

# 基因视野中的中华民族共同体上古起源

李辉



李辉  
复旦大学现代人类学教育部重点实验室主任  
复旦大学大同市中华民族寻根工程  
研究院院长

中华民族的起源即中华文明的上古史部分，因时代久远而材料稀少，从不同的学科视角往往看到巨大差异，尤其因无文字材料，而科学研究极其缺乏，所以在学术界争议颇多。事物的真相只有一个，如有争议，最多只有一方正确，甚至往往都不对。所以目前的上古史认知必然存在很多误区。我们如何走出误区，接近真相，无非是摆事实讲道理，也就是整理科学逻辑和科学证据。上古史的科学逻辑是人类进化和社会发展的客观规律，科学证据必须综合遗传学、考古学、语言学、民族学、神话学等所有相关领域和相关数据。把所有证据都通过逻辑串起来，没有不可解释的证据，这样的假说才能称为科学假说，才可能最接近真相。

## 人类命运共同体中的中华民族

对于中华民族起源的认识，要基于全球现代人起源的科学研究进展的基础上。全世界的现代人属于智人物种的一个亚种。从基因、体质、语言、考古等方面的证据综合分析证实，现代人亚种 20 万年前源于非洲，约 7 万年前因为多峇巨灾的推动而扩散到了非洲之外，在世界各地适应不同地理气候条件而陆续形成了 8 个地理种。全球现代人的基因都非常接近，构成一个紧密的人类命运共同体。世界范围内文明与民族起源，都是 1.2 万年前末次盛冰期结束以后的事件，与 7 万年前现代人走出非洲的自然史无关。

中华民族是 7 万年前出非洲人群的分支之一。大约 4 万

年前进入东亚大地，渐渐萌发出中华文明。目前对于中华民族和文明的起源虽然研究不断深入，但在民众中甚至在相关学界还存在两种极其错误的观点：1) 文明西来说，有观点认为中华文明起源于古巴比伦，甚至古埃及，很晚的时候才从西北进入中国；2) 中华民族多源说，认为中国的民族有很多独立起源，各民族是不同来源的，中华民族是近代构建的，这一观点在社会科学界几乎成为主流。实际上这两个观点既缺乏科学证据，又严重影响了民族团结。

大量研究证明，中华文明的主体核心都是中国本土孕育，并且传承有序的。与其他文明发生的交流互鉴，只是中华文明的部分补充。中华民族的同源性始终是主流，各个民族的形成伴随着不断的重组，因此形成了“你中有我”的遗传基因与语言文化形态。外来的成分从来没有在中华民族形成过程中占过主流，都是中华民族的少量补充血液。

1990 年代的分子人类学研究定出了东亚人类起源的大框架。基因组数据说明东亚人主要源于非洲晚期智人，极少部分来自欧亚大陆的早期智人（尼人和丹人）。2010 年代以来，复旦大学牵头的团队详细推演了现代人进入东亚，以及各民族类群起源迁徙的时间和路线，开拓了东亚人群史前人口史精细分析的新领域（李辉、金力，《Y 染色体与东亚族群演化》，上海科学技术出版社）。

通过采集东亚数十万样本基因组，分析其中 Y 染色体、线粒体等的多态性后，我们发现东亚现代人是由不同时期进



图1 现代人起源模式图。图中字母代表Y染色体单倍群类型，数字表示分支距今万年数，人像为八个地理种代表

入东亚的至少两批现代人混合而成的，呈“两阶段两路线”。第一批约7万年前出非洲后沿海岸线东进，约5万年前进入东亚，在现代东亚基因库中比例较小，而在澳洲土著中比例较高。第二批约5万年前从西亚扩张并于约4万年前进入东亚，形成了现代东亚人的主体。两批人进入东亚的路线分别有东西两条主线。

第一批，部分人群约5万年前从缅甸溯着青藏高原边缘的河谷进入东亚西部。另一部分从中南半岛沿着冰期海岸线进入东亚东侧，并一直北上抵达东亚北部。

第二批，4万年前进入东亚的两条主线分别是滇川、两广，其后代构成了现代东亚人口的主体。其中滇川的人群涉及孟高棉、苗瑶、汉藏三个族群的起源。从基因谱系看，苗瑶和汉藏族群的早期人口都源于孟高棉族群的祖先。两广主线进入中国的人群奠定了现在侗傣-南岛族群的主体人口，成为中华民族另一支来源。在迁徙过程中，东南亚丛林气候形成遗传过滤效应，使得东亚人群的遗传、体质和生理病理特征得以特化。

### 东亚族群起源于原始人群的聚合

早期狩猎采集人群并未形成民族，而是大量文化关联疏松的散在部落。我们深入研究了“新石器转型期”的遗传结构变化，发现民族是冰期结束农业起源以后，由散在人群向



图2 4万年内的人群扩散与8000年内的民族聚合模式图。O1与O2是中国主要的两种Y染色体单倍群，O2之下有三个主要的扩张节点，以希腊字母 $\alpha$ 、 $\beta$ 、 $\gamma$ 命名。五个早期族群的来源自西南顺时针分别是高庙、仰韶、红山、龙山、良渚等考古文化。数字为早期政权建立的距今年数

农业核心聚合形成的。民族是人口与文化稳定积累的产物。文化的稳定积累需要建立在足够的人口基础上，而人口的大规模增长依赖于农业生产。冰期的寒冷气候不支持农业的起源，所以只有冰期结束以后，生物资源大量增长，才促进了人口初步增长。初步增长的人口发明了农业并稳定地传承，从而进一步刺激人口大量增长，在农业核心上聚合形成了原始民族群体。充分的剩余时间使得新石器工具出现，并渐渐孕育了具有相对完善的社会规则的早期国家雏形。

按照语言谱系，中国现代各个民族可以归为若干语系，即汉藏、苗瑶、侗傣（新名仡佬，Kra-Tai）、阿尔泰（满蒙为主），以及边缘的南亚、南岛、印欧等。苗瑶和侗傣因为与汉藏关系紧密，也会被归入汉藏中。这些语系族群中，中国的主体族系——汉藏语系族群，具有最高多样性的Y染色体O-M134单倍群。通过古人骨DNA分析发现，此谱系起源于海河流域粟作农业形成的早期磁山-裴李岗文化。最早的驯化小米发现于北京郊区永定河边的东胡林遗址。汉语族与藏缅语族经历了两次分化：1) 7000多年前南下黄河流域的人群后来形成仰韶文化，成为藏缅语族的源头；约6200年前部分人群北上西辽河流域，形成红山文化，成为汉语族的源头。2) 约5300年前红山文化人群南下，迫使仰韶文化部分人群西迁成为藏缅语族，而中原混合的人群成为初具规模的汉语族（李辉、金雯俐，《人类起源与迁徙之谜》，上海科技教育出版社）。

处于汉藏族群北侧的满蒙族群，虽然不以农业为主，但是在早期汉族先民北上的影响下，聚合形成了文化共同体。黑龙江的小南山文化等可能与其有关。但东北地区的早期文化多样性很高，存在多个族群系统的起源，如芬兰-乌拉尔语系人群可能起源于辽西赵宝沟文化。赵宝沟文化被红山文化覆盖，或是芬乌族群北迁的原因。而汉族也在从桑干河流域南来人群和遗留的赵宝沟人群的混合中形成，所以现代汉语体现出明显的藏缅语和芬乌语混合的形态，现代汉族 Y 染色体也兼有藏缅的 O 系和芬乌的 N 系。

因为气候变冷，5 300 年前汉族人群南下，影响到了湖广的苗瑶祖先群体和江浙的侗傣祖先群体，并发生了大规模人口融合。广布于台湾、东南亚和太平洋的南岛语系人群与侗傣语系人群以 Y 染色体 O1-M119 谱系为特征，精细进化分析发现，他们同源于江浙新石器时代人群。南岛祖先离开江浙约为 5 900 年前，当地马家浜文化结束，其上层建筑在来自长江以北的大汶口文化驱使下南迁。这为太平洋岛屿数亿人口找到了源头。而 5 300 年前汉族祖先的到来，使马家浜之后的崧泽文化终结，在当地混合形成了良渚文化，是侗傣语系壮侗语族人群的起源。而崧泽文化上层的西迁可能形成了侗傣语系仡黎语族群体。崧泽同时南方最发达的是凌家滩文化，很可能是一个原始政权中心，有苗瑶文化的背景，这可能与仡黎语族兼有侗傣语和苗瑶语特征有关。而壮侗语族兼有侗傣语和汉语的特征，也可能与良渚文化的混合起源有关。

大约 5 300 年前，中国境内的五个考古区系大多发生了文化变革。东北的红山文化南下。中原的仰韶文化大体西迁，西北的大地湾文化变成马家窑文化。南方中心的凌家滩文化被摧毁。江浙的崧泽文化灭亡，良渚文化开始。湖广的大溪文化灭亡，屈家岭文化开始。只有山东的大汶口文化基本稳定。这说明中国发生了一场全国范围内的巨大政治变革，来自东北的力量几乎影响了全国。此变革使中国人群的遗传结构发生大混合。这可能与传说中蓟地轩辕黄帝的兴起有关，也与汉族成为中华民族融合核心有关。

以往认为民族的形成主要是不断分化的过程。通过遗传谱系分析我们发现，各个民族之间共享的基因类型非常普遍，而这些共享类型的分化时间大多在新石器时代的民族起源期。所以证实民族类群主要是聚合形成的，族群之间的交流融合多于分化，中国各民族的形成过程伴随着不断的人口重组。

## 聚合与派生中形成的中华民族共同体

文明是民族文化的积累和升华。中华文明的形成恰恰是中华民族共同体形成的产物，或者说两者互为因果，协同发生。从遗传学、语言学、考古学、历史学、神话学、民俗学等等领域综合证据分析，可以清楚地看到中华文明的主体是完全根植于中国，是本土自生的。冰期结束过程中，中国南方就开始了世界最早的农业萌芽，长江中游的人群开始驯化水稻。江浙和燕云的农业萌芽也随后迅速萌发，这奠定了中华文明的基础。同时，中华文明海纳百川，在数千年的发展与其他文明交流互鉴，很好地补充和促进了中华文明的发展。但如果因为部分的交流互鉴而提出所谓“中华文明西来说”则是非常荒谬的，各个学科都有明确的证据予以反驳。一是遗传学证据非常清晰地呈现出中国人群的主体基因谱系都是本土起源的，数万年前已经进入中国，5 000 ~ 7 000 年前就完成了文明起源过程中的家族谱系扩张；二是经过谱系分析年代计算，现代人群中累计少于 8% 的各种外来成分，都是晚于距今 4 000 年的。这就证明，外来人口进入之前，中华文明就起源了。而文明的起源点，应该就是父系谱系的扩张原点，因为父权社会中财富积累和继承决定了父系谱系的扩张。我们通过古代人骨 DNA 检测，发现最早的三个扩张原点是湖南和河北的农业起源点，证明中华文明起源于本土最早的农业人群，而不可能是西亚文明的衍生。

中华民族的主体起源于中国，那么中国境内数个族群是否完全独立起源的呢？现在学界流行一种由“多元一体”理论过度异化而来的“中华民族多源说”。多元一体理论应该指的是农业和文化的起源有多元素，而后发展成一体化的民族集团。并不是民族有多个起源，更不是各个族群起源不同。对于“中华民族”和“中华民族共同体”的认识，这里有一个“名”和“实”谁先谁后的问题。社会团体可以先有“名”再构建“实”，而自然事物基本上是先有“实”体再据此赋予“名”称。我们有大量遗传学数据可以证明中华民族“实”先于“名”。很多研究者认为，“中华民族”这一“名”，既然是民国初期提出的，是基于反侵略反殖民的需求构建起来的，因此那以后才有了中华民族的“实”体，甚至有人认为中华民族的实体至今仍在构建中，这是非常唯心主义的观点。包括部分学者认为汉族叫作“汉”族，必然是汉朝以后的，历史考证是晋代外族压迫才使中原人群自称“汉人”，因此认为汉族是晋代形成的。这就把主观认识的“名”凌驾于客观存在的“实”之上了。从历史唯物主义角度看，汉族是一个有稳定的文化和遗传传承的族群，无论他被叫作华人、夏人、唐人或者其他什么，只要稳定地传承发展汉语，以天

人合一等世界观为族群文化标记，以三皇为共同的血缘认同，这一群体就是汉族。所以无论是秦始皇，还是周文王、尧、舜、禹，实际上都属于汉族，不然历史逻辑就荒诞了。

中华民族共同体提出也是基于这一实体的客观存在。我们从人类学角度很容易证实这一点。一方面，中国各民族之间有大量共享谱系和同源谱系，远大于域外人群，中华民族有共同的遗传背景。中国目前的 56 个民族都与汉族有共享遗传世系，例如藏族、蒙古族、维吾尔族都有高频的汉族主体 Y 染色体 O-M134 类型。另一方面，中国各语系民族都是较晚历史中分化形成的，在新石器时代早期转型中，没有民族概念的散在采集狩猎人群渐渐聚合成了中华民族的雏形，没有现代式的民族分化和文化差距。后来部分人群迁到了边缘地区，渐渐形成少数民族。而留在中部的主体人群融合成了汉族，这才有了民族的分化。所以中华民族是先融合再分化而形成的。从人群的角度分析，中华民族不是多元的，而是一元的。考古中看到的文化区系只是次要的地方特色，是民族形成之前的族群初步聚合。中华民族因为先有“实”体，才能被命“名”，实在名前，是被广泛接受的基础。

## 中华民族的超级遗传祖先

上古族群结构精细研究的基础，在于对中国人群父系 Y 染色体谱系的精细研究和遗传精准估年方法的建立。利用这一技术，复旦团队在中国人群的大规模分析中，发现中国人群 Y 染色体非重组区的近 2 万个新的 SNP 位点。用这一系列高质量序列重构了中国人群相关的 Y 染色体 C、D、N 和 O 的精细演化树，并重新计算了谱系分化时间。在演化树上观察到，距今 6 000 年前后，Y 染色体 O $\gamma$ -F11、O $\beta$ -F46、O $\alpha$ -F5 三大个人节点出现了迅速扩张，在数代内这三人分别产生上千后代，其直系后裔在现代中国人中超过 40%，成为中国人的三个超级遗传祖先。三大节点之后，还有三个略小的迅速扩张节点。这六个祖先的直系后代占现在中国 Y 染色体的 70% 以上。这应该是文明早期国家成型和领袖产生的一种迹象，也就是中国民族形成的第一个阶段。

超级祖先的时间和地点与古史传说中的帝王谱系、考古发现的早期大型陵墓都吻合。三大祖先的年代估算分别为距今约 6 800 年、6 500 年、5 400 年。有些学者认为不应该把传说的上古人物与遗传谱系和考古遗址对号入座，这种观点并不符合科学逻辑。首先，上古传说虽然因为流传太久而出现版本多样性，但并不能简单否定其历史背景的真实性，而是需要科学验证，与遗传学和考古学对照就是一种科学验证。

其次，历史真相是必然能够从各个角度检验都达到对应的。不能“对号入座”，就不可能是历史真实，可能是某些学科角度的观察发生了扭曲。实际上我们观察到的中华民族上古史，在遗传学、神话学、考古学、语言学、民族学各个领域都能完美对应。尤其是三个最早的超级祖先，是现代汉族群体的核心成分，也分别是中国其他各个族群的主要类型，他们的起源发现于不同的考古区系。

第一，距今约 6 800 年是中国已发现的第一个城市始建的年代，在稻作起源的湖南沅江流域澧县发现了城头山古城。我们突破了南方人骨腐败程度高的技术难点，检测了城中年代最早规格最高的墓葬主人的基因，发现其 Y 染色体正是中国三大谱系中的第一个 O $\gamma$ -F11 的早期类型。结合其所属的高庙文化出现的大量“凤鸟”“八角星”图案等文化特征，这一谱系与传说中的中华人文始祖“伏羲”高度吻合。

第二，距今约 6 500 年，在河南濮阳西水坡遗址留下一个巨大的陵墓，属于仰韶文化鼎盛时期。这一墓葬中围绕墓主人用蚌壳排布着青龙、白虎、北斗、南方七宿等天象，象征着墓主人的崇高地位，疑似传说中的“神农”。仰韶文化大量高等级墓葬人骨基因都属于 O $\beta$  谱系。

第三，距今约 5 400 年，在辽宁、内蒙古交界处建平县的牛河梁山岗上，开始建造金字塔形积石冢、神庙和祭坛，已发现的 16 个积石冢排布成轩辕星座的造型。我们检测了最早而且最大的积石冢墓主遗骸 DNA，发现其为 O $\alpha$  谱系的源头。这一遗址开始于红山文化的鼎盛期，出现大量龙凤玉雕，与传说中的“轩辕黄帝”高度吻合。



图3 藏族的吉祥八宝图案在仰韶彩陶中都能找到可能的源头

更重要的是，这三个考古区系和遗传谱系及其代表的族群文化，与现代民族群体的文化也明显有传承关系。高庙文化的陶器上出现大量的凤鸟、八角星、建木、夔龙造型。这些造型都是典型的苗瑶族群文化，也大量融入汉族文化中。高庙-大溪遗址人骨中高频检出的 O-M7 类型恰恰是现代苗瑶族群的特征。

仰韶文化的彩陶纹样中最典型的是太阳、双鱼、螺旋纹、网纹、太极、花朵等，这是现在的藏缅族群的典型文化符号。例如藏族的传统图案吉祥八宝，在仰韶文化彩陶中都能找到源头，藏族 Y 染色体最高频的 O 和 D，分别可以追溯到仰韶居民和大地湾居民。藏缅的圆圈舞习俗，在仰韶后续的马家窑文化彩陶中也清晰绘制着。所以有理由相信，仰韶-马家窑文化是藏缅族群的源头。同时太极、中国结等文化也完全融入了汉文化中。牡丹也长期成为中国人最喜欢的国花。

汉族虽然融合了大量远古族群的文化因素，但其文化核心应该来自红山文化。红山文化以大量玉器的使用为特征，玉器代表造型是龙和风、“道人”、玉瑁冠冕、玉螭玉龟、熊头燕子、终葵斫刀……这些是最典型的汉文化特征。红山文化分布于涿鹿之北，与传说中的轩辕黄帝部族发源地一致。轩辕黄帝无疑是汉族公认的最重要的祖先。所以汉族主体起源于东北是遗传学、语言学、考古学和神话学都可以一致对应的。

以上三个文化和族群的扩张是中华民族最早的起源。此外齐鲁的后李-北辛-大汶口区系和江浙的上山-马家浜-崧泽-良渚区系也是关键的源头。在红山扩张之后，约 4 600 年前大汶口文化开始西扩，与中原文化融合形成龙山文化，可能代表了少昊时代的开启，成为汉族人口的主要来源之一。从遗传谱系上看，大汶口的大量谱系来自高庙文化，如果高庙文化属于太昊伏羲氏，那么少昊与太昊有传承关系也是合理的。一般认为大汶口文化派生出古东夷族，后来完全融入汉族。

良渚文化于约 4 500 年前北扩，大汶口文化遗址中出现了良渚贵族的遗存。这可能就是传说中颛顼时代的开始。约 4 400 年前，颛顼时代结束，族群分两个方向迁徙，中原部分迁到西北覆盖马家窑文化形成齐家文化，江浙部分的上层建筑南迁到两广发展成侗傩族群。良渚文化遗骸和现代侗傩族群中大量检出的 Y 染色体 O1-M119 类型证实这一遗传关系。齐家与良渚一样用大量的玉琮玉璧，而疑似齐家统治中心的陕北石峁古城甚至出土了大量直接来自良渚的精美玉器。石峁文化大约 4 000 年前的南下可能就是夏朝的起源，

也是汉族人群的主要构成成分之一。

这一系列研究，拨开了中国上古史的迷雾，可以证实中华文明与其他三大文明一样有六七千年历史，也反驳了否定中国古史、否定三皇五帝真实性的观点，为重建中国古史奠定科学基础。基于此，中华民族共同体的起源也就非常清晰了。中国的五个农业区形成了五个考古文化区系，也形成了数个语系的源头。这五个区系陆续兴起和扩张，构建原始政权，可能对应了传说中最早的上古帝王：伏羲、神农、轩辕、少昊、颛顼。此间部分人群迁到偏远地区，渐渐成为少数民族甚至外国民族，而大部分人口融合成为汉族。在历史发展中，中华民族继续不断地发生语言文化和遗传基因的交流。这就是为何中华各族之间存在大量的语言、文化、基因的共享性。可以说中国各民族起源于上古的中华民族共同体。

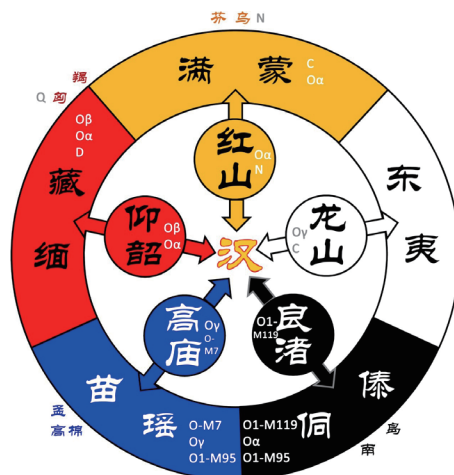


图4 中华民族共同体形成模式——融合与派生。字母为Y染色体类型

历史学的研究，以往只能通过传世文献的梳理、考古发现的比较研究来推进，其模式为“文本→文本”，解析精度和客观性都有限。而精细遗传学分析的介入，创建了“文本→科学证据→文本”的历史人类学新模式，使得历史分析的精度大大提高。进一步打破社会科学与自然科学之间的界限，这正是“中华民族共同体”研究的重要守正创新。

本文根据笔者在上海市科学技术普及及志愿者协会主办的“海上科普论坛”上的报告撰写而成