

现代人起源的分子模型

黄石

中南大学生命科学学院医学遗传学中心, 长沙 410078

摘要 20世纪60年代初, 研究者们根据物种间蛋白序列的遗传距离现象开创了分子进化研究领域, 并提出了不同于自然选择理论的分子进化理论, 即分子钟假设和中性理论。该理论认为可以通过基因的遗传距离判断不同物种之间的亲缘关系或分化时间, 遗传分子距离越大, 则亲缘关系越远, 分化时间越长。研究者们随后以该理论为前提推导出了一个现代人起源的分子模型, 即近期走出非洲假说, 遗传多样性最高的非洲人群被认为较其他人群有更远古的最近共同祖先, 现代人被认为最早出现在非洲并随后迁徙扩散到欧亚地区, 基本取代了当地的古老人群。中性理论虽然是一个非常有价值的零假设, 但并不能完美解释遗传多样性现象, 这就导致了出非洲说必然具备不确定性。一个发表于2008年的新的分子进化理论——遗传多样性上限理论, 对遗传距离现象重新进行了解读: 遗传距离大并非必然意味着分化时间久, 也可能反映的是表型差距较大, 遗传距离随时间增加并非如中性理论所隐含的是无限的, 而是有个上限, 上限值主要是由物种的表型复杂性决定的。多重检验分析显示, 目前现实里的遗传距离或遗传多样性基本都处在上限值。有研究者依据遗传多样性上限理论对人类遗传多样性现象重新进行了分析, 独立重新做出了最早于1983年就被发现的现代人走出东亚的单亲染色体进化树, 还从常染色体序列分析得出了支持多地区进化假说的结论, 并从多个角度包括古DNA分析论证了出东亚说是一个更接近真相的学说。随着对进化基础理论的深入探索, 人类对自身起源的认识也必将会变得更加清晰和完善。

关键词 分子钟; 中性理论; 遗传多样性上限理论; 出非洲说; 出东亚说

1 引言

最早的直立人出现在非洲和欧亚大陆, 之后如何进化为现代智人, 即现代人起源问题, 目前学界争议仍然较大。现代人就是与现生人没有显著差异的人, 是否各个主要地区的现代人都是来自本土, 独立连续进化而形成的? 目前流行的模型是近期出非洲说, 主要依据是分子人类学研究, 认为现代人最早起源于非洲南部, 然后迁徙扩散到欧亚地区。但近些年来, 一个被普遍遗忘的分子模型——出东亚说, 因为以新进化理论为基础的推演而重新返回进入了人们的视野。这两个分子模型最早都可以追溯到1983年的一篇文章^[1]。为何会出现不同的分子模型? 哪个模型更接近真相? 本文试图对这两个重要问题进行简要论述。

收稿日期: 2023-09-25; 接受日期: 2024-09-14

基金项目: 国家自然科学基金(81171880); 国家基础研究项目(2011CB51001)

作者简介: 黄石, 教授, 主要从事分子人类学、进化树和分子进化理论的研究。E-mail: huangshi@sklmg.edu.cn

2 遗传等距离现象与分子进化理论

分子人类学研究依赖对DNA序列数据的解读，这种解读必须以分子进化理论为基础，流行的理论是中性理论，近年来又出现了一个竞争理论——遗传多样性上限理论，这两个理论的提出都受到了遗传等距离现象的启发。

2.1 遗传等距离现象

遗传距离特指不同物种DNA序列或蛋白质序列的差异而不是亲缘距离，可以用序列相似度的高低来表示，即两个物种的同源序列之间有差别的位点数目占该序列位点总数的百分比。

遗传多样性可以用种群内每两个个体的平均核苷酸差异数值来代表，也可以用某一个体的基因组的杂合子数量来衡量。若某个物种的某一个代表性个体的杂合子数量多，则意味着该物种群体的遗传多样性水平高。遗传多样性与遗传距离是等价概念，都代表不同个体或物种基因序列的差异值。遗传多样性通常指的是同一物种群体内部的遗传多态现象，但也可以代表不同物种之间的遗传差异现象。

对任意三个或更多直观看来复杂性不同的物种的某个同源基因，可以做两种不同序列比对。第一种是用某一复杂物种如人与各个简单物种比序列差异，比如青蛙和鱼，会发现人与青蛙近，与鱼远。大多数人可轻易猜到这个比对的正确结论。第二种比对与第一种反过来，是用某一简单物种与各个复杂物种比，这时会看到简单物种如鱼与所有复杂物种如人和青蛙的距离是大致一样的，这一结果基本无人能够预料，这就是遗传等距离现象（图1）。等距离并不意味着完全相等，而是近似相等，等距离现象的真实性可以通过统计方法来论证^[2]。

这个示例把三个不同复杂等级的四足动物物种（T1—T3：人、鸟、青蛙）与另一个低等物种（X：鱼）进行比较，这些四足物种具有共同的最近祖先（T）。物种X是外群物种且与物种T1—T3具有较远的亲缘关系，而物种T1—T3是群内物种。物种T1—T3的进化谱系在同一点V与物种X的谱系分离。此外，物种T1—T3是进化过程的产物，该过程自V以来已经运行了完全相同的时间。从T这个共同祖先点开始，人、鸟、青蛙各自都进化了同样长的时间，它们共享一个最近祖先V。它们与鱼在V这个点分开，各自与鱼分开的时间是相同的，根据前述遗传等距离现象，它们与鱼的遗传距离也都相同，进化速率等于遗传距离除以进化时间，所以它们的进化速率也相同。

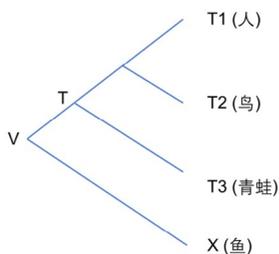


图1 遗传等距离现象及其对该现象的分子钟解读

Fig.1 Illustration of the genetic equidistance result and its interpretation known as the molecular clock

1962年，通过对不同物种的血红蛋白进行

上述第一种比对, 祖卡坎德尔和鲍林非正式地首次提出分子钟^[3]。1963年, 马戈利来希对多个物种的细胞色素c序列同时进行了以上所提的两种比对后, 首次发现了遗传等距离现象, 并首次明确正式表述了分子钟假说^[4,5], 认为不同物种置换速率大致相同。等距离现象被后续研究发现是极其普遍的^[2,6], 这直接导致了后续分子进化理论的提出, 可以说, 若这个现象是假象, 则会直接导致整个分子进化领域的坍塌。

2.2 分子钟假说

分子钟假说对遗传等距离现象的解释如图1所示, 但遗传等距离现象后来被发现与置换速率无关^[7], 被解读为分子钟其实是一个不谨慎犯的错误。如果假定遗传距离一直都与时间成正比, 则通过简单的初中数学换算, 就会从距离现象推导出各个不同物种的突变速度是一样的, 也即分子钟。但这里需要质疑的是: 能否进行这种简单的数学换算? 进行这种换算隐含了序列差异没有饱和上限的预设假定, 也隐含了不同物种对遗传距离的贡献是相等的, 但这些假定是不是真实的呢?

大量研究显示分子钟只有有限的存在, 只适用于同类物种或近似物种, 不同脊椎动物物种的突变速率差别可高达40倍^[8]。今天, 几乎没有专业人士会认同存在一个普遍的 (universal)、适用于所有物种的分子钟, 但他们通常并没有意识到否定“普遍分子钟”的巨大代价, 不知道还有一个被埋没在其阴影下的、被其曲解并替代的遗传等距离现象, “普遍分子钟”的瓦解意味着遗传等距离现象重新成为一个无解的谜。

2.3 中性理论

分子钟仅是对一个现象的假设解释, 却被许多学者认为是一个普遍的自然现象, 因此导致他们去努力解释分子钟, 其中, 木村提出的中性理论得到了较大认同^[9], 他认为分子钟是中性理论最好的证据^[10]。1966年, 勒旺廷和哈比发现很大比例的酶编码位点都具有多态性^[11], 他们对比了自然选择和中性变异两种解释, 但认为都不能完美解释大量遗传多样性的存在。中性理论认为, 虽然一小部分突变是有利的并被自然选择固定, 也有很多是有害的并被自然选择淘汰, 但绝大多数被固定的突变实际上是中性的, 是因为遗传漂变而固定的。

中性理论最重要的一个推论是, 置换速率以世代为单位恒定, 突变的固定速率等于突变速率, 并保持恒定。这是“普遍分子钟”的 (近似) 理论基础, 可以解释类似物种有相同的、以年为单位的分子钟, 但很难解释世代年数差异较大的物种也同样有相同的、以年为单位的分子钟。从长时间尺度上看, 替代的速率不依赖于种群的大小: 如果种群很大, 每个突变漂向固定的速率比较低, 但是较低的固定速率被更多的新突变弥补。

如果两个物种在 t 代以前由共同祖先分化而来, 而且每个物种每代经历了 u_0 个替代 (相对于其共同祖先的等位基因), 那么这两个物种间序列差异的数量应该是 $D=2u_0t$, 因为两个谱系各自都累积了 u_0t 个替代。因此, 如果估计出了种群经历的世代数, 中性突变

速率就可以按以下公式估计：

$$u_0 = D/2t$$

这一公式是有条件的：假定两个物种中固定下来的所有突变对遗传距离的贡献是相等的。如果一个谱系的 u_0 已经被准确估算出来，根据关系式 $t = D/2u_0$ ，可以估算相关物种从其共同祖先分化的时间。但是，这个公式只适用于用慢变序列观测的遗传距离，过去的分子进化树构建都忽略了快变序列的饱和上限问题，虽然认识到了部分饱和问题并做了计算修正，但从未意识到上限距离是不可能被任何算法修正的。对于处在未到饱和进化阶段的中性变异，中性理论和分子钟基本成立，这些变异随时间的积累，可以用来计算物种分化的时间。

分子钟假说的失败自然意味着中性理论作为解读自然的解释性理论也失败了^[12,13]，但中性假定仍然是一个有价值的零假设，用来检验自然选择。但很多依据中性假定推理得出的结论，比如进化树，理应是未知的，却经常被当作真相宣传，这里就隐含了把中性理论当成解释自然的理论。

中性理论的另一个推论是遗传多样性与有效群体大小成正比，有效群体大小的实际数值是无法观测的，都是通过遗传多样性水平来换算的，但如果忽略了上限概念，则这类换算就有可能是错的。实际群体规模大的群体理应有有效群体也大，即遗传多样性也高，但实际情况是，实际群体数目相差很大的群体，在遗传多样性水平上的差异却不大，这是一个长期未解的谜，所谓陆文亭悖论，显示中性理论与事实的脱节^[14]。分子生物学家基本无人认可中性理论，他们都在努力寻找所谓中性“垃圾”DNA的功能，并不断获得成功^[15,16]。

2.4 遗传多样性上限理论

遗传多样性或遗传距离是否有上限？先验逻辑推理是必须有上限，因为若没有，则不同物种同源基因的相似性会随时间的推移下降到没有任何相似性，这就意味着一个基因的功能可以被两个或多个乃至无限多个不相干的序列编码，而这是荒谬的。若有上限，则它的决定因素是什么？长期以来，业内一直没有讨论这个问题，这就隐含了没有上限的假定。

2.4.1 遗传多样性上限理论的基本概念

遗传距离会随着时间的推移而增加，但不能达到100%的非同一性，将在某上限水平上停止增加。某生物的不同物种之间序列差异的百分比上限称为该生物的最大遗传多样性或遗传多样性上限，某物种内个体之间序列差异的百分比上限称为该物种的遗传多样性上限。

生物表型是由DNA以及表观遗传编程决定的，每种细胞都代表一种独特的表观遗传编程。DNA可以被认为是建材，表观遗传编程是建构图纸，决定了DNA如何被利用。生物体的复杂性可以定义为它的表观遗传编程的复杂性，可通过参与表观遗传编程的基因数目和细胞种类数来估算。对于神经细胞，很多情况下，一个单独的神经细胞就代表一

个独特的细胞种类, 具有独特的基因表达图谱和突触链接方式。因此, 生物复杂性可通过神经系统来反映, 人是最高级物种就是因为其脑细胞种类最多。

表观遗传复杂性与遗传多样性之间存在反相关关系, 这是遗传多样性上限理论的核心理念, 简单物种能容忍更多的DNA变异, 或者说有较高的遗传多样性上限^[17]。这反映了一个简洁自明的直觉: 一个系统越复杂, 它能容忍的建材的随机变异误差就越小。新达尔文主义的创始人之一费希尔第一个论述了这一原理^[18], 知名学者奥尔曾通俗地表达过该原理: “费希尔进一步论证了, 对于任意一个具有一定效力的突变, 其在复杂生物中比在简单生物中会更不可能是有利的。原因很直观: 对表型能产生某种程度影响的随机突变更有可能破坏(而不是改进)复杂的而不是简单的生物体, 同理, 对性状能产生某种程度影响的随机变化更有可能破坏(而不是改进)复杂的机器而不是简单的机器。例如, 将任意机械零件的长度改变一英寸, 更有可能破坏显微镜而不是锤子的功能。”^[19]这个原理近百年来没有受到过怀疑, 但中性理论隐含了对这一原理的违背, 在遗传多样性上限理论之前也一直无人用该原理系统地解读自然现象。

除了上述核心理念外, 遗传多样性上限理论还包含以下三条直观概念: 第一, 任何由物质构建的系统都会容忍一定的零件误差, 这些误差对系统性状随时间推移或环境变化可以是有利、有害或中性的, 这其实就是流行理论合理部分的基石。第二, 物质随机运动与意识思维能力互为对立, 随机运动会破坏意识思维, 反过来意识思维也能够压缩随机运动, 因此, 遗传多样性会与意识思维能力成反比。第三, 对性状有害的变异可以通过其他突变来挽救^[20], 具有较高遗传多样性的群体较易挽救有害突变, 这不利于自然选择对性状质量的维持和对有害变异的淘汰, 因此, 抑制遗传多样性对于维持性状和消除有害变异是必要的。

遗传多样性上限理论有如下四点主要内容: 第一, 表观遗传复杂性限制了物种可以容忍的基因突变的范围, 而突变范围反过来也限制了表观遗传复杂性改变的范围, 突变范围过宽, 不可能构建复杂物种。第二, 复杂物种能容忍的突变范围较小或遗传多样性上限较低。第三, 虽然大进化与小进化都涉及基因突变及表观遗传复杂性的微小改变, 但从简单到复杂的大进化是表观遗传复杂性的台阶式突然提高, 以及相应的遗传多样性上限的降低, 而在小进化中表观遗传复杂性没有大的改变, 其线性阶段基本上可以用自然选择理论和中性理论来正确描述。第四, 遗传多样性只要是在上限范围内, 就是越多越好, 因为它能够增强适应能力, 所以它会受到正选择而加速到达上限。基因组不同区域突变率的最大差别可能超过一万倍, 而基因组的大部分是属于突变率较快的, 因此, 今天观测到的遗传距离大部分是上限距离, 对大部分经历过长期进化的物种来说, 它们的基因组今天大部分处在遗传多样性上限平衡态, 只有小部分慢变序列例外^[17,21-23]。

2.4.2 遗传多样性上限理论与有性生殖之谜

有性生殖为什么成为真核生物的主要生殖方式仍然是生物学重大未解之谜之一^[24]。为何有性生殖的真核生物能够随时间的推移进化出越来越复杂的高级物种而无性生殖的

原核生物却不能? 常见的猜想有两种。其一, 有性生殖可以增加遗传多样性, 有利于适应环境。然而, 细菌等无性生物通常比真核生物具有更高的遗传多样性。其二, 有性生殖可以消除染色体和表观遗传异常或其他有害突变。然而, 这种异常也可以通过自然选择淘汰异常表型来消除。遗传多样性上限理论为有性生殖的普遍性之谜提供了一个解决方案, 从简单到复杂的物种进化需要减少遗传多样性(在核苷酸水平上), 类似于精密机器需要减少零件的误差范围。简单物种某个个体的遗传多样性的减少, 对于该个体成为更复杂新物种的初始个体是必要的。由于后代某个个体遗传变异的总体水平主要遗传自父母携带的单核苷酸变异的组合, 有性生殖可以增加也可以减少后代相对于父母的遗传变异水平, 而无性生殖只能增加后代的遗传变异水平。因此, 有性生殖对于减少遗传多样性至关重要, 对于高级物种的进化也至关重要。原核生物尽管进化时间长于真核生物数倍, 并成功进化出了大量不同物种, 但却基本都是简单物种, 缺乏有性生殖可能是原核生物未能进化出多细胞复杂物种的原因。

2.4.3 遗传等距离现象的重新解谜

对进化分离时间较长的物种或快变序列而言, 可以给出只与极限距离相关的遗传等距离现象的定义, 姑且称之为极限遗传等距离: 各个物种或姐妹物种与另一复杂程度较低或复杂性相同的组外物种之间的遗传距离大致相等, 且不随时间的推移而改变。共享一个共同祖先的两个物种叫作姐妹物种, 与姐妹物种亲缘关系较远的物种称为组外物种(outgroup)。比低等物种高等的不同物种, 可以包含两个姐妹物种, 也可以是成千上万个不同物种, 这些物种之间也可以有复杂性的差别, 比如所有真核生物都比细菌高等, 任何某同源蛋白的序列比对, 都是与细菌等距离。在复杂物种中可容忍突变的位点, 在简单物种里一般也是, 但反之不然。各物种与一简单物种的极限遗传距离大部分是由该简单物种的遗传多样性上限决定的, 简单物种的更多变异会覆盖复杂物种的变异, 所以必然导致等距离现象。

简单物种的遗传多样性上限决定了它与更复杂物种之间的遗传距离, 这可以通过细胞色素 c 的例子来说明。该蛋白质序列在细菌内的最大多样性大约有 70% 的差异, 例如副百日咳鲍特菌(*Bordetella parapertussis*)和善变副球菌(*Paracoccus versutus*)这两种细菌所显示的距离, 而这近似于细菌和哺乳动物之间的最大距离65%, 例如副百日咳鲍特菌和黑猩猩之间的距离。真菌的最大多样性约为米曲霉(*Aspergillus oryzae*)和解脂亚罗酵母(*Yarrowia lipolytica*)之间的距离40%, 而这近似于真菌和哺乳动物之间的最大距离, 例如米曲霉和黑猩猩之间的距离是43%^[17]。

一个简单的思想实验可以通俗地解释遗传等距离^[17]。如果能通过使用相同的某一基因序列创造三种不同生物群体——酵母、鱼和人, 然后让这三种生物分化很长时间或大约5亿年, 并假定它们那时还是各自与原先表型差异不大的物种, 酵母中的一个基因会改变很多, 比如50%, 而它在鱼中的对应基因也会改变较多但会低于酵母, 比如30%, 它在人类中的对应基因则会发生很少变化, 比如1%。任何超过50%的变化对酵母都是致命

的, 任何超过30%的变化对鱼类都是致死的, 任何超过1%的变化对人类都是有害的。因此, 酵母与人或鱼之间的遗传距离主要由酵母中的突变决定, 酵母中 50% 的变化将解释酵母和人之间50%的距离数值, 也将解释酵母和鱼之间50%的距离数值, 以及酵母与酵母之间50%的距离数值, 这就是遗传等距离现象。同理, 鱼的30%的变化决定了鱼与人之间的距离数值是30%。相比之下, 中性理论则认为人和鱼都可以像酵母一样变化高达50%, 因此, 它们之间的遗传距离都是50%或更多, 不管是人与酵母、人与鱼, 还是酵母与鱼。

所谓分子钟其实真正反映的现象是物种复杂性随时间的推移而匀速提升, 这种匀速提升的直接证据就是遗传等距离现象。

2.4.4 遗传距离或遗传多样性是不是在上限的检验

遗传多样性上限理论已经被写入七本中外学术书籍^[25-31], 包括中美教材各一本^[25,26]。该理论与中性理论直接相反的主张有如下列: 第一, 大多数常见遗传变异不是中性的; 第二, 遗传距离或遗传多样性大多处于上限水平; 第三, 无限多位点模型不现实, 重复突变很普遍^[32]; 第四, 大多数DNA序列处于被选择状态, 因此不能用来构建系统发育树; 第五, 所有突变都有有害的一面; 第六, 非保守序列不是没有功能^[33]; 第七, 同义替换也是功能性的^[34,35]; 第八, 两个复杂性不同的物种之间的分子距离并不是由两个谱系中的突变同等贡献的, 而主要是由低等物种贡献的; 第九, 复杂性的增加需要抑制遗传多样性; 第十, 等位基因的进化或更新非常快而不是慢^[36]。

检验遗传距离是不是在上限的方法主要有两种。第一种检验方法是分析重叠突变位点特征^[37]。在两个姐妹物种与一个简单的组外物种的某同源基因的三条序列比对中, 某些位点在三个物种各自序列中都具备独特的氨基酸, 这说明至少有两个物种都在这些位点发生了突变, 这一现象被称为遗传等距离中的突变重叠特征。如果两个物种分离后, 各自按中性理论随机积累突变位点, 则某一位点碰巧同时在这两个物种发生突变的概率应符合概率论的计算, 这一概率与可变位点总数目成反比, 与突变速度成正比。由于遗传多样性上限理论估计的可变位点数目较少, 且复杂物种的可变位点与简单物种有重叠, 因此它预测的重叠位点数比中性理论预测的多。分析结果显示, 重叠位点的数目符合遗传多样性上限理论的预测, 远多于中性理论的预测^[37]。

第二种检验方法是分析“遗传不等距离”现象^[38], 对进化分离时间较长的物种或快变序列而言, 当有两个以上的复杂性不相同、同属某进化枝的物种与某一复杂性更高的组外物种相比时, 则出现“遗传不等距离”现象, 即其中复杂性最低的物种与组外物种的距离会大于其中复杂性较高物种与组外物种的距离。例如, 同是非脊椎动物的章鱼和鸟蛤, 与人比序列时, 鸟蛤与人的距离大于章鱼与人的距离, 而章鱼是智力最高的非脊椎动物。但分子钟认为遗传距离只是与时间有关系, 因此会得出人与章鱼和鸟蛤在序列差异上相等的结论。这一“遗传不等距离”现象非常普遍, 在所有被分析的物种中都得到了验证, 包括章鱼和鸟蛤、同属爬行动物的鸟和蛇、同属软体动物的穿孔贝和舌形贝、昆虫类的蜻蜓和虱子、真菌类的黑穗病真菌和酵母, 以及新大陆猴类的枭猴和狨猴^[25,38-40]。

其他验证研究还包括如下几类：第一，个体遗传多样性整体水平与受教育程度（即智力）呈反相关关系^[41]，大部分遗传变异对人脑有害^[42]，蛋白质断裂变异总数的负载与智力成反比^[43]；第二，许多复杂疾病患者群体的遗传多样性水平高于正常对照组，说明正常人的遗传多样性水平处在上限状态^[44-46]；第三，低等物种有较多的常见遗传变异，对人是有害的^[47,48]；第四，很多人类性状受到稳定化选择，也即一个数量性状有个最佳度^[49]；第五，动物模型研究发现遗传多样性与多种复杂性状包括智力和免疫功能的关联^[50]；第六，功能基因组学研究发现基因组大部分是有功能的^[15,16]，例如，短串联重复序列（STR）占人类基因组的5%，几乎都具备调控基因表达的功能^[51]；第七，复杂物种抑制随机序列的表达，但简单物种却正常表达，说明人类基因组的75%表达成RNA这个现象不是中性论者所预期的随机噪声，同时也论证了遗传多样性上限理论的核心原理，即简单物种能容忍更多的随机噪声（表达）^[52]。最后，细菌对突变的鲁棒性不受进化代数的影响，这不符合流行理论预期，但支持遗传多样性上限理论，因为物种的鲁棒性基本上取决于物种自身的建构图纸，物种构造越简单，鲁棒性越强，对突变的容忍度越高^[53]。

3 出非洲说和出东亚说分别来自对立的分子进化理论

分子进化树需要挑出中性变异来构建，分子钟合理部分的一般原理是，中性序列的遗传距离在未达到上限之前，反映了亲缘关系或分化时间。

3.1 分子建树方法

构建分子进化树时，必须把达到上限饱和的快变序列筛选去掉，慢变序列里发生的能稳定存在的氨基酸变异才是中性的，而快变序列里发生的能稳定存在的氨基酸变异有更多的功能性改变^[54]。现实里大部分能较稳定存在的变异是非中性的，反映的是生理构造选择和自然选择，这些变异的相似性不一定反映的是亲缘关系，而是类似的生理特征或生存环境。当然，亲缘关系近，一般生理特征也类似，因此，用非中性变异构建的进化树有时候也会反映真实的亲缘关系。

建树方法有多种，基本可以分为两类：距离法，即基于相似距离的方法，如UPGMA（非加权配对法）、NJ法（邻接法）、ME法（最小进化法）；特征法，即基于特征变化的方法，如ML法（最大似然法）、MP法（最大简约法）、Bayes法（贝叶斯法）。遗传多样性上限理论不否认有少部分变异属于中性，如果是真正用了慢变序列的线性距离，避免了重复突变，则应该是距离法比较可靠。如果各分枝的进化速率是一致的，则UPGMA是比较适宜的。UPGMA是根据不同物种同源序列的相似程度，分别计算出相似系数，然后将最相似的两个序列先归为一枝，用一个序列来代表，这样就使原始序列数目少了一个，再循环往下，就可以得出一个分类图，这种进化树也适合按照分子钟原理计算分化时间。

树根的确定有四种方法：一是按照分子钟把分化时间最长、遗传多样性最高的群体

定为树根；二是根据与组外物种的距离，也就是与一个已知关系较远的物种进行比对，与其较近的物种为树根；三是根据与古代DNA的距离，古人是祖先或更接近祖先，把距离古DNA较近的群体定为树根；四是根据中心型祖先型，相对于较后出现的其他衍生型，祖先型应该分布更加广泛。这四种方法理应给出同一个结论，但前两种方法如果误用了饱和态的遗传多样性数据，则会给出错误的结论。

3.2 人与类人猿的关系

在分子进化研究出现之前，形态分类学家通常把类人猿类归为一个分枝——pongid，包括猩猩、大猩猩和黑猩猩，而人是另外一枝。人与类人猿的分离时间大约是1800万年前，这两个不同的进化枝在早期的性状差别包括肩部是否具备吊挂结构，以及牙釉质厚度^[27]。人与类人猿的最近共同祖先起源自非洲，这个祖先的后代们在1500万年前左右曾经迁徙进入欧亚地区，并随后返回非洲。亚洲1450万年前的腊玛古猿化石曾长期被判定为属于人分枝，后来受分子钟分类结论的影响才被修改为属于大猿分枝。分子钟分类结论把上限距离当成线性距离来处理，得出了黑猩猩是人类最近物种的结论，认为人与黑猩猩的分离时间仅是450万年前^[55]，这个时间被后来发现的700多万年前能直立行走的乍得沙赫人化石修正^[56]。

通过全基因组测序发现，其实只有70%的基因组是黑猩猩与人类最近，而有30%的基因组是大猩猩与人最近，另外有1%的基因组是猩猩与人最近^[57,58]。这些现象可能来自序列的趋同进化，而不是所谓的种群不完全谱系分选。后者应与突变速度无关，但事实是快变的非编码序列、重组率高的区域、和免疫相关的序列更容易出现这种所谓的不完全谱系分选^[59]。相比同属类人猿类分枝的其他物种，比如大猩猩，黑猩猩与人序列更近，这其实属于前述的“遗传不等距离”现象，类似新大陆猴类的某种猴（泉猴）比其他猴（绒猴）与人序列更近^[39]。把黑猩猩从类人猿类中挑出来与人组成一枝，其荒谬性等于把泉猴从新大陆猴类中拣出来与人组成一枝。

研究者用慢变序列重新分析了人猿关系，发现黑猩猩和大猩猩与猩猩的序列相似度高于人与猩猩的序列相似度，显示人与类人猿类分开后，类人猿才分化出了黑猩猩、大猩猩和猩猩，计算显示人与类人猿类分开的时间是1760万年前^[38]。猿类化石在1450万年前就分出了两类，而不是在700万年前才分，是支持这一新分类结论的直接证据。另外，近期研究把非洲稀树草原出现的时间从700万—1000万年前提前到1700万—2000万年前^[60]，按照稀树草原假说，由于在这样的环境中需要长时间行走来寻找食物和水源，人类便逐渐演化出了直立行走模式。因此，稀树草原出现时间的大幅度提前也支持了类人猿类与人是不同分枝的结论。

3.3 现代人起源的分子模型

关于现代人起源，目前学界有三个主要模型，即多地区进化说、近期出非洲说和近

期出东亚说。

3.3.1 多地区进化说

该学说认为,近代的人群进化现象是由更新世早中期的直立人分布情况造成的^[61,62],地缘环境等差异导致了各个地区的人群各自进化成现代人。东亚人的连续进化特征比非洲人群更为显著,直立人的某些形态特征一直保留在今天的东亚人中,比如铲形门齿、鼻梁较低等^[63]。多地区进化说拥有较充足的化石和旧石器时代文化遗存等证据的支持^[64],但长期以来一直缺少分子学方面的证据。

3.3.2 近期出非洲说

中性理论对于撒哈拉以南非洲人拥有最高的遗传多样性这一现象的解读,直接为出非洲说提供了分子学证据支持。其基本逻辑是把进化树树根定在遗传多样性最高的种群里,或与组外物种序列最近的种群里^[65]。近期出非洲说将组外物种黑猩猩的基因型作为祖先位点来确定人类的位点是否突变过,与黑猩猩相同的为祖先型,不同的则为衍生型,树的分叉都是用衍生型等位基因来决定的,这些都需要中性理论的无限多位点假定作为前提^[66]。非洲南部的桑人恰好是遗传多样性最高且与组外物种序列最近的^[67],因此,这些人群的祖先被认为是所有现代人的祖先。

卡恩等的线粒体研究得出了现代人出自非洲的结论^[68],但他们未给出必要的警告提醒人们注意中性理论前提的不确定性。后续研究也建立了Y染色体进化树,用类似的原理把树根定在了非洲^[66,69]。但是,出非洲说的单亲染色体进化树不能做到自洽,建树的前提假定是没有回复突变和重复突变,结果却包含大量的这类突变^[70,71]。所有中国人的Y染色体都带有M168T等位基因^[69],但若有回复突变,就无法区分新生型与祖先型,则M168T可能是欧亚人生理特征使然,而M168G有可能是亚洲现代人进入非洲与非洲古人杂交后所发生的非洲某些人群特异的变异,因为杂交的偶然性,大部分非洲人没有这个变异。

出非洲说有一个显著特征,那就是变得越来越复杂,因为就事论事的假设越来越多,需要被迫化解越来越多与出非洲说模型对立的新发现。比如,所有欧亚人的Y染色体祖先是F,来自东亚,是无争议的数据^[70,72],而出非洲说理应预测F在距离非洲最近的西亚中东,那么出非洲学派为了能自圆其说,就必须引入一个就事论事的假设:来自非洲的东亚人取代了同样来自非洲的中东人和欧洲人^[72]。另外,出非洲说学派基本不科普中性理论前提,因此造成了一个奇怪现象:虽然分子生物学学者基本都不认同中性理论假定,但他们并不了解其与出非洲说的因果关系,因此都没有反对出非洲说和其他基于中性理论假定的推论如种族无遗传基础。比如,知名学者、ENCODE(DNA元件百科全书)项目负责人之一伯尼(Birney)就不认可中性理论^[73],但却认可它的推论^[74]。

3.3.3 近期出东亚说

卡瓦利-斯福札团队在1983年利用线粒体构建了现代人出自东亚的模型^[1],他们提出了一种判断最古老人群的方法,依据线粒体最古老型或中心型应该具备高频率、在多个不同种族中存在等特征,得出了现代人出自东亚的结论。他们若假定分子钟成立,就可

以得出非洲起源的结论, 但作者明确认为分子钟不成立^[75]。卡瓦利-斯福札后来在1994年的一部专著里也肯定了东亚起源模型的数据和逻辑都没有问题^[76]。但该学说自从提出后就被长期埋没了, 在中文文献里没有被提及或造成任何影响^[77], 卡瓦利-斯福札后来也变成了出非洲说的支持者。

从遗传多态性角度研究人类起源, 必须在完全理解其所代表含义的前提之下。桑人的遗传多样性最高其实意味着上限平衡态水平与较低心智能力^[41]、较强免疫力^[78-81]、对有害突变的较高容忍度等因素有关^[70,82]。大部分基因组序列(95%)是受到自然选择的^[83], 因此不能被用来进行进化树分析。

人类遗传多样性有一个广为人知的特征, 即大部分的人类遗传变异(85%)是不同种族所共享的^[84]。这被中性理论解读为现代人起源的时间是较短的, 不同种族有着大量的混血杂交。但新研究显示, 不同种族的共享变异主要集中在快变序列, 反映的是趋同进化、独立突变和快变序列的遗传多样性的饱和平衡态, 与混血杂交关系不大, 因为杂交不会导致快变序列与慢变序列的显著差别^[70]。另外, 非洲人在快变序列上与不同种族的遗传距离呈现出等距离现象(图2), 即非洲人与欧洲人的遗传距离等于非洲人与东亚人的遗传距离, 并同时等于非洲人内部不同个体之间的遗传距离。但慢变序列则完全不同, 显示了正常预期结果, 非洲人内部不同个体间的遗传距离小于非洲人与其他族群的遗传距离^[70]。这一研究说明, 人群在快变序列的遗传多样性处于上限状态, 只有用慢变序列才能得到真实的遗传谱系树。

近年来, 有研究团队以遗传多样性上限理论为基础重新推演了现代人的进化, 重新发现了出东亚说, 并为多地区进化说找到了常染色体的分子证据(预印本论文第一版发表于2017年)^[70]。他们选择了慢变位点进行人群统计学分析^[28,70], 常染色体分析结果显示现代人的起源更接近于多地区进化说的结论, 然而不同的是, Y染色体和线粒体都起源于同一个地区, 即东亚南方。这些东亚现代人后来部分迁徙到其他地区并与当地人混血杂交, 导致当地人的古老型单亲染色体被现代型取代, 但东亚人的常染色体则被当地人部分替换, 导致了表型的本土化。杂交也促使这些其他地区人群的遗传多样性高于东亚人, 并因为回复突变而具备更多与大猿共享的等位基因。

出东亚说单亲染色体进化树的树根确定来自常理的推导, 认为原初的单倍型应该是现今人群中存在最多的, 并认为与古DNA较近的现生人单倍型比较接近树根^[1,70]。出东亚说模型认为, 单亲染色体的祖先型分别是线粒体单倍型R0和Y染色体单倍型F(图3)^[70]。新研究显示F在中国西南(云南、贵州、广西、四川)汉族男性人口中有一定比例(1.5%左右)的存在^[85]。

3.3.4 从第一性原理判断模型的合理性

从第一性原理或公理推导出的结论, 理应是最合理的结论。出东亚说较出非洲说有如下几个方面的合理性, 第一, 东亚环境复杂多样, 挑战性适中, 较适合人类进步性进化^[86]。第二, 东亚环境广大又相对隔离, 经常的混血杂交不利于进化出高级的新性状,

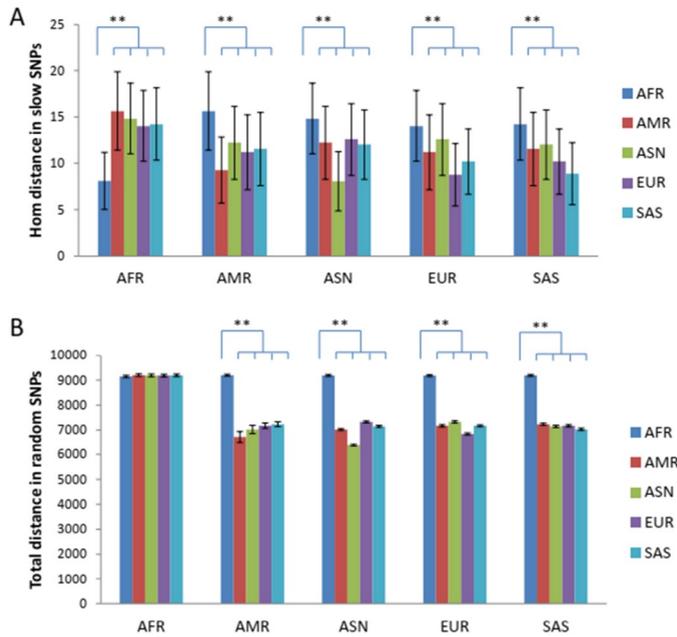


图2 非洲人与其他种族的遗传等距离现象^[70]

Fig.2 Africans are genetically equidistant to non-Africans in fast-evolving genes

A.用慢变变异SNPs计算的遗传距离（仅包括纯合子的差异导致的遗传距离）；B.用随机选择的变异SNPs（快变）计算的遗传距离（包括纯合子和杂合子的差异导致的遗传距离）。图中英文缩写AFR指非洲人，AMR指美洲人，ASN指东亚人，EUR指欧洲人，SAS指南亚人。**指 $p < 0.05$ 。

因为这种性状会被杂交湮没。第三，适合直立人进化成现代人的地区，也应适合现代人的进步进化，今天落后的族群所在地区也应是较不适合现代人最早出现的地区。出非洲说存在一定的自相矛盾之处，需要既假设欧亚不如非洲适合人类进步（在200万—5万年前这个时段），又假设欧亚比非洲更适合人类进步（在近5万年前这个时段），导致了欧亚地区的现生人在基因上更加远离大猿类。第四，最先成为现代人的群体应该有较低的遗传多样性，并保持至今，除非有意外事件的发生。第五，热带地区有较多病原因子，需要人群有较强的免疫力，因此不利于压缩人群的遗传多样性。第六，秩序来自压缩混杂混乱无序。来自不同种族的男女组合有较高的不育症发生率^[87]，显示混杂是受到了负选择，进化方向是朝向较低的遗传多样性和较纯的族群。第七，相较其他古人群体，最先成为现代人的群体的基因组应该是最远离黑猩猩并保持至今的。第八，心智能力等表型领先的群体，理应是最先跨入现代人门槛的群体，并维持至今，除非有意外事件的发生。

研究人类起源于哪个地区或种族是个重大问题，必然离不开探讨族群间和古人类化石之间的各种遗传和表型差异，但西方流行的种族主义定义却把相信种族有差异看作种族主义，这种定义既违背事实和科学，又干扰对现代人起源的探索，必须被修正^[88]。

3.3.5 现代人起源分子模型的验证

分子模型都需要一些前提假定，这些假定通常都带有一定的不确定性，因此需要被

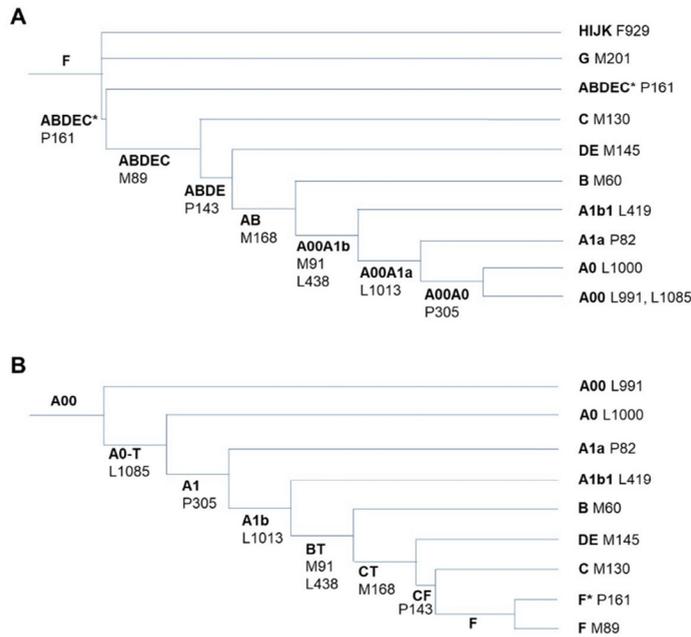


图3 人类Y染色体进化树模型

Fig.3 The Y chromosome phylogenetic tree of modern humans

A.出东亚说模型；B.出非洲说模型。仅显示主要分支和代表性 SNV（特定的核糖核酸序列），分支长度未按比例绘制。出非洲说模型里F单倍群内的所有单倍型的拓扑结构，包括G、H、I、J和K，在出东亚模型里也是相同模式，这些单倍型和所有末端单倍型在两个模型中具有相同的遗传关系和分支模式。

独立验证才能真正成立。目前可以用来做验证的公开数据有两套，一套是古DNA数据，另一套是Y染色体完整全测序数据。

3.3.5.1 古DNA验证

现代人起源模型的建立基本上来自对现生人DNA变异的分析，是对过去历史事件的一个猜想推测，而古DNA才真正见证了古代事件，才是对古代进化事件最有发言权的。出非洲说学派一直未用古DNA来验证出非洲说，而仅是用古DNA与现生人DNA进行合并构建了一个父系进化树^[89]，但这一方法所依赖的假定都是出非洲说一直依赖的中性理论假定，而具备说服力的验证需要数据和逻辑的独立性。

古DNA显示，古代样本应只是突变了一个单倍型的全部位点中的一部分，不管是基底单倍型也即树根或树干位置的单倍型，还是末端单倍型也即树叶位置的单倍型。这意味着末端单倍型的分化其实发生在基底单倍型完全完成所有相关位点的突变之前，而不是发生在中性理论预测的之后。这也意味着基底单倍型由部分突变到完全突变的过程是与末端单倍型的完全形成几乎同步的，且必须有趋同突变的参与，也即不同末端单倍型的群体发生了在同一位点出现同样突变的情况。这类似一棵活树的生长，树叶或树杈的生长与树干的生长是同步的。趋同突变在出非洲说模型中是不被容许发生的，但在出东亚说模型中是可以普遍发生的。另外，出非洲说模型中早期基底单倍型所含的突变，被

认为是在现代人出现后逐步分化完成的。但在出东亚说模型里,这些突变已经发生在第一个现代人的基因组里了。因此若出东亚说模型正确,这些出非洲说模型中早期基底单倍型的所有位点就会在古代样本里总是呈现全部突变了的状态。而若出非洲说模型正确,则古代样本就应该被发现在这些单倍型里是部分突变的状态。分析结果显示,出非洲说模型的基底单倍型在古代样本里都是全部突变的状态,而出东亚说模型的几乎所有基底单倍型都被发现是部分突变的状态^[90],这就意味着出东亚说模型里的基底单倍型是真实存在的。

古DNA研究领域较为常见的发现就是杂交混合普遍,以及古人与当地现生人的遗传连续性较弱^[91-93]。这些以中性理论为基础的研究基本无法验证出非洲说或任何单地起源说模型,因为这种验证需要证明起源地的古人与当地现生人是遗传连续的。如果遗传变异是中性变异,则按照哈代-温伯格定律,祖先群体的基因型频率应该与后代的无差异。但如果变异受到了自然选择,则祖先与后代因为环境不同就会呈现有较大的基因型频率差异,这种基因频率随时间的改变可以是非常快速的^[36]。目前大部分文献把功能变异假定为中性变异,但如果研究者用真正的中性变异进行分析,就会发现东亚地区的古人与东亚现生人有遗传连续性,但类似的情况在其他地区则没有,基本支持出东亚说的预测(黄石等,论文准备中)。其他研究者也发现,欧洲45000年前的古人与现生欧洲人没有关系,而是带有东亚人基因^[94],即便目前还没有直接证据说明这些欧洲古人出自东亚,这些发现也与出东亚说不矛盾。如果说东亚人来自这些西方古人的东迁,则需要一个比较牵强的假定:这些欧洲古人全部迁徙到东亚,没有在欧洲留下后裔。

3.3.5.2 Y染色体完整全测序验证

直到最近,人类Y染色体仅是进行了部分测序(53.8%),Y染色体的系统发育树是通过使用位于非重组区域的遗传变异构建的(图3)。然而,一项突破性进展是最近对43条人类Y染色体进行了完整测序^[95],揭示了大量来自假常染色体区域的变异,这就为人类起源模型的独立验证提供了宝贵的资源。有研究者探讨了不同单倍型之间共享这些变异的情况^[96],他们通过确认许多无争议的单倍型关系来验证了他们的方法,并揭示了一个明确的模式:出东亚说特有的单倍型得到了新数据的支持,出非洲说特有的单倍型则与新数据不吻合。例如,A0b和A1a彼此共享最多的变异显示彼此关系最近,这与两者同属于出东亚说的A00A1a单倍型一致,但在出非洲说模型里,A1a与A0b关系较远而与其他单倍型如B、C等较近(图3)。

3.3.5.3 证据汇总

支持出东亚说的各种独立证据包括如下十个方面。第一,化石证据链^[97,98],例如,3万多年前的北京山顶洞人是最具备现代人特征的全新世之前的古人化石^[99]。第二,东亚石器文化的本土连续证据链^[64]。第三,卡瓦利-斯福札的线粒体结论^[1]。第四,在遗传多样性上限理论指导下对现今DNA常染色体和单亲染色体的分子研究结论^[70]。第五,利用古DNA单亲染色体数据对出东亚说的验证^[71,90]。第六,利用Y染色体全测序数据对出东亚说

的验证^[96]。第七, 早期欧洲保加利亚古人主要是东亚基因的结论^[94]。第八, 欧洲和西亚的Y染色体来自东亚的结论^[72]。第九, 欧洲和西亚的X染色体可能来自东亚的结论^[100]。第十, 在过去的300万年内, 东亚环境更为复杂多样, 更适合人类的进步进化^[86]。

相比之下, 与出非洲说貌似一致的数据仅有两类。第一, 非洲有早期的部分现代的人类化石, 虽然北非有30万年前的杰贝尔依罗人(Jebel Irhoud)化石^[101], 但撒哈拉以南非洲没有早于25万前的现代人化石。而北非从古到今都是以欧亚人迁入为主, 已发现的古DNA揭示至少从15000年前就是如此^[102], 因此, 杰贝尔依罗人化石不能成为出非洲(南方)说的证据。其实, 几乎所有全新世之前的非洲古人化石都不是完全现代^[99,103], 非洲的古人类连续进化成现代人的化石证据链并不强, 没有被用来支持多地区进化说。第二, 在中性理论指导下对现生人群遗传多样性的解读。

4 结语

现代人起源的分子模型主要有两个不同学说, 即出非洲说和出东亚说, 最早都是于1983年被提出的。出非洲说于1987年被重新发现, 出东亚说则是于2017年被重新发现。前者根据中性理论推导而来, 后者以遗传多样性上限理论为基础。中性理论作为自然的解释性假说已经被大部分学者抛弃, 这对出非洲说意味着釜底抽薪。遗传多样性上限理论作为有实力取代中性理论的、能更好地解释自然的新进化学说, 已经获得了多项分析实验的直接验证。与此同时, 出东亚说也获得了多项独立检验的支持。

未来研究会从如下几个方面进一步验证遗传多样性上限理论及出东亚说。第一, 大规模的功能基因组学研究可能会表明几乎所有碱基都是有功能的。第二, 对更多种群或物种进行研究, 以进一步表明遗传多样性处于上限水平。第三, 通过研究更多的物种、更多的基因来进一步确立遗传等距离现象。第四, 通过确认基于遗传多样性上限理论的系统发育树, 比如出东亚说, 该理论本身也得到了验证。第五, 对认知能力的更好认识会进一步确立随机无序对秩序或认知的有害性。第六, 需要在分子水平上更好地理解大量遗传变异的集合效应是如何影响性状的。

总之, 对进化基础理论和现代人起源谜底的正确认识必然会逐步深入完善, 也必定会大力促进人类准确把握自身命运共同体的今天和未来。

致谢 谨以此文纪念遗传等距离现象被发现60周年。

参考文献

- [1] JOHNSON M J, WALLACE D C, FERRIS S D, et al. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns[J]. *Journal of Molecular Evolution*, 1983, 19(3/4): 255-271.
- [2] LUO D, HUANG S. The genetic equidistance phenomenon at the proteomic level[J]. *Genomics*, 2016, 108(1): 25-30.

- [3] ZUCKERKANDL E, PAULING L. Molecular disease, evolution, and genic heterogeneity[M]//KASHA M, PULLMAN B. Horizons in biochemistry. New York: Academic Press, 1962: 189-225.
- [4] MARGOLIASH E. Primary structure and evolution of cytochrome c[J]. Proceedings of the National Academy of Sciences, 1963, 50(4): 672-679.
- [5] KUMAR S. Molecular clocks: four decades of evolution[J]. Nature Reviews Genetics, 2005, 6(8): 654-662.
- [6] YUAN D J, HUANG S. Genetic equidistance at nucleotide level[J]. Genomics, 2017, 109(3/4): 192-195.
- [7] HUANG S. The genetic equidistance result of molecular evolution is independent of mutation rates[J]. Journal of Computer Science and Systems Biology, 2008, 1: 92-102.
- [8] BERGERON L A, BESENBACHER S, ZHENG J, et al. Evolution of the germline mutation rate across vertebrates[J]. Nature, 2023, 615(7951): 285-291.
- [9] KIMURA M. Evolutionary rate at the molecular level[J]. Nature, 1968, 217(5129): 624-626.
- [10] KIMURA M. The rate of molecular evolution considered from the standpoint of population genetics[J]. Proceedings of the National Academy of Sciences, 1969, 63(4): 1181-1188.
- [11] HUBBY J L, LEWONTIN R C. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*[J]. Genetics, 1966, 54(2): 577-594.
- [12] KREITMAN M. The neutral theory is dead. Long live the neutral theory[J]. Bioessays, 1996, 18(8): 678-683.
- [13] KERN A D, HAHN M W. The neutral theory in light of natural selection[J]. Molecular Biology and Evolution, 2018, 35(6): 1366-1371.
- [14] LEFFLER E M, BULLAUGHEY K, MATUTE D R, et al. Revisiting an old riddle: what determines genetic diversity levels within species?[J]. PLoS Biology, 2012, 10(9).
- [15] DUNHAM I, KUNDAJE A, ALDRED S F, et al. An integrated encyclopedia of DNA elements in the human genome[J]. Nature, 2012, 489(7414): 57-74.
- [16] MATTICK J S. A Kuhnian revolution in molecular biology: most genes in complex organisms express regulatory RNAs[J]. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology, 2023, 45(9).
- [17] HUANG S. Inverse relationship between genetic diversity and epigenetic complexity[J]. Nature Precedings, 2009.
- [18] FISHER R A. The genetical theory of natural selection[M]. Oxford: Clarendon Press, 1930.
- [19] ORR H A. Adaptation and the cost of complexity[J]. Evolution, 2000, 54(1): 13-20.
- [20] PARTS L, BATTÉ A, LOPES M, et al. Natural variants suppress mutations in hundreds of essential genes[J]. Molecular Systems Biology, 2021, 17(5).
- [21] HUANG S. New thoughts on an old riddle: what determines genetic diversity within and between species? [J]. Genomics, 2016, 108(1): 3-10.
- [22] 胡涛波, 龙孟平, 袁德健, 等. 遗传等距离现象: 分子钟和中性理论的误读及其近半世纪后的重新解谜[J]. 中国科学(生命科学), 2013, 43(4): 275-282.
- [23] 黄益敏, 夏梦颖, 黄石. 遗传多样性上限假说所揭示的进化历程[J]. 遗传, 2013, 35(5): 599-606.
- [24] BARTON N H. The “New Synthesis” [J]. Proceedings of the National Academy, 2022, 119(30).
- [25] BICKEL D. Testing hypotheses of molecular evolution[M]//BICKEL D R. Phylogenetic trees and molecular evolution. Cham: Springer, 2022: 71-78.
- [26] 黄石, 朱作斌. 系统生物学与进化理论[M]//朱作斌, 张潇, 王亮. 系统生物学. 南京: 东南大学出版社, 2022.
- [27] MARTIN M K. The wrong ape for early human origins: the chimpanzee as a skewed ancestral model[M]. Lanham, MD: Lexington Books, 2023.
- [28] 黄石. 用与种系发生相关的 DNA 序列解读人类起源[G]//席焕久, 刘武, 陈昭. 21 世纪中国人类学的发展. 北京: 知识产权出版社, 2015.
- [29] DENTON M. Evolution: still a theory in crisis[M]. Seattle, WA: Discovery Institute Press, 2016.
- [30] HO M W. Development and evolution revisited[M]//HOOD K E, HALPERN C T, GREENBERG G, et al. Handbook of developmental science, behavior, and genetics. Malden, MA: Wiley Blackwell, 2010.
- [31] HUANG S. Histone methylation and the initiation of cancer[M]//TOLLEFSBOL T. Cancer Epigenetics. Boca Raton: CRC Press, 2008: 109-150.

- [32] KUDERNA L F K, GAO H, JANIAC M C, et al. A global catalog of whole-genome diversity from 233 primate species[J]. *Science*, 2023, 380(6648): 906-913.
- [33] LOGSDON G A, ROZANSKI A N, RYABOV F, et al. The variation and evolution of complete human centromeres[J]. *Nature*, 2024, 629:136-145.
- [34] LU J, WU C I. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee[J]. *Proceedings of the National Academy of Sciences*, 2005, 102(11): 4063-4067.
- [35] SHEN X K, SONG S L, LI C, et al. Synonymous mutations in representative yeast genes are mostly strongly non-neutral[J]. *Nature*, 2022, 606(7915): 725-731.
- [36] RUDMAN S M, GREENBLUM S I, RAJPUROHIT S, et al. Direct observation of adaptive tracking on ecological time scales in *Drosophila*[J]. *Science*, 2022, 375(6586).
- [37] HUANG S. The overlap feature of the genetic equidistance result—a fundamental biological phenomenon overlooked for nearly half of a century[J]. *Biological Theory*, 2010, 5(1): 40-52.
- [38] HUANG S. Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers[J]. *Science China Life Sciences*, 2012, 55(8): 709-725.
- [39] MAO Y F, HARVEY W T, PORUBSKY D, et al. Structurally divergent and recurrently mutated regions of primate genomes[J]. *Cell*, 2024, 187(6): 1547-1562.
- [40] BICKEL D R. A generalization of null hypothesis significance testing with applications to replication failures, molecular evolution models, and bounded parameter spaces[EB/OL]. (2021-07-22). <https://doi.org/10.5281/zenodo.5123388>.
- [41] WANG M R, HUANG S. The collective effects of genetic variants and complex traits[J]. *Journal of Human Genetics*, 2023, 68(4): 255-262.
- [42] WAINBERG M, FORDE N J, MANSOUR S, et al. Genetic architecture of the structural connectome[J]. *Nature Communications*, 2024, 15(1): 1962.
- [43] CHEN C Y, TIAN R Y, GE T, et al. The impact of rare protein coding genetic variation on adult cognitive function[J]. *Nature Genetics*, 2023, 55(6): 927-938.
- [44] ZHU Z B, YUAN D J, LUO D H, et al. Enrichment of minor alleles of common SNPs and improved risk prediction for Parkinson's disease[J]. *PLoS ONE*, 2015, 10(7).
- [45] LEI X Y, YUAN D J, ZHU Z B, et al. Collective effects of common SNPs and risk prediction in lung cancer[J]. *Heredity*, 2018, 121: 537-547.
- [46] HE P, LEI X Y, YUAN D J, et al. Accumulation of minor alleles and risk prediction in schizophrenia[J]. *Scientific Reports*, 2017, 7(1): 11661.
- [47] GAO H, HAMP T, EDE J, et al. The landscape of tolerated genetic variation in humans and Primates[J]. *Science*, 2023, 380(6648).
- [48] GIBBS R A, ROGERS J, KATZE M G, et al. Evolutionary and biomedical insights from the rhesus macaque genome[J]. *Science*, 2007, 316(5822): 222-234.
- [49] SANJAK J S, SIDORENKO J, ROBINSON M, et al. Evidence of directional and stabilizing selection in contemporary humans[J]. *Proceedings of the National Academy of Sciences*, 2017, 115: 151-156.
- [50] YUAN D J, ZHU Z B, TAN X H, et al. Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms[J]. *Science China Life Sciences*, 2014, 57(9): 876-888.
- [51] HORTON C A, ALEXANDARI A M, HAYES M G B, et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression[J]. *Science*, 2023, 381(6664).
- [52] CAMELLATO B R, BROSH R, ASHE H J, et al. Synthetic reversed sequences reveal default genomic states[J]. *Nature*, 2024, 628: 373-380.
- [53] COUCE A, LIMDI A, MAGNAN M, et al. Changing fitness effects of mutations through long-term bacterial evolution[J]. *Science*, 2024, 383(6681).
- [54] WANG M R, WANG D P, YU J, et al. Enrichment in conservative amino acid changes among fixed and standing missense variations in slowly evolving proteins[J]. *PeerJ*, 2020, 8.
- [55] WILSON A C, SARICH V M. A molecular time scale for human evolution[J]. *Proceedings of the National*

- Academy of Sciences, 1969, 63(4): 1088-1093.
- [56] DAVER G, GUY F, MACKAYE H T, et al. Postcranial evidence of late Miocene hominin bipedalism in Chad[J]. *Nature*, 2022, 609(7925): 94-100.
- [57] SCALLY A, DUTHEIL J Y, HILLIER L W, et al. Insights into hominid evolution from the gorilla genome sequence[J]. *Nature*, 2012, 483(7388): 169-175.
- [58] Hobolth, A., et al., Incomplete lineage sorting patterns among human, chimpanzee and orangutan suggest recent orangutan speciation and widespread selection[J]. *Genome Research*, 21: 349-356.
- [59] RIVAS-GONZÁLEZ I, ROUSSELLE M, LI F, et al. Pervasive incomplete lineage sorting illuminates speciation and selection in primates[J]. *Science*, 2023, 380(6648).
- [60] PEPPE D, COTE S, DEINO A, et al. Oldest evidence of abundant C₄ grasses and habitat heterogeneity in eastern Africa[J]. *Science*, 2023, 380: 173-177.
- [61] WOLPOFF M H, WU X Z, THORNE A G. Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from East Asia[G]/SMITH F H, SPENCER F. The origins of modern humans: a world survey of the fossil evidence. New York: Alan R. Liss, 1984.
- [62] WU X Z. On the origin of modern humans in China[J]. *Quaternary International*, 2004, 117(1): 131-140.
- [63] 吴新智. 中国和欧洲早期智人的比较研究[J]. *人类学学报*, 1988, 7(4): 287-293.
- [64] 高星. 更新世东亚人群连续演化的考古证据及相关问题论述[J]. *人类学学报*, 2014, 33(3): 237-253.
- [65] 雷晓云, 袁德健, 张野, 等. 基于DNA分子的现代人起源研究35年回顾与展望[J]. *人类学学报*, 2018, 37(2): 270-283.
- [66] UNDERHILL P A, SHEN P, LIN A A, et al. Y chromosome sequence variation and the history of human populations[J]. *Nature Genetics*, 2000, 26: 358-361.
- [67] MALLICK S, LI H, LIPSON M, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations[J]. *Nature*, 2016, 538(7624): 201-206.
- [68] CANN R L, STONEKING M, WILSON A C. Mitochondrial DNA and human evolution[J]. *Nature*, 1987, 325(6099): 31-36.
- [69] KE Y, SU B, SONG X, et al. African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes[J]. *Science*, 2001, 292(5519): 1151-1153.
- [70] YUAN D J, LEI X Y, GUI Y Y, et al. Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA[EB/OL]. <https://doi.org/10.1101/101410>.
- [71] 张野, 黄石. 古DNA的新发现支持现代人东亚起源说[J]. *人类学学报*, 2019, 38(4): 491-498.
- [72] HALLAST P, AGDZHOYAN A, BALANOVSKY O, et al. A Southeast Asian origin for present-day non-African human Y chromosomes[J]. *Human Genetics*, 2021, 140(2): 299-307.
- [73] BHATTACHARJEE Y. The vigilante[J]. *Science*, 2014, 343(6177): 1306-1309.
- [74] KOZLOV M. ‘All of Us’ genetics chart stirs unease over controversial depiction of race[J]. *Nature*, 2024.
- [75] 黄石. 对“新证据下的现代人起源模型”一文的商榷[EB/OL]. (2022-10-01). <https://mp.weixin.qq.com/s/FOgic4wrHN3FY44j69M12g>.
- [76] CAVALLI-SFORZA L L, MENOZZI P, PIAZZA A. The history and geography of human genes[M]. Princeton, NJ: Princeton University Press, 1994.
- [77] 黄石. 上世纪两篇论证现代人出东亚的论文在中国的影响[EB/OL]. (2023-02-22). <https://mp.weixin.qq.com/s/9FwGStmx6k8MfgnumjnL1A>.
- [78] HARRISON G F, SANZ J, BOULAIS J, et al. Natural selection contributed to immunological differences between hunter-gatherers and agriculturalists[J]. *Nature Ecology & Evolution*, 2019, 3(8): 1253-1264.
- [79] AJAYI T, RAI P, SHI M, et al. Race-specific association of an *IRGM* risk allele with cytokine expression in human subjects[J]. *Scientific Reports*, 2023, 13(1).
- [80] HAWLEY D M, SYDENSTRICKER K V, KOLLIAS G V, et al. Genetic diversity predicts pathogen resistance and cell-mediated immunocompetence in house finches[J]. *Biology Letters*, 2005, 1(3): 326-329.
- [81] O’NEILL M B, QUACH H, POTHLICHER J, et al. Single-cell and bulk RNA-sequencing reveal differences in monocyte susceptibility to influenza A virus infection between Africans and Europeans[J].

- Frontiers in Immunology, 2021, 12.
- [82] LEK M, KARCEWSKI K J, MINIKEL E V, et al. Analysis of protein-coding genetic variation in 60, 706 humans[J]. Nature, 2016, 536: 285-291.
- [83] POUYET F, AESCHBACHER S, THIÉRY A, et al. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences[J]. eLife, 2018, 7.
- [84] LEWONTIN R C. The apportionment of human diversity[G]//DOBZHANSKY T, HECHT M K, STEERE W C. Evolutionary biology: vol.6. New York: Springer, 1972: 381-398.
- [85] WANG Z Y, WANG M G, LIU K J, et al. Paternal genomic resources from the YanHuang cohort suggested a Weakly-Differentiated Multi-source Admixture model for the formation of Han's founding ancestral lineages[EB/OL]. (2023-11-13). <https://doi.org/10.1101/2023.11.08.566335>.
- [86] ZELLER E, TIMMERMANN A, YUN K S, et al. Human adaptation to diverse biomes over the past 3 million years[J]. Science, 2023, 380(6645): 604-608.
- [87] QIAN Z C, LICHTER D T. Racial pairings and fertility: do interracial couples have fewer children?[J]. Journal of Marriage and Family, 2021, 83(4): 961-984.
- [88] 黄石. 质疑美西方对种族主义的定义[EB/OL]. (2023-03-05). https://mp.weixin.qq.com/s/1BpVfA_XLH8qCVaKzOHTdQ.
- [89] WOHNS A W, WONG Y, JEFFERY B, et al. A unified genealogy of modern and ancient genomes[J]. Science, 2022, 375(6583).
- [90] XIA Z Y, CHEN H Y, ZHANG Y H, et al. Ancient uniparental DNAs in distinguishing the competing theories of molecular evolution and modern human origins[EB/OL]. (2021-11-17). <https://doi.org/10.21203/rs.3.rs-1087730/v1>.
- [91] LAZARIDIS I, PATTERSON N, MITTNIK A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans[J]. Nature, 2014, 513(7518): 409-413.
- [92] BRAMANTI B, THOMAS M G, HAAK W, et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers[J]. Science, 2009, 326(5949): 137-140.
- [93] SKOGLUND P, MATHIESON I. Ancient genomics of modern humans: the first decade[J]. Annual Review of Genomics and Human Genetics, 2018, 19: 381-404.
- [94] HAJDINJAK M, MAFESSONI F, SKOV L, et al. Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry[J]. Nature, 2021, 592(7853): 253-257.
- [95] HALLAST P, EBERT P, LOFTUS M, et al. Assembly of 43 human Y chromosomes reveals extensive complexity and variation[J]. Nature, 2023, 621(7978): 355-364.
- [96] HUANG S. Examining models of modern human origins through the analysis of 43 fully sequenced human Y chromosomes[EB/OL]. (2023-11-13). <https://doi.org/10.1101/2023.11.09.566475>.
- [97] 刘武, 吴秀杰. 中更新世晚期中国古人类化石的形态多样性及其演化意义[J]. 人类学学报, 2022, 41(4): 563-575.
- [98] WU X J, PEI S W, CAI Y J, et al. Morphological and morphometric analyses of a late Middle Pleistocene hominin mandible from Hualongdong, China[J]. Journal of Human Evolution, 2023, 182.
- [99] HOWELLS W W. Skull shapes and the map: craniometric analyses in the dispersion of modern *Homo*[M]. Cambridge, Mass.: Peabody Museum Press, 1990.
- [100] SKOV L, COLL MACIÀ M, LUCOTTE E A, et al. Extraordinary selection on the human X chromosome associated with archaic admixture[J]. Cell Genomics, 2023, 3(3).
- [101] HUBLIN J-J, BEN-NCER A, BAILEY S E, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*[J]. Nature, 2017(7657), 546: 289-292.
- [102] VAN DE LOOSDRECHT M, BOUZOUGGAR A, HUMPHREY L, et al. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations[J]. Science, 2018, 360(6388): 548-552.
- [103] PEARSON O M. Statistical and biological definitions of "anatomically modern" humans: suggestions for a unified approach to modern morphology[J]. Evolutionary Anthropology: Issues, News, and Reviews, 2008, 17(1): 38-48.

The molecular models of modern human origins

HUANG Shi

Center for Medical Genetics, School of Life Sciences, Central South University, Changsha 410078

Abstract In the early 1960s, researchers started the field of molecular evolution based on the genetic distance phenomenon of protein sequences among species, and proposed a molecular evolutionary theory different from the natural selection theory, the molecular clock hypothesis and the neutral theory. This theory holds that the genetic distance of gene sequences can be used to infer the phylogenetic relationship between different species. The greater the genetic distance, the farther the phylogenetic relationship and the longer the evolutionary time. Researchers then deduced a molecular model of the origin of modern humans based on this theory, the recent out of Africa hypothesis. African populations have the highest genetic diversity among all racial groups, which was interpreted to mean that Africans have a more ancient most recent common ancestor than other populations. Modern humans are considered to have first appeared in Africa who then migrated to Eurasia and largely replaced the indigenous populations. Although the neutral theory is a very valuable null hypothesis, it cannot fully explain the phenomenon of genetic diversity, which therefore deems the out-of-Africa model, at best, uncertain. In 2008, a new theory of molecular evolution, known as the maximum genetic diversity theory, was published, offering a reinterpretation of the phenomenon of genetic distances. A large genetic distance does not necessarily mean a long evolutionary time, but can also reflect a large phenotypic difference. The increase in genetic distance over time is not infinite, as implied by the neutral theory, but has an upper limit, which is mainly determined by the phenotypic complexity of the species. Several lines of tests show that the genetic distances or genetic diversities are largely at the upper limit levels. Based on the maximum genetic diversity theory, new research has independently re-discovered the out of East Asia model of modern humans that was first proposed in 1983. It also provides autosomal DNA support for the multiregional hypothesis. Multiple lines of tests, including ancient DNA analyses, lend robust support to the out of East Asia model as the more accurate representation of modern human origins. As exploration of the fundamental theories of evolution deepens, humanity's understanding of its own origins is destined to become clearer and more refined.

Keywords molecular clock; neutral theory; maximum genetic diversity theory; out of Africa; out of East Asia