



中华民族共同体：藏在染色体里的 民族融合史诗

李 辉



李辉，复旦大学生命科学学院教授、博士生导师，现代人类学教育部重点实验室第二届主任，民族研究院副院长，大同市中华民族寻根工程研究院院长，亚洲人文与自然研究院（澳门）院士，中国人类学民族学研究会常任理事，中华民族共同研究会常务理事，上海人类学学会常务副会长；主要研究现代人类学，从脱氧核糖核酸（DNA）探索人类起源与文明肇始，追寻中华民族与中华文化的生物学起源；被《科学》（Science）以“复活传奇”为题专版报道；曾应邀在联合国总部作报告；在《自然》（Nature）、《科学》、美国国家科学院院刊（PNAS）等期刊发表论文 300 多篇；出版《Y 染色体与东亚族群演化》《人类起源和迁徙之谜》《茶道经》《傜俵话：世界上元音最多的语言》《自由而无用的灵魂》等科学、哲学、史学、文学专著数十部。



中华民族共同体虽然是当代提出的名称，但其实早在中华文明起源之初就萌发了。末次冰期结束以后，中国出现了五个农业区，聚合成了五个考古区系：湖广、中原、辽西、齐鲁、江浙。这五个考古区系陆续兴起并产生原始政权，在后来的群体记忆中可能演化成了伏羲、神农、黄帝、少昊、颡顼的传说。遗址骨骼和当代族群的基因分析证明中国各语系族群与不同考古文化高度相关，各族之间也有大量遗传类型和词汇语素的共享。综合遗传学、考古学、语言学、神话学、民族学等领域的证据，我们认为发源于这五个区系的上古政权的原始族群在不断互动中形成中华民族共同体，其中部分人口迁到边远地区成为不同语系的少数民族，中部地区的多数人口融合形成汉族。所以中国各民族实际上源于上古的中华民族共同体。中华民族共同体是先有实体，后有名称的。

中华民族的起源，也就是中华文明的上古史部分，因为时代久远而材料稀少，从不同的学科视角往往看到大量差异，特别是不可能文字材料，而科学研究极其缺乏，所以在学术界争议非常多。对于一个事物，真相只有一个，如果有不同的争议观点，其中只有一个是对的。所以目前对上古史的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的所有相关数据。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。

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。首先，基因、语言、体质、考古等诸方面证据的综合分析都证实，全世界的现代人都 在 20 万年前走出非洲，大约 7 万年前因为多峇巨灾的推动而扩散到了非洲之外。在多峇巨灾引发的末次冰期中，现代人类扩散到了世界各地，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，在自然界中构成紧密的人类命运共同体。我们知道，在全世界范围内文明与民族起源，都是 1.2 万年前末次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。

在认识中华民族在人类命运共同体中的起源过程之后，我们需要研究中华文明如何在中华大地上起源。但是，虽然目前对于中华民族和文明的起源研究不断深入，但是在民众中甚至在相关学界还是存在两种极其错误的观点：一是文明西来说，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。二是中华民族多源说，认为中国的民族有很多独立起源，各民族的来源不同，中华民族的概念是近代构建的，这一观点在社会科学界几乎成为主流。实际上这两种观点既缺乏科学证据，又严重影响民族团结。

大量研究证明，中华文明的主体核心都由中国本土孕育，并且传承有序的，虽然与其他文明发生过交流互鉴，但这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。

东亚族群和语系起源的聚合假说

对于东亚人群起源和民族形成过程，在 20 世纪 90 年代到 21 世纪最初十年的相关研究中定出了大框架，但是缺乏细节。基因组数据说明东亚人主要源于非洲，极少部分来自欧亚大陆的早期智人（尼安德特人和丹尼索瓦人）。21 世纪第二个十年以来，由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。

通过采集东亚数十万样本基因组，分析其 Y 染色体、线粒体等的多态性后，我们发现东亚人群起源整体呈现“两阶段两路线”特点^[1]。东亚现代人是由不同时期进入东亚的至少两批现代人混合而成的。第一批约 7 万年前沿海岸线走出非洲，约 5 万年前进入东亚，在现代东亚人群中比例较小，而在大洋洲比例较高。这一批现代人迁徙属于早期扩散。第二批约 5 万年前从西亚扩张并于约 4 万年前进入东亚，形成现代东亚人的主体。四五万年前从西亚的人口扩张是世界早期主要的人口扩张^[2]。两批人进入东亚的路线分别有东西两条主线。

第一批，部分人群于约 5 万年前从缅甸沿着青藏高原边缘的河谷进入东亚西部，这是第一条主线。另一条主线从中南半岛沿着冰川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云南—四川、广西—广东，其后代构成了现代东亚人口的主体。其中云南—四川线的人群涉及孟—高棉、

苗瑶、汉藏三个语族的起源。苗瑶和汉藏语族的早期人口都起源于孟—高棉语族的祖先。广西—广东主线进入中国的人群奠定了现在侗傣—南岛语族的主体人口，成为中华民族另一支来源^[3]。在迁徙过程中，东南亚和东亚之间的丛林形成遗传过滤效应，使得东亚人群的遗传、体质和生理病理特征得以特化。东亚人群最特异的体质特征包括直发、铲形门齿、纤细汗毛、增多的汗孔，都是外胚层总控基因 *EDAR1* 上的一个突变造成的，这一突变发生于三四万年前。这种有利于大量排汗散热的机制，应该源于东南亚湿热气候的自然选择，也促使东亚人的身体对环境变化更敏感，从而可能激发了中华文明崇尚天人合一的鲜明特点。

早期狩猎采集人群的迁徙并没有形成民族，而是形成了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，自然资源才能大量增长，从而促进了人口的初步增长。人口初步增长后，人类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早期的原始民族群体。充分的剩余时间使得新石器工具出现，并渐渐形成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族，主要是汉藏、苗瑶、侗傣 [新名仡佬 (Kra-Dai)]、阿尔泰 (满蒙为主)，以及相关的南亚、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。这些语族中，中国的主体民族系统——汉藏语族，起源于华北桑干河和滹沱河流域



4 万年内的人群扩散与 8 000 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 3 个主要的扩张节点以希腊字母 α 、 β 、 γ 命名。5 个早期族群的来源自西南按顺时针方向分别是高庙文化、仰韶文化、红山文化、龙山文化、良渚文化）另外 3 个小的类群来自后续的王湾三期文化、陶寺早期文化和陶寺中期文化

的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上的东胡林遗址。汉语族与藏缅语族经历了两次分化^[4]：①7 000 多年前南下黄河流域的人群后来形成仰韶文化，成为藏缅语族的起源；约 6 200 年前部分人群北上西辽河流域，形成红山文化，成为汉语族的起源。②约 5 300 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。

处于汉藏族群北方的满蒙族群，虽然不以农业为主，但是在早期汉语族先民北上的影响下，聚合形成了文化共同体。这一群体可能最早源于东北的

细石器文化，以及在此基础上发展的早期玉器文化。特别是黑龙江的小南山文化等都可能与其有关。但是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，例如芬兰-乌拉尔语系人群可能起源于辽西的赵宝沟文化。赵宝沟文化被红山文化覆盖，可能是芬兰-乌拉尔语族离开北迁的原因。而汉族也在从桑干河流域以南来的人群和遗留的赵宝沟文化人群的混合中形成，所以汉语体现出明显的汉藏语和芬兰-乌拉尔语混合的形态^[5]。

更北方的匈奴语族，在东西方人群扩张的影响下产生聚合。这个族群的核心 Y 染色体类型是 Q，是东亚人群的祖先四五万年前离开西亚以后，在青藏高原西侧分道北上进入北亚的分支，也就是过去说的蒙古人种北支。但是在长期发展过程中族群与从缅甸进入东亚的南支发生大量遗传交流。通过基因组进化分析发现这一类群大约在 1.6 万年前的西伯利亚境内发生过一次急剧的扩张，可能与气候回升过程中西伯利亚出现大片草原以及大规模狩猎人群的形成有关。而气候继续回升使得这一环境不再，人群四散迁徙，其遗传扩张影响了整个欧亚大陆和美洲大陆，也影响了东亚大多数族群^[6]。

5 300 年前汉语族人群南下，影响了湖广地区的苗瑶祖先群体和江浙地区的侗傣祖先群体，并发生了大规模的人口融合。我们研究发现广布于中国台湾地区、东南亚和太平洋的南岛语系人群与侗傣语系人群同源源于江浙地区。精细谱系分析发现南岛祖先离开江浙地区的时间为 5 900 年前，当地马家浜文化结束，部分统治阶级在来自长江以北的大汶口文化的驱使下南迁。这为我国的台湾岛，远一点的菲律宾、马来西亚、印度尼西亚、太平洋岛屿数亿人口找到了源头。而汉族祖先的到来，使得马家浜文化之后的崧泽文化终结，在当地混合形成了良渚文化，这正是侗傣语系壮侗语族族群的起源。而崧泽文化上层建筑的西迁可能形成了侗傣语系仡黎语族群体^[7]。崧泽时期南方最



发达的是凌家滩文化，很可能是一个原始政权中心。凌家滩文化有苗瑶文化的背景，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 5300 年前，中国境内的五个考古区系大多发生了文化变革。东北的红山文化大规模南下。中原的仰韶文化大体西迁，覆盖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。这种变革在中国人群的遗传结构中也可以观察到。这可能与传说中蜀地的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形成期，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

文明和民族是分不开的，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、民俗学等领域的综合证据分析，可以清楚地看到中华文明的主体完全根植于中国，是本土自生的。在冰期结束的过程中，中国南方萌发了世界最早的农业，湖广地区的人群开始驯化水

稻。江浙和燕云的农业也随后迅速萌发。这奠定了中华文明的基础。同时，中华文明海纳百川，在千百年的发展中与其他文明发生交流互鉴，很好地促进了中华文明的发展。但是如果因为部分的交流互鉴而提出所谓的“中华文明西来说”，则是非常荒谬的，各个学科领域都有明确的科学证据予以反驳。遗传学的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，在盛冰期结束之前数万年已经进入中国，在 5 000~7 000 年前就完成了文明起源过程中的家族父系谱系扩张。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^[8]。这就证明，在外来人口进入之前，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，发现最早的一个扩张原点是湖南的水稻起源点^[9]，长江流域的水稻驯化可能是世界上最早的农业起源。这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文明。

中华民族的主体起源于中国，那么在中国境内是否由数个族群完全独立起源的呢？现在学界流行一种“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，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”，而自然事物基本上应该是先有“实”体再据此赋予一个“名”称。我们有大量遗传学数据可以证明中华民族“实”先于“名”。很多民族学研究者认为，中华民族这一“名”，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，从而才开始有中华民族的“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。部分学者认为汉族叫作“汉”族，必然是汉朝以后的，历史考证是五胡十六国时期外族压迫才使中原人群自称“汉人”，因此认为汉族是晋代以后形成的。这就把主观认识的“名”凌驾于客观存在的“实”之上了。从历史唯物主义的角度看，汉族是一个有稳定的文化和遗传传承的族群，无论被叫做华人、夏人、唐人或者其他什么，只要稳定地传承发展汉语，以天人合一等思想的“道”为族群文化标记，以三皇五帝为共同的血缘认同，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，都不可能排除出汉族的范围，不然历史逻辑就荒诞了。这样的道理很简单。举个更近的例子，云南的德昂族过去被叫作崩龙族，如果按照某些以“名”为准的观点，在改名之前德昂族就不存在了？实际上改名前后德昂族没有任何其他变化，所以“名”没有害“实”，谁都不能否认德昂族已经存在了上千年。

中华民族和中华民族共同体也是一样的。中华民族共同体虽然是当代的习近平新时代中国特色社会主义思想中提出的，但是这一名称的提出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。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 55 个少数民族都与汉族有共享遗传世系，人口越多，共享成分越多，例如藏族、蒙古族、维吾尔族都有很高比例的汉族主体的 Y 染色体 O-M134 类型。中国各语系、各民族都是在较晚的历史中分化形成的，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隔离。由于各地领导家族之间的利益冲突，部分人群迁到了边缘地区，从主流人群中分离出来，渐渐形成了少数民族；而留在中部地区的主要人群融合形成了汉族。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。中华民族从人群的角度分析，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。

对上古族群结构开展精细研究的基础，在于对中国人群父系 Y 染色体遗传谱系的精细研究和遗传精准估年方法的建立。精确的遗传估年可以定位遗传谱系中特殊事件的发生年代，从而解答相关历史问题。Y 染色体是男性特有的遗传材料，所以在家系中的男性间直系传承，不受混血影响。家系中 Y 染色体序列的传承是相对稳定的，是用来分析突变率，构建人类进化分子钟的最好材料之一。对 Y 染色体进行全序列分析可用以寻找新单核苷酸多态性 (SNP) 位点，计算突变率。为进一步节约、合理利用资源，复旦大学的分子人类学团队改进了原有测序方法，利用 DNA 标签序列以区分不同样本来源，再合并样本，自主设计序列捕获诱饵阵列获得 Y 染色体片段并用 Solexa 进行高通量测序，提高了 Y 染色体测序效率，极大降低了成本，已达到世界最前沿领先水平。以此方法，我们对多个有记录的大跨度深结构家系进行 Y 染色体全序列分析，得出的 Y 染色体点突变率为 $2.02 \times 10^{-8} \sim 3.8 \times 10^{-8}$ 突变/代，这是迄今最准确的突变率。利用 Y 染色体的分子钟从后代的基因序列多样性估算祖先人物的生活时间，结果误差从原来的大于 1000 年缩小到小于 50 年，为遗传学应用于历史研究，开创历史人类学领域，提供了一种十分有效且可信的研究方法。

利用这一技术，复旦大学的分子人类学团队在中国人群的大规模分析中，发现了中国人群 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 Y 染色体主要分支节点和相关东亚族群的分化时间。团队在演化树上观察到, 6000 年前左右, Y 染色体 O γ -F11、O β -F46、O α -F5 三个个人节点出现了迅速扩张, 这三个人分别在数代内产生了上千个后代, 其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^[10]。在三个大节点之后, 还有三个迅速扩张的相对较小节点。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, 也就是中华民族形成的第一个阶段。和超级祖先相关的时间和地点与古史传说中的帝王、考古发现的早期大型陵墓都吻合。三个超级遗传祖先的年代估算分别为距今约 6800 年、6500 年、5400 年。有些学者认为不应该把传说中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严重扭曲。实际上我们观察到的中华民族上古史和现代中国各族的共同上古史, 在遗传学、神话学、考古学、语言学、民族学等领域都能完美对应。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, 他们的起源被发现于不同的考古区系。

(1) 距今约 6800 年是中国已发现的第一座城市始建的年代, 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组, 发现其 Y 染色体正是中国三大谱系中的第一个 O γ -F11 的早期类型。结合这一古城所属的高庙文化出现的大量“凤鸟”“八角星”图案等文化特征, 这一个体与传

说中的中华人文始祖“伏羲”高度吻合^[11]。

(2) 距今约 6500 年，在河南濮阳西水坡遗址留下了一个巨大的陵墓，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天象，东侧是青龙，西侧是白虎，北侧是北斗，南侧是南方七宿动物造型，象征着墓主人的崇高地位，疑似传说中的“神农”。后续通过检测墓主人的基因组，可以确认是否为第二个谱系的源头。通过对仰韶文化中年代略晚于面水坡大墓的贵族墓葬开展检测分析，我们发现这些墓主的基因组主要归属第二个谱系。基于这一谱系在时间序列上的延续性特征，可治理推断：面水坡大墓的墓主极有可能同样属于该谱系。

(3) 距今约 5400 年，在辽宁、河北、内蒙古交界处辽宁建平县的牛河梁山岗上，红山文化先民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。我们检测了最早且最大的积石冢墓主遗骸 DNA，发现其为第三个谱系的源头。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^[12]。

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有明显传承关系。高庙文化的高级陶器上往往用点刻方式绘制大量图案，主要是巨目曲喙凤鸟、八角星、建木、夔龙造型。其中八角星是现存水书《连山》中的八卦造型，在今天的苗瑶族群中广泛使用。这些造型反映了典型的苗瑶族群文化，也大量融入汉族文化。

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。藏缅族群传统的文化习俗圆圈舞，在仰韶文化的后续——马家窑文化彩陶上也清晰绘制着。所以有理由相信，仰韶文

化-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。



藏族的吉祥八宝图案在代表仰韶文化的彩陶中都能找到可能的源头

汉族虽然融合了大量远古族群的文化要素，但是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，玉器的代表造型是龙和凤、“道人”造型、玉瑁冠冕、玉鼈玉龟、熊头燕子、终葵砭刀……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。《山海经》所述的北山经的北次三经即太行山大兴安岭一线，牛河梁陵区正是北次三经中的轩辕坟所在。而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。此外山东一带的后李-北辛-大汶口文化区系和江浙一带的上山-马家浜-崧泽-良渚文化区系也是关键的源头。在红山文化扩张之后,大约4600年前大汶口文化开始向西扩张,与中原融合形成龙山文化,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,大汶口文化的大量谱系来自高庙文化,如果高庙文化是太昊伏羲氏族群,那么少昊与太昊有传承关系也是合理的。大汶口文化是否派生出历史上的东夷族,是否有现代民族作为主要后裔,目前还有待研究,一般认为东夷族后来完全融合入汉族。瑶族称与高辛帝有关,高辛就是帝喾,帝喾、帝舜都是来自山东少昊一脉,苗族瑶族之分或与太昊、少昊相关。

良渚文化起源于5300年前的大变革,族群于约4500年前北扩。此后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。较早的良渚文化和较晚的进一步融合的石家河文化,都以夸张的人头像造型为典型特征。而“项”字就是“玉首”。约4400年前,颛顼时代可能被帝喾时代取代^①,颛顼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,江浙部分的统治阶级南迁到两广的石硖文化而后发展成侗傣族群。良渚文化遗骸和现代侗傣族群中大量检出的Y染色体O1-M119类型证实了这一发展关系。齐家文化与良渚文化一样,以大量使用玉琮、玉璧为显著特征,而疑似齐家文化统治中心的陕北石峁古城甚至出土了大量直接来自良渚文化的精美玉器。石峁人群大约4000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。

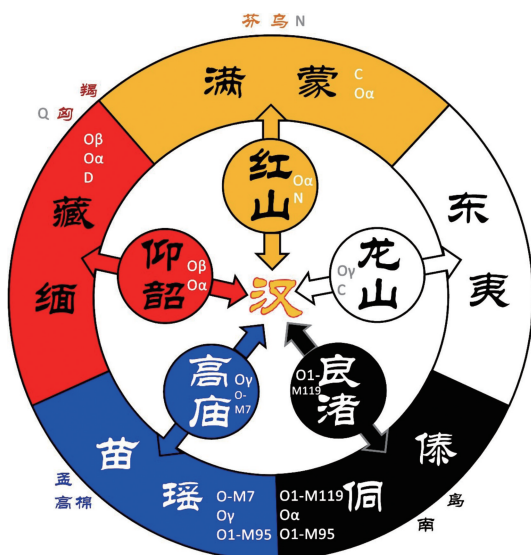
^① 传世文献对于早期历史的时间误差比较大,战国的文本对此事件的时间就有若干不同版本。民国初年学界曾对此有过探讨。文中所写时间是目前考古证据支持的一个年份。

时代	伏羲	神农	轩辕	少昊	颛顼
考古文化	高庙	仰韶	红山	龙山	良渚
派生族群	苗瑶	藏缅	满蒙	东夷	侗傣
融合	汉				

中华民族共同体的上古起源

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的历史，而不是西方某些研究者用双重标准宣称的只有三千年历史。对历史学界和考古学界部分人所持的否定中国古史、否定三皇五帝真实性的观点，也是有利的反驳，为重建中国古史奠定科学基础。基于中国古史的科学重建，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，对应了中国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。中华民族共同体的起源实际上与早期民族的聚合形成是同步的，甚至可以说中国各民族起源于上古的中华民族共同体。

我们构建的精细的基因分子钟也可以研究不同历史时期家族的变迁和人物间的关系。通过对现代家系与古代人物遗骸 Y 染色体进行比对，我们确定了汉、魏、晋、宋、金、元、清等皇室家族的血统起源，解答了大量历史谜



中华民族共同体形成模式——融合与派生

团。例如，通过对曹操后代家系的遗传学梳理，确定了曹操的基因标记和身世，证实曹操父亲是家族内过继，而不是异性收养，解决了历史学遗留两千年的问题。有一些现象是历史上没有记载而且不可能知道的，但是通过基因的分析却能突破普通人的认知，发现历史真相。例如，我们的基因研究发现辽代的耶律氏的 Y 染色体类型 N-M128 来自西周的黎国君主，金代部分完颜氏样本与晋代王羲之所属的琅琊王氏有关。中华民族共同体内的交流强度是超出想象的。

结语

我们的研究开创了用分子生物学手段研究历史疑难问题的历史人类学领域。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。随着人类学各分支领域的深化发展，尤其是大数据运算与人工智能技术的赋能支撑，一系列系统性、全面性的研究方法应运而生——基因组学、表型组学等组学研究已成为人类学的核心方向。这种以组学为纽带的研究范式，通过整体观的构建打破了学科壁垒，使不同领域的学术视野得以重叠交织，进而推动多维度证据的协同聚焦。在此过程中，历史的脉络逐渐从模糊走向清晰，那些曾被视作“无法分辨”的上古史细节，正依托跨学科的技术手段被逐一解析。正是这一学术背景的革新，推动中国早期历史研究真正步出“疑古时代”，曾经笼罩在“三皇五帝”身上的传说迷雾，正随研究的深入愈发消散，历史真实性逐渐显现。

参考文献

- [1] Wang C C, Li H. Inferring Human History in East Asia from Y Chromosomes [J]. *Investigative Genetics*, 2013, 4(1):11-21.
- [2] Wang C C, Ding Q L, Tao H, et al. Comment on Phonemic diversity supports a serial founder effect model of language expansion from Africa [J]. *Science*, 2012, 335(6069):657.
- [3] Li H, Wen B, Chen S J, et al. Paternal genetic affinity between Western Austronesians and Daic populations [J]. *BMC Evolutionary Biology*, 2008, 8:146.
- [4] Wang L X, Lu Y, Zhang C, et al. Reconstruction of Y-chromosome phylogeny reveals two neolithic expansions of Tibeto-Burman Populations [J]. *Molecular Genetics and Genomics*, 2018, 293:1293-1300.
- [5] 李尧, 王晓斌, 刘慧. 随园文心 李葆嘉先生七秩同乐文集[M]. 南京: 河海大学出版社, 2021: 148-161.
- [6] Huang Y Z, Pamjav H, Flegontov P, et al. Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia [J]. *Molecular Genetics and Genomics*, 2017, 293(1):107-117.

- [7] Sun J, Wei L H, Wang L X, et al. Paternal gene pool of Malays in Southeast Asia and its applications for the early expansion of Austronesians [J]. American Journal of Human Biology, 2020,33(3):e23486.
- [8] Xu D, Li H. Languages and genes in Northwestern China and adjacent regions [M]. Singapore: Springer Singapore Pte. Limited, 2017.
- [9] Yu X E, Li H. Origin of ethnic groups, linguistic families, and civilizations in China viewed from the Y chromosomes [J]. Molecular Genetics and Genomics, 2021,296(4):783 – 797.
- [10] Yan S, Wang C C, Zheng H X, et al. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers [J]. PLoS One, 2014, 9 (8):e105691.
- [11] 贺刚. 湘西史前遗存与中国古史传说[M]. 长沙:岳麓书社,2013.
- [12] 郭大顺. 红山文化[M]. 北京:文物出版社,2005.