古DNA研究35年回顾与展望

盛桂莲*, 赖旭龙, 袁俊霞, 侯新东, 宋凌峰

中国地质大学生物地质与环境地质国家重点实验室, 武汉 430074

* E-mail: glsheng@cug.edu.cn

收稿日期: 2016-03-30; 接受日期: 2016-09-27; 网络版发表日期: 2016-11-03

国家自然科学基金项目(批准号: 41672017, 41472014, 41202013)资助

摘要 自1980年湖南医学院学者尝试从马王堆汉代古人类遗存获取核酸分子开始, 古DNA研究至今已经历了35年时间. 近年来, 随着高通量测序技术平台的开发应用及对小片段古DNA分子捕获能力的加强, 古DNA研究走出了单纯依靠线粒体基因组及少量核基因的时代, 在全基因组测序方面取得重要突破, 古基因组数据在人类起源、大型绝灭哺乳动物的迁移演化、动植物家养驯化过程及其对人类社会生活模式的启示、生物与环境的相互关系等诸多方面发挥了独特作用. 本文以第二代测序技术问世之后古DNA研究流程为主线, 系统回顾古DNA各阶段的进展以及近年来取得的突破, 阐明古DNA数据的应用领域, 并基于现有研究实例对古DNA将来可能的发展前景提出展望.

关键词 古DNA, 高通量测序, 全基因组

1 引言

DNA作为生物遗传信息的分子载体, 能准确地从分子水平反映生物表观性状的异同, 进而反映生物类群之间的谱系演化关系. 在已死亡或绝灭的生物中, 有些个体或群体由于保存环境条件较理想, 其组织中有少量遗传信息会以高度片段化、微量的古DNA(ancient DNA, aDNA)形式保存下来, 从而为死亡或绝灭生物的遗传组成提供实时分子数据. 古DNA研究诞生至今, 在生物分子系统演化、分子水平的种群遗传学与谱系地理学、分子演化速率、人类的起源和演化、动植物的家养驯化过程等方面发挥着越来越重要的作用, 对演化生物学、古生物学、地质

学、考古学和人类学等学科产生了越来越显著的影响. 本文主要介绍第二代测序技术问世以来的古DNA研究方法, 回顾古DNA的发展历史及近些年来的突破性进展, 并对未来可能的发展前景进行预测.

2 古DNA研究简史

1980年我国湖南医学院的研究者对约2000年前长沙马王堆汉代女尸进行了古核酸提取(湖南医学院, 1980). 尽管该项研究因没有对提取的古核酸进行序列测定而未得到国际公认, 但这是从古代生物中提取古DNA的最早尝试. 本文以此为古DNA研究起始, 依据古DNA序列数据的多少及研究者对序列数据的应用, 将古DNA研究划分为三个阶段(图1).

中文引用格式: 盛桂莲, 赖旭龙, 袁俊霞, 侯新东, 宋凌峰. 2016. 古DNA研究35年回顾与展望. 中国科学: 地球科学, 46: 1564–1578, doi: 10.1360/N072015-00539
英文引用格式: Sheng G L, Lai X L, Yuan J X, Hou X D, Song L F. 2016. Ancient DNA: Achievements in the first 35 years and its prospective (in Chinese). Scientia Sinica Terrae, 46: 1564–1578, doi: 10.1360/N072015-00539

图 1 古DNA研究阶段划分及重要研究实例

2.1 古DNA萌芽阶段(1980~1995年)

在前述我国马王堆汉代女尸古核酸研究之后,至Mullis于1986年提出聚合酶链式反应(Polymerase Chain Reaction, PCR)技术之前,仅有两例古DNA研究:其一为绝灭动物斑驴229bp线粒体DNA研究(Higuchi等, 1984),国际上一般认为该研究标志着古DNA的起始;其二为埃及木乃伊3.4kb的核酸研究(Pääbo, 1985),后经证实为现代DNA的污染(Pääbo等, 2004)。PCR技术诞生之后,众多学者通过对目标片段的靶向PCR扩增和短片段拼接方法,先后从博物馆干化标本、洞穴沉积物和永久冻土中保存的古代不同动植物材料及古人类标本中提取得到古DNA(Doran等, 1986; Pääbo 1989; Golenberg等, 1990; DeSalle等, 1992; Cano等, 1993),研究成果引发了分子生物学家、古生物学家、考古学家的极大兴趣,在国际上掀起了化石古DNA的研究和讨论热潮(Pääbo, 1991)。在萌芽阶段,学者以从不同生物材料中成功提取到古DNA为目的。

2.2 古DNA发展阶段(1996~2005年)

随着古DNA提取方法的改进以及规范的古DNA研究标准的建立和完善(杨东亚, 2003; Gilbert等, 2005; Kalmar等, 2000; Taylor, 1996),古DNA研究成果所涉及

的生物种类逐渐呈现出多样化特征,如:绝灭生物地獾(Poinar, 1998)、冰川时期的棕熊(Leonard等, 2000)、新西兰恐鸟(Cooper等, 2001)等。在此阶段,研究者开始侧重于将获得的线粒体基因或少数核基因古DNA序列用于解决传统古生物学研究中存在争议的科学问题,所取得的一系列进展主要是基于古DNA提取方法的改进和高灵敏度的PCR技术。然而,正是由于PCR的高灵敏度,外源DNA因其分子拷贝数远远超出高度片段化、微量存在的古DNA分子易被优先扩增,部分古DNA研究成果不能得到重复性实验的验证(图1),学界对序列结果(尤其是涉及古人类及微生物古DNA研究内容时)可靠性的质疑曾一度使古DNA研究陷入困境(Hofreiter 等, 2001a; Wayne等, 1999)。

2.3 古DNA成熟阶段(2006年至今)

以测序技术的革新为引领,伴随着古DNA提取、测序文库构建、DNA富集等方法的不断优化,2006年之后的古DNA研究进入到古基因组时代,研究者开始拼装并解析绝灭生物的线粒体及核基因组。在此阶段,冻土保存的猛犸象是率先被用于第二代测序(Poinar等, 2006)及核功能基因多态性分析的绝灭生物(Römpler等, 2006),而博物馆保存的4000年前爱斯基

摩人头发样品核基因组单核苷酸多态性位点的分析,则是第二代测序技术在古人类基因组测序中的应用先例(Rasmussen等, 2010)。之后的几年中,古人类研究与第二代测序技术的融合令人瞩目,除尼安德特人基因组草图(Green等, 2010)和全基因组(Prufer等, 2014)外,旧石器时代晚期丹尼索沃人(Reich等, 2010)、21世纪早期澳洲土著人(Rasmussen等, 2011)、4万年前中国田园洞人(Fu等, 2013)、旧石器时代晚期西伯利亚中南部古人类(Raghavan等, 2014)等古基因组的研究,为丰富世界各地古人类的迁移及是否存在基因交流提供了必不可少的分子依据。尤为重要的是,古DNA提取及测序文库建立方法的改进,使得古DNA研究突破了10万年这一理论保存年限:研究者先后从洞穴保存的材料中得到40万年前洞熊及古人类线粒体全基因组(Dabney等, 2013; Meyer等, 2014),而在冻土环境下,已得到核基因组的马科动物材料则已回溯到70万年前的中更新世早期(Orlando等, 2013)。

3 古DNA保存的影响因素

DNA损伤修复机制随着生物死亡而崩解,水解、氧化及微生物作用将直接造成DNA双螺旋主链的断裂及分子降解(Hofreiter等, 2001b)。因此,样品中古DNA分子的保存状况与样品年代并无直接联系,埋藏环境的温度、pH值、离子强度、腐植酸含量等直接影响古DNA的降解速度。一般而言,低温尤其是冻土环境有利于古DNA的保存。Allentoft等(2012)曾用158个有详尽测年数据的新西兰恐鸟骨骼样品,对相同pH值、不同温度条件下的古DNA半衰期进行了研究,发现在13.1℃的埋藏温度下,DNA的降解速度比理论推算值慢近400倍。结合全球各地温度历史记录和土壤深度矫正,Hofreiter等(2014)建立了模型以显示全球范围内地表和埋藏环境下不同长度(150和25bp)的古DNA保存概率,经与部分地点已有的DNA降解速率数据对比,发现运用模型推算出的古DNA保存情况与实际相符,为古DNA的样品采集提供了一定参考。此外,近年来有研究表明,埋藏以后骨骼石化过程中晶体的形成,有利于保护内源古DNA分子免受环境因素影响,从而减缓其降解过程,显示出材料的孔隙度大小对古DNA保存的影响(Ginolhac等, 2012)。正因如此,对古代骨骼材料而言,对首次提取物的再次EDTA消化,可以释放

骨骼晶体中的古DNA分子,从而提高内源古DNA序列在全部测序结果中的比例。

鉴于材料的珍稀性,实验前如何判定博物馆馆藏标本或考古发掘材料中是否有古DNA分子残留尤为必要。对于动物骨骼及牙齿材料,曾有学者认为天冬氨酸的外消旋过程与DNA分子降解速率相近,因此将发生外消旋作用的天冬氨酸量用作古DNA含量的可信赖指标(Poinar等, 1996)。但近年来基于多达91个骨骼和牙齿样品的研究发现:在样品中左旋/右旋天冬氨酸的比值普遍较低(0.02~0.17)的情况下,该比值与古DNA的扩增成功率并无线性关系(Collins等, 2009);对于沉积物及植物材料,目前尚无确定指标能指示其古DNA保存情况。因此,在实验材料的选择上,研究者更多地是借鉴地层学、伴生动物群测年信息及工作经验而对材料进行取舍。

4 古DNA序列数据的获取

在古DNA研究的前两个阶段(2006年以前),学者们采用“样品采集-古DNA提取-PCR扩增-分子克隆-序列测定-序列拼接-序列分析”这一研究流程来获取古DNA数据。第二代测序技术(Next Generation Sequencing, NGS)及第三代单分子测序技术(true Single Molecule Sequencing, tSMS)问世之后,除“古DNA提取”在技术和方法上相对而言有一定延续性外,其他环节中有些有了新的控制标准,如“样品采集”的污染控制可借由后续分析来弥补;有些不再是必备环节,如“分子克隆”;有些在原理上得到了革新,如“PCR扩增”、“序列测定”;此外,测序前新引入了“测序文库建立”等环节。关于第二代测序技术问世之前的古DNA研究流程和污染控制方法,笔者之前已有综述文章对其进行系统介绍和分析(盛桂莲等, 2009a),本文主要介绍古DNA研究第三阶段序列数据的获取步骤、方法及其应用。

4.1 古DNA提取

古DNA提取方法主要有两类:基于有机溶剂的“萃取法”和基于硅粒的“吸附法”。两类方法都包含用加入蛋白酶K的EDTA缓冲液进行脱钙质、去蛋白质的预处理过程。在后续提取操作中,萃取法利用了核酸分子在有机溶剂和水相的溶解度差异,先用苯酚、氯仿和异丁醇去除蛋白质,回收水相后用乙醇沉淀DNA,

最后用带滤膜的过滤浓缩柱对古DNA进行纯化; 吸附法利用了SiO₂颗粒能够吸附核酸分子的特性, 不同研究者根据材料特性分别选用了高浓度的异硫氰酸胍(GuSCN)、氯化钠、或二者结合使用的吸附缓冲液, 将裂解后释放出的古DNA吸附到SiO₂颗粒上, 高速离心并洗脱后即可用于后续研究. SiO₂吸附法可以排除材料中的PCR抑制剂, 在古DNA提取的同时达到纯化古DNA的目的. 继2007年Rohland和Hofreiter(2007a)发表对动物骨骼和牙齿样品尤为有效的SiO₂吸附法以后, Dabney等(2013)在2013年从三方面着手对原方案进行了改进: (1) 改变吸附缓冲液的成分, 在原有基础上加入了醋酸钠(90mM, pH5.2)和苯酚(40%, vol/vol); (2) 提高吸附缓冲液与提取缓冲液的体积比(原为4:1, 改为13:1); (3) 用商业化生产的硅胶膜离心吸附柱替代原有SiO₂悬浮液. 改进后的提取方法, 使得短至35bp的短片段具备与长片段古DNA分子同等的提取效率, 能用于非冻土环境或保存年限较长、DNA降解严重的古代材料.

即使是保存情况良好的材料, 其中的目标古DNA分子数量也极其有限, 如何最大限度地提高对已有材料的提取效率, 在古DNA实验中具有举足轻重的作用. 然而, 有助于现代DNA从材料中释放从而提高提取效率的方法, 如高温处理、去离子试剂的使用等, 对古DNA而言却会造成其分子的进一步损伤, 降低古DNA的获取量(Rohland和Hofreiter, 2007b). 此外, 古代材料中通常含有PCR抑制剂, 如果在古DNA提取物中不能被有效去除, 对后续PCR特异性扩增也会有一定的阻滞作用(Hänni等, 1995; King等, 2009). 有研究表明, 基于古人类或古代动物的毛发或骨骼/牙齿材料的已有研究中, 测序结果显示大部分研究得到的内源古DNA模板只占总DNA的一小部分(Der Sarkissian等, 2015; Green等, 2010; Orlando等, 2013), 针对目标基因组的提取方法及技术优化仍有待提升.

4.2 古DNA测序文库建立

第二代测序技术开始应用于古DNA研究之后, “测序文库建立”替代“PCR扩增”成为提取之后的必要步骤: 通过随机打断基因组获得DNA文库片段, 并在片段两端连上接头序列, 经变性成单链DNA分子, 以用作DNA富集模板(Margulies等, 2005).

相对于现代材料, 长度一般不超过500bp的古DNA

样品无需超声波或酶切将基因组打断成为短片段, 可直接连接具样品识别特征的接头, 制备古DNA文库. 追踪2013年之前的高通量古DNA测序研究不难发现: 其全基因组数据组装都是建立在>40bp的序列读长基础上. 而在理论上, 古DNA分子的拷贝数会随着序列长度的降低而增加(Allentoft等, 2012), 小片段DNA如能得到有效提取且能被测序仪识别, 应该是所得测序数据中的主体; 其在测序数据中的缺失, 主要是由于在测序文库建立环节由于方法的限制而被剔除. 基于此, Meyer等(2012)针对古DNA分子高度片段化、普遍存在碱基损伤等特征, 优化出专用于古基因组的单链DNA文库建立方法. 该方法优于双链DNA文库建立方法的原因在于: 首先, 通过在古DNA模板分子3'-端接上生物素化的接头, 将目标古DNA分子吸附到含生物素的磁珠上, 避免了以去除多余接头及引物为目的的清洗步骤带来的小片段分子损失; 其次, 古DNA分子模板通常在双链的不同位置含有碱基损伤, 单链DNA文库建立方法为这类分子提供了变性后被接上接头并得到扩增的可能, 避免了双链DNA文库建立方法对这类分子的剔除; 最后, 双链DNA分子中的一条链上存在的末端修饰, 通常会阻止接头与目标分子的结合, 从而影响了文库中对这类分子的捕获, 而单链DNA文库建立过程中, 双链中末端具修饰的那条链本身虽仍然受限, 但其互补链却能被接头连上并被磁珠捕捉到(Gansauge和Meyer, 2013). 单链DNA文库建立方法先被Meyer等用于获取已绝灭的古人类丹尼索沃人高覆盖率的基因组数据, 之后又先后被用于西班牙同一洞穴中的中更新世(40万年)洞熊及古人类线粒体全基因组研究(Dabney等, 2013; Meyer等, 2014), Orlando等(2013)更是将此方法用于冻土环境下中更新世早期(70万年)马科动物的核基因组获取, 由此证明这一针对受损严重的短片段DNA的提取及文库建立方法不啻为古基因组测序中的一次重要技术革新(Dabney等, 2013; Der Sarkissian等, 2015; Meyer等, 2014).

4.3 古DNA富集

古DNA一般带脱氨基化学修饰, 存在外源环境DNA污染, 加上DNA分子交联、DNA-蛋白质大分子缩合反应等因素的综合作用, 限制了内源的目标古DNA分子在下游分析中的应用, 必须经过富集使其达到一定的拷贝数、提升其在总DNA分子中的比例. 第

一代和第二代测序技术富集古DNA的基础都是PCR扩增。

用第一代测序技术对古DNA进行序列测定的研究者, 借助GenBank中研究物种本身或其近缘物种的基因组序列设计引物, 通过引物与古DNA模板的特异性结合, 在DNA聚合酶作用下实现模板链的延伸以达到对目标序列的富集目的。理论上, 如果设计的引物具备物种或种群级别的特异性, 能够被扩增得到的唯有目标片段, 而环境外源DNA会因为没有能与之匹配的引物而被PCR过程屏蔽。

在第二代测序中, 如前所述经变性制备单链DNA文库, 通过接头的互补性将单链分子锚定到固体支持介质(微阵列芯片或微珠)作为PCR扩增模板, 再通过桥式扩增或乳液PCR等方式实现对模板的大量原位复制, 在芯片或微珠表面形成数以百万计的单分子阵列集合以达到富集目的(Margulies等, 2005; Shendure等, 2005)。第二代测序技术用于古DNA序列测定后, 首先发展起来的富集方法为引物延伸捕获(Primer Extension Capture, PEC): Briggs等(2009)在测序文库中加入5'端被生物素化的引物作为探针, 与古DNA模板特异结合并延伸, 产生的双链分子被对生物素亲和的磁珠或其它支持介质吸附, 而外源DNA分子及多余的引物被洗脱, 从而提高了内源古DNA在测序文库中的含量; Maricic等(2010)随后提出了长片段PCR产物杂交捕获法(Long-Range PCR Products Capture): 先利用长片段PCR扩增出待测基因组, 然后将PCR产物打断成150~850bp短片段并变性成单链, 再接上生物素化的接头制得探针, 最后用对生物素亲和的磁珠吸附探针分子并从构建的单链测序文库中捕获目标古DNA分子。两类富集方法的建立都有赖于探针和测序文库中DNA分子的碱基互补杂交特性; 不同之处在于, 前者可以依照不同基因组大小制备满足要求的探针序列和组合, 但对探针的制备及实验仪器有特定要求, 富集成本较高; 而后者在常规实验室即可完成全部富集探针的制备和富集过程, 但只适用于如线粒体基因组等较小的目标文库。为了克服上述富集方法对样品中目标基因组含量要求的限制, Carpenter等(2013)提出了能提高内源性DNA在古DNA测序文库中比例的全基因组溶液捕获法(Whole-genome In-Solution Capture, WISC): 给被打断的目标生物现代基因组加上含T7 RNA聚合酶启动子的接头, 通过体外转录得到能覆盖该生物完整

基因组、含生物素的RNA“诱饵”文库, 将RNA“诱饵”放入样品的古DNA测序文库, 从中特异性钓取目标生物古DNA片段, 利用微珠对生物素的识别将目标片段吸附到微珠上, 洗掉非特异性的古DNA后, 将目标古DNA从吸附珠上洗脱并完成PCR扩增。该方法能从内源性DNA含量<1%的古代样品中得到目标基因组, 是大区域片段的关键富集技术。

运用PCR技术对古DNA进行富集, 曾为绝灭的大型哺乳动物及古人类线粒体基因组、核基因、部分染色体、甚至核基因组序列测定奠定了必备分子基础。但也存在以下问题: 首先, PCR的单分子起始、指数级递增特性, 使得最终的扩增产物不能代表原始模板的组成类型及丰度; 其次, DNA聚合酶的保真度通常较低, PCR扩增过程中可能会引入不配对的错误碱基, 从而影响序列的原生性。第三代测序技术真正实现了基于单分子的边合成边测序(Sequencing by Synthesis, SBS), 不再依赖PCR对模板进行重复扩增, 不需要DNA富集过程。

4.4 古DNA序列测定

4.4.1 第一代测序技术

第一代测序技术的基础是Sanger等(1977)在20世纪70年代中期发明的末端终止法, 也被称作Sanger双脱氧链终止法。该方法运用DNA聚合酶使结合到特定模板上的引物得到延伸, 每一次测序反应由四组单独的反应组成, 每组反应中加入了延伸必须的四种脱氧核糖核酸(dNTPs)及一种双脱氧核糖核酸(ddNTP)。由于ddNTP缺乏延伸所需要的3'-羟基, 延伸中的寡聚核苷酸会因加入的双脱氧核糖核酸种类而选择性地终止在G、A、T或C处, 四组反应分别得到3'-端以ddG、ddA、ddT和ddC残基结尾的四组长短不一的片段混合物。通过高分辨率聚丙烯酰胺变性凝胶电泳或毛细管电泳平行分离各组大小不同的片段, 经放射自显影或非同位素标记所得图谱, 即可读得DNA碱基序列(图2)。

在用第一代测序技术对古DNA进行测序时, 样品埋藏、采集及实验操作过程中来自环境或试剂的外源DNA污染是研究者必须面临的问题。古DNA的高度片段化特征, 使得对较完整基因序列的获取依赖于多对引物的分段扩增和短序列拼接。此外, 分子克隆技术在此阶段是古DNA实验中的重要环节。因此, 第一

图2 第一代测序技术原理

代测序技术阶段对古DNA实验材料和投入实验工作量有一定要求. 无论是从对实验材料的量的需求, 还是从实验工作量而言, 要从古代生物材料中借助PCR分段扩增的方法得到生物完整的核基因组几乎是不可能完成任务. 以大小为3GB的哺乳动物核基因组为例, 即使是保存情况良好的冻土材料, 要从中得到完整的核基因组, 大约需要约180kg的实验材料和超过103百万次的PCR扩增, 再加上PCR前古DNA模板提取及PCR后扩增产物的电泳、纯化、分子克隆、测序等实验工作, 即便不考虑试剂耗材成本, 也鲜有研究对象和实验者能满足上述要求.

尽管如此, 第一代测序技术在古DNA研究中仍有用武之地, 其应用与新的测序平台并存, 在古DNA样品筛选、特定PCR产物检测等方面将继续发挥重要作用.

4.4.2 第二代测序技术

早在2000年, Brenner等(2000)即介绍了一种不再依赖凝胶电泳、基于微珠芯片体外扩增及荧光标签附加标记的大规模平行信号测序(Massively Parallel Signature Sequencing, MPSS), 从理论上启动了第二代测序技术. 2004年, 454生命科学公司推出第二代焦磷酸测序仪, 问世之初其测序成本即比Sanger测序法降低了6倍, 2005年Margulies等率先将其用于支原体基因组测序(Margulies等, 2005). 至此, 以高通量、低成

本为特色的第二代测序技术开始了在生物全基因组测序及分析、转录组、蛋白质相互作用组、重测序等方面的应用. 其在古DNA研究中的应用, 始于Poinar等(2006)对猛犸象核基因组测序, 之后Green等(2010, 2006)将其用于尼安德特人基因组分析, 使古DNA研究进入古基因组(paleogenomic)研究时代.

关于第二代测序技术平台及原理, 周晓光等(2010)曾撰文进行过全面介绍, 图3对比呈现了常见第二代测序技术平台涉及原理: 经过如前所述的DNA文库建立及富集过程, 在引物3'-端每新加入一个dNTP的聚合反应与各碱基对应的荧光信号释放相偶联, 在得到固体介质上(微吸附珠或芯片)数以百万计的单分子阵列集合的同时, 增强了测序过程中的光学检测灵敏度. 通过检测荧光的释放和强度, 达到实时检测DNA序列的目的(图3). 已有第二代测序技术平台都采用了合成测序法, 只是在测序生化反应标记、光学信号的产生原理方面存在差异, 并因此在序列读长上有所差别. Loman等(2012)对Junior (Roche)、MiSeq(Illumina)和Torrent PGM(Life Technologies)三类第二代测序技术平台进行了对比, 发现Illumina公司的MiSeq平台在单次测序反应中具备最高测序通量的同时兼备最低的错误率.

第二代测序技术摒除了用细菌载体对目标条带进行克隆并对单个克隆进行筛选和测序的操作, 且无需电泳, 极大减少了实验者的工作量. 其单次测序反应只需要4小时, 序列读长约250bp, 由于其测序反应在皮升(pL)级的微反应器上批量进行, 一次测序可得到20Mb序列信息, 比传统测序方法速度快100倍; 其碱基读取准确率为99%, 低于传统Sanger毛细管电泳测序法, 但由于可通过同一片段的多拷贝测序校正误差, 因而具有高效快速、准确和灵敏度高等特点. 近几年来, 研究者致力于第二代测序技术各环节的改进. Meyer等(2007)将同一样品来源的DNA片段用相同标签标记, 一次测序反应在理论上可涵盖250个哺乳动物线粒体基因组. Green等(2010)在构建用于第二代测序技术平台Illumina的DNA文库时, 加入限制性酶切消化, 有效去除了文库中环境微生物来源的DNA模板、从而提升了目标古DNA模板在测序文库中的比例. 此外, 针对不同的基因组测序需求, Illumina平台已研发出MiniSeq、MiSeq及HiSeq等不同测序体系, 其中MiniSeq是最简单方便的测序体系; MiSeq主要用于线

图3 第二代测序技术原理

据Margulies等(2005)和Shendure等(2005)修改绘制

粒体等小基因组测序,数据通量可达15Gb; HiSeq则是全基因组测序中最经济的选择,能获得的数据通量高达1500~1800Gb.

4.4.3 第三代测序技术

第三代测序技术依原理可分作单分子荧光测序和纳米孔测序两类,二者都不需要经过PCR扩增,真正实现了对每一条DNA分子进行测序,因此第三代测序技术也称单分子测序(true Single Molecule Sequencing, tSMS). 单分子测序理念早在1989年即已被提出,2008年Harris等(2008)用单分子荧光测序法获得了M13病毒基因组:将待测基因组打断变性成单链后加上多聚A尾巴,运用A-T互补,将待测分子锚定到固体支持介质;当用荧光标记的dNTP被掺入DNA链时,由显微镜实时探测其荧光信号;新掺入的dNTP与DNA形成化学键以后,其荧光基团被DNA聚合酶切除,荧光消失,进入下一个碱基掺入循环. 纳米孔测序则是采用电泳技术,借助电场驱动力让单个DNA分子逐一通过纳米

孔,由于纳米孔直径非常细小,仅允许单个核酸聚合物通过,四种碱基带电性质有差异,因此可通过电信号的差异检测所通过的碱基类别,从而实现DNA测序(图4).

相比于第二代测序技术,第三代测序技术避开了PCR扩增的弊端,提高了测序精度,单次反应能够测得的序列更长(1~100kb),减少了后续基因组拼接和注释工作量,能够直接测RNA序列,并且能够识别甲基化位点,在鉴定新的病原体和细菌的基因组测序方面得到广泛应用;不足之处在于:由于其光学组件增加了仪器成本,使得运行成本较高. 在古基因组研究方面,Orlando等(2011)首次尝试运用美国螺旋生物公司HeliScope单分子测序平台对冻土保存的古马古基因组进行测序,将模板的变性温度从95℃调整到80℃,并用磷酸酶处理提取物,得到的序列中内源DNA含量较高、序列读长较长、GC含量较低且胞嘧啶去氨基作用水平更高. 究其原因,古DNA模板在埋藏过程中自发生胞嘧啶去氨基作用,尿嘧啶的形成使得DNA双

图4 第三代测序技术原理

据Harris等(2008)修改绘制

链间的氢键作用力减弱,在较低的变性温度下,目标古DNA分子相对外源DNA更容易变性成单链并用作测序模板;而磷酸酶的加入,可以去除古DNA分子中普遍存在的3'-端磷酸化修饰,从而使测序文库中的古DNA模板在结构上满足单分子测序的要求。

第三代测序技术目前还处在初步发展阶段,在记录数据所涉及的工程学、光学以及测序反应所涉及的化学和生物工程领域都面临着多方面的挑战(Au等, 2013; Treffer和Deckert, 2010),不仅不是古DNA测序的主流,也不是大量产生现代DNA数据的主要测序方法。

4.5 古DNA序列拼接与基因组组装

由于古DNA的高度片段化特征及测序技术本身在序列读长上的限制,经测序所得古DNA序列长度极少能超过500bp,需要进行至关重要的序列拼接或基因组组装过程,得到完整的基因或基因组用于后续分析。

第一代测序技术运用于古DNA测序时,研究者一般将PCR目标产物进行分子克隆,对多个克隆子进行序列测定,得到测序图及序列数据后,用软件剪除两端的载体及引物序列,寻找相邻片段的重合区域,实现序列的拼接。

高通量测序技术开始应用于古DNA研究之后,研究者通常将研究对象定位在染色体或基因组水平,测序后产生海量短序列(35~150bp),基因组组装显得尤为重要。基因组组装通常包含以下步骤:测序数据纠

错-构建连续共识序列(contig)-由contig构建不连续的脚手架序列(scaffold)-填补scaffold空白区域(gap)(Fan和Li, 2012)。由于高通量测序技术的序列读长比传统Sanger测序法序列读长短得多,所以在基因组组装步骤中最具难度的是“共识序列的构建”。影响基因组组装效果的因素较多,如:在较大的基因组中普遍存在的重复序列的存在、测序策略及序列数据质量、基因组不同区域测序深度的不均匀性及所使用的组装软件类型等。目前各类组装软件都需要在高性能工作站、计算机集群甚至大型计算机上运行,在所采用的算法、短序列重叠区的查找索引创建等方面有所差别,并对不同测序技术平台有所侧重,如SHORTY是用于ABI SOLiD测序平台产生的序列组装,而华大基因开发的SOAPdenovo,则特异性针对Illumina测序技术平台所得短序列而设计,不依靠参考序列即能组装得到大型哺乳动物的基因组(Li等, 2010)。

随着测序技术的发展,基因组组装是一个快速发展的领域。在未来的几年里,测序成本的降低及测序效率的提升,将极大促进古代生物基因组相关研究的发展,效果和性能更好的组装软件的开发,必将为序列数据质量相对较低的古基因组研究提供更好的辅助。

5 全基因组时代古DNA序列数据的应用

自2006年始,古DNA跨入全新的古基因组(paleogenomics)研究时代,在以下方面取得突破性进展。

首先,解决了多年来争论不休的长鼻目动物演化问题. 2006年,先是Rogaev等(2006)运用传统PCR方法从西伯利亚距今3万多年的猛犸象标本中获得其线粒体全基因组序列;随后Poinar等(2006)运用结合乳液PCR和焦磷酸测序(pyrosequencing)的第二代测序方法从猛犸象化石中获得了1300万bp核基因组序列. 两类基因组数据构建的系统树结果都支持猛犸象与亚洲象亲缘关系最近, 2006年因此被称作“猛犸象之年(The Year of the Mammoth)”(Cooper, 2006), 以古DNA研究为主的“从化石中提取DNA”、“从猛犸象到癌症患者更快、更廉价的基因组测序”分别被美国《科学》杂志选为2006和2008年度世界十大科学进展. 不仅如此, Rohland等(2007)运用新一代测序技术获得了已绝灭美洲乳齿象的全线粒体基因组并将其用作长鼻目系统发育分析的外类群, 结果认为现生非洲象的祖先于距今24~28Ma前与猛犸象及亚洲象的共同祖先发生分离, 并认为猛犸象与亚洲象的分歧时间为距今6.7Ma; 三年之后, Rohland等(2010)将研究从线粒体基因组向核基因组推进, 借助第二代测序技术得到涵盖375个基因座、共计39763bp的美洲乳齿象和猛犸象核基因组序列, 并将之用于更为细化的长鼻目分类研究, 认为草原非洲象和森林非洲象的演化分歧时间与猛犸象和亚洲象的分歧时间等同, 应该被看作两个不同的种, 而非不同生境中的两个不同种群.

其次,谱写了人类演化历史的新篇章,更新了尼安德特人、丹尼索沃人等古人类与现代人的遗传演化关系,并丰富了人类迁移、演化的细节知识. 第二代测序技术用于古人类研究,始于博物馆保存的头发样品中4000年前古爱斯基摩人核基因组测序及单核苷酸位点多态性分析,该研究为古人类5500年前从西伯利亚向北美大陆的迁移提供了证据(Rasmussen等, 2010). 在第二代测序技术应用于古人类研究之前,学界一致认为尼安德特人是人类演化历史中走向绝灭的一个支系,对现代人的基因组没有任何贡献(Serre等, 2004);随着尼安德特人基因组草图的完成,研究者得以对尼安德特人和5个现代人(法国人、中国汉族人、太平洋西南岛国巴布亚新几内亚人、非洲南部的布须曼人和非洲西部的优鲁巴人)基因组进行对比,发现尼安德特人与现代欧亚人在基因组的某些区域的相似性程度(1~4%),既高出尼安德特人与现代非洲人,也高出同为现代人的欧亚人与非洲人,研究者认

为这是非洲以外的现代人共同祖先在向欧亚大陆及太平洋岛屿迁移扩散并发生演化分歧之前,在中东与尼安德特人发生基因交流的结果,由此对现代人起源的“走出非洲”模式提出了初步质疑图(Green等, 2010);从阿尔泰山区尼安德特人的高质量基因组所蕴含的遗传信息中, Prufer等(2014)不仅发现晚更新世各古人类群体都存在近亲婚配,而且找到了尼安德特人、丹尼索沃人及早期现代人之间基因交流的证据,确定与尼安德特人、丹尼索沃人祖先发生演化分歧之后现代人基因组中发生并被固定下来的遗传突变. 4000年前旧石器时代晚期(3~5万年前)丹尼索沃人基因组的测定,显示其与尼安德特人亲缘关系较近,并对现代马来西亚人基因组有4~6%的贡献(Reich等, 2010). 20世纪早期澳洲土著人古基因组数据分析显示:人类分别在6.2~7.5万年前、2.5~3.8万年前两个时间段曾在亚洲东部发生过两次迁移扩张高峰,第二次迁移的古人类沿迁移路径与第一次迁移居住到当地的古人类之间有基因交流,澳洲土著人是第一次迁移高峰期经南线迁移到澳洲的古人类后裔,并发现该古人类支系在迁移过程中与丹尼索沃人有过基因交流(Rasmussen等, 2011);4万年前中国田园洞人核DNA和线粒体DNA显示其在表现更多早期现代人的基因特征的基础上,也携带有少量尼安德特人DNA(Fu等, 2013);旧石器时代晚期西伯利亚中南部古人类基因组草图的完成,显示出该样本代表的古人类和现代西欧人及现代美洲人群的遗传相似性较高,说明现代西欧人的祖先在2.4万年前曾在欧亚大陆东北部有更为广泛的分布,并对美洲原住民基因组有高达14~38%的贡献(Raghavan等, 2014);4.5万年前西伯利亚西南部古人类的高通量测序结果显示该古人类所属群体与现代欧亚大陆人群共同祖先同祖,并预示早期现代人走出非洲并非仅有南部大洋洲路线(Fu等, 2014).

第三,启动了“全基因组-大数据-多物种”尺度研究生物对气候变化的分子响应. 通过对涉及猛犸象、披毛犀、野马、驯鹿、野牛及麝牛等多个第四纪大型哺乳动物及古人类化石的古DNA研究, Lorenzen等(2011)指出气候变化是过去5万年来动物种群发生变化的主要诱因,并认为不同物种在应对气候变化及人类活动等生态压力时存在种间差异. Thuiller等(2011)考量了气候变化对欧洲植物、鸟类及哺乳动物等多个生物类群的影响,认为相对于自然状态下的随机绝

灭模型, 那些对气候变化敏感的物种的绝灭并不会使生物谱系树上物种的消失加速. 通过对北极圈沉积物大规模系统取样, Willerslev等(2014)用古DNA条形码方法研究了极地5万年来的植物多样性, 通过线虫多样性研究以指示植被覆盖率及土壤环境, 并结合近1万年来大部分已经绝灭的食草动物的食性分析, 认为北极圈的主要植被为指示干燥的草原-苔原环境的非禾本类, 其多样性在末次冰期有显著降低, 并在距今1万年之后发生植被更替, 出现了指示湿润苔原环境的木本植物及禾本植物.

最后, 刷新了古DNA研究涉及的样品年代. 早先一般认为古DNA只能保存10万年, 而目前已经得到全基因组数据材料早已突破了这一年限, 如: 得到核基因组的冻土保存的中更新世马测年结果为70万年(Orlando等, 2013); 得到线粒体基因组的非冻土保存材料、西班牙北部发掘的中更新世古人类年代约为40万年(Meyer等, 2014).

6 展望

古DNA研究领域近五年来取得了令人瞩目的成就, 不断被优化的古DNA提取、文库建立及测序技术, 为该领域的发展带来了革命性的变化, 预计将来可能的发展主要体现在以下方面.

(1) 群体水平的古基因组研究. 第二代及第三代测序技术在古基因组中的运用, 目前已通过个体水平的研究实例被证实其在技术方法是可靠的, 当测序成本降低到可广泛用于不同群体, 将有望从古基因组群体遗传学角度探讨人类起源、演化、迁移历史中的关键问题. 对人类群体自身基因组的解析是提供人类演化历史的最直接的方法; 对其他物种的研究, 也可间接为此提供佐证. 如: 伴生动、植物能指示人类迁移; 对家养动植物古基因组的研究, 将有望找到人工选择模式下家养驯化过程中动植物从野生到家养的关键基因, 了解现生家养动植物遗传多样性变化之基础; 对古细菌、古病毒基因组的分析, 可揭示流行病及其曾经对人类健康、人口数量产生的重大影响.

(2) 大型脊椎动物的古基因组及全基因组SNP位点分析. 很多大型哺乳动物在中更新世都经历过辐射及灭绝事件, 相对于晚更新世出现的末次冰期, 中更新世有多次“冰期-间冰期”循环. 因此, 对中更新世以来大型脊椎动物全基因组SNP位点分析, 将为绝灭或

濒危生物对气候环境变化的应对策略提供分子数据支持, 从而为目前全球变暖大环境下的濒危物种保护提供借鉴.

(3) 古蛋白质及古RNA等其他分子介质研究. 地质载体中的生物小分子, 如类脂化合物, 比DNA更容易保存, 保存年限相对较长, 且是很好的环境和气候变化指标. 对中更新世早期古马样品的分析发现, 甘氨酸、脯氨酸和丙氨酸在骨胶原蛋白中含量丰富(Orlando等, 2013); Fordyce等(2013)对古代玉米研究发现, RNA分子的保存状况优于DNA分子, 为检测生物发育关键时期具有演化意义的基因表达变化提供了可能. 因此, 古DNA研究若能结合类脂等生物小分子及其他生物大分子研究, 将为全面揭示古代生物随环境和气候变化发生的演化趋势变化提供更为全面的分子依据.

在我国, 古DNA研究起步相对较晚, 近几年逐渐出现了一些具有一定影响力的研究实例. 如: 我国东北地区大型脊椎动物的系统演化(盛桂莲等, 2009b; Sheng等, 2009; Sheng等, 2014; Yang等, 2006; Yuan等, 2013), 相关考古点古代人类及伴生动物遗骸鉴定、遗传分析(谢承志等, 2007; 许月等, 2006; Cai等, 2007; Fu等, 2013; Fu等, 2008; Shi等, 2008; Yu等, 2006; Yu等, 2007), 动植物的家养驯化历史等(Cai等, 2009; Han等, 2010; Hou等, 2014; Li等, 2014; Xiang等, 2014; Zhang等, 2013). 随着国内外合作的加强, 一方面, 由我国学者参与甚至主导的对中国境外材料的研究, 近年来逐渐在国际权威期刊上呈现, 显示出我国古DNA研究的上升态势(Fu等, 2014; Krause等, 2010; Lorenzen等, 2011); 另一方面, 深圳华大基因公司近年来经市场开拓与国外顶级古DNA实验室建立了合作关系, 在测序技术方法上显示出我国古基因组测序与国际前沿的同步. 例如: 与哥本哈根大学合作, 对古爱斯基摩人核基因组进行了测序并对其单核苷酸位点多态性进行了分析(Rasmussen等, 2010). 2014年5月, 华大基因研究院杨焕明研究员、中国科学院古脊椎动物与古人类研究所周忠和研究员、西北大学舒德干教授三位院士担任执行主席, 组织了以“古今对话—分子遗传和古生物信息的整合研究”为主题的香山会议, 其中“基因组学与古DNA研究”为三大中心议题之一.

我国分布着众多独具特色的第四纪动植物群及考古遗址, 留下了我国史前人类及很多具全球影响的已绝灭生物的遗体 and 遗迹, 为我国研究者进行该方面

研究提供了得天独厚的资源条件。本文结合分子生物学领域的技术革新,以古DNA保存、序列数据的获取及运用为主线,对古DNA领域过去35年来的发展历史及研究成果进行概括和回顾,并对该领域今后的发展进行展望,以期相关领域研究者能打破学科壁垒,加强交流与合作,充分利用新技术、新方法,促进我国古DNA研究的蓬勃发展。

参考文献

- 湖南医学院. 长沙马王堆一号汉墓古尸研究. 北京: 文物出版社. 1980. 184–187
- 盛桂莲, 赖旭龙, 侯新东. 2009a. 古DNA实验体系及技术. 中国生物化学与分子生物学报, 25: 116–125
- 盛桂莲, 袁俊霞, 伊剑, 金昌柱, 刘金毅, 侯新东, 赖旭龙. 2009b. 河北秦皇岛灵仙洞斑鬣狗化石的古DNA初步分析. 地球科学——中国地质大学学报, 34: 877–883
- 谢承志, 李春香, 崔银秋, 蔡大伟, 王海晶, 朱泓, 周慧. 2007. 新疆山普拉古代居民线粒体DNA研究. 自然科学进展, 17: 871–876
- 许月, 张小雷, 张全超, 崔银秋, 周慧, 朱泓. 2006. 古代契丹与现代达斡尔遗传关系分析. 吉林大学学报(理学版), 44: 997–1000
- 杨东亚. 2003. 古代DNA研究中污染的控制和识别. 人类学学报, 22: 163–173
- 周晓光, 任鲁凤, 李云涛, 张猛, 俞育德, 于军. 2010. 下一代测序技术: 技术回顾与展望. 中国科学: 生命科学, 40: 23–37
- Allentoft M E, Collins M, Harker D, Haile J, Oskam C L, Hale M L, Campos P F, Samaniego J A, Gilbert M T P, Willerslev E, Zhang G, Scofield R P, Holdaway R N, Bunce M. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc R Soc B-Biol Scis*, 279: 4724–4733
- Au K F, Sebastiano V, Tootoonchi Afshar P, Durruthy Durruthy J, Lee L, Williams B A, van Bakel H, Schadt E E, Reijo-Pera R A, Underwood J G, Wong W H. 2013. Characterization of the human ESC transcriptome by hybrid sequencing. *Proc Natl Acad Scis*, 110: E4821–E4830
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd D H, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams S R, Moon K, Burcham T, Pallas M, DuBridge R B, Kirchner J, Fearon K, Mao J, Corcoran K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*, 18: 630–634
- Briggs A W, Good J M, Green R E, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Schmitz R, Doronichev V B, Golovanova L V, de la Rasilla M, Fortea J, Rosas A, Pääbo S. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, 325: 318–321
- Cai D W, Tang Z, Han L, Speller C F, Yang D Y, Ma X, Cao J, Zhu H, Zhou H. 2009. Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J Archaeol Sci*, 36: 835–842
- Cai D W, Han L, Zhang X-L, Zhou H, Zhu H. 2007. DNA analysis of archaeological sheep remains from China. *J Archaeol Sci*, 34: 1347–1355
- Cano R J, Poinar H N, Pieniazek N J, Acra A, Poinar G O. 1993. Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature*, 363: 536–538
- Carpenter M L, Buenrostro J D, Valdiosera C, Schroeder H, Allentoft M E, Sikora M, Rasmussen M, Gravel S, Guillén S, Nekhrizov G, Leshtakov K, Dimitrova D, Theodossiev N, Pettener D, Luiselli D, Sandoval K, Moreno-Estrada A, Li Y, Wang J, Gilbert M T P, Willerslev E, Greenleaf W J, Bustamante C D. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am J Human Genets*, 93: 852–864
- Collins M J, Penkman K E, Rohland N, Shapiro B, Dobberstein R C, Ritz-Timme S, Hofreiter M. 2009. Is amino acid racemization a useful tool for screening for ancient DNA in bone? *Proc Biol Sci*, 276: 2971–2977
- Cooper A. 2006. The Year of the Mammoth. *PLoS Biol*, 4: e78
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature*, 409: 704–707
- Dabney J, Knapp M, Glocke I, Gansauge M T, Weihmann A, Nickel B, Valdiosera C, Garcia N, Pääbo S, Arsuaga J L, Meyer M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Scis*, 110: 15758–15763
- Der Sarkissian C, Allentoft M E, Avila-Arcos M C, Barnett R, Campos P F, Cappellini E, Ermini L, Fernandez R, da Fonseca R, Ginolhac A, Hansen A J, Jonsson H, Korneliusson T, Margaryan A, Martin M D, Moreno-Mayar J V, Raghavan M, Rasmussen M, Velasco M S, Schroeder H, Schubert M, Seguin-Orlando A, Wales N, Gilbert M T, Willerslev E, Orlando L. 2015. Ancient genomics. *Philos Trans R Soc Lond B Biol Sci*, 370: 1–10
- DeSalle R, Gatesy J, Wheeler W, Grimaldi D. 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science*, 257: 1933–1936
- Doran G H, Dickel D N, Ballinger W E, Agee O F, Laipis P J, Hauswirth W W. 1986. Anatomical, cellular and molecular analysis of 8,000-yr-old human brain tissue from the Windover archaeological site. *Nature*, 323: 803–806
- Fan W, Li R. 2012. Test driving genome assemblers. *Nat Biotechnol*, 30: 330–331
- Fordyce S L, Ávila-Arcos M C, Rasmussen M, Cappellini E,

- Romero-Navarro J A, Wales N, Alquezar-Planas D E, Penfield S, Brown T A, Vielle-Calzada J P, Montiel R, Jørgensen T, Odegaard N, Jacobs M, Arriaza B, Higham T F G, Ramsey C B, Willerslev E, Gilbert M T P, Fuller D Q. 2013. Deep Sequencing of RNA from Ancient Maize Kernels. *PLoS ONE*, 8: e50961
- Fu Q, Li H, Moorjani P, Jay F, Slepchenko S M, Bondarev A A, Johnson P L F, Aximu-Petri A, Prufer K, de Filippo C, Meyer M, Zwyns N, Salazar-Garcia D C, Kuzmin Y V, Keates S G, Kosintsev P A, Razhev D I, Richards M P, Peristov N V, Lachmann M, Douka K, Higham T F G, Slatkin M, Hublin J J, Reich D, Kelso J, Viola T B, Pääbo S. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, 514: 445–449
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano H A, Kelso J, Pääbo S. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Scis*, 110: 2223–2227
- Fu Y, Xie C, Zhao H, Xu X, Cui Y, Zhang Q, Zhou H, Zhu H. 2008. Molecular genetic analysis of Dongzhou-period ancient human of Helingeer in Inner Mongolia, China. *Front Biol China*, 3: 9–12
- Gansauge M T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc*, 8: 737–748
- Gilbert M T P, Bandelt H J, Hofreiter M, Barnes I. 2005. Assessing ancient DNA studies. *Trends Ecology Evolution*, 20: 541–544
- Ginolhac A, Vilstrup J, Stenderup J, Rasmussen M, Stiller M, Shapiro B, Zazula G, Froese D, Steinmann K E, Thompson J F, AL-Rasheid K A, Gilbert T M, Willerslev E, Orlando L. 2012. Improving the performance of True Single Molecule Sequencing for ancient DNA. *BMC Genomics*, 13: 177
- Golenberg E M, Giannsi D E, Clegg M T, Smiley C J, Durbin M, Henderson D, Zurawski G. 1990. Chloroplast DNA sequences from a Miocene Magnolia species. *Nature*, 334: 656–658
- Green R E, Krause J, Briggs A W, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz M H Y, Hansen N F, Durand E Y, Malaspinas A S, Jensen J D, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano H A, Good J M, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander E S, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev V B, Golovanova L V, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz R W, Johnson P L F, Eichler E E, Falush D, Birney E, Mullikin J C, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Paabo S. 2010. A Draft Sequence of the Neandertal Genome. *Science*, 328: 710–722
- Green R E, Krause J, Ptak S E, Briggs A W, Ronan M T, Simons J F, Du L, Egholm M, Rothberg J M, Paunovic M, Pääbo S. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature*, 444: 330–336
- Han L, Yu H X, Cai D W, Shi H L, Zhu H, Zhou H. 2010. Mitochondrial DNA analysis provides new insights into the origin of the Chinese domestic goat. *Small Ruminant Res*, 90: 41–46
- Hänni C, Brousseau T, Laudet V, Dominique S. 1995. Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleic Acids Res*, 23: 881–882
- Harris T D, Buzby P R, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, DiMeo J, Efcavitch J W, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake S R, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z. 2008. Single-Molecule DNA Sequencing of a Viral Genome. *Science*, 320: 106–109
- Higuchi R, Bowman B, Freiburger M, Ryder O A, Wilson A C. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312: 282–284
- Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res*, 29: 4793–4799
- Hofreiter M, Paijmans J L, Goodchild H, Speller C F, Barlow A, Fortes G G, Thomas J A, Ludwig A, Collins M J. 2014. The future of ancient DNA: Technical advances and conceptual shifts. *Bioessays*, 37: 284–293
- Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001b. Ancient DNA. *Nat Genet*, 2: 7
- Hou X D, Sheng G L, Yin S, Zhu M, Du M, Jin C Z, Lai X L. 2014. DNA analyses of wild boar remains from archaeological sites in Guangxi, China. *Quaternary Int*, 354: 147–153
- Kalmar T, Bachrati C Z, Marcil A, Rasko I. 2000. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Res*, 28: 67e
- King C, Debruyne R, Kuch M, Schwarz C, Poinar H. 2009. A quantitative approach to detect and overcome PCR inhibition in ancient DNA extracts. *Biotech*, 47: 941–949
- Krause J, Fu Q, Good J M, Viola B, Shunkov M V, Derevianko A P, Pääbo S. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, 464: 894–897
- Leonard J A, Wayne R K, Cooper A. 2000. Population genetics of Ice Age brown bears. *Proc Natl Acad Scis*, 97: 1651–1654
- Li M, Tian S, Yeung C K, Meng X, Tang Q, Niu L, Wang X, Jin L, Ma J, Long K. 2014. Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. *Sci Rep*, 4: 4678
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z, Zhang Y, Wang W, Li J, Wei F, Li H, Jian M, Li J, Zhang Z, Nielsen R, Li D, Gu W, Yang Z, Xuan Z, Ryder O A, Leung F C C, Zhou Y, Cao J, Sun X, Fu Y, Fang X, Guo X, Wang B, Hou R, Shen

- F, Mu B, Ni P, Lin R, Qian W, Wang G, Yu C, Nie W, Wang J, Wu Z, Liang H, Min J, Wu Q, Cheng S, Ruan J, Wang M, Shi Z, Wen M, Liu B, Ren X, Zheng H, Dong D, Cook K, Shan G, Zhang H, Kosiol C, Xie X, Lu Z, Zheng H, Li Y, Steiner C C, Lam T T Y, Lin S, Zhang Q, Li G, Tian J, Gong T, Liu H, Zhang D, Fang L, Ye C, Zhang J, Hu W, Xu A, Ren Y, Zhang G, Bruford M W, Li Q, Ma L, Guo Y, An N, Hu Y, Zheng Y, Shi Y, Li Z, Liu Q, Chen Y, Zhao J, Qu N, Zhao S, Tian F, Wang X, Wang H, Xu L, Liu X, Vinar T, Wang Y, Lam T W, Yiu S M, Liu S, Zhang H, Li D, Huang Y, Wang X, Yang G, Jiang Z, Wang J, Qin N, Li L, Li J, Bolund L, Kristiansen K, Wong G K S, Olson M, Zhang X, Li S, Yang H, Wang J, Wang J. 2010. The sequence and de novo assembly of the giant panda genome. *Nature*, 463: 311–317
- Loman N J, Misra R V, Dallman T J, Constantinidou C, Gharbia S E, Wain J, Pallen M J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*, 30: 434–439
- Lorenzen E D, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske K A, Ugan A, Borregaard M K, Gilbert M T P, Nielsen R, Ho S Y W, Goebel T, Graf K E, Byers D, Stenderup J T, Rasmussen M, Campos P F, Leonard J A, Koepfli K P, Froese D, Zazula G, Stafford T W, Aaris-Sørensen K, Batra P, Haywood A M, Singarayer J S, Valdes P J, Boeskorov G, Burns J A, Davydov S P, Haile J, Jenkins D L, Kosintsev P, Kuznetsova T, Lai X, Martin L D, McDonald H G, Mol D, Meldgaard M, Munch K, Stephan E, Sablin M, Sommer R S, Sipko T, Scott E, Suchard M A, Tikhonov A, Willerslev R, Wayne R K, Cooper A, Hofreiter M, Sher A, Shapiro B, Rahbek C, Willerslev E. 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, 479: 359–364
- Margulies M, Egholm M, Altman W E, Attiya S, Bader J S, Bemben L A, Berka J, Braverman M S, Chen Y J, Chen Z. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437: 376–380
- Maricic T, Whitten M, Pääbo S, Fleischer R C. 2010. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE*, 5: e14004
- Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga J L, Martinez I, Gracia A, de Castro J M, Carbonell E. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*, 505: 403–406
- Meyer M, Kircher M, Gansauge M T, Li H, Racimo F, Mallick S, Schraiber J G, Jay F, Prüfer K, de Filippo C, Sudmant P H, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green R E, Bryc K, Briggs A W, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer M F, Shunkov M V, Derevianko A P, Patterson N, Andres A M, Eichler E E, Slatkin M, Reich D, Kelso J, Pääbo S. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, 338: 222–226
- Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M. 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res*, 35: e97
- Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, Magnussen K, Steinmann K E, Kapranov P, Thompson J F, Zazula G, Froese D, Moltke I, Shapiro B, Hofreiter M, Al-Rasheid K A S, Gilbert M T P, Willerslev E. 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res*, 21: 1705–1719
- Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson P L F, Fumagalli M, Vilstrup J T, Raghavan M, Korneliusen T, Malaspina A S, Vogt J, Szklarczyk D, Kelstrup C D, Vinther J, Dolocan A, Stenderup J, Velazquez A M V, Cahill J, Rasmussen M, Wang X, Min J, Zazula G D, Seguin-Orlando A, Mortensen C, Magnussen K, Thompson J F, Weinstock J, Gregersen K, Roed K H, Eisenmann V, Rubin C J, Miller D C, Antczak D F, Bertelsen M F, Brunak S, Al-Rasheid K A S, Ryder O, Andersson L, Mundy J, Krogh A, Gilbert M T P, Kjaer K, Sicheritz-Ponten T, Jensen L J, Olsen J V, Hofreiter M, Nielsen R, Shapiro B, Wang J, Willerslev E. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499: 74–78
- Pääbo S. 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, 314: 644–645
- Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci USA*, 86: 1939–1943
- Pääbo S. 1991. Amplifying DNA from archeological remains: A meeting report. *Genome Res*, 1: 107–110
- Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. Genetic Analyses from Ancient DNA. *Annu Rev Genet*, 38: 645–679
- Poinar H N, Hofreiter M, Geoffrey Spaulding W, Martin P S, Stankiewicz B A, Bland H, Evershed R P, Possnert G, Pääbo S. 1998. Molecular Coproscopy: Dung and Diet of the Extinct Ground Sloth *Nototheriops shastensis*. *Science*, 281: 402–406
- Poinar H N, Hoss M, Bada J L, Pääbo S. 1996. Amino Acid Racemization and the Preservation of Ancient DNA. *Science*, 272: 864–866
- Poinar H N, Schwarz C, Qi J, Shapiro B, MacPhee R D E, Buigues B, Tikhonov A, Huson D H, Tomsho L P, Auch A, Rampp M, Miller W, Schuster S C. 2006. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science*, 311: 392–394
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant P H, de Filippo C. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505: 43–49
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke

- I, Rasmussen S, Stafford T W Jr, Orlando L, Metspalu E. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, 505: 87–91
- Rasmussen M, Guo X, Wang Y, Lohmueller K E, Rasmussen S, Albrechtsen A, Skotte L, Lindgreen S, Metspalu M, Jombart T, Kivisild T, Zhai W, Eriksson A, Manica A, Orlando L, De La Vega F M, Tridico S, Metspalu E, Nielsen K, Avila-Arcos M C, Moreno-Mayar J V, Muller C, Dortch J, Gilbert M T P, Lund O, Wesolowska A, Karmin M, Weinert L A, Wang B, Li J, Tai S, Xiao F, Hanihara T, van Driem G, Jha A R, Ricaut F X, de Knijff P, Migliano A B, Gallego Romero I, Kristiansen K, Lambert D M, Brunak S, Forster P, Brinkmann B, Nehlich O, Bunce M, Richards M, Gupta R, Bustamante C D, Krogh A, Foley R A, Lahr M M, Balloux F, Sicheritz-Ponten T, Villems R, Nielsen R, Wang J, Willerslev E. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science*, 334: 94–98
- Rasmussen M, Li Y, Lindgreen S, Pedersen J S, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert M T P, Wang Y, Raghavan M, Campos P F, Kamp H M, Wilson A S, Gledhill A, Tridico S, Bunce M, Lorenzen E D, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre T L, Gornnow B, Meldgaard M, Andreasen C, Fedorova S A, Osipova L P, Higham T F G, Ramsey C B, Hansen T V O, Nielsen F C, Crawford M H, Brunak S, Sicheritz-Ponten T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*, 463: 757–762
- Reich D, Green R E, Kircher M, Krause J, Patterson N, Durand E Y, Viola B, Briggs A W, Stenzel U, Johnson P L F, Maricic T, Good J M, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler E E, Stoneking M, Richards M, Talamo S, Shunkov M V, Derevianko A P, Hublin J J, Kelso J, Slatkin M, Pääbo S. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468: 1053–1060
- Rogaev E I, Moliaka Y K, Malyarchuk B A, Kondrashov F A, Derenko M V, Chumakov I, Grigorenko A P, Penny D. 2006. Complete Mitochondrial Genome and Phylogeny of Pleistocene Mammoth *Mammuthus primigenius*. *PLoS Biol*, 4: e73
- Rohland N, Hofreiter M. 2007a. Ancient DNA extraction from bones and teeth. *Nat Protoc*, 2: 1756–1762
- Rohland N, Hofreiter M. 2007b. Comparison and optimization of ancient DNA extraction. *Biotech*, 42: 343–352
- Rohland N, Malaspina A S, Pollack J L, Slatkin M, Mathews P, Hofreiter M, Penny D. 2007. Proboscidean Mitogenomics: Chronology and Mode of Elephant Evolution Using *Mastodon* as Outgroup. *PLoS Biol*, 5: e207
- Rohland N, Reich D, Mallick S, Meyer M, Green R E, Georgiadis N J, Roca A L, Hofreiter M, Penny D. 2010. Genomic DNA Sequences from *Mastodon* and Woolly Mammoth Reveal Deep Speciation of Forest and Savanna Elephants. *PLoS Biol*, 8: e1000564
- Römpler H, Rohland N, Lalueza-Fox C, Lalueza-Fox C, Wilerslev E, Kuznetsova T, Rabeder G, Bertranpetit J, Schöneberg T, Hofreiter M. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science*, 313: 62
- Sanger F, Nicklen S, Coulson A R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*, 74: 5463–5467
- Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Menecier P, Hofreiter M, Possnert G, Pääbo S, David Penny S. 2004. No Evidence of Neandertal mtDNA Contribution to Early Modern Humans. *PLoS Biol*, 2: e57
- Shendure J, Porreca G J, Reppas N B, Lin X, McCutcheon J P, Rosenbaum A M, Wang M D, Zhang K, Mitra R D, Church G M. 2005. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science*, 309: 1728–1732
- Sheng G, Wu L, Hou X, Yuan J, Cheng S, Zhong B, Lai X. 2009. Short sequence effect of ancient DNA on mammoth phylogenetic analyses. *Front Earth Sci China*, 3: 100–106
- Sheng G L, Soubrier J, Liu J Y, Werdelin L, Llamas B, Thomson V A, Tuke J, Wu L J, Hou X D, Chen Q J, Lai X L, Cooper A. 2014. Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, *Crocuta crocuta*. *Mol Ecol*, 23: 522–533
- Shi H, Zhong H, Peng Y, Dong Y, Qi X, Zhang F, Liu L, Tan S, Ma R, Xiao C, Wells S, Jin L, Su B. 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol*, 6: 45
- Taylor P G. 1996. Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. *Mol Biol Evol*, 13: 3
- Thuiller W, Lavergne S, Roquet C, Boulangeat I, Lafourcade B, Araujo M B. 2011. Consequences of climate change on the tree of life in Europe. *Nature*, 470: 531–534
- Treffer R, Deckert V. 2010. Recent advances in single-molecule sequencing. *Curr Opin Biotech*, 21: 4–11
- Wayne R K, Leonard J A, Cooper A. 1999. Full of Sound and Fury: History of Ancient DNA. *Annu Rev Ecol Syst*, 30: 457–477
- Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards M E, Lorenzen E D, Vestergaard M, Gussarova G, Haile J, Craine J, Gielly L, Boessenkool S, Epp L S, Pearman P B, Cheddadi R, Murray D, Bråthen K A, Yoccoz N, Binney H, Cruaud C, Wincker P, Goslar T, Alsos I G, Bellemain E, Brysting A K, Elven R, Sønstebo J H, Murton J, Sher A, Rasmussen M, Rønn R, Mourier T, Cooper A, Austin J, Möller P, Froese D, Zazula G, Pompanon F, Rioux D, Niderkorn V, Tikhonov A, Savvinov G, Roberts R G, MacPhee R D E, Gilbert M

- T P, Kjær K H, Orlando L, Brochmann C, Taberlet P. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, 506: 47–51
- Xiang H, Gao J, Yu B, Zhou H, Cai D, Zhang Y, Chen X, Wang X, Hofreiter M, Zhao X. 2014. Early Holocene chicken domestication in northern China. *Proc Natl Acad Sci USA*, 111: 17564–17569
- Yang S, Lai X, Shi S, Cheng S, Zhou X, Yang X, Yang H. 2006. New ancient DNA sequences suggest high genetic diversity for the woolly mammoth (*Mammuthus primigenius*). *Prog Nat Sci*, 16: 379–386
- Yu C, Xie L, Zhang X, Zhou H, Zhu H. 2006. Genetic analysis on Tuoba Xianbei remains excavated from Qilang Mountain Cemetery in Qahar Right Wing Middle Banner of Inner Mongolia. *FEBS Lett*, 580: 6242–6246
- Yu L, Li Y W, Ryder O A, Zhang Y P. 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *BMC Evol Biol*, 7: 198
- Yuan J, Sheng G, Hou X, Shuang X, Yi J, Yang H, Lai X. 2014. Ancient DNA sequences from *Coelodonta antiquitatis* in China reveal its divergence and phylogeny. *Sci China Earth Sci*, 57: 388–396
- Zhang H, Paijmans J L, Chang F, Wu X, Chen G, Lei C, Yang X, Wei Z, Bradley DG, Orlando L, O'Connor T, Hofreiter M. 2013. Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nat Commun*, 4: 2755