论文

中国知名大学及研究院所专栏 复旦大学生命科学学院专辑

新石器时代农业驱动男性人口扩张

王传超^①, 黄韵之^①, 俞雪儿^①, 陈淳^②, 金力^①, 李辉^{①*}^① 复旦大学生命科学学院, 现代人类学教育部重点实验室, 上海 200438;^② 复旦大学文物与博物馆学系, 上海 200433

* 联系人, E-mail: LHCA@fudan.edu.cn

收稿日期: 2016-01-18; 接受日期: 2016-03-01

国家自然科学基金优秀青年科学基金(批准号: 31222030)、教育部科学技术研究项目(批准号: 113022A)和上海市教育委员会曙光计划(批准号: 14SG05)资助

摘要 农业的出现被认为驱动了大范围的群体人口增长。然而, 来自母系线粒体基因组的遗传证据表明, 主要的人群扩张开始于农业发生之前。因此, 农业在最初的人群扩张中扮演的角色仍然存在争议。本文分析了来自千人基因组计划的全球分布的 526 个男性样本的 Y 染色体和线粒体基因组集合。发现大多数主要的父系扩张均可追溯至新石器时代。本研究估测的相应时间的有效群体大小, 有力地证明了农业出现使得男性人口增长了 10~100 倍。在农业发生之前, 男性因为从事狩猎而死亡率较高。农业的发生使得男性从狩猎的危险中解放出来, 死亡率降低, 从而在女性人口扩张之后男性人口也扩张了, 造成了新石器时代人口扩张时间上的男女性别差异。

关键词 群体遗传学, Y 染色体, 人类进化, 农业起源

大约 11000~12000 年前, 在西亚的新月沃土的定居者们驯化了少量野生植物和动物物种, 同时创造了农业。在中国的长江和黄河流域, 人们也发现约 9000 年前水稻(*Oryza sativa*)和谷子(*Setaria italica*)的种植遗迹。在随后的 9000~4000 年前之间的这段时间里, 农业在新几内亚高地(9000~6000 年前)、西非(5000~4000 年前)、墨西哥中央和南美洲北部(5000~4000 年前)和美国东部(4000~3000 年前)独立起源^[1]。在 9000~6000 年前, 农业传播到欧洲, 而不是在那里独立起源^[2]。与狩猎和采集相比, 农业带来的食物产量可以支持远高于前的人群密度, 因此, 主要的人口扩张长期以来被认为发生于农业出现之

后^[3,4]。然而基于全球人群的线粒体(mtDNA)分析发现, 大多数主要的母系扩张发生于末次盛冰期(last glacial maximum, LGM, 约 15000 年前)之后, 却在农业第一次出现之前, 并且人口规模的增加可能是导致农业出现的驱动因素^[5~8]。迄今为止同样有众多关于人口扩张的研究, 一是使用父系的 Y 染色体单核苷酸多态性(single nucleotide polymorphism, SNP)构建系统树, 二是使用迅速突变的短串联重复序列(short tandem repeat, STR)^[9]。不过, 对 SNP 的非随机取样可能造成确定偏倚, 而在 Y 染色体测年中选择 STR 的进化速率还是系谱速率也存在争议, 因为二者的结果可能有将近 3 倍的差异^[9]。随着第二代测序

引用格式: 王传超, 黄韵之, 俞雪儿, 等. 新石器时代农业驱动男性人口扩张. 中国科学: 生命科学, 2016, 46: 468–474

Wang C C, Huang Y Z, Yu X E, et al. Agriculture driving male expansion in neolithic time. Sci Sin Vitae, 2016, 46: 468–474. doi: 10.1360/N052016-00126

技术的出现, 人类 Y 染色体的完全测序在最近才成为可能^[10-14]. 千人基因组计划对超过 500 个男性个体的整条 Y 染色体进行测序^[14], 这提供了绝好的机会, 可以从一个没有确定偏倚的全球群体集合中评估群体规模随时间的变化, 从而有助于解决新石器时代农业与人口最初扩张孰先孰后的长期争端.

1 材料与方法

1.1 群体与样本

群体分析包含了千人基因组计划中来自 3 个非洲群体、5 个欧洲群体、3 个亚洲群体和 3 个美洲群体的 526 个男性个体的测序结果(网络版附表 1)^[14]. 非洲群体样本来自: 美国西南部非洲裔(ASW, 24 例); 肯尼亚韦布耶的卢亚人(LWK, 48 例); 尼日利亚伊巴丹的约鲁巴人(YRI, 43 例). 欧洲群体样本来自: 犹他州北欧与西欧裔居民(CEU, 45 例); 芬兰人(FIN, 35 例); 不列颠人(英格兰与苏格兰)(GBR, 41 例); 西班牙的伊比利亚人(IBS, 7 例); 意大利的托斯卡尼人(TSI, 50 例). 亚洲群体样本来自: 中国北京师范大学的汉族人(CHB, 44 例); 中国湖南与福建的南方汉族人群(CHS, 50 例); 东京的日本人(JPT, 50 例). 美洲群体样本来自: 哥伦比亚麦德林的哥伦比亚人(CLM, 29 例); 加州洛杉矶的墨西哥裔(MXL, 31 例); 波多黎各人(PUR, 28 例). 唯一的 Y 染色体单倍群 A 个体 NA21313 来自肯尼亚 Kinyawa 的马赛人. 欲知详细人群信息, 可访问千人基因组计划主页(<http://www.1000genomes.org>).

1.2 数据分析

从欧洲生物信息学研究所(<ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/>)和美国国家生物技术信息中心(<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/>)可公开连接的 FTP 站点获取了千人基因组计划一期来自 514 个男性个体的 Y 染色体独特区域和线粒体全基因组约 8.9 Mb 序列数据^[14].

对 Y 染色体, 根据 ISOGG 于 2013 年 9 月 6 日版的系统发育树(<http://www.isogg.org/>), 使用修订的 WHY.pl 和 AMY-tree.pl 脚本对单倍群进行分类^[15]. 提取 Y 染色体 SNP 放入伪随机序列中, 使用 PhyML (版本 20120412)计算最大似然树, 100 遍二次抽样估

计置信度^[16]. 构建系统树时去除了杂合子信号. 使用 BEAST 计算系统发育树上每个节点的分歧时间^[17,18], 使用 jmodeltest-2.1.4^[16,19]为随后的贝叶斯 MCMC 分析拟定近似的 DNA 取代模型. 在贝叶斯 MCMC 分析中, 由 BEAST1.8.0 评估得出每一个聚类的年代^[17,18]. 每个 MCMC 样本由 3000 步迭代产生的 3000 万个状态组成, 最初 300 万个状态作为初始状态. 为验证该树的分子钟假设, 使用了基于 GTR (generalised time reversible)模型的 PAML package v4.4^[20]. 比较各个模型后不能拒绝分子钟的零假设($P>0.05$), 因此它可以应用于该项分析. 在有严格分子钟的条件下, 使用 jmodeltest-2.1.4^[16,19]确定的核苷酸置换的 GTR 模型, 单核苷酸置换速率设定为 1×10^{-9} 核苷酸⁻¹ 年⁻¹. 使用程序 Tracer v1.5.0 分析每次循环参数评估的独立性和 MCMC 链的稳定性^[17,18]. 每个人群的 BSP (bayesian skyline plot)也是使用 BEAST v1.8.0 和 Tracer v1.5.0 所作, 基于如上的相似设定.

分析线粒体数据时, 依据 2012 年 9 月 30 日版本的 mtDNA PhyloTree Build 15 (<http://www.phylotree.org/>), 使用 MitoTools 对单倍群进行分类^[21]. 为验证该树的分子钟假设, 使用了基于 HKY+G 模型的 PAML package v4.4^[20]. 不能拒绝分子钟的零假设($P>0.05$). 在程序 BEAST v1.8.0 中使用 MCMC 抽样方法, 使用线粒体 DNA 编码区(位点 577~16023)产生 BSP^[17,18]. 每个 MCMC 样本由 3000 步迭代产生的 3000 万个状态组成, 最初 300 万个状态作为初始状态. 在有严格分子钟的条件下, 假定先验的置换速率服从均值为 2.038×10^{-8} 置换 位点⁻¹ 年⁻¹, 标准差为 2.064×10^{-9} 置换 位点⁻¹ 年⁻¹ 的正态分布^[7,8]. 使用程序 Tracer v1.5.0 分析每次循环参数评估的独立性和 MCMC 链的稳定性^[17,18].

1.3 人群增长速率计算

使用 BEAST 生成的 BSP 以计算人口年增长率. 每张 BSP 由 100 个平滑的数据点组成, 约有 300~800 个年段. 在即将发生人口增长之前的时段, 最初的人口规模设定为最小的人口规模. 本文同样评估了数据中人口增长最快时的增长速率(表 1). 在这项分析中, 根据文献[5]的建议选择了指数增长方程: $r = \ln(N_t/N_0)/t$. 其中“ r ”代表每年的人口增长速率, “ N_0 ”是最初的人口规模, t 是增长开始后的时间长度.

表 1 依据 BSP 计算的人口增长速率

	农业起源 时间(年)	群体	Y 染色体				线粒体			
			增长时段 (年)	增长速率 (%)	最快增长时 段(年)	该时段最大 增长速率(%)	增长时段 (年)	增长速 率(%)	最快增长时段 (年)	该时段最大 增长速率 (%)
非洲	5000~4000	ASW	5700~470	0.0472%	3820~3340	0.1842%	25000~7000	0.0052%	20380~19600	0.0111%
		LWK	6480~3990, 2990~500	0.0196%, 0.1247%	2000~1500	0.4004%	29600~9600	0.0013%	22400~21600	0.0025%
		YRI	6200~340	0.0678%	3100~2760	0.2626%	12100~0	0.0178%	7290~6480	0.0385%
欧洲	9000~6000	CEU	4120~310	0.1002%	2220~1900	0.4358%	13300~650	0.0258%	10400~10080	0.0967%
		FIN	4480~750	0.0946%	2240~1870	0.4812%	27200~3300	0.0061%	10200~9920	0.0177%
		GBR	4600~1060	0.1300%	3180~2830	0.4801%	12400~1500	0.0207%	8000~7700	0.0722%
		IBS	6060~2240	0.0362%	4490~4260	0.0967%	15110~2350	0.0088%	12500~12250	0.0236%
		TSI	3480~870	0.1405%	2610~2180	0.5424%	14700~1670	0.0318%	11030~10700	0.1616%
亚洲	9000	CHB	36300~8890, 8520~5560, 5180~370	0.0025%, 0.0185%, 0.0488%	2960~2590	0.2485%	35500~720	0.0141%	33750~33390, 11500~11130	0.0619%, 0.0328%
		CHS	6460~650	0.0432%	1940~1610	0.3434%	14810~3170	0.0133%	34210~33860, 10580~10230	0.0556%, 0.0471%
		JPT	4200~840	0.0896%	2520~2100	0.5368%	6600~350	0.0148%	33350~33000, 4170~3820	0.0366%, 0.0295%
美洲	5000~4000	CLM	4240~1410	0.1071%	3300~2830	0.3121%	70000~29100, 1500~0	0.0022%, 0.0278%	1490~750	0.0281%
		MXL	6400~460	0.1554%	5030~4570	0.3021%	11150~3500	0.0371%	9240~8920	0.1684%
		PUR	3900~430	0.05117%	3030~2600	0.1206%	55000~23400, 10330~6200	0.0016%, 0.0057%	36490~35800	0.0305%

2 结果

分析了千人基因组计划提供的来自非洲、欧洲、亚洲和美洲的共 526 个男性个体 Y 染色体的独特区域和线粒体全基因组上的大约 8.9 Mb 对的测序结果(网络版附表 1)^[14].

使用 Y 染色体 SNP 构建了最大似然树(图 1). 该树包含来自单倍群 A, B, C, D, E, G, I, J, N, O, Q, R 和 T 的样本, 因而很好地代表了基于地理位置的父源谱系. 为推断扩张时间, 使用基于 1×10^{-9} 置换 碱基⁻¹ 年⁻¹ 的常数突变速率的贝叶斯方法计算了整棵树上每个分歧事件的发生时刻^[22]. 主要的单倍群之中, 大部分(C, I, J, O, E 和 R)在早于 LGM 的 25000~30000 年前产生. 几乎主要的单倍群的所有初级亚群都在晚于 LGM 而早于新石器时代的时间内(10000~15000 年前)分化. 在那些初级亚群里, 较长的内部分支长度暗示了旧石器时代晚期阶段的低人口增长和频繁的瓶颈效应期. 然而, 新石器时代发生了新兴谱系的大爆发(图 1). 本文还识别出系统发育树里的许多星状结构, 如 I1a, O2b, O3a2c1 (O α 和 O β), O3a1c1 (O γ),

R1a, R1b1a2a1, E1b1a1a1g 和 E1b1a1a1f1a (图 1). 星状分支中对应其他单倍群的短内部分支长度是这些分支最近人口迅速扩增的有力证据. 所有星状谱系也都在新石器时代(5000~8000 年前)汇合. 最近发表的全 Y 染色体测序的文章也揭示了 R1b 与 O $\alpha\beta\gamma$ 在新石器时代的扩张^[10,13]. 这些在新石器时代扩张的分支占有现今人口的很大比例: R1b1a2a1 在西欧父源谱系中为主流, 占总人口的比例超过 1/2^[23]; O α , O β 与 O γ 在中国人中的基因频率达到 40%^[13,24]; O2b 在日本与韩国占比超过 20%^[25]; E1b1a1a1g 与 E1b1a1a1f1a 组成了非洲撒哈拉以南的大部分地区超过 80%的人口^[26]. 因此, 对谱系聚合的时间估算表明, 父系人口增长可能主要归结于新石器时代的人群扩张.

随后通过联合 BSP^[17,18]评估相应时间的有效群体大小来推测从 LGM 到农业出现的这段时间里人口规模的变化(图 2). 所有图表均证明存在人口增长, 也反映出父系与母系人口历史的迥然差异. 推断在 LGM 之后(12000~8000 年前)大部分欧洲、亚洲母源谱系经历了迅速、汹涌的人口增长, 而在随后一段时间内(8000~7000 年前)增长极为缓慢. 非洲的 YRI 与

图1 人类Y染色体的系统发育树

该树由千人基因组计划的526个样本测序结果构建, 分支长度与该分支上SNP位点数目成正比。

标红的数字代表分化时间(按年计算)和节点的95%置信区间

美洲的MXL展现出与欧洲、亚洲人群相似的人口增长模式。若从父系追踪, 所有的人群都在6500~2000年前经历了主要的明显扩张, 人口规模提高了10~100倍。除了CHB与JPT在13000年前轻微增加, 全球人群最初的男性扩张都开始于新石器时代, 而不是LGM后立即发生。该类男性人口增加阶段的估测时间与各个地区农业起源的考古学证据完美地吻合(表1)。随后计算了人口年增长率^[5]。对大部分欧洲、亚洲和MXL母源谱系而言, 最高的人口年增长率出现在11000~9000年前(图2和表1), 时间与LGM后开始的一段稳定的全球温暖期相一致。同样发现35000~33000年前亚洲群体中母系人口急剧增加, 而这可能反映出人们最初定居这片大陆后的人口爆炸。然而, 全球男性人口增长最快的时段却是3300~1500年前, 比农业起源要晚1000~3000年(图2和表1)。在此期间, 农业开始走向成熟与集约, 因而可以支撑迅猛增长的人口。此外, 除亚洲之外的所有BSP(图2)都显示出农业变迁的明显瓶颈, 这可能是因为男性主导的人群扩散才导致农业扩散。在著名的欧洲案例中, 新石器时代时, 来自中东的农人扩散并遍及欧

洲, 并逐渐取代土生的狩猎采集者^[27,28]。

能否对时间进行合理估算的关键是突变率是否准确。现有Y染色体突变率主要是通过人类与黑猩猩的分歧时间、测序大跨度家系和基于奠基迁徙事件校正等来得到的。前期已使用千人基因组数据, 以考古学上的现代人走出非洲的时间对应Y染色体CF单倍群的共祖时间来选取最适突变率, 发现由大跨度家系和以进入美洲时间为基准得出的突变率较可靠^[29], 因此本文选用由家系得出的 1×10^{-9} 核苷酸⁻¹年⁻¹作为突变率。另外, 千人基因组发布了第三期数据^[30], 对其进行后续分析, 可进一步验证本文结论。

大部分父源谱系的聚合时间与BSP中观察到的人口增长时段相一致, 这表明, 最初的男性人口扩张开始于新石器时代, 其原因可能是农业的起源与传播。与狩猎、采集相比, 农业提供了更稳定的食物供给, 从而提高了人群繁殖力与婴儿存活率。更为重要的是, 因为农业使男性远离危险的狩猎活动, 所以狩猎相关的男性死亡率的降低可能是新石器时代性别偏向的人口扩张的主因。

图2 Y染色体与线粒体的BSP显示了相应时间的男性与女性的有效群体大小

Y轴是有效大小和世代时间的乘积, X轴是以年为单位距今的时间. IBS 由于样本规模太小而不在图中. 群体缩写: ASW: 美国西南部非洲裔; LWK: 肯尼亚韦布耶的卢亚人; YRI: 尼日利亚伊巴丹的约鲁巴人; CEU: 犹他州北欧与西欧裔居民; FIN: 芬兰人; GBR: 英格兰与苏格兰人; IBS: 西班牙的伊比利亚人; TSI: 意大利的托斯卡尼人; CHB: 中国北京师范大学的汉族人; CHS: 中国湖南与福建的南方汉族人群; JPT: 东京的日本人; CLM: 哥伦比亚麦德林的哥伦比亚人; MXL: 加州洛杉矶的墨西哥裔; PUR: 波多黎各人

参考文献

- 1 Diamond J, Bellwood P. Farmers and their languages: the first expansions. *Science*, 2003, 300: 597–603
- 2 Pinhasi R, Fort J, Ammerman A J. Tracing the origin and spread of agriculture in Europe. *PLoS Biol*, 2005, 3: e410
- 3 Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*, 2002, 418: 700–707
- 4 Jobling M, Hurles M, Tyler-Smith C. *Human Evolutionary Genetics: Agricultural Expansions*. New York: Garland Publishing, 2004. 299–306
- 5 Gignoux C R, Henn B M, Mountain J L. Rapid, global demographic expansions after the origins of agriculture. *Proc Natl Acad Sci USA*, 2011, 108: 6044–6049
- 6 Atkinson Q D, Gray R D, Drummond A J. MtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol Biol Evol*, 2008, 25: 468–474
- 7 Zheng H X, Yan S, Qin Z D, et al. Major population expansion of East Asians began before Neolithic Time: evidence of mtDNA genomes. *PLoS One*, 2011, 6: e25835
- 8 Zheng H X, Yan S, Jin L, et al. MtDNA analysis of global populations support that major population expansions began before Neolithic Time. *Sci Rep*, 2012, 2: 745
- 9 Wang C C, Li H. Inferring human history in East Asia from Y chromosomes. *Invest genet*, 2013, 4: 11
- 10 Wei W, Ayub Q, Chen Y, et al. A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res*, 2013, 23: 388–395
- 11 Poznik G D, Henn B M, Yee M C, et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science*, 2013, 341: 562–565
- 12 Francalacci P, Morelli L, Angius A, et al. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-Chromosome phylogeny. *Science*, 2013, 341: 565–569
- 13 Yan S, Wang C C, Zheng H X, et al. Y Chromosomes of 40% Chinese are descendants of three Neolithic super-grandfathers. *PLoS One*, 2014, 9: e105691
- 14 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 2012, 491: 56–65
- 15 Van Geystelen A, Decorte R, Larmuseau M H. AMY-tree: an algorithm to use whole genome SNP calling for Y chromosomal phylogenetic applications. *BMC Genomics*, 2013, 14: 101
- 16 Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, 2003, 52: 696–704
- 17 Drummond A J, Nicholls G K, Rodrigo A G, et al. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, 2002, 161: 1307–1320
- 18 Drummond A J, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*, 2007, 7: 214
- 19 Darriba D, Taboada G L, Posada D, et al. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*, 2012, 9: 772
- 20 Yang Z H. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, 2007, 24: 1586–1591
- 21 Fan L, Yao Y G. An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. *Mitochondrion*, 2013, 13: 360–363
- 22 Xue Y, Wang Q, Long Q, et al. Human Y chromosome base-substitution mutation rate measured by Direct Sequencing in a deep-rooting pedigree. *Curr Biol*, 2009, 19: 1453–1457
- 23 Balareshque P, Bowden G R, Adams S M, et al. A predominantly Neolithic origin for European paternal lineages. *PLoS Biol*, 2010, 8: e1000285
- 24 Wang C C, Yan S, Qin Z D, et al. Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *J Syst Evol*, 2013, 51: 280–286
- 25 Ding Q L, Wang C C, Farina S E, et al. Mapping human genetic diversity on the Japanese archipelago. *Adv Anthropol*, 2011, 1: 19–25
- 26 Sims L M, Garvey D, Ballantyne J, et al. Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Hum Mutat*, 2007, 28: 97
- 27 Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic maps of human gene frequencies in Europeans. *Science*, 1978, 201: 786
- 28 Skoglund P, Malmström H, Raghavan M, et al. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, 2012, 336: 466–469
- 29 Wang C C, Gilbert M T, Jin L, et al. Evaluating the Y chromosomal timescale in human demographic and lineage dating. *Invest Genet*, 2014, 5: 1
- 30 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 2015, 526: 68–74

Agriculture Driving Male Expansion in Neolithic Time

WANG Chuan-Chao¹, HUANG YunZhi¹, YU Xue-Er¹, CHEN Chun², JIN Li¹ & LI Hui¹

1 MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China;

2 Department of Cultural Heritage, Fudan University, Shanghai 200433, China

The emergence of agriculture is suggested to have driven extensive human population growth. However, genetic evidence from maternal mitochondrial genomes suggests major population expansions began before the emergence of agriculture. Therefore, the role of agriculture in initial population expansions remains controversial. Here, we analyzed a set of globally distributed whole Y chromosome and mitochondrial genomes of 526 male samples from 1000 Genome Project. We found that most major paternal lineage expansions coalesced in Neolithic Time. The estimated effective population sizes through time revealed strong evidence for 10- to 100-fold increase in population growth of males with the advent of agriculture. This sex-biased Neolithic expansion might result from the reduction in hunting-related mortality of males.

population genetics, Y chromosome, human evolution, agriculture origin

doi: 10.1360/N052016-00126