群分支合并为一个单倍群 G 大类。

聚的很近划为 A 聚类 (▲表示); 各地汉族、哈萨克族、俄罗斯族、回族、锡伯族、藏族聚的很近划为 B 聚类 (●表示)。A 聚类人群的 R 单倍群及其下游亚单倍群频率很高, 但是 O 单倍群及其下游亚单倍群频率很低; B 聚类人群的这两大主单倍群及其下游亚单倍群频率正好与 A 聚类人群相反。两个聚类中 O 和 R 两大单倍群频率差异极显著 ($t=8.42$, $p<0.01$; $t=6.41$, $p<0.01$), 提示这两个单倍群频率的差异是将两个聚类分开的主要因素。罗布人稍微偏离 A 聚类是因为含有极高频率的 J(59.37%) 特别是 J2a1(43.75%), 远高于 A 聚类中其他人; 蒙古族在 A 和 B 聚类之间, 一是因为含有频率较高的 C (27.13%) 和 D (34.11%), 二是 O, R 两大单倍群及其下游亚单倍群频率处于两个聚类中间水平。

3.3 系统进化分析

依据测序 SNP 突变位点信息利用 Mega7.0 软件构建人群的邻接 (NJ) 系统进化树图 4。结果显示, 所有人群可分为三大支。LB, DL, KA 最先聚在一起, 后与 HTW, TLFW 聚在一起, 再与 MGZ, SLZ, HZ 共同组成第一大支, 说明他们的亲缘关系近, 尤其是三个人群与维吾尔族的亲缘关系最近; 各地汉族、ZZ, ELSZ, HSKZ, XBZ 处在一起组成第二大支, 他们的亲缘关系近; TJKZ, TTEZ, KEKZZ, WZBKZ 处在一起聚为第三大支, 他们有着较近的亲缘关系。

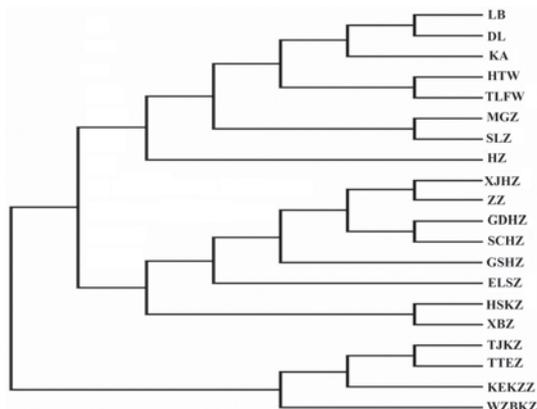


图 4 用邻接方法 (NJ) 构建的人群系统进化树
Fig.4 Using neighbor joining method (NJ) constructed the population system phylogenetic tree

综上所述, 我们运用了多种遗传统计分析方法在不同的侧面对三个隔离人群的单倍群类型和频率进行了分析, 揭示三个人群的 Y 染色体遗传多样性和遗传结构与维吾尔族差异性最小, 亲缘关系最近。

4 讨论

4.1 三个隔离人群 Y 染色体的遗传结构

通过 Y 染色体全序列测序, 克里雅人总共发现了 27 个 SNP 突变位点, 共划分出 12 个单倍群类型。J2a1b1 (25.64%), R1a1a1b2a (20.51%), R2a (17.95%), R1a1a1b2a2 (15.38%) 频率相对较高, 对应的 SNP 高频突变位点分别为 (M260, Page14, M92), (L342.2, S278.2, Z94), (P249, L266, PF6108, P267, PF6109, M124), (Z2124)。R 和 J 两个主单倍群及其下游亚单倍群频率高达 87.16%, R 群支占 58.96%, J 群支占 28.20%, 为克里雅人群特征单倍群。罗布人总共发现了 41 个 SNP 突变位点, 共划分出 16 个单倍群类型。J2a1 (43.75%), J2a2 (14.06%), R2 (9.38%), L1c (7.81%), O3a2c (4.69%) 频率相对较高, 对应的 SNP 高频突变位点分别为 (L26, PF5110, Page55, S57, L27, PF5111, S396), (L581, S398), (M479, PF6107), (M357, L1307), (F130, F131, P164, CTS4723, F422, F299, F422, F427, CTS11109, CTS12099)。R 和 J 两个主单倍群及其下游亚单倍群频率就高达 76.56%, J 群支占 59.37%, R 群支占 17.19%, 为罗布人群特征单倍群。其次 O 单倍群及其下游单倍群频率为 7.81%。刀郎人总共发现了 145 个 SNP 突变位点, 共划分出 40 个单倍群类型。没有特别高频的单倍群, R1b1a1a1 (9.21%), R1a1a1b2a1a (7.89%), R1a1a1b2a2b (6.58%), R1a1a1b2a2 (5.26%), C3c1 (6.58%) 频率相对较高, 对应的 SNP 高频突变位点分别为 (M478), (Y7), (Z2122), (Z2124), (M77, M86)。R, C, J 主单倍群及其下游亚单倍群频率高达 68.46%, R 群支占 47.39%, C 群支占 11.85%, J 群支占 9.22%, 为刀郎人群特征单倍群。C, J, O, R 单倍群及其下游亚单倍群在

表 3 人群主要单倍群比较
 Tab.3 The main haplogroups of population comparison

人群	主单倍群种类-频率										
	C	D	E	G	H	J	L	N	O	Q	R
KA	2.56%			2.56%	5.12%	28.20%			2.56%		58.96%
LB	1.56%					59.37%	7.81%	1.56%	7.81%	1.56%	17.19%
DL	11.85%	3.95%		6.59%	1.32%	9.22%		5.28%	5.27%	7.91%	47.39%
HTW ^[21]	8.37%	2.72%	2.51%	5.65%	3.56%	15.69%	3.56%	5.02%	14.23%	7.74%	28.24%
TLFW ^[21]	16.78%	2.80%	2.80%	0.70%		12.59%	0.70%	4.90%	10.49%	15.38%	32.87%
MGZ ^[21]	27.13%	34.11%		0.78%	4.65%	1.55%		2.33%	7.75%	5.43%	15.50%

注: 三个隔离人群、维吾尔族、蒙古族遗传关系最接近, 其主单倍群能更加直观的比较各人群的遗传结构的来源组成。单倍群 C, D, O 主要分布在亚洲为东亚特征单倍群; 单倍群 G, J, R 主要分布在欧洲大陆, 高频分布与欧洲、中亚。

三个隔离人群中都含有, D, G, H, L, N, Q 单倍群及其下游亚单倍群在个别人群中不存在或者由于样本限制而未检出(表 1)。三个隔离人群单倍群类型与陆艳^[20]中国西部人群遗传混合中维吾尔族的遗传结构相符, 也与热比亚木·巴克^[24]的维吾尔族 Y 染色体遗传多样性发现的两大谱系单倍群相吻合, 说明三个隔离人群与维吾尔族的遗传结构类似。已有报道单倍群 J 主要分布在欧洲大陆(阿拉伯、犹太人等高频), 在中亚人群(维吾尔族 14.43%, 回族 11.43%)也有广泛的分布^[25]。J2 现在分布在北非、中东、欧洲南部, 意大利频率为 20%, 西班牙南部频率为 10%。单倍群 R 属印欧语系的主要群体, 高频分布于欧洲至中亚、南亚, 汉族 2% 左右, 中国部分少数民族较高(维吾尔族 32.87%, 撒拉族 44.44%, 塔吉克族 64.52%)^[24]。R2 现存于印度北部、巴基斯坦、亚洲中南部地区, 频率在 5%-10%。O 单倍群较晚期到达东亚的人群, 广泛高频分布在东亚, 占汉族 75% 以上^[6]; O2 是中国最常见的单倍群, 遍及整个东亚和东南亚, 占汉族 50%-60% 左右^[7]。由此看出三个隔离人群单倍群类型具有典型的亚欧混合现象。Zhong 等^[26]研究得出东亚北方人群 Q, R 单倍群约 1.8 万年前末次盛冰后期由中东进入中亚和东亚北部的迁徙, 徐智对新疆地区古 DNA 样本的研究认为, 亚欧大陆东、西部人群在中国新疆地区开始混合的时间大概为 2500 年前^[27]。

4.2 亚欧谱系单倍群在三个隔离人群中的遗传混合

单倍群 C 在东亚、东南亚、大洋洲有着广泛的分布, 被认为是南亚起源, 沿着非洲海岸线前进, 一部分人渡海抵达澳大利亚, 一部分人留在了东南亚沿海逐步向内陆迁移进入亚洲东部、蒙古、西伯利亚^[6, 28-30]。高频分布于阿尔泰语系的蒙古族、满族、哈萨克族及澳大利亚土著等, 汉族约为 5%-10%。单倍群 G 源于中东东部地区, 巴基斯坦、印度的喜马拉雅山脚下, 在新石器时代随着农业的发展而扩散^[25]; G2 出现在中东受冰封期的影响人口发展缓慢, 到冰封期结束 G2 各个血统从中东向欧洲的北部和东部扩散, 有些进入亚欧大陆的西部^[25]。现在高频分布于土耳其、高加索、哈萨克斯坦西部, 中国西部一些少数民族有分布(维吾尔族 9.52%)^[24]。单倍群 J 约 1.5 万年前诞生于新月沃土, 现在频率最高的地区包括中东—北非—埃塞俄比亚; J 在欧洲开始仅仅分布在地中海附近, 新石器革命之后, 随着农业的扩张, J 开始扩散, 一部分留在中东一部分北上进入欧洲西部^[31]。

有研究表明, J2 起源于近东, 随着近东农业文明的发展, 早期从事农业的生产者由黎凡特走廊将这一谱系带进了欧洲, 西亚、中亚地区有一定频率分布^[32, 33]。单倍群 O 及下游单倍群 O1a, O1b1a1a, O2 是东亚现今人群中分布最广泛的谱系单倍群, 频率约占 50% 以上, 其中 O2 的频率在 40% 以上^[34]。思远等^[35]证明单倍群 O 为东亚南方起源, 并从东亚由南向北扩散。单倍群 R 由两个下游单倍群 R1 和 R2 组成, 广泛分布在亚欧大陆, 遍布欧洲、中亚、西亚等地区^[36-38]。随着人类文明和人群的扩张, 亚欧人群最原始的 R1a1a 单倍群随之扩张, 广泛高频分布在欧洲、中亚、南亚、东欧、印度西北、巴基斯坦^[23]。有研究报道^[37], 现代欧洲人单倍群 R1b1a1a2 最常见, 特别在西欧, 频率约为 40%-80%。

农业的扩展、经济贸易的往来大都伴随着人类的交流与迁徙, 南方起源的 C, D, O 单倍群是东亚特征单倍群^[34, 39, 40]; G, J, R 单倍群广泛分布在欧亚大陆, 高频分布与欧洲、中亚、南亚、西亚^[31, 36-38, 41]。“古丝绸之路”作为欧亚大陆的交通枢纽, 加上新疆又处于丝绸之路的核心位置经济贸易、人类交流频繁, 在这条路上东亚人群自东向西迁徙, 欧洲人群自西向东迁徙, 经过多次文化更替、民族聚散融合和基因交流, 随着时间的积淀而形成了现今的人群遗传结构。

4.3 三个隔离人群的父亲起源推断

克里雅人的单倍群构成及频率与两地的维吾尔人的最接近, 且通过主成分分析和系统进化树分析都显示他们的亲缘关系最近, 我们初步推断克里雅人是当地沙漠土著居民, 其与维吾尔人族源有渊源。Chen^[42]等研究的克里雅人的父系及母系遗传分析也证明其不是西藏藏族族群的后裔, 克里雅人是东西欧混合的群体, 可能起源于当地的维吾尔群体。段然慧^[43]等研究说明克里雅人群的核苷酸变异度和平均核苷酸差异都介于所报道的东方人群和西方人群之间。系统发育树也说明中亚的各人群均处于东方人群的亚洲谱系和西方人群的欧洲谱系之间, 并且克里雅人与维吾尔人遗传距离最近, 表明他们之间有很密切的亲缘关系。

关于罗布人的起源, 普尔热瓦夫斯基根据罗布人的体质特征判断其是蒙古人和雅利安人的混血; 1934 年, 中瑞考察团在孔雀河故道发现了著名的古墓沟舟形墓葬群, 经过研究判断他们是楼兰时期之前的土著人; 中国社会科学院文物研究所杨镰研究员经过实地考察认为罗布人是古“楼兰遗民”的后裔, 然而, 楼兰人操印欧语系, 而罗布人使用夹杂方言的维吾尔族语言与其相矛盾; 我国史料《西域地理图说》、《西域闻见录》等记载 17 世纪末、18 世纪初居于焉曹的维吾尔族人, 由于准噶尔贵族的强占土地与压迫, 他们四散逃亡, 其中的一部分人被迫迁往罗布泊地区。类似的考古发现、史料记载还有不少, 我们没有办法确定它们的准确性, 但是他们处在不同时期, 在同一地域的所见所闻有很大差别, 这种可能性是不容置疑的。再加上罗布地区地处东西方的交通枢纽, 为古丝绸之路核心位置, 各地区人群在此交流密切。结合本研究罗布人 Y 染色体遗传结构特点, 我们认为: 楼兰城郊墓葬中发现有欧洲人种和蒙古人种; 普氏看到部分罗布人具有蒙古人种的特征等等。这些都会在人群长期不断融合中出现。现代罗布人的遗传结构特点就是维吾尔人、蒙古人、欧洲人等长期民族融合的结果, 这也与本研究中罗布人的 Y 染色体遗传结构与维吾尔族和蒙古族的相近相符。如: J 单倍群频率高达 59.37%, 而 J 及其下游亚单倍群主要分布在欧洲西部、西亚、中亚地区; O, R 单倍群的频率与蒙古族的接近。

《明史·西域传》记载, 蒙古人起源于漠北高原, 后统治亚洲、东欧, 建立了蒙古帝国, 南疆及中亚大片地区就包含在内, 后因蒙古宗王争权夺利, 陷入了长期的武力争斗。当地的人们为躲避战乱纷纷外逃, 荒无人烟的叶尔羌河下游平原自然成了栖身的好地方。来自各地的人聚集在一起自称“杜拉”后改为“刀郎”。进入南疆的蒙兀儿人、“杜拉”长期与当地的维吾尔族交流逐渐被同化^[44]。据记述, 当地人穿着蒙古式衣服, 特别是妇女依然保持着蒙古族的面貌^[45]。1976 年阿瓦提县刀郎人古时驻地发现的长袍、腰巾与蒙古人的很相像。刀郎木卡姆为刀郎文化的重要组成部分, 更是维吾尔族十二木卡姆不可分割的一部分, 且较多地保留了古代西域文化, 但也蕴含了较大成分的蒙古文化和萨满教、祆教遗迹。以上历史记载说明刀郎人应出现在察合台汗国之后, 由蒙古、维吾尔等民族融合而成。通过我们的研究同样发现刀郎人的单倍群类型在维吾尔族和蒙古族中都含有, 且主要单倍群都相同, 所含频率有一定的差异(如表 3), 主成分分析和 NJ 树分析同样得到刀郎人与维吾尔族和蒙古族亲缘关系很近, 这就在分子遗传学上为“混成说”确立了新的证据。

综上所述, 我们初步得出以下结论: 三个隔离人群 Y-SNP 位点多态性极其丰富, 具有典型的中亚人群中维吾尔族的特征, 适用于法医遗传学; 三个隔离人群的单倍群类型和频率显示出明显的亚欧混合现象。

总之, 民族的起源本来就是复杂漫长的过程, 加上人群的迁徙、融合、自然灾害、民族婚配等都对人群的遗传结构产生影响。本研究中三个人群的 Y-SNP 多态性极其丰富能为法医遗传学提供依据, 但人群的遗传结构有待进一步深入研究。人群族源问题不能只从历史记载、语言学、考古学、体质人类学、分子遗传学单个方面进行研究, 需要综合所有学科精准地作出结论。

参考文献

- [1] 韩建利. 云南白族男性人群 Y-SNP 遗传多态性研究 [D]. 昆明: 昆明医科大学, 2014:6-14
- [2] 高静. 云南汉族人群 Y 染色体 SNP 位点的遗传多态性研究及法医学意义 [D]. 昆明: 昆明医科大学, 2013:5-11
- [3] 陈雪云. 云南景颇族人群 Y 染色体遗传多态性研究 [D]. 昆明: 昆明医科大学, 2015: 6-16
- [4] YCC. A nomenclature system for the tree of human Y-chromosomal haplogroups[J]. *Genome Res*, 2002. 12(2): 339-348
- [5] Karafet TM, Mendez FL, Meilerman MB, et al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree[J]. *Genome Res*, 2008, 18:830-838
- [6] Su B, Xiao JH, Jin L. Y chromosome evidence for a northward migration of modern humans in East Asia during the last Ice Age[J]. *Am J Hum Genet*, 1999, 65: 1718-1724
- [7] Jin L, Su B. Natives or Immigrants: modern human origin in East Asia[J]. *Nature Genetics Reviews*, 2000, 11(1):126-132
- [8] 邓琼英, 王晓庆, 王传超, 等. 茶洞话群体的 Y 染色体遗传结构及其父系起源研究 [J]. *人类学学报*, 2014, 23(1): 118-124
- [9] 姚伊人, 赵兴春, 白雪, 等. 新疆和田地区维吾尔族人群 18 个 STR 基因座遗传多态性 [J]. *中国法医学杂志*, 2015, 30(3): 295-296
- [10] 热比亚木·巴克, 阿瓦古丽·阿布都克热木, 伊力哈木·乃扎木, 等. 维吾尔族人群 LPL 基因 Pvu II 多态性与身体测量和血生化指标的相关性研究 [J]. *环境卫生学杂志*, 2015, 5(6): 498-503
- [11] 孟已琦, 武培松, 李晋新, 等. 新疆维、汉两民族 GNAS1 基因 T393C 单核苷酸多态性与非瓣膜性心房颤动相关性的研究 [J]. *中国心脏起搏与心电生理杂志*, 2015, 29(1): 38-41
- [12] 彭向东. 秘境中的克里雅人 [M]. 北京: 中国民族摄影艺术出版社, 2014: 06
- [13] 王铁男, 花雕. 发现达里雅布依: 对一个世界秘境的百年追踪 [J]. *环球人文地理*, 2011(2): 84-95
- [14] 明西, 蒋君, 陈亚强. 克里雅人死亡之海中的孤独传奇 [J]. *中国科学探险*, 2012(12): 104-129
- [15] 艾比不拉·卡地尔. 罗布人: 绿洲文化变迁的人类学研究 [M]. 北京: 社科文献出版社, 2014: 05

- [16] 艾比不拉·卡地尔. 罗布人: 绿洲文化变迁的人类学研究——以喀尔曲尕乡个案为例 [D]. 厦门: 厦门大学, 2012: 12
- [17] 艾比布拉·阿布都沙拉木. 人文视野中的刀郎文化 [M]. 乌鲁木齐: 新疆人民出版社, 2007: 04
- [18] 阿迪力·穆罕默德. 麦盖提历史文化 [M]. 乌鲁木齐: 新疆大学出版社, 2008: 09
- [19] 蔡晓云. Y 染色体揭示的早期人类进入东亚和东亚人群特征形成过程 [D]. 上海: 复旦大学, 2009: 74-99
- [20] 陆艳. 中国西部人群的遗传混合 [D]. 上海: 复旦大学, 2011: 1-84
- [21] 寿维华. 中国西北丝绸之路地区民族群体的 Y 染色体多态性研究 [D]. 云南: 云南大学, 2008: 28-45
- [22] 温有锋. 西藏藏族起源初探 [D]. 沈阳: 中国医科大学, 2007: 9-54
- [23] 乔恩发. 丝绸之路中国区域少数民族 Y 染色体多态性的研究 [D]. 云南: 云南大学, 2006: 20-33
- [24] 热比亚木·巴克. 新疆维吾尔族人群 Y 染色体非重组区遗传多样性分析 [D]. 乌鲁木齐: 新疆大学, 2016: 11-83
- [25] 张振. 人类六万年 [M]. 合肥: 安徽人民出版社, 2013: 01
- [26] Zhong H, Shi H, Qi XB, et al. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route[J]. *Mol Biol Evol*, 2011, 28(1): 717-727
- [27] 徐智. 中国西北地区古代人群的 DNA 研究 [D]. 上海: 复旦大学, 2008: 95-116
- [28] Xue Y, Zerjal T, Bao W, et al. Male demography in East Asia: A north-south contrast in human population expansion times[J]. *Genetics*, 2006, 172(4): 2431-2439
- [29] Hudjashov G, Kivisild T, Underhill PA, et al. Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis[J]. *Proc Natl Acad Sci USA*, 2007, 104 (21): 8726-8730
- [30] Kayser M, Brauer S, Cordaux R, et al. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific[J]. *Mol Biol Evol*, 2006, 23(11): 2234-2244
- [31] Semino O, Magri C, Benuzzi G, et al. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area[J]. *Am J Hum Genet*, 2004, 74(5): 1023-1034
- [32] Zerjal T, Wells RS, Yuldasheva N. A genetic landscape reshaped by recent events Y-chromosomal insights into central Asia[J]. *Am J Hum Genet*, 2002, 71(3): 466-482
- [33] Shou WH, Qiao EF, Wei CY, et al. Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians[J]. *J Hum Genet*, 2010, 55(5): 314-322
- [34] Shi H, Dong YL, Wellet B, et al. Y-chromosome evidence of southern origin of the East Asian. specific haplogroup O3-M122[J]. *Am J Hum Genet*, 2005, 77(3): 408-419
- [35] 思远, 王升启. 单核苷酸多态性基因分型技术原理与进展 [J]. *生物技术通讯*, 2003, 14(3): 218-231
- [36] Semino O, Passarino G, Oefner PJ, et al. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y-chromosome perspective[J]. *Science*, 2000, 290(5494): 1155-1159
- [37] Sengupta S, Zhivotovsky LA, King R, et al. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists[J]. *Am J Hum Genet*, 2006, 78(2): 202-221
- [38] Wells RS, Yuldasheva N, Ruzibakiev R, et al. The Eurasian heartland: a continental perspective on Y-chromosome diversity[J]. *Proc Natl Acad Sci USA*, 2001, 98(18): 10244-10249
- [39] Zhong H, Shi H, Qi XB, et al. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia[J]. *J Hum Genet*, 2010, 55(7): 428-435
- [40] Shi H, Zhong H, Peng Y. Y-chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations[J]. *BMC Biol*, 2008, 6: 45
- [41] Cinnioglu C, King R, Kivisild T, et al. Excavating Y-chromosome haplotype strata in Anatolia[J]. *Hum Genet*, 2004, 114(2): 127-148
- [42] Chen KX, Ablimit A, Ling F, et al. Paternal and maternal genetic analysis of a desert keriyan population: keriyan are not the descendants of guge Tibetans[J]. *Plos ONE*, 2014, 9(6): e100479
- [43] 段然慧, 崔银秋, 周慧, 等. 塔克拉玛干沙漠腹地隔离人群线粒体 DNA 序列多态性分析 [J]. *遗传学报*, 2003, 30(5): 437-442
- [44] 王治来. 新疆简史 [M]. 乌鲁木齐: 新疆人民出版社, 1980: 8
- [45] 余太山. 新疆各族历史文化词典 [M]. 上海: 中华书局出版, 1996: 9