



突厥和乌拉尔族群东亚起源的 Y 染色体线索

兰海

中国语言论坛, 苏州 215000

评论文献: Rootsi S, Zhivotovsky LA, Baldovič M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumäe AM, Khusnutdinova EK, Voevoda MI, Osipova LP, Stoneking M, Lin AA, Ferak V, Parik J, Kivisild T, Underhill PA, Villems R (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet* 15:204-211.

摘要: Rootsi 等于 2007 年初发表了一篇研究 Y 染色体单倍群 N 的论文。这是至今关于单倍群 N 的最全面、最详细的研究。论文中尚未完美地解释的东欧的 N3 多样性比西伯利亚高的结果。本文应用 Derenko 等新发表的数据, 更好地解释了这个结果。同样, 应用 Derenko 等的 Y 单倍型分类, 对之前文献报道的土耳其的 N* 个体进行深入分型, 发现他们大多是 N2-A, 由此得到了突厥东亚起源的强有力证据: 东亚起源单倍群 N2-A 是突厥的特征单倍型, 而 N2-A2 几乎只存在于突厥的后裔中。经分析土耳其人中的 R1b1b-M73, 认为它的分布与中国古代部落月支的迁徙有关。同时探讨了 R1a1、Q 在突厥形成过程中的演变。并推断突厥祖先丁零—铁勒诸部的主要 Y-SNP 类型为 N3a1、N3a2 和 N2-A。乌拉尔族群的起源也与突厥群体有很大关系。

关键词: 单倍群 N; 迁徙; 突厥; 铁勒; N2-A; R1b1b-M73

Y chromosome Clues to the East Asian Origins of Turki and Uralic populations

LAN Hai

Linguistic Forum of China, Suzhou 215000 China

ABSTRACT: The paper on the evolution of Y chromosome haplogroup N published by Rootsi et al. in 2007 may be the most comprehensive and detailed review of this haplogroup. However, the higher diversity of sub-haplogroup N3 in East Europe than that in Siberia was not well explained. In this paper, we explained this phenomenon by analyzing the new data published by Derenko et al. Moreover, using the sub-haplogroup classification by Derenko et al, the previously reported Turkey samples of N* haplogroup were reclassified into N2-A mostly, and then the East Asian origin of Turki populations was well proved. The East Asian originated haplogroup N2-A is the characteristic haplogroup of Turki populations, and the subgroup N2-A2 is almost restricted in the descendants of ancient Turki. Another specific haplogroup R1b1b-M73 in modern Turks was traced back to an ancient Chinese tribe Yüeh-chin / Scythian. We also discussed the evolution of the haplogroups R1a1 and Q in the formation of Turki, and deduced that the haplogroups N3a1, N3a2 and N2-A might be the major Y haplogroups of Dingling-Tiele, the ancestor of Turki. The origin of Uralic populations was largely related to the Turki populations.

Key words: Haplogroup N; Migration; Turki; Tiele; N2-A; R1b1b-M73

本文翻译并评论了 Rootsi 等于 2007 年初发表的一篇研究 Y 染色体单倍群 N 的论文。这是至今关于单倍群 N 的最全面的、最详细的研究。文章指出, N 在东欧的高频是源于约 1.2~1.4 万年前的来自西伯利亚南部的迁徙。频率最高的下游单倍群 N3, 可能起源于今天的中国, 然后在西伯利亚经历多次瓶颈效应最后扩散到东欧。另一个下游单倍群 N2, 形成了亚洲 (N2-A) 和欧洲 (N2-E)

两个不同的 STR 分支, N2-E 现在主要分布在芬兰—乌戈尔语人群以及相关人群。此文得到了 N 的下游支系的起源地以及分化年代, 给出了 N 在世界范围内各个人群中的分布。

一、单倍群 N 的起源与扩散

Rootsi 的论文提供了全世界范围内 N 的支系分布, 但中国境内的分布却不够清晰。

收稿日期: 2007 年 12 月 19 日 **修回日期:** 2007 年 12 月 26 日 **联系人:** 兰海 ranhaer@gmail.com

表 1 中国的各个群体中单倍群 N 的各分支的分布

Tab.1 Distributions of the haplogroup N subgroups in the Chinese populations

群体	样本量	NO*-M214	N*-LLY22g	N1-M128	N2-P43	N3-TAT	N3a1-M178
布依族 Buyi	35	2		2			
布努瑶族 Yao Bama	35	1					
鄂温克族 Ewenki	26	1					
达斡尔族 Daur	39	1					3
韩国朝鲜族 Korean Korean	25	1	1				
满族 Manchu	35		3	2	1		
锡伯族 Xibe	41		4	1			2
乌鲁木齐维族 Uygur Urumqi	31		1		2		
赫哲族 Hezhen	45		1		8		
哈尔滨汉族 Han Harbin	35		1				1
伊犁维族 Uygur Yili	39		1				2
藏族 Tibetan	35		3				
哈尼族 Hani	34		4				
兰州汉族 Han LanZhou	30		2				
成都汉族 Han ChenDu	34		1				
梅县汉族 Han MeiXian	35		1				
回族 Hui	35		1				
中国朝鲜族 Chinese Korean	25		1				
日本人 Japanese	47		2				
外蒙古 Outer Monglain	65				2	1	4
鄂伦春族 Oroqen	31				2		
内蒙古 Inner Monglian	45						6

而 Xue[1]的文献提供了目前报道的最详细的划分（表 1）。

与 Rootsi 的论文中彝族有 30.2%的 N* 一样，生活在云南省的同属于彝语支的哈尼族也有较多的 N*。更值得注意的是，N1 出现在南方少数民族布依族中，再一次提示 N、N1 的南方起源。从上表还可以看出，N 也普遍存在于通古斯语支中，N2 在赫哲族中达到 17.8%。文献 [2] 报道了西鄂温克人 (W.Ewenks) 中有 27.5% (11/40) 的 N2，文献 [3] 报道了鄂霍次克鄂温克 (Okhotsk Evenk)、叶尼塞鄂温克 (Yenisey Evenk)、沃且/那乃 (Ulchi/Nanai)、涅吉达尔 (Negidal) 分别有 37.5%、9.7%、9.4%、35.2% 的 N3。这些数据说明，N 深度分布于通古斯语支之中。

根据 Rootsi 文,结合 Derenko 发表的最新数据[4]计算的 N 各个分支的年代,从现有的各个支系的分布推断,单倍群 N 的迁徙大致是这样的:沿藏缅走廊北上一河套地区—贝加尔湖以东地区。至少在 12,000 年前的时候,已经达到贝加尔湖与大兴安岭之间。在这里, N3a1 发展壮大,迅速向东北、西北扩散,成为乌拉尔语系民族的重要单倍群。

后来约于 5000 年前 N3a1 又诞生了 N3a2, 主要存在于贝加尔湖以东地区, N3a2 也不断迁徙进入东欧。此前, 一部分 N2-A 向西迁徙到达乌拉尔地区, 途中经历了强烈的瓶颈效应, 约于 7000 年前诞生了 N2-E, 然后随着乌拉尔语而扩散到科米 (Komi)、玛里 (Mari)、维普 (Vepsa)、鞑靼 (Tatar) 等人群中。一部分 N2-A 向西南进入萨彦岭的草原地带, 约于 2000 年前, 诞生了仅仅存在于突厥后裔的图瓦人 (Tuva) 和图法拉人 (Tofalars) 中 N2-A2。

二、N3 的下游分支与多样性

N3 的 STR 多样性是之前的研究关注的重点。至今所有的研究都表明, 欧洲的 N3 的 STR 多样性要比西伯利亚的要高[5,6,7], 因此 N3 有可能起源于东欧然后向西伯利亚传播[8]。但是, Rootsi 的研究应用更大的样本量计算显示, 尽管 N3 在中国北部的频率很低并局限在少数人群中, 那里的 STR 方差却很高, 比阿尔泰和伏尔加—乌拉尔地区 (分别是 0.16、0.17) 都高, 因此说明了 N3 扩散的起源地是中国北部而不是南部西伯利亚。

表2 土耳其群体单倍群 N 下游分支的重新分类

Tab.2 Reclassification of the haplogroup N subgroups in Turkey population.

单倍型	N3a1	N3a2	N2-A1	N2-A2	N*	合计
样本量	3	2	11	3	1	20

STR 结构图同时得到了两个不同的分支,但是未能提供进一步的数据。文献[4]为此提供了详细的证据,估算了 N3 各个支系在不同地区的年代。引用如下:

“N3 的 STR 结构图显示了两个不同的支系(我们可以看到, N3a2 诞生于 N3a1)。N3a1 的第一次扩张约 10000 年前发生在南部西伯利亚,然后向欧洲扩散,在那里它的年代约为 8000 年。而更年轻的支系 N3a2,大约于 4000 年前诞生在南部西伯利亚(可能是贝加尔地区)。中接网络结构图和方差分析显示 N3a2 进入东欧的过程中经历了多次的瓶颈效应。

……

N3a1 在西伯利亚的年代比在欧洲的更古老(分别为 1 万年和 0.82 万年)。支系 N3a2 的 STR 有不同的拓扑结构。它在南部西伯利亚地区有星状扩张图形,主要出现在贝加尔湖地区的布里亚特人中。但是,它进入东欧的部分在系统发生树上显得十分复杂,主要出现在伏尔加—乌拉尔地区的芬兰—乌戈尔语以及突厥语人群中。比起西伯利亚人群,在伏尔加—乌拉尔人群中观察到了较高的 N3a2 的方差(0.66 万年对 0.37 万年),可能是源于以下原因:多次瓶颈效应,或东欧的 N3a2 样本量太少。另外, N3a2 的 STR 结构图中的东欧部分显示了明显的非星状拓扑的结构特征,这会降低基因年代计算的准确性(Saillard et al. 2000)。”

上文中最后一句话的更确切的含义是: N3a2 起源于西伯利亚的 N3a1,且西伯利亚的 N3a1 的方差要比欧洲的方差高。而欧洲的 N3a2 的西伯利亚方差要比西伯利亚的高,这是由于多次复杂的迁徙叠加造成的假象。综合起来,对于 N3 的总体的方差,欧洲的也就会比西伯利亚的高。但在 STR 结构图中就会表现出来:这些单倍型彼此离散,没有明显的联系。正如 Rootsi 在论文中提到的,“这么高的 STR 方差可能不是当地的奠基者类型长期演化的结果,而是多个不同的奠基

者类型混合的结果。”关于挪威、瑞典、芬兰的 Y-SNP 的研究都说明了这一点[9,10,11]。

三、N2 的下游分支与土耳其 N 的 STR

这篇文章指出, N2 的 STR 结构图区分了的亚洲支系 N2-A 和欧洲支系 N2-E,更古老的 N2-A 的年代反而小一些,与 N3a2 的情况类似,又一次显示了西伯利亚人群中有更强烈的瓶颈效应,使得很多的原始单倍型经过漂变而消失了。文中提到,土耳其的 14 个体属于亚洲支系 N2-A。而文献[4]提供了 N2 下更详细分类的奠基单倍型(founder haplotypes),并指明 N2-A2 几乎只存在于突厥的后裔中(Tuva、Tofalars),N2-E 主要存在东北欧的乌拉尔语以及相邻人群中。

由于文献[12]提供了土耳其 N 的样本的所有 STR 数据,因此我利用这个奠基单倍型对这些数据进行了分类(表 2)。

对照文献[4]提供的 STR 表(见引文附表 1),找到了土耳其样本的匹配单倍型(表 3)。

总的说来,从 STR 单倍型可以看见,12 例 N_xN₃ 的土耳其样品中,有 7 例(58.3%)能找到匹配的样本,除 1 例在卡尔梅克(Kalmyks)外,其中 6 例(50%)都在阿尔泰地区的突厥人后裔中(Khakassians、Tuva、Tofalars)。在样本量这么少的情况下,如此高比例的匹配是相当令人惊讶的。这意味着土耳其的这部分人的祖先,直接来源于阿尔泰地区突厥人。另外,即便没有找到匹配类型的个体,同样可以找到大量的“一步突变相似型”(one-step neighborhood),这意味着以后随着阿尔泰地区样本量的增大,这些个体也可能找到匹配类型。

因此可以说, N2-A 是突厥的特征单倍型,而 N2-A2 几乎只存在于突厥的后裔中。

表 3 土耳其样品与 Derenko 等报道样品的 STR 单倍型匹配分析

Tab.3 Matching of the STR haplogroups between Turkey samples and those published by Derenko et al.

土耳其样品	与文献[4]样品匹配	对应单倍群	群体
59	109;120;122;131	N3a2	Russian,Vepsa,Tuva,Chukchi
60	NO	N3a2	
61	NO	N3a1	
62	NO	N3a1	
63	23	N3a1	Russian
88;90	64;65;66;67;68	N2-A1	Evenks,Mongolians,Buryats,Altaians, Tofalars,Khakassians,Shors
89	69,73	N2-A1	Khakassians
91;93	74	N2-A1	Khakassians
94	NO	N2-A1	
94	75	N2-A1	Kalmyks
95	NO	N2-A1	
96	NO	N2-A1	
97	81,82	N2-A2	Tuva, Tofalars
98	NO	N2-A2	
99	NO	N2-A2	

注：由于缺少数据，其中一例 N2-A1 虽然匹配，但实际应为 N*。但无法确指是哪一例。

四、土耳其人 Y 染色体中的罕见类型 R1b1b-M73

R1b1b-M73 是一个罕见的单倍型，至今还没有专门的描述。它定义为 Y 染色体 *JARID1D* 基因内含子 11 中两个碱基对 (GT) 缺失，refSNP ID: rs2032634，首先被 P. J. Oefner 发现。Underhill 等于 2000 年 [13] 首次报道了中亚/西伯利亚的 6 例样本，但没有提供 STR 数据。文献 [12] 报道了土耳其有 4 例，文献 [14] 报道了亚洲的 19 例样本，两篇文章都提供了 10 个 STR 位点的数据。此外，Italy DNA Project (familytreedna.com) 显示意大利有 2 例，但也是欧洲唯一发现的两例。也就是说，目前全世界仅发现了 31 个样本 (表 4、表 5)。我们期待 Genographic Project 能够分析到更多的样本。

下面我们就这个单倍群做一些分析。

首先，R1b 是一个广泛分布的类型，在西欧、高加索地区有较高比例的分布。但那些地区的类型都属于 R1b1c-M269 下的分支，与 R1b1b-M73 不同。根据研究，R1b 在末次冰期之前就已经扩散。末次冰期到来的时候，群体生存空间被压缩到“避难绿岛 (refuge)”中。在末次冰期之后，R1b1c-M269 的下游支系分别由伊比利亚 [15] 和高加索 [12]

扩散开来。这个结果提示，R1b1b-M73 是 R1b 在末次冰期之前扩散过程中，遗留在中亚的一部分。但是 M73 的样本量过少，分布也很离散，我们只能在较大范围内猜测，它在末次冰盛期的避难绿岛可能在帕米尔至阿尔泰山一带。

其次，R1b1b-M73 现在的频率很低，但在某些彼此隔离人群中却很高，比如中国的纳西族和巴基斯坦的千户 (Hazara) 人。这两个群体之间的联系实在是难以想象。

巴基斯坦的 Hazara 人在阿富汗是一个人口众多的民族，有蒙古人种的体质特征。文献 [14] 显示，在巴基斯坦的 Hazara 人中有 41.7% (比例确实很高) 的 C3，33.3% 的 R1b1b，Q1 和 O3 分别为 8.3%，R2 和 J2a 分别为 4.1%。基本上可以确定 R2 和 J2a 为迁徙到阿富汗以后融入的部分，而 C3，Q1 和 O3 来自东亚。R1b1b 则应当来自中亚，或者更加东面的地方。需要分析 STR 的数据，才能得到更深入的理解。而中亚的群体历史上主要为贵霜王朝的人群。公元前 177 至前 176 年间，匈奴冒顿单于遣右贤王大败月氏。大月氏西迁侵入塞人地域伊犁河流域，后又遭乌孙的侵入继续西迁，最终在中亚阿姆河流域定居立国，成为强大的贵霜王朝的

表4 文献公布的 M73 突变型在各人群中的比例

Tab.4 Reported frequencies of M73 derived allele in the world populations

人群	总样本	发现数	比例	参考文献
Hazara	24	8	33.3%	[14]
Han	18	1	5.5%	[14]
Mongola	7	1	14.3%	[14]
Naxi	8	4	50.0%	[14]
Japanese	23	1	4.3%	[14]
Tu	7	1	14.3%	[14]
Uygur	8	3	37.5%	[14]
Turky	523	4	0.7%	[12]
Central Asia/Siberia	/	6	/	[13]

表5 文献公布的 M73 的 STR 数据

Tab.5 STR haplotypes of M73 group reported

文献中 样品号	群体	国家	微卫星位点重复数										个 体 数
			DYS19	DYS388	DYS389AB	DYS389CD	DYS390	DYS391	DYS392	DYS393	DYS439	DYSA7.2	
754	Hazara	Pakistan	14	12	18	13	23	11	13	14	12	9	4
755	Hazara	Pakistan	14	12	17	13	23	11	13	14	12	9	3
766	Hazara	Pakistan	14	12	18	12	24	11	13	12	13	10	1
927	Han	China	13	12	15	14	24	9	14	14	11	9	1
972	Mongol	China	13	12	17	13	23	10	14	13	13	11	1
976	Naxi	China	13	12	16	14	23	9	14	14	11	10	1
979	Naxi	China	13	12	17	14	24	9	14	14	10	10	3
1057	Japanese	Japan	13	12	16	13	23	10	14	13	13	10	1
1001	Tu	China	16	12	18	12	25	11	11	13	10	9	1
1014	Uygur	China	15	12	17	13	25	10	11	13	10	9	1
1016	Uygur	China	15	12	16	14	24	9	13	13	11	10	1
1017	Uygur	China	15	12	19	14	25	11	11	13	10	10	1
442	Turk	Turkey	14	12	17	13	19	11	13	13	12	9	1
443	Turk	Turkey	14	12	16	14	19	11	13	13	12	9	1
444	Turk	Turkey	14	12	15	13	25	10	13	12	11	9	1
445	Turk	Turkey	15	13	16	13	24	11	11	14	12	10	1
	Italy	Italy	14	12	15	14	25	11	13	12	11	/	1
	Italy	Italy	14	12	15	14	26	10	13	12	13	/	1

居民的一部分(公元45~250年)。但大月氏中的一部分仍然留在伊犁河一带,即《汉书·西域传》载:“故乌孙氏有塞种、大月氏种云”[16],成为现今哈萨克族的族源的一部分[17]。

土耳其民族源于中亚西突厥乌古斯人的游牧联盟,他们此时的活动地域正好包含了哈萨克斯坦大部[18]。

据史料记载,纳西族原为游牧在青海省黄河、湟水谷地的古旄牛羌人。约公元前200

年后,牦牛羌陆续向南迁徙至四川西南、云南西北一带,并分化成不同的民族,其中就包括纳西族[19]。遗传学[20]和语言学[21]的证据都说明,纳西族与藏族的关系非常密切。

《史记·大宛列传》记载:当月氏离弃河西时,有一小部分越祁连山,“保南山羌,号小月氏”[22]。这部分月氏人日后长期留住该地,与青海羌人逐渐融合。《后汉书·西羌传》载,到汉武帝时出山与汉人杂居,分为七个大种,住在湟中令居一带的称“湟中月氏

胡”，另有一部居张掖的称“义从胡”，后融入羌族，语言服饰均与羌人相似，但仍然与羌族区别明显，故仍名“胡”。《后汉书·西羌传》即云：“湟中月氏胡，其先大月氏之别也”。并且明确说明：“其羸弱者南入山阻，依诸羌居止，遂与共婚姻。” [23]

小月支保南山羌与纳西祖先在湟中在时间，地点上切合无间。因此推断，纳西族中的 R1b1b 可能来自小月支。

再次，分析 STR 数据，在 DYS19 的重复数上，形成了明显的簇。甘肃以东的样本全部为 13；Hazara 的样本全部为 14；新疆维族的样本全部为 15（暂不考虑土族特殊的 16）。土耳其人中有 DYS19 为 14、15 的样本，主要是 14，而没有 13，与以下事实是相符的：土耳其直接起源于中亚地区的乌古斯部落，而不是新疆的。而 Italy 的数据属于土耳其/Hazara 这个簇。

以上分析显示，古代月支人的 Y-SNP 中，有一部分是 R1b1b-M73。R1b1b-M73 的分布与古代部落月支的迁徙有关。

五、与突厥起源有关的其它单倍群

1) R1a1-M17

《周书·突厥传》称：“突厥之先出於索国” [24]。突厥西迁到中亚以后，融合了众多的当地居民。研究者检测叶尼塞河中游米努辛斯克盆地—克拉斯诺亚尔斯克地区 5500 至 1800 年前的古代 DNA [25]，结果显示有高比例的 R1a1-M17（11 例中有 10 例）。这些地区，现在是突厥语人群聚居的地方（Altaians-Kizhi、Shors、Tofalars、Khakassians）。在这些人群中，同样可以检测到一定比例的 R1a1 [26]。因此，R1a1-M17 也应该是突厥语人群中的重要组成部分。但是，M17 下尚未能发现有效的区分人群的突变。按照最新的分类 [26]，也只能大致区分出，东欧的 R1a1 与南西伯利亚的 R1a1 在 STR 上形成不同的簇，还无法与晚近的历史事件联系起来。

2) Q

我们还可以看到，文献 [12] 在土耳其发现了 10 例 Q。Q 被认为起源于中亚—阿尔泰

地区 [27]，现在在 Kets（93.8%）和 Selkups（66.4%）中比例较高 [28]。Kets 和 Selkups 被认为是来自目前居住地的南方——萨彦岭斜坡地区 [29,30]（Kets 所说的语言是一种孤立语，又有分类法称之为叶尼塞语系）。萨彦岭地区现在正是图瓦人聚居的地方。从 Seielstad 等 [31] 报道的欧亚大陆各人群中 Q 的比例看，Q 在突厥语人群中的比例也是不小的，在图瓦人（Tuvinian）、土库曼人（Turkmen）和乌兹别克（Uzbek/Tashkent）中分别达到 17%、10%、14%。特别是土库曼人，历史上与九姓乌古斯（九姓突厥）、塞尔柱突厥人的关系是密切的。而 Uzbek 族在形成的过程中，也融合了大量的突厥语部落，如克烈部、乃蛮部等 [32]。而在属于乌拉尔语系的匈牙利人 [33] 的 113 个样本中，Q 的数量（3 例）多过 N（1 例 N3）。甚至在同样源自东欧的 Ashkenazi Jews [34] 中，Q 也达到 5.2%。可见，N 从南西伯利亚向西迁徙扩散的过程中，伴随着一定比例的 Q。其中细节的故事必定更加引人入胜。

3) N3

根据“突厥之先出於索国”看来，突厥汗国的王族来自中亚的塞种。但是，突厥在早期就融合了铁勒部落阿史德氏，并且在降服了高车人五万部落之后才发展壮大。后来，异姓突厥也融合进来，使得“突厥人”几乎称为所有铁勒人的统称。突厥汗国的民众，大部分来源于丁零—铁勒诸部。

考古学者，人类学者已经对雅库特人（Yakuts）和布里亚特人（Buryats）的起源做了研究。俄罗斯的考古证据证实 [35]，公元五世纪中叶至六世纪初，贝加尔湖地区出现了一支灿烂的古代文化——骨利干（Kurykan，火儿罕），它的居民是操突厥语的游牧民，在公元第一个千年为文献所记录。基于多种文化因素的相似性，比如岩画、陶器、骑马的服饰、马缰的装饰以及民俗等等，人类学家认为，在公元 11 至 13 世纪蒙古人扩张之后，有一支骨利干人沿着勒拿河向北迁徙，融合了当地的土著居民，形成雅库特人 [36,37]。研究表明，雅库特人有极高比例的 N3（94%） [2]，全部都是 N3a1 [4]（以及本文评论的文献）。

根据历史记载[38], 古铁勒部落之一拔野古部(巴尔虎)最初居住在漠北, 其后迁入贝加尔湖两岸。此时, 强大的骨利干人已被豁里-秃马特(Polecat-tumat)打败, 向更北方迁徙。此后, 拔野古(巴尔虎)逐渐融合为蒙古部落, 成为布里亚特族的一部分。文献[4]的数据表明, 布里亚特人中的N几乎全部都是N3a2(97.6%), 而完全没有与骨利干人有联系的N3a1, 说明布里亚特地区的N3a2是N3a1消失以后来到那里的。文献[39]用强有力的数据说明了, 布里亚特人的mtDNA类型是蒙古语群体和突厥语群体的类型的混合, 从另一方面支持了以上融合的说法。

以上论证说明, 丁零-铁勒诸部的主要Y-SNP类型N3a1、N3a2和N2-A。

遗憾的是, 至今仍未有萨莫耶德语族和楚瓦什人中N2的STR分析。萨莫耶德语族民族中有很高比例的N2。楚瓦什语有原始突厥语的特征。更深入的研究有助于更清楚地了解N2在乌拉尔语人群和突厥语人群中的作用。另外, Karafet等[40]的论文显示, 维吾尔族有高比例的R1b(17.6%)。中亚的其他人群也有高比例的R1xR1a1[41], 比如Uzbek/Kashkadarya的21%、Tajik/Khojant的32%。因为高加索类型的R1b1c的扩散范围很小[12], 因此这些个体都很有可能是R1b1b-M73。希望学者们在以后的研究中, 更加关注R1b1b-M73。

参考文献

- Xue YL, Zerjal T, Bao WD, Zhu SL, Shu QF, Xu JJ, Du RF, Fu SB, Li P, Hurler ME, Yang HM, Tyler-Smith C (2006) Male Demography in East Asia: A North-South Contrast in Human Population Expansion Times. *Genetics* 172:2431-2439.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, Stoneking M. (2006) Investigating the effects of prehistoric migrations in Siberia- genetic variation and the origins of Yakuts. *Hum Genet* 120:334-353.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, and Wallace DC (2002) The Dual Origin and Siberian Affinities of Native American Y Chromosomes. *Am J Hum Genet* 70: 192-206.
- Derenko M, Malyarchuk B, Denisova G, Wozniak M, Grzybowski T, Dambueva I, Zakharov I (2007) Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J Hum Genet* 52:763-770.
- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjida D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Gene* 60: 1174-1183.
- Villems R, Adojaan M, Kivisild T, Metspalu E, Parik J, Pielberg G, Rootsi S, Tambets K, Tolk HV (1998) Reconstruction of Maternal lineages of Finno-Ugric speaking people and some remarks on their Paternal inheritance. In: Julku K, Wiik K (Toim.) *The Roots of Peoples and Languages of Northern Eurasia I*. pp180-200.
- Rootsi S, Kivisild T, Tambets K, Adojaan M, Parik J, Reidla M, Metspalu E, Laos S, Tolk HV, Villems R(2000)On the phylogeographic context of sex-specific genetic markers of Finno-Ugric populations. In: Kiinnap A (ed.) *The Roots of Peoples and Languages of Northern Eurasia II and III*, Tartu. pp148-164.
- Villems R, Rootsi S, Tambets K, Adojaan M, Orekhov V, Khusnutdinova E, Yankovsky N (2002) Archaeogenetics of Finno-Ugric speaking populations. In: Julku K (ed.) *The Roots of Peoples and Languages of the Northern Eurasia*. pp271-284.
- Dupuy BM, Stenersen M, Lu TT, Olaisen B(2005)Geographic heterogeneity of Y-chromosomal lineages in Norway. *Forensic Sci Int* 164:10-19.
- Karlsson AO, Wallerstrom T, Gotherstrom A, Holmlund G (2006)Y-chromosome diversity in Sweden-A long-time perspective. *Eur J Hum Genet* 14:963-970.
- Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, Lahermo P(2006) Regional differences Among the Finns: a Y-chromosomal perspective. *Gene* 376:207-215.
- Cinniöglü C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA(2004)Excavating Y-chromosome Haplotype Strata in Anatolia. *Hum Genet* 114:127-148.
- Underhill PA, Shen P, Lin AA, Passarino G, Yang WH, Karuffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oerner PJ (2000) Y Chromosome Sequence Variation and the History of Human Populations. *Nat Genet* 26:358-361.
- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA (2006) 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. *Am J Hum Genet* 78:202-221.
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y-chromosome perspective. *Science* 290:1155-1159.
- 班固(汉) 汉书. 卷九十六下.西域传第六十六下.

17. 王锺翰(2001) 中国民族史(增订本).北京:中国社会科学出版社.
18. 敬东(1996) 试论乌古斯突厥蛮塞尔柱克人的联系与区别. 西北民族研究 19:227-234.
19. 杨福泉(2005) 纳西族与藏族历史关系研究. 北京:民族出版社.
20. 石宏, 董永利, 李卫翔, 杨洁, 李开源, 咎瑞光, 肖春杰 (2002)中国云南 25 个少数民族 Y 染色体 DYS287 位点的地理多态性.中国科学(C 辑) 32:373-378.
21. 李英(1992)从语言探索纳西族与古氏羌人的渊源关系.纳西族研究论文集.北京:民族出版社,41-48.
22. 司马迁(汉) 史记 卷一百二十三 大宛列传第六十三. 大月氏.
23. 范晔(南朝宋) 后汉书.卷八十七.西羌传第七十七.湟中月氏胡.
24. 令狐德棻(唐) 周书.卷五十.列传第四十二.异域下.突厥吐谷浑高昌鄯善焉耆龟兹于阗嚙哒粟特安息波斯.
25. Bouakaze C, Keyser C, Amory S, Crubézy E, Ludes B(2007)First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. Int J Legal Med 121:493-499.
26. Derenko M, Malyarchuk B, Denisova G, Wozniak M, Dambueva I, Dorzhu C, Luzina F, Miścicka-Śliwka D, Zakharov I (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. Hum Genet 118:591-604.
27. Underhill PA, Passarino G, Lin AA, Shen P, Mirazon LM, Foley R, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43-62.
28. Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, Hammer MF (2002) High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. Hum Biol 74:761-789.
29. Popov AA, Dolgikh BO (1964) The Kets. In: Levin MG, Potapov LP (ed.) The peoples of Siberia. Chicago: University of Chicago. pp 587-606.
30. Prokof'yeva ED (1964) The Selkups. In: Levin MG, Potapov LP (ed.) The peoples of Siberia. Chicago: University of Chicago Press. pp 607-619.
31. Seielstad M, Yuldasheva N, Singh N, Underhill PA, Oefner PJ, Shen P, Wells RS (2003) A Novel Y-Chromosome Variant Puts an Upper Limit on the Timing of First Entry into the Americas. Am J Hum Genet 73: 700-705.
32. 拉施特(波斯). 余大钧,周建奇译 (1997)史集[中译文] 第一卷.第一分册. 北京:商务印书馆.
33. Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli EL, Tolk HV, Reidla M, Metspalu E, Pliss L, Balanovsky O, Pshenichnov A, Balanovska E, Gubina M, Zhadanov S, Osipova L, Damba L, Voevoda M, Kutuev I, Bermisheva M, Khusnutdinova E, Gusar V, Grechanina E, Parik J, Pennarun E, Richard C, Chaventre A, Moisan JP, Barác L, Perić M, Rudan P, Terzić R, Mikerezi I, Krumina A, Baumanis V, Kozziel S, Rickards O, De Stefano GF, Anagnou N, Pappa KI, Michalodimitrakis E, Ferák V, Füredi S, Komel R, Beckman L, Villems R (2004) The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. Am J Hum Genet 74:661-682.
34. Behar DM, Garrigan D, Kaplan ME, Mobasher Z, Rosengarten D, Karafet TM, Quintana-Murci L, Ostrer H, Skorecki K, Hammer MF (2004) Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. Hum Genet 114:354-365.
35. Okladnikov AP (1955) The history of the Yakut ASSR. Volume 1: Yakutia before its incorporation into the Russian state [in Russian]. Moscow: Izdatel'stvo Akademii Nauk SSSR.
36. Alekseev AN (1996) Ancient Yakutia: the Iron Age and the Medieval epoch [in Russian]. Novosibirsk: Izdatel'stvo Instituta Arkheologii i Etnografii SO RAN.
37. Konstantinov IV (1975) The origins of the Yakut people and their culture [in Russian]. Yakutia and her neighbors in antiquity. Publications of the Prilenskaya Archaeological Expedition. Yakutsk: Yakutskiy filial SO AN SSSR. pp 106-173.
38. 宝敦古德.阿毕德(1982) 布里亚特蒙古简史. 内蒙古:内蒙古文化出版社.
39. Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. (2003) Mitochondrial DNA Evidence for Admixed Origins of Central Siberian Populations. Am J Phys Anthropol 120:211-224.
40. Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal Population History of East Asia: Sources, Patterns, and Microevolutionary Processes. Am J Hum Genet 69:615-628.
41. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian heartland- a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA. 98:10244-10249.

原文翻译 兰海 译

从东南亚到欧洲 Y 染色体单倍群 N 的顺时针北方路线

摘要：东欧和东亚人群的很大一部分 Y 染色体属于单倍群 NO。NO 由两个姊妹单倍群组成：N-231 和 O-M175。O 相当古老(约为 30ky)，是东亚和东南亚男性谱系的最主要的单倍群，也是大洋洲男性的显著成分。另外，我们对于单倍群 N 的详细信息显示，N 在东欧的高频是缘于很晚近的迁徙，这次迁徙约从 1 万 2

千到 1 万 4 千年前开始, 由内亚/南西伯利亚出发, 走一条逆时针的北部路线。N 在西伯利亚广泛分布却没有在美洲土著发现, 这意味着它是在首次前往美洲的迁徙发生之后才开始扩散的。频率最高的下游单倍群 N3, 可能起源于今天的中国, 然后在西伯利亚经历多次瓶颈效应最后扩散到东欧。另一个下游单倍群 N2, 形成了亚洲 (N2-A) 和欧洲 (N2-E) 两个不同的 STR 分支, N2-E 现在主要分布在芬兰—乌戈尔语人群以及相关人群。这些系统地理分布状况提供的证据支持以下的事实: 在晚更新世到全新世之间, 东亚的旧石器时代谱系中, 有多次以男性为主的沿逆时针路线一直到达西北欧洲的迁徙。

介绍

男性特有的 Y 染色体非重组区差异在世界范围内的分布, 已经成为重建早期人类迁徙的重要手段。Y 染色体的系统结构的分辨率随着新的双等位多态位点 (大部分是单核苷酸多态—SNP) 的发现变得越来越高。这些 SNP 不同的地理分布暗含着现有人群的历史关系的证据[1-5]。

单倍群 NO 是欧亚大陆上分布最广频率最高的 Y 染色体系统分支之一, 它由 SNP 标记 M-214 定义 (由 Cinnioglu 等[6]正确定义, 如图 1)。它包含没有可辨认的其他下游突变 SNP 标记的低频的 NO* (见图 2a) 和两个高频的姊妹分支 N 和 O, 分别由 SNP 标记 M231 和 M175 定义 (见图 2b, c)。尽管 O 分支的系统地理分布已经被描述得相当清楚[7-10], 但是关于 N 的知识还相当贫乏, 特别是它的起源、系统结构以及人口统计学的重要意义。

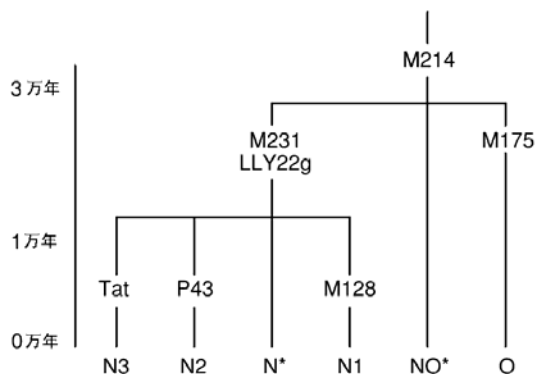


图 1 NO 的系统结构 NO 及它的下游支系的系统结构以及定义它们的 SNP 标记。突变标记遵从 YCC 的命名规则。在系统结构的左边有各支系分离的年代。

材料与方法

样本与 DNA 分型

在这项研究中, 根据最新的系统结构, 来自不同地区 58 个人群的总共 5389 个样本

被定型或重定型 (92R7 位点无突变而 M9 位点突变的样本都分析了 M214、M231、M128、P43、TAT、M175)。这些样本和相关文献报道的 8019 个个体的数据一起进行了分析 (数据见 EJHG 网上附表 1)。突变标记遵从 YCC 的命名规则[1,4]。在已知的系统发生关系中, 标记点 M128、P43 和 TAT 分别定义了 N1、N2、N3 三个分支[2,11,12], 图 1 的 Y 染色体分化树中仅有 M231 是最近被发现的 (它定义整个 N 分支)。标记点 M231 在系统发生关系上等同于结构更复杂的 LLY22g 多态, 后者以前被用来定义单倍群 N。

标记点 M128 和 M124[2,5]由点突变分型而定型, 标记点 M175[2]、P43[11]和 TAT[12]通过 RELP 技术 (内切酶限制性片段长度多态性) 分别用限制酶 *Mbo* II、*Nla*III、*Tai* I 来分型。M231 最早由 Cinniolu 等[6]描述, 最初是由高效液相色谱方法进行变性来分型。它的位点状态现在可以用 RELP 技术很快完成分型 (切割酶 *Tai* I 切割 G 等位型, 得到 223bp 和 108bp 片段, 而不会切割 C 等位型从而得到 331bp 的片段)。

STR 用 Y-filer 系统 (荧光标记复合扩增系统, Applied Biosystems, Foster City, CA, USA) 进行分析。PCR 分析是在 ABI 3100Avant 基因分析仪上用标准片段分析程序进行的。每个样本中都加入了 GeneScan 500LIZ 的长度标准 (Applied Biosystems) 以便进行长度分级, 使用 GeneMapper3.5 记录等位基因长度。等位基因长度被记录为重复数目。

数据分析

应用 Network4.112, 我们构建了单倍群 NO 的 STR 单倍型的中接 (MJ) 网络结构图。其中用到了 17 个 STR 的数据 (DYS19、DYS385a,b、DYS389 I, II、DYS390、DYS391、DYS392、DYS393、DYS437、

DYS438、DYS439、DYS448、DYS456、DYS458、DYS635、Y GATA) 以及 58 个个体的双等位标记 (M231、M128、P43、TAT) 的数据 (见附表 2)。单倍型的系统发生关系由中接网络结构(Median Joining Network)方法决定,在此之前,数据需要经过 Bandelt[14]等描述的 RM(reduced-median)法应用

Network4.112 处理。

年代是基于 STR 的多态并根据报道的方法计算[15,16]得到的。

NO 以及下游支系的突变型频率分布图(黑点代表人群)是基于附表 1 的数据应用 Surfer 软件 (版本 7, Golden Software, Inc) 分析得到的。

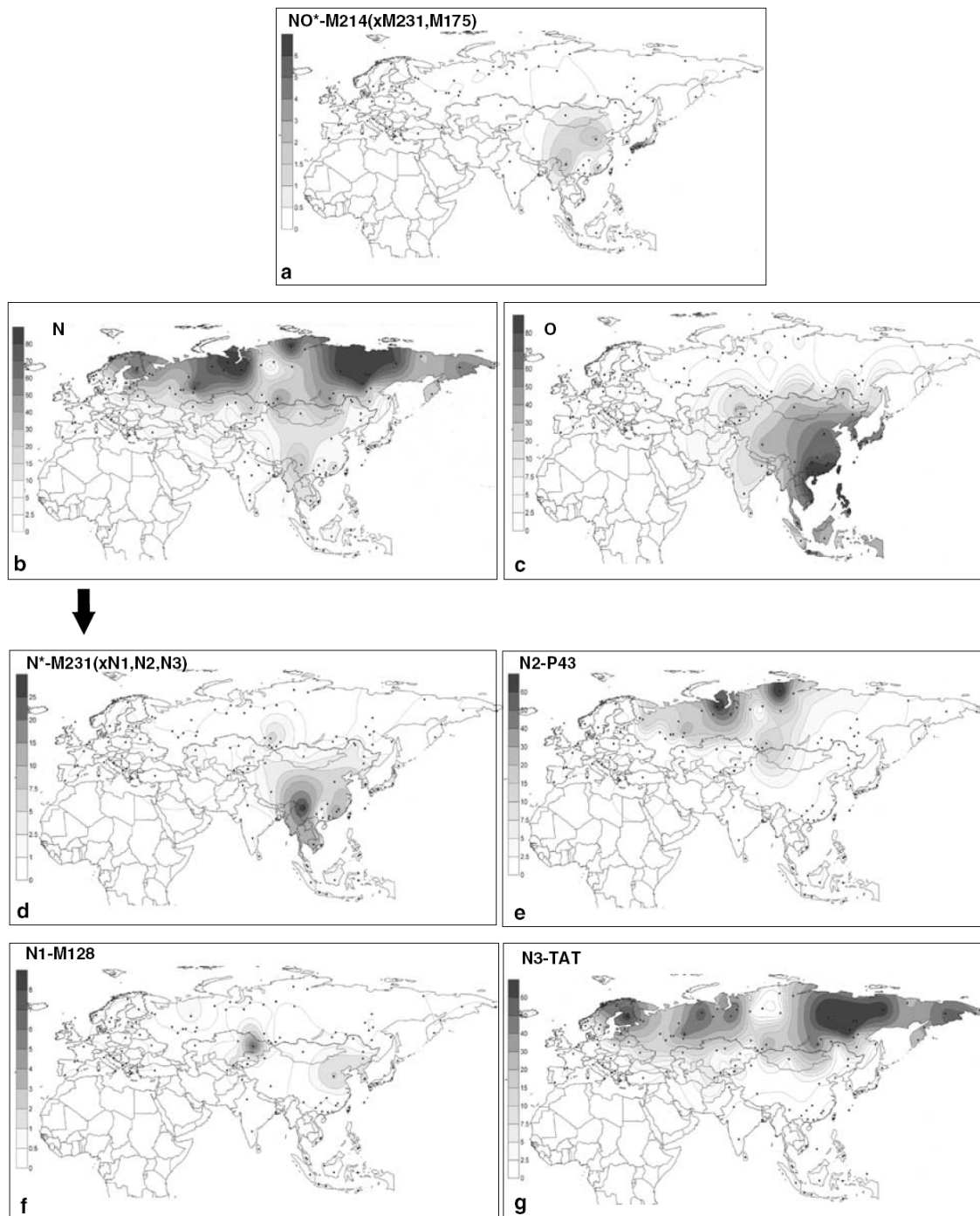


图 2 NO 系的地理分布 (a-g) NO 各支系的空间频率分布: NO*、N(合计), O(合计), N*、N1、N2、N3。图示是基于附表 1 的数据。我们按照 YCC 的命名规则, 在各个地图中标明了单倍群, 并在其后注明了相应的突变标记。

结果与讨论

N2、N3 与 NO*、N、O 的不同分布状态

单倍群 N 在欧亚大陆北部有独特且广泛的分布, 从远东一直到东欧的高纬度地区都有高频的分布[11,12,17]。我们对来自欧洲、东亚/东南亚以及大洋洲等不同地区的样本, 确定单倍群 N、其下游支系以及姊妹单倍群 O 的 SNP 标记, 并通过详细的谱系分析, 研究了单倍群 N 的历史。分析表明, 尽管 N 很早就与 O 系分离, 但单倍群 N 的下游支系显示了更晚近的人口扩散分布, 与姊妹单倍群 O 有着完全不同的接近逆时针的迁徙路线。

尽管有不同的频率数值, 祖先型超单倍群 NO*-M214、N-M231*、以及高频的单倍群 O-M175 都一致而且明显指出, 东南亚是最有可能的这些单倍群产生的地方。尽管超单倍群 NO* 比单倍群 N* 的频率要低得多[18,19], 但它大致与单倍群 N* 分布在东南亚的同一个区域 (Kayser et al [19] 中的数据在这里经过重新分型)。更值得注意的事实是, 整个单倍群 N 与 O 的空间分布方向是非常不同的。N* 与 O 分离的年代确定为 3.47 ± 0.47 万年前。东南亚单倍群 O 的 STR 年代大约超过 2.6 万年[10], 它的数个下游单倍群现在在东亚、东南亚人群占主导, 一致延展到华北、中国东北和南西伯利亚人群[7,9,11,20,21], 并向西进入印度次大陆东部, 向东达到大洋洲。

单倍群 N 下游支系的分布与扩散

单倍群 N 体现了很晚近的人口扩散事件。祖先型 N* 尽管频率很低, 但广泛分布于斐济、文莱、柬埔寨、华南以及日本, 一直到南部西伯利亚 (附件 1), 但在印度半岛显然没有发现[10]。应用 Zhivotovsky et al[15] 的方法计算 STR 变异得到的年代指向更新世晚期到全新世早期之间 ($1.19 \pm 0.25 \sim 1.26 \pm 0.31$ 万年), 随分析涉及的 SNP 数量与 STR 基因座个数而变化 (表 1)。但是, 我们也要注意, N* 的频率是很低的。

考虑到这些, 综合 N1、N2、N3 的 15 个 STR 位点的数据, 计算单倍群 N 的 STR 变异得到年代为 1.94 ± 0.48 万年。但是正如

下面将要提到的, 欧洲的 N2 下游支系和雅库特的 N3 有可能分别来自一个单独的奠基者, 后者可能在多个位点经历了多级跳跃突变, 导致基于逐级突变模式的统计方式的计算结果有所偏差。排除这些染色体单倍群, 单倍群 N 的 STR 分化程度就要小一些, 年代为 1.42 ± 0.1 万年。

基于进化差距估计和在家系中的观察得到的两种突变速率进行计算, 得到的年代差异相当大 (表 1)。因为并没有在所有遗传学家中达成一致, 因此本文中两种方法都作了估算。最近的模拟检验显示, 家系速率并没有考虑人群动力学的进化结果, 比如新产生的微卫星位点的快速消亡 (Zhivotovsky, Underhill and Feldman[23])。因此, 基于家系速率计算得到的年代过于年轻, 与考古学的证据不符。与此相关的其他因素有: 1、调查偏差 (在家系中没有观察到突变的研究较少发表); 2、位点之间的进化速率差异, 家系速率得到的是进化速度最快的位点的平均速率; 3、饱和度, 进化速率的计算忽略了回复突变。

基于 17 个 STR (附表 2) 与 SNPs 位点绘制的网络结构图 (图 3) 显示了单倍群 N 的各下游单倍群的分布差异。尽管 N2、N3 已经出现在楚科奇人以及楚科奇半岛的尤匹克人等西伯利亚人群中, 它们在美洲土著中却完全不存在[24-28]。这个发现说明, 在大约 1.2~1.7 万年前, 旧石器时代的西伯利亚人移民美洲的时候, 单倍群 N 在他们中并不是占主要或是普遍的单倍型。但是也有可能单倍群 N 由于奠基者效应或者遗传漂变而在美洲消失, 这种可能不能排除。

单倍群 N3 是 N 系最普遍的支系 (图 2g, 附表 1), 在几乎所有居住于欧亚大陆北部的人群中是最主要的染色体类型 [11,12,17,26,30-35], 而在中国、韩国、文莱和日本频率很低[18,19]。尽管在亚洲北部人群中普遍存在, 单倍群 N 在欧洲局限在北部和东部人群, 在斯堪的那维亚以及立陶宛与波兰之间显示出快速的减少趋势 (附表 1 以及 [17] 中的数据)。NO* 和 N* 的人口地理分布以及 N* 在东南亚人群 (华南、柬埔寨, 见附表 1) 中的存在, 表明这个区域应该是单倍

表 1 各单倍群的分子钟时间

单倍群	进化速率计算结果 (万年)	家系速率计算结果 (万年)
N*/O	3.46±0.47	1.04±0.14
N* (5条染色体)	1.26±0.31	0.38±0.09
N* (中国和柬埔寨11条染色体)	1.19±0.25	0.36±0.07
N* (16条染色体)	1.26±0.31	0.38±0.09
N(综合N1-N3数据)	1.94±0.48	0.58±0.14
含雅库特N3和N2-E		
N 不含雅库特N3和N2-E	1.42±0.40	0.42±0.12
N2-A	0.62±0.20	0.18±0.06
N2-E	0.68±0.29	0.20±0.09
N3(华北)	1.18±0.68	0.35±0.20
N 和 I 分化	5.41±3.18	1.62±0.59

说明从略。

群 N 的起源地。这样看来,阿尔泰山/萨彦岭/南西伯利亚地区就是单倍群 N 向西迁徙的中转站,因为现在还可以在那里找到 N 系的所有下游支系。

尽管 N3 在中国北部的频率很低并局限在少数人群中,但那里的 STR 方差却很高 (0.26, 8 个位点的平均值: DYS19、DYS389I&II、DYS391、DYS392、DYS393、DYS439, 数据来自 Sengupta et al [10]), 比阿尔泰和伏尔加—乌拉尔地区都高 (分别是 0.16、0.17), 因此再一次说明, N3 扩散的起源地是中国北部而不是南部西伯利亚。计算中国北部的 N3 的 STR 方差得到的年代为 1.18 ± 0.68 万年, 处在更新世与全新世之间。不过这一结果需要谨慎对待, 因为由于这里 N3 的样本太少可能会导致很大的标准差。

根据我们的推论, 在由西伯利亚迁往东欧的途中, N3 类型的人群经历了奠基者效应或强烈的瓶颈效应。东北欧可以认为是 N3 二次扩张的地方。事实上, N3 确实相当高频地存在于伏尔加—乌拉尔地区的乌戈尔语人群以及相关的芬兰人、萨米人和爱沙尼亚人中。应该注意到, 伏尔加—乌拉尔地区人群的 STR 方差很低, 而某些北欧人群的却很高 (例如, 芬兰人的 0.32, 数据来自[36], 不包含 DYS385ab)。后者这么高的 STR 方差可能不是当地的奠基者类型长期演化的结果, 而是多个不同的奠基者类型混合的结果。

东欧人群中有相当一部分的 N3 来自单一的祖先, 他们处在 N3 扩散的西部边界, 倾向于有较低的 STR 方差。比如, 在波罗的海地区的立陶宛/拉脱维亚人中, STR 方差分

别是 0.12 和 0.09(5 个位点的数据, [37])。一些欧洲人群 N3 的频率较低, 同时却有较高的 STR 方差, 比如非萨米人的挪威人 (0.27, 数据不包括 DYS439)。这很可能是由于晚近来自毗邻的芬兰—乌戈尔人群的基因流动 [36,38]。类似的情形在最近关于瑞典人 Y 染色体基因库的研究中也有报道[39]。

STR 变异的系统结构分析得到了 N3 的两个重叠的支系 (图 3)。其一包含伏尔加—乌拉尔地区人群, 说芬兰语以及突厥语的人群, 还有阿尔泰人。另外一个包含波罗的一芬兰语人 (爱沙尼亚人、卡里亚拉人、维普人), 东斯拉夫人 (俄罗斯人、乌克兰人) 以及西斯拉夫的斯洛伐克人中的 N3 单倍型。雅库特人的染色体形成了自己的分支, 他们总是与别的支系不同, 与之前的研究一致 [12,40]。

单倍群 N2 在西伯利亚人群中的分布没有规则, 向西扩散到东欧, 一直到波罗的海边的维普人和卡里亚拉人中。N2 的最高频率在西北部西伯利亚人群中观察到: 恩加纳桑人的 92%, 埃内茨人的 78% 和苔原涅涅茨人的 74% [11]。在欧洲, N2 在伏尔加—乌拉尔地区人群到达最高的 20% [17]。N2 扩散分布的最西边界是芬兰, 在那里它的频率极低, 只有 0.4% [34]。有意思的是, N2 在维普人中频率很高 (17.9%)。维普人是一个操芬兰语的小人群, 生活在芬兰人、卡里亚拉人与爱沙尼亚之间。

N2 单倍型的网络结构图展示了两个清晰的 STR 分支 (图 3), 区分了欧洲和西伯利亚的支系, 分别定义为 N2-A 和 N2-E。可以

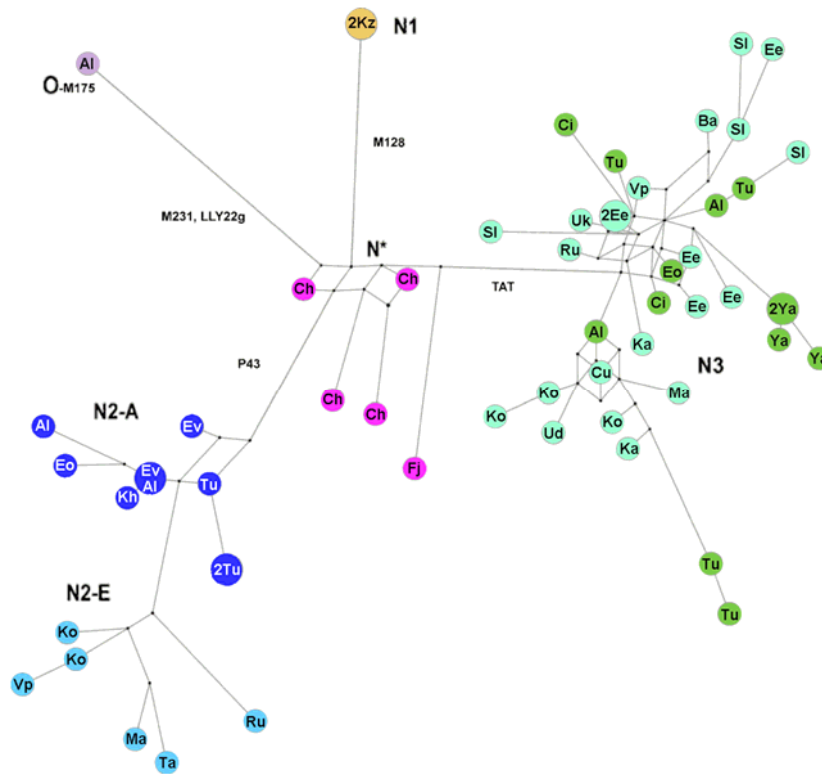


图3 单倍群N0的中接网络结构图 我们应用Network 4.112, 基于17个STR和58个个体的双等位标记的数据, 构建了单倍群N0的中接网络结构图。每一个圆圈表示一个由一组STR定义的单倍型。圆圈的大小表示单倍型的比例, 参见附表2。单倍型的标识如下: Al-阿尔泰人, Ba-巴什基尔人, Ch-中国人, Ci-楚科奇人, Cu-楚瓦什人, Eo-爱斯基摩人, Ee-爱沙尼亚人, Ev-鄂温克人, Fj-斐济人, Ka-卡累利阿人, Kh-哈卡斯人, Ko-科米人, Ma-玛里人, Ru-俄罗斯人, Si-斯洛伐克人, Ta-塔塔尔人, Tu-图瓦人, Ud-乌德穆特, Uk-乌克兰, Vp-维普人, Ya-雅库特人。颜色表示单倍群下的细分结构: 淡绿色表示欧洲的N3单倍型, 深绿表示亚洲的N3单倍型, 淡蓝表示欧洲的N2单倍型, 蓝色表示亚洲的N2单倍型, 棕色表示N1的单倍型, 粉红色表示N*的单倍型, 紫色表示O的单倍型。

推知, 最接近根部的亚洲支系 N2-A 首先产生, 然后才产生了欧洲支系 N2-E。尽管 N2-A 在多个 STR 位点上的重复数与 N3 的相同, 但 N2-E 的重复数与 N3 差别很大 (表 2), 说明欧洲的 N2 染色体来自一个单独的奠基者类型。我们甚至可以推测, 可能有某个双等位点多态还没有被发现, 而它可能是 N2-E 支系独有的。N2-A 和 N2-E 都很年轻, STR 方差计算得到的年代分别为 0.62 ± 0.20 万年和 0.68 ± 0.29 万年。更古老的 N2-A 的年代反而小一些, 可以解释为在西伯利亚人群中有更强烈的瓶颈效应以及较少的样本量。事实上, 西伯利亚土著居民的人口规模与东欧人群相比是相当小的; 即便是人口众多的农业人口, 雅库特人和布里亚特人只有数十万,

而东欧的人群都有数百万以上[41,42]。

表 2 单倍群 N 各支系中信息量最大的 STR 标记点的重复数的中值

单倍群	高信息含量标记				
	DYS19	DYS389b ^a	DYS392	DYS385b	DYS448
N2					
N2-E	13	18	12	12	18
N2-A	14	16	14	13	19
N3	14	16	14	13	19
N1	14	16	15	13	19
N*	14	17	14	12	19

^a DYS389b 的长度是由 DYS389II 减去 DYS389I 得到的。

在我们的样本中, N2-E 主要局限于伏尔加—乌拉尔地区, 那里极可能是 N2 向北向东扩散的起源地 (汉特人、曼西人, 数据来自 Stepanov et al [43])。与西伯利亚西北 N2-A 的 STR 类型不同, 西部的类型更多地属于 N2-E。有意思的是, Cinniolu et al[6]中土耳其的 14 个个体属于亚洲支系 N2-A (数据在本文经过更新)。这表明, N2 由西伯利亚向西扩散至少有两个途径: 其一向西北, 穿过伏尔加---乌拉尔地区, 产生了 N2-E, 可能主要通过芬---乌戈尔语人群扩散; 其二 N2-A, 伴随突厥语人群向西南迁徙。因此, N2-A 与 N2-E 的 STR 数据的不同, 以及雅库特人 STR 数据的一致性, 显示了冰后期的多重奠基事件的结果, 特别是在人口密集的中亚。这一点与中亚被称为“瓶颈之地”的观点是一致的。

N 系最少的支系是 N1(图 2f), 低频分布于一些中亚人群、韩国人、华北汉族以及中国东北的鄂温克人。要想更多地了解这个支系的扩散和分布状况, 我们需要按照现在系统树的分辨率进行更大规模的研究 (早期的研究往往没有提供足够的数据)。

总之, Y 染色体单倍群 N 说明了向东欧的基因流动, 而它最初的起源地是东亚。单倍群 N 的迁徙故事里并没有伴随着类似的 mtDNA: 东南亚人群特异的 mtDNA 类型出现在东波罗的语人群中总的频率低于 1% [17,45,46]。只有一些亚洲 mtDNA 的小分支, 比如在阿尔泰/中亚多样性很高的 Z1、D5, 在一些北欧人群中超过 1%, 比如萨米人、芬兰人[17,45,46]。

原文参考文献

1. Jobling MA & Tyler-Smith C (2003) *Nat Rev Genet* 4:598-612.
2. Underhill PA et al (2000) *Nat Genet* 26: 358-361.
3. Hammer MF et al(2001) *Mol Biol Evol* 18:1189-1203.
4. Y Chromosome Consortium (2002) *Genome Res* 12: 339-348.
5. Underhill PA et al (2001) *Ann Hum Genet* 65: 43-62.
6. Cinniolu C et al (2004)*Hum Genet* 114: 127-148.
7. Su B& Jin L (2000)*Nat Rev* 1: 126-133.
8. Wen B et al(2004) *Nature* 431: 302-305.
9. Shi H et al(2005)*Am J Hum Genet* 77: 408-419.

10. Sengupta S et al(2006) *Am J Hum Gen* 78: 202-221.
11. Karafet TM et al(2002)*Hum Biol* 74: 761-789.
12. Zerjal T et al(1997)*Am J Hum Genet* 60: 1174-1183.
13. Hurles ME et al(2005)*Am J Hum Genet* 76: 894-901.
14. Bandelt HJ et al(2000)*Mol Phylogenet Evol* 16:8-28.
15. Zhivotovsky LA, Underhill PA, Cinniolu C et al(2004)*Am J Hum Genet* 74: 50-61.
16. Dupuy BM et al(2004)*Hum Mutat* 23: 117-124.
17. Tambets K et al(2004) *Am J Hum Genet* 74: 661:682.
18. Hammer MF et al(2006)*J Hum Genet* 51: 47-58.
19. Kayser M et al(2003)*Am J Hum Genet* 72: 281-302.
20. Su B et al(1999) *Am J Hum Genet* 5: 1718-1724.
21. Tajima A et al(2002)*Hum Genet* 110: 80-88.
22. Rootsi S et al(2004)*Am J Hum Genet* 75: 128-137.
23. Zhivotovsky LA et al(2006) *Mol Biol Evol* 23:2268-2270.
24. Bergen AW et al(1999)*Ann Hum Genet* 63: 63-80.
25. Karafet TM et al(1999)*Am J Hum Genet* 64: 817-831.
26. Lell JT et al(2002)*Am J Hum Genet* 70: 192-206.
27. Seielstad M et al(2003)*Am J Hum Genet* 73: 700-705.
28. Bortolini MC et al(2003)*Am J Hum Genet* 73: 524-539.
29. Zegura SL et al (2004)*Mol Biol Evol* 21: 164-175.
30. Semino O et al (2000)*Science* 290: 1155-1159.
31. Zerjal T et al (2001)*Mol Biol Evol* 18: 1077-1087.
32. Rootsi S et al (2000) in Künnap A (ed): *The roots of peoples and languages of Northern Eurasia II and III.* Tartu: University of Tartu. Division of Uralic Languages/Societas Historiae Fenno-Ugricae, pp148-164.
33. Rosser ZH et al(2000)*Am J Hum Genet* 67: 1526-1543.
34. Villems R(2002) in Julku K (ed): *The Roots of Peoples and Languages of Northern Eurasia IV.* Oulu: Societas Historiae Fenno-Ugricae, pp 271-284.
35. Laitinen V et al (2002)*Hum Hered* 53: 68-78.
36. Lappalainen T et al(2006)*Gene* 376: 207-215.
37. Kasperaviciute D et al(2004)*Ann Hum Genet* 68: 438-452.
38. Dupuy BM et al(2006)*Forensic Sci Int* 164(1):10-16.
39. Karlsson AO et al(2006)*Eur J Hum Genet* 14:963-970.
40. Pakendorf B et al(2002)*Hum Genet* 110: 198-200.
41. Kolga M et al (2001) *The Red Book of the Peoples of the Russian Empire.* Tallinn, NGO Red Book.
42. Russian Population Census (2002) www.perepis2002.ru.
43. Stepanov VA et al(2006)*Vestnik VOGiS* 10: 57-73.
44. Zerjal T et al(2002)*Am J Hum Genet* 71: 466-482.
45. Richards M et al(2000)*Am J Hum Genet* 67: 1251-1276.
46. Pliss L et al(2005) *Ann Hum Genet* 69: 1-20.
47. Finnila S et al(2001)*Am J Hum Genet* 68: 1475-1484.
48. Torroni A et al(1993)*Am J Hum Genet* 53: 591-608.
49. Comas D et al(2004)*Eur J Hum Genet*12: 495-504.
50. Bermisheva M et al(2002)*Mol Biol* 36: 990-1001.
51. Yao YG et al(2002)*Am J Hum Genet* 2002; 70: 635-651.
52. Kivisild T et al(2002)*Mol Biol Evol* 2002; 19: 1737-1751.
53. Kong QP et al(2006)*Hum Mol Genet* 2006; 15: 2076-2086.