

古蛋白质分析在东亚古人类演化中的应用前景

饶慧芸^{1,2}

1. 中国科学院脊椎动物演化与人类起源重点实验室, 中国科学院古脊椎动物与古人类研究所, 北京 100044;

2. 中国科学院生物演化与环境卓越创新中心, 北京 100044

摘要: 东亚古人类演化是学术界关注的热点科学问题, 国内外学者对此进行了多学科的相关研究, 取得了很多重要进展, 但仍然存在许多尚未解决的问题。古蛋白质分析近年来成为古生物演化领域的又一个前沿和热点方向, 取得了一系列重要突破。较之古 DNA, 古蛋白质的保存优势使其可以在时间上和地域上突破古 DNA 的限制, 在古人类演化领域大有可为。东亚古人类化石丰富且时段大致连续, 但更新世或更早期时期的分子证据非常缺乏。本文从古蛋白质分析的发展史、研究潜力、难点与挑战以及思考与展望等几方面, 对古蛋白质分析在东亚古人类演化研究中的应用前景进行梳理与思考。相信随着更多分子证据的积累, 古蛋白质分析可为东亚古人类的演化脉络提供更多关键性的线索, 极大地促进人类演化研究。

关键词: 生物人类学; 系统发育; 古蛋白质组学; 质谱; 东亚

An application prospect of paleoproteomic analysis in the evolution of East Asian populations

RAO Huiyun^{1,2}

1. Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044; 2. CAS Center for Excellence in Life and Paleoenvironment, Beijing 100044

Abstract: The evolutionary history of East Asian populations has involved several multidisciplinary approaches, but this research will bring in a different type of analysis, that of paleoproteomics (proteome [or protein?] analysis). Compared with ancient DNA, ancient proteins preserve in (sub)tropical areas, thus offering significant potential in tracking human evolutionary history. Abundant human fossils have been discovered in East Asia, including *Homo erectus*,

收稿日期: 2021-12-01; 定稿日期: 2022-03-14

基金项目: 国家重点研发计划 (2022YFF0903800); 中国科学院 B 类战略性先导科技专项 (XDB26000000); 中国科学院稳定支持基础研究领域青年团队计划 (YSBR-019); 教育部人文社会科学研究一般项目 (17YJAZH107); 中国科学院青年创新促进会项目 (2020076)

作者简介: 饶慧芸, 副研究员, 主要从事古蛋白质和有机残留物研究。E-mail: raohuiyun@ivpp.ac.cn

Citation: Rao HY. An application prospect of paleoproteomic analysis in the evolution of East Asian populations[J]. Acta Anthropologica Sinica, 2022, 41(6): 1083-1096

archaic *Homo sapiens*, early modern humans and late *Homo sapiens*, which generally form a continuous sequence. However, molecular evidence from the Pleistocene or earlier are scarce. In order to promote the application of paleoproteomic analysis to East Asia, this paper reviews its history, potential, challenges and future prospects. Key highlights are listed below. 1) Although ancient protein analysis goes back to 1954, its development was quite slow until the advent of soft-ionization mass spectrometry in 2000. Advances in high-resolution and high-throughput instrumentation now allows researchers to study ancient proteomes; a study that has been well employed in archaeological and evolutionary research. 2) Enamel, dentine and bone are three main substrates for protein preservation in human fossils, with enamel possessing the highest potential for deep time. The earliest enamel proteome sequences were retrieved from a 1.9 MaBP *Gigantopithecus* molar from South China, with a thermal age of 11.8 MaBP at 10°C. 3) Different taxonomic groups and protein types have diverse amino acid substitution rates. Based on current research, there are certain mutation positions along the main proteins of bones and teeth from human fossils that could help construct human evolutionary history in broader space and deeper time. 4) Paleoproteomic analysis faces the challenges of low endogenous protein content in fossil samples and a lack of full reference database. 5) Further methodological exploration should focus on three aspects, i.e., optimization of extraction methods, automatic species identification of ZooMS spectra and in-depth interpretation of mass spectrometry data. With an accumulation of more molecular evidence, especially from earlier human fossils which are out of reach for ancient DNA, paleoproteomic analysis could provide more clues in helping to draw a more accurate and clear phylogenetic tree on human evolutionary history in East Asia.

Keywords: biological anthropology; phylogeny; palaeoproteomics; mass spectrometry; East Asia

1 引言

人类的起源和演化一直是学术界关注的重大课题，正确认识人类的过去有利于我们客观地审视现在和规划未来。古人类化石是探讨人类演化的最直接证据，随着这些化石的不断发现，人类演化的谱系图也变得更为复杂。自 20 世纪初在内蒙古萨拉乌苏和北京周口店发现古人类化石以来，迄今在中国境内至少发现有 70 余处更新世时期的古人类化石地点，年代跨越更新世早期到晚期，包括直立人、古老型智人、早期现代人和晚期智人，以及一些在演化分类上存在争议的人类化石成员^[1]，这些珍贵的化石材料为揭示东亚古人类演化的奥秘提供了证据。

古人类学研究发现，中国直立人的体质特征内部变异非常大，且不同地区直立人的演化速率不一致，可能存在不同的直立人支系或者隔离人群；中国境内的古老型智人则表现出较大的复杂性，具有东亚古人类和欧洲尼安德特人镶嵌型的体质特征¹⁾，并呈现出向

1) 或许与欧洲人群存在基因交流

早期现代人过渡的特征；距今 10 万年左右早期现代人出现在中国的华南地区，并向华北扩散^[2-7]。最近发表的华龙洞古人类相关成果却表明，距今 30 万年前东亚地区人类已经发生了从古老形态向现代形态的演化过渡，可将华龙洞人定义为东亚地区最早的准现代人^[8]。由此可见，更新世以来的东亚古人类群体复杂多样，目前对于直立人、古老型智人、早期现代人和晚期智人之间的关系尚不十分清楚。

二十多年来，分子生物学在探索东亚古人类演化中同样发挥了重要作用。2001 年，国内学者基于现代人群 Y 染色体的遗传学分析，支持东亚的现代人群起源于非洲^[9]，然而单亲遗传标记会丢失掉大量的祖先信息。近年来，东亚的古基因组研究获得了众多突破性进展，但受限于保存状况，目前大多数研究集中在探索万年以来现代人群的迁徙与演化，更新世古人类的相关研究较少^[10-16]。最新研究成果表明，以田园洞人和 AR33K 个体为代表的古东亚人群曾在东亚北部广泛存在，但在末次盛冰期之后可能消失，而东亚古北方人群开始出现，并且伴随着 EDAR 基因的适应性突变^[14]；而距今约 1.1 万年前的广西隆林个体则是一个未知的古老人群，对现代东亚人群无明显的遗传贡献^[15]。由此可见，东亚人群的遗传历史非常复杂，目前古基因组的相关研究仅初步理清了部分演化历史，东亚的古人类演化仍然存在很多疑问，这些问题的解决有赖于对更多早期古人类化石的遗传学研究。

华南是东亚现代人形成的中心区域，出土了丰富的早期现代人化石。由于华南地处热带亚热带区域，化石的 DNA 难以保存，利用先进的捕获技术获得的古基因组数据目前最早也只达到距今一万年左右^[15,17,18]。同时，直立人和古老型智人阶段的许多化石由于年代久远，其中的古 DNA 可能已经消耗殆尽，无论从地域上还是时间上都对探讨东亚古人类演化的完整脉络带来了巨大的挑战。幸运的是，在 DNA 保存不佳的情况下，古蛋白质可以作为一种替代手段，用以探讨不同人群之间的系统发育关系及其演化历史。不同类群的同源蛋白质序列中存在氨基酸的变异位点，其来源于相应基因编码区的核苷酸替换，因此同源蛋白质氨基酸序列的相似程度同样可以反映不同人群之间的亲缘关系，可以利用蛋白质序列构建系统发育树，探讨古老人群与现代人群之间的演化关系。

值得一提的是，古蛋白质比古 DNA 降解速率慢，可在更长的时间尺度（百万年以上）和更广阔的地域范围（热带、亚热带区域）探索早期人类的演化历程。目前古蛋白质的研究已经应用到距今 380 万年前的非洲热带地区，学者从鸵鸟蛋壳中成功提取和检测到相关蛋白质，并提出蛋白质与无机矿物结合可保存更长时间^[19]。学者也成功地从距今 190 万年前的广西吹风洞（亚热带地区）出土的巨猿牙釉质上提取到相关古蛋白质并进行系统发育分析，解决了学术界长期争论的巨猿演化之谜^[20]。近五年来发表了一系列古蛋白质研究的重要成果^[20-23]，尤其是 2019 年夏河人的发现以及 2020 年先驱人的系统演化研究，引起学术界广泛关注；通过古蛋白质序列分析不仅发现了除丹尼索瓦洞以外的首例丹尼索瓦人群相关化石，探讨了青藏高原最早的人类活动及其环境适应^[24]，还从分子角度确定先驱人是尼安德特人、丹尼索瓦人和现代人的最近共同祖先的姐妹谱系^[25]。由此可见，古蛋白质研究在早期人类、尤其是更为古老的人类群体的演化领域具有非凡的潜力。目前，国际上古蛋白质研究已经兴起，然而国内刚刚起步，为了加快古蛋白质分析在东亚古人类演化这一重大科学问题上取得重要进展，下面将从古蛋白质分析的发展史、研究潜力、难点与挑战以及思考与展望等几方面进行综述和介绍。

2 古蛋白质分析的发展史

蛋白质是生物的重要组成部分, 主要由 20 种常见氨基酸组成, 由于氨基酸序列由相应基因编码而成, 其中同样蕴含着丰富的生物种属和演化信息。相较古 DNA 研究的蓬勃发展, 古蛋白质分析的发展明显滞后。2014 年, Cappellini 等总结了古蛋白质分析的发展史并指出其未来的发展潜力^[26]。在长时间尺度上, 古蛋白质不仅比古 DNA 保存更好(抗降解)^[27,28], 而且不同的组织、发育阶段和生理过程(如疾病状态)都可能具有特定的蛋白质表达, 通过对考古和古生物化石样本的蛋白质组分析亦可得到个体的生理学信息^[29]。

对古蛋白质的分析可追溯到 1954 年, Abelson 首次分析了化石骨骼和贝壳类中的氨基酸^[30,31]。组成生物体的常见氨基酸除最简单的甘氨酸外都含有不对称碳原子, 在结构上可以分为 L 型和 D 型旋光异构体; 生物体产出的氨基酸基本都为 L 型, 在个体死亡之后随着时间的推移逐渐向 D 型转化, 直到两者等量达到平衡, 从而失去旋光性, 这一过程称为外消旋作用^[32]。由于 D/L 型的转化率是时间的函数, 而转化速度则主要由温度决定, 因此可以通过测定考古样品中 D/L 型氨基酸的比率, 即氨基酸外消旋法, 来建立系列年代, 耦合古环境和古气候^[33]。也可以此评估不同埋藏环境中的蛋白质保存状况^[19], 为进一步的古蛋白质分析提供预判依据。

抗原抗体的特异性免疫反应, 为引入免疫法进行古蛋白质的鉴定奠定了基础。常用的方法有酶联免疫吸附测定法、免疫荧光显微技术、交叉免疫电泳法、消化捕获免疫测定法等, 目前已应用于陶器残留物^[34]、石器残留物^[35,36]、绘画所用粘合剂^[37,38]以及古代丝织物^[39]的鉴定和分析, 在先民食谱、器物用途、古代彩绘工艺以及古代织物鉴定等领域具有重要意义。近年来, 这一方法也被应用到古生物领域, 探讨恐龙骨骼或软骨中胶原蛋白等的存在^[40,41], 并通过对亿万年前羽毛化石中角蛋白的鉴定, 来探讨羽毛分子的演化以及间接确定色素体的存在^[42,43]。该方法只针对特定的蛋白质, 灵敏度高, 所需样本量小, 非常适用于考古样品分析和文物保护研究, 然而对亿万年前样品的有效应用尚有争议。

2000 年, 随着软电离质谱技术的问世, 古蛋白质的研究步伐加快^[44]。2009 年, Buckley 等学者建立了一套通过质谱法进行动物种属鉴定的技术, 即 ZooMS (Zooarchaeology by Mass Spectrometry) 技术。其主要流程是将碎骨中提取的胶原蛋白经过酶解得到肽段混合物, 点靶后放入基质辅助激光解析电离飞行时间质谱仪 MALDI TOF MS(Matrix-Assisted Laser Desorption/ Ionization Time of Flight Mass Spectrometry) 进行检测, 可快速获得其质谱图, 通过谱图数据库比对即可准确鉴定其物种^[45,46]。除骨骼外, 这一方法同样可用于牙齿、皮毛、角甚至蛋壳的鉴定^[46-49]。该方法简单高效, 且可实现无损或者微损分析, 目前在考古和古生物领域已经有了广泛的应用。有学者通过该技术对丹尼索瓦洞出土的 2000 多片碎骨进行分析, 最终找到一片人骨^[50], 进一步的古基因组分析发现她是一名尼安德特人与丹尼索瓦人的第一代混血儿^[51]。由此可见, ZooMS 手段对于筛选古人类化石与指导进一步的分析, 如测年、食谱分析、古蛋白质和古 DNA 测序等, 具有重要价值。该方法同样可用于建立比较全面的遗址动物群, 对动物考古和古环境研究是一个很好的补充, 通过与传统形态学鉴定结果的比较亦可在一定程度上反映古人类的行为^[52]。通过对骨角器、

皮质物、装饰品等进行 ZooMS 分析, 鉴别其原料来源, 可反映古代人群对于动物资源的选择和利用, 并进一步探讨其物质、精神及艺术审美层面的需求^[53-55]。

近年来, 高分辨率高通量质谱技术的发展使得探索古蛋白质组成为可能, 目前在考古残留物研究中应用广泛, 可以反映先民食谱(如陶器残留物^[56]、牙结石^[57]、胃食^[58]等)、材料文化(如皮革^[59]、丝绸^[60]、羊毛^[61]等)、粘合剂使用(如彩绘^[62]、建筑地仗层^[63]、其他工具^[64]等)、性别判断(如牙釉质^[65,66]等)、古病理(如牙结石^[67]、肺组织^[68]等)等各方面的信息。由此可见, 蛋白质组学方法为有机残留物分析开辟了一个新领域。2013 年, 古蛋白质分析应用到古生物演化领域^{[69] 2)}, 在此发挥了它较古 DNA 保存更佳的优势, 不仅探讨了热带亚热带区域各类群间的系统发育关系^[22,69,70], 还突破年代的限制, 研究早更新世甚至上新世时期的生物演化^[71,72]。2019 年, 随着牙釉质蛋白质组应用到该领域, 解决了史蒂芬犀、巨猿的演化等一系列重要的科学问题^[20,23], 古蛋白质分析已经成为古生物领域的一个前沿和热点方向。早在 2018 年, Welker 就探讨了古蛋白质分析在古人类演化方向的巨大潜力^[73]。2019 年, 学者运用古蛋白质分析手段发现了首个丹尼索瓦洞之外的丹尼索瓦相关人群, 即夏河人, 该发现为丹尼索瓦人的地史分布以及形态特征提供了更多信息, 并且将人类登上青藏高原的时间提前到距今 16 万年前^[24]。鉴于这一成果的重大影响, 有学者提出在古人类演化领域, 继古 DNA 后古蛋白质的时代已经到来^[74]。2020 年, 参与夏河人研究的相关团队又报道了先驱人的牙釉质蛋白质组, 探讨先驱人与其他中更新世和晚更新世古人类的系统发育关系^[25]。随着更多数据的积累, 未来有望绘制出更加清晰的古人类演化图谱。

3 研究潜力

关于古蛋白质分析在古人类演化方向的研究潜力, 这里主要关注两方面内容: 1) 古蛋白质能够保存多久? 2) 相近物种的蛋白质序列差异多大, 即系统发育分析的数据集丰富度如何?

3.1 古蛋白质的保存潜力

骨骼是有机分子与无机矿物的复合体, 胶原蛋白与羟磷灰石交错排列, 这一体系有利于蛋白质保存^[75]。牙本质的蛋白质组体系与骨骼相似, 但前者更为致密的结构和更高的矿物密度使其中的蛋白质保存更好, 这在夏河人的古蛋白质研究中已经得到证实^[24]。牙釉质是脊椎动物中最坚硬的组织, 其矿物含量高达 90% 以上。相较于牙本质和骨骼, 牙釉质组成了一个更为封闭的体系^[76], 利于釉质蛋白质的保存, 近年来发表的牙釉质相关成果同样证实了这一规律^[20,23,25]。随着样品年代的增加, 牙本质和骨骼中的蛋白质组复杂度逐渐下降, 到最后只有 I 型胶原蛋白保存下来, 但是牙釉质蛋白质组基本保持不变^[73]。

自 2007 年学者报道了距今 6800 万年前的霸王龙骨骼化石中的胶原蛋白序列以来, 学术界对于蛋白质能否保存到白垩纪一直争论不断^[77-82]。目前无争议的最早的胶原蛋白序列来自距今 350 万年前的一个古老巨骆驼胫骨, 其发现在加拿大东部北极地带的努纳武特(Nunavut)地区^[72]; 而最早的牙釉质蛋白质组来自于广西田东吹风洞的一枚距今 190 万年

2) 因为存在争议, 这里没有将 2007 年以来恐龙蛋白质分析的相关工作包括在内。

前的步氏巨猿臼齿^[20]。除了与保存体系本身的性质有关,蛋白质的保存和降解同样受到很多外界因素的影响,其中最主要的是温度和湿度^[83]。为了便于不同遗址点间的比较,学者引入了热年龄(Thermal Age)的概念,即结合各遗址点的外界保存环境因素,将其绝对年龄转换成10度恒温下的相对年龄。因此,发现于北极地带的距今350万年前的古老巨骆驼胫骨,其热年龄并不高;相反,发现于广西亚热带地区的距今190万年前的巨猿釉质,其热年龄达到了距今1180万年^[20]。理论上来说,如果在10度左右的遗址开展工作,可以成功获取到千万年以上的古蛋白质序列。当然,目前还没有比较确切的获取到千万年以上的古蛋白质序列的报道,但古蛋白质能够保存百万年以上是毋庸置疑的。

3.2 相近物种的蛋白质序列差异

这里涉及到氨基酸的突变速率,突变速率越大,两个相近物种在同样的分化时间内即可积累更多的突变位点,利于两者的区分和系统发育分析。骨骼和牙本质中含量最高的蛋白质是I型胶原蛋白(col1),它由三条多肽链缠绕成三螺旋结构。在大多数脊椎动物(包括哺乳类、鸟类、爬行类和两栖类等)中,I型胶原蛋白由两条 $\alpha 1$ 链(col1 $\alpha 1$)和一条特异性更高的 $\alpha 2$ 链(col1 $\alpha 2$)组成,而在许多硬骨鱼类中,还含有一条特异性最高的 $\alpha 3$ 链(col1 $\alpha 3$),构成异三聚体($\alpha 1\alpha 2\alpha 3$)的I型胶原蛋白^[84]。I型胶原蛋白是一种比较保守的蛋白质,在不同的脊椎动物类群中表现出不一样的突变速率,其中突变速率最高的是鱼类,其次为爬行类和两栖类,接着是哺乳类,而鸟类的I型胶原蛋白序列最为保守^[85]。除I型胶原蛋白外,骨骼和牙本质中还含有其他胶原蛋白以及非胶原蛋白,如胎球蛋白A、双糖素等,这些蛋白质的氨基酸突变速率一般比I型胶原蛋白更高^[73]。虽然牙釉质蛋白质组较小,但其保存时间更长且保存更完整,具有更高的蛋白质序列覆盖率和氨基酸突变速率,同样是构建系统发育树的理想数据来源^[20,23,25,73]。

具体到古人类的分子证据,目前测有尼安德特人和丹尼索瓦人的高覆盖率的基因组^[86-91],可通过相关基因的外显子序列转录、翻译得到蛋白质的氨基酸序列,发现其中的突变位点。通过比对距今70万年以来分化的尼安德特人、丹尼索瓦人和现代人三者的蛋白质序列,发现前两者有各自特有的突变位点,也有前两者共有但与现代人有差异的位点,这些突变位点分布在许多蛋白质上,包括一些在骨骼和牙齿中常见的蛋白质,如I型胶原蛋白 $\alpha 2$ 链、X型胶原蛋白 $\alpha 1$ 链、XVII型胶原蛋白 $\alpha 1$ 链、釉质素、成釉蛋白等,这在蛋白质水平实现三者的鉴定提供了依据^[90,91]。除此之外,学者应用蛋白质组学方法直接测定了夏河人、先驱人和德马尼西直立人的蛋白质序列。在距今16万年前的夏河人牙本质中发现了8种胶原蛋白,其中有两个突变位点,分别为丹尼索瓦人特有和夏河人特有的突变位点^[24]。在距今77-95万年前的先驱人牙釉质中,学者发现了7种蛋白质,其中有两个突变位点,一个与其他人属成员不一致而与*Pongo/Gorilla/Pan*属一致,另外一个为先驱人特有的突变位点,与目前已有序列的所有其他人科成员都不一致;学者报道的德马尼西直立人(距今177万年前)牙釉质蛋白质组中含有6种蛋白质,其中也有一个突变位点,与其他人属成员不一致,而与*Pongo/Gorilla/Pan*属一致^[25]。可见,虽然一些古人类化石的年代久远,保存的蛋白质种类较少,序列覆盖也不完整,其中仍然可以发现一定的突变位点,补充了古人类演化史上一些关键性的线索,有望在更长的时间尺度上探讨这一重要的科学问题。

4 难点与挑战

虽然古蛋白质分析在古生物和古人类演化领域表现出巨大的研究潜力，目前仍存在一定的难点与挑战。这主要表现在两方面：1) 随着样品年代增加，其中的内源性蛋白质含量不断下降；2) 参考数据库严重缺乏。

4.1 古老样品内源性蛋白质含量低

从生物个体死亡、埋藏到被发现并分析得到数据的这一过程中，伴随着各种各样的外源性污染。古生物化石一经埋藏，其基底会迅速被细菌和真菌占领，内源性的蛋白质成分不断被消耗；在发掘、整理和储存过程中，人为的接触和处理，如加入蛋白质类的固化剂，也可能产生额外的污染；在样品分析前处理过程中，有目的地添加胰蛋白酶等将引入一个确定的污染源；在提取和质谱分析期间，来自实验室的交叉污染也可能发生；整个过程中，化石中内源性蛋白质成分不断下降，而外源性的污染蛋白质成分则不断增加，其蛋白质组体系变得更为复杂^[92]。对于年代古老的样本而言，如何在繁杂的数据中获取到更多的内源性蛋白质信息是一个难点。

为了得到更为准确可靠的数据，古蛋白质领域的相关学者在 2018 年提出了一套古蛋白质分析的操作规范，涉及样品选择、实验室提取、质谱分析、数据搜库、解释、验证和共享等一系列流程，为相关分析提供了一些有益的参考建议^[92]。在实验室的规范操作下获取的蛋白质组数据本身可用于确定其内源性，如鉴定的蛋白质组成、肽段分布、氨基酸修饰等是否符合古蛋白质组的特征；同时，其他手段也可提供辅助证据，如热年龄评估、氨基酸外消旋比例、C 和 N 等有机元素含量、分子谱图（红外光谱、拉曼光谱等）、形态学观察（电镜、组织学观察）等^[19,20,23,24,73,93-97]。在确定古蛋白质内源性的一系列手段中，计算脱酰胺修饰比例是目前应用最为广泛的一种方法。脱酰胺修饰发生在蛋白质序列中含量较多的谷氨酰胺和天冬酰胺这两种氨基酸上，通过 MALDI 谱图和串联质谱数据的搜库结果都能计算其比例，学者还开发了相应的 R 包 (q2c) 和 deamidATE 工具实现自动计算^[98,99]。然而，除样品年代外，脱酰胺修饰的比例还受到很多其他因素影响，如样品部位、蛋白质本身的一些性质（谷氨酰胺和天冬酰胺的差异、其邻近氨基酸、蛋白质的三维结构等）、埋藏环境（PH、温度、湿度等）、实验条件（样品处理方式、仪器状态等）、计算方法等^[73,94,100-105]，还需进一步探索这一方法的更为准确的利用模式。

4.2 参考数据库严重缺乏

以哺乳动物为例，截止到 2021 年 8 月底，在 Uniprot 网站上记录的总物种数达 8546 种，然而有蛋白质组的仅仅 108 种，有胶原蛋白序列或文献发表有肽质量指纹图谱的也仅 300 多种，其中有 95% 以上的空白。参考数据库的缺乏不仅会影响蛋白质的鉴定及获取序列的准确性^[22,97,106]，而且不利于后续的系统发育树构建。具体到古人类领域，目前在 Uniprot 和 NCBI(National Center of Biotechnology Information) 网站上可以下载到的现代人蛋白质组中包括 2 万多个基因。除此之外，目前测有蛋白质组的古人类化石较少，包括夏河人（牙本质蛋白质组，发现 8 种胶原蛋白）^[24]、先驱人（牙釉质蛋白质组，发现 7 种蛋白质）和德马尼西

直立人（牙釉质蛋白质组，发现 6 种蛋白质）^[25]。早期也有学者运用 MALDI TOF/TOF 质谱仪对尼安德特人的骨钙蛋白和牙釉质蛋白质进行部分测序，但得到的序列与现代人并无差异^[107,108]。目前，高通量基因组测序技术已经获取到了尼安德特人、丹尼索瓦人和一些早期现代人的核基因组数据，可通过相关基因的外显子序列转录、翻译得到相应的蛋白质组数据。综上可知，目前古人类化石的蛋白质组数据仍然极度缺乏，未来有大量工作需要开展。

5 思考与展望

前文总结了古蛋白质分析的研究潜力及挑战，下面将讨论未来需要继续探索的三个主要方向及其在东亚古人类演化方面的应用前景：1）古蛋白质提取方法的优化；2）ZooMS 谱图的自动种属鉴定；3）质谱数据的深度解读；4）在东亚古人类演化方面的应用前景。

5.1 古蛋白质提取方法的优化

由于考古和古生物样品，尤其是古人类化石都非常珍贵，并且具有唯一性，为了减少样品损失，需要探讨如何从少量的样品中尽可能多地获取其中的内源性蛋白质信号。目前学者已经采用一系列无损的方法，如近红外光谱、拉曼光谱等，首先评估样品中蛋白质的保存状况，筛选出保存较好的样品进行下一步的分析，以提高成功率和减少不必要的样品损失^[95,96]。针对不允许有损分析的羊皮纸、艺术作品、骨器等样品，学者发明了聚氯乙烯橡皮擦摩擦取样的手段^[109]，甚至从盛放骨器的塑料袋和塑料盒里成功地提取到了相关肽段信号^[55,110]，还针对性地发明了整合有蛋白酶和 C18 纯化膜的相关应用耗材，以达到微损提取和一站式酶解纯化的目的^[111-113]。此外，相关研究也评估了不同提取方法和不同酶切手段对获取到的蛋白质信号的影响^[114,115]。然而这些研究尚不够系统，且大多是针对蛋白质保存较好的样品，在提取方法上作了一定的优化。为了突破年代的限制，更加切合古生物化石分析的要求，未来需要结合埋藏环境，针对年代久远蛋白质保存不佳的样品开展更系统的探索。

5.2 ZooMS 谱图的自动种属鉴定

ZooMS 方法简单高效，能够快速获得样品的质谱图，然而，对谱图的进一步解析和鉴定严重滞后。目前主要依靠有经验的研究人员进行手动鉴定，这个过程耗时较长且鉴定结果无法量化，尤其对于一些质量不高的谱图，无法显示鉴定结果的可信度以及其他备选鉴定物种。为了解决这些问题，2018 年学者尝试采用机器学习（随机森林的集成学习）的方法进行谱图鉴定，选取了 14 个物种每个物种各 5 个样本的数据作为训练集，并在 6805 个谱图上进行验证，结果表明这一方法可鉴别到科或亚科，而亲缘关系更近的属种一级的区分仍然需要人工处理^[116]。2020 年，学者建立了 bacollite（R 包）的方法用于实现三个相近物种（同属牛科的牛、绵羊和山羊）的自动种属鉴定，与机器学习不同，这一方法有清晰的分类标准，解释性更好，具体流程如下：首先，输入不同物种的肽段标记物序列；然后，将未知样品的谱图与所有肽段标记物的理论谱进行匹配，设定一系列的相关度阈值，计算每个物种不同阈值下的肽段标记物匹配数目；最后，转化为得分，得分最高的物种即为鉴定结果^[117]。然而，这个方法目前仅适用于少数几个物种间的鉴定，使用的肽段标记物也不多。如果需要在古人类活动遗址中进行大批量的 ZooMS 分析，筛选可

能的古人类化石，未来或许可以探索机器学习和 *bioinformatic* 方法的兼容性，结合两者的优势，实现对未知样品的快速而准确的自动种属鉴定。

5.3 质谱数据的深度解读

得到原始质谱数据之后，对数据的深度解读也值得探索。首先是肽段的准确鉴定，机器学习近年来被应用到多肽鉴定中，涉及从头测序、谱图预测、保留时间预测^[118-122]等，并且这一手段也被成功地应用到古代样品中，进一步验证一些含有突变位点的多肽的鉴定结果^[25]。其次，肽段中特定氨基酸的修饰也可在一定程度上反映样品的生理功能和成岩作用^[21,29]。最后，对样品中检测到的古蛋白质组的功能分析，可以获取个体的生理学、生物学过程和疾病等相关信息^[21,29]。比如，牙釉质中釉原蛋白的 X 型和 Y 型表达可以反映个体的性别^[123]，骨骼中胎球蛋白 A 的定量表达与个体年龄相关^[124]，一些免疫蛋白可以反映个体的疾病状态^[124]。

除此之外，特殊的样品如牙结石、胃容物、粪化石等的蛋白质组分析，不仅可以提供生物个体本身的信息，还可揭示个体的食谱、口腔健康、肠道菌群等一系列信息，反映出丰富的个体生活史^[57,58,67,125]。同时，蛋白质组是基因组功能的最终执行者，不同组织的细胞中虽然含有同一套遗传物质，但是表达的基因有所不同，这就导致了不同组织中蛋白质组成的差异，进一步在生命过程中发挥不同的功能^[126]。由于肌肉、奶、血液等不同组织的蛋白质组成相差很大，学者可通过相关样品（如牙结石、残留物等）的蛋白质组分析来探讨先民对于驯化动物的次级产品利用^[57,127]。同为个体硬组织的骨骼和牙本质的蛋白质组成较为相似，但其中不同蛋白质的相对含量亦有差异，牙釉质与牙本质更是拥有两个截然不同的蛋白质组体系^[73,76]。在不同组织或不同物种中，蛋白质的差异化表达亦可帮助我们推测灭绝物种的生理学过程。例如，胎球蛋白 A 是骨骼中常见的蛋白质，可调控骨骼的矿化过程，研究者通过对广西吹风洞出土的距今 190 万年前的巨猿牙齿进行古蛋白质分析，首次在牙釉质蛋白质组中发现胎球蛋白 A，由此推测，巨猿需要额外的胎球蛋白 A 来促进其大白齿的生物矿化作用从而形成厚重的釉冠^[20]。值得注意的是，古生物领域目前在这些方面的研究还较为缺乏，未来需要更多的探索。

古蛋白质分析作为一个前沿和热点方向，未来还需进行更多方法学上的探索。在这个长期的过程中，可以结合埋藏学、机器学习、功能蛋白质组学等领域的最新研究进展，如果能在以上三个方面实现方法学上的再突破，可帮助我们高效地获取古生物化石中的遗传学和生理学信息，未来必将会极大地推动古生物的演化研究。

5.4 在东亚古人类演化方面的应用前景

东亚古人类化石丰富，前人的研究已经取得了一系列原创性的重要成果，但仍然存在许多未解之谜。比如，中国的直立人是否存在多个支系，其与后期人类之间的关系如何？丹尼索瓦人是一个以遗传物质确立的古老型人类，目前关于其体质特征、时空分布特点等信息知之甚少，引发了学术界广泛的关注。甘肃夏河人的发现验证了其在东亚广泛存在的推测^[24]，龙人被认为与夏河人属于同一支系^[128]，许昌人头骨化石的体质特征与古基因组甲基化预测的丹尼索瓦人特征相吻合^[129]，许昌人、龙人、乃至中国的其他古人类化石是否同样属于丹尼索瓦人？能否从分子层面找到东亚古人类与欧洲人群基因交流的证据？中

国是否存在一些未知的古人类？哪几个类群是东亚现代人的直系祖先？如果我们可以获取到足够多东亚古人类化石的分子证据，有望为上述问题的解答提供一些关键性的线索。

本文梳理了古蛋白质分析的发展史、研究潜力、难点与挑战以及思考与展望等，表明蛋白质可以在热带亚热带的埋藏环境中保存至百万年以上，这基本涵盖了东亚所有古人类化石的时空范围，无论在时间上还是地域上，都可以突破古 DNA 的限制，在人类演化研究中表现出巨大的应用潜力。相信随着古蛋白质提取和分析方法的改进以及古人类化石分子数据库的完善，未来有望绘制出更为清晰完整的东亚古人类演化脉络。

致谢：中国科学院大学杨益民教授为本文的写作提出重要建议，作者谨此致谢。

参考文献

- [1] 刘武, 吴秀杰, 邢松, 等. 中国古人类化石 [M]. 北京: 科学出版社, 2014
- [2] 刘武, 吴秀杰, 邢松. 现代人的出现与扩散——中国的化石证据 [J]. 人类学学报, 2016, 35(2): 161-171
- [3] 刘武, 吴秀杰, 邢松. 更新世中期中国古人类演化区域连续性与多样性的化石证据 [J]. 人类学学报, 2019, 38(4): 473-490
- [4] 刘武, 邢松, 吴秀杰. 中更新世晚期以来中国古人类化石形态特征的多样性 [J]. 中国科学: 地球科学, 2016, 46(7): 906-917
- [5] 吴秀杰. 中国古人类演化研究进展及相关热点问题探讨 [J]. 科学通报, 2018, 63(21): 2148-2155
- [6] 吴秀杰. 中国发现的主要直立人头骨化石 [J]. 科学, 2019, 71(3): 20-24
- [7] 刘武, 邢松, 张银运. 中国直立人牙齿特征变异及其演化意义 [J]. 人类学学报, 2015, 34(4): 425-441
- [8] Wu XJ, Pei SW, Cai YJ, et al. Morphological description and evolutionary significance of 300 ka hominin facial bones from Hualongdong, China[J]. Journal of Human Evolution, 2021, 161: 103052
- [9] Ke YH, Su B, Song XF, et al. African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes[J]. Science, 2001, 292(5519): 1151-1153
- [10] Zhang DJ, Xia H, Chen FH, et al. Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau[J]. Science, 2020, 370(6516): 584-587
- [11] Wang CC, Yeh HY, Popov AN, et al. Genomic insights into the formation of human populations in East Asia[J]. Nature, 2021, 591(7850): 413-419
- [12] Zhang F, Ning C, Scott A, et al. The genomic origins of the Bronze Age Tarim Basin mummies[J]. Nature, 2021, 599(7884): 256-261
- [13] Robbeets M, Bouckaert R, Conte M, et al. Triangulation supports agricultural spread of the Transeurasian languages[J]. Nature, 2021, 599: 616-621
- [14] Mao XW, Zhang HC, Qiao SY, et al. The deep population history of northern East Asia from the Late Pleistocene to the Holocene[J]. Cell, 2021, 184(12): 3256-3266.e3213
- [15] Wang TY, Wang W, Xie GM, et al. Human population history at the crossroads of East and Southeast Asia since 11,000 years ago[J]. Cell, 2021, 184(14): 3829-3841.e3821
- [16] Yang MA, Fan XC, Sun B, et al. Ancient DNA indicates human population shifts and admixture in northern and southern China[J]. Science, 2020, 369(6501): 282-288
- [17] Bai F, Zhang XL, Ji XP, et al. Paleolithic genetic link between Southern China and Mainland Southeast Asia revealed by ancient mitochondrial genomes[J]. Journal of Human Genetics, 2020, 65(12): 1125-1128
- [18] Liu YL, Wang TY, Wu XC, et al. Maternal genetic history of southern East Asians over the past 12,000 years[J]. Journal of Genetics and Genomics, 2021, 48(10): 899-907
- [19] Demarchi B, Hall S, Roncal-Herrero T, et al. Protein sequences bound to mineral surfaces persist into deep time[J]. Elife, 2016, 5: e17092
- [20] Welker F, Ramos-Madrigal J, Kuhlwil M, et al. Enamel proteome shows that *Gigantopithecus* was an early diverging pongine[J]. Nature, 2019, 576: 262-265
- [21] Welker F, Hajdinjak M, Talamo S, et al. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne[J]. Proceedings of the National Academy of Sciences, 2016, 113(4): 11162-11167
- [22] Presslee S, Slater GJ, Pujos F, et al. Palaeoproteomics resolves sloth relationships[J]. Nature Ecology & Evolution, 2019, 3: 1121-1130
- [23] Cappellini E, Welker F, Pandolfi L, et al. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny[J]. Nature, 2019, 574: 103-107

- [24] Chen FH, Welker F, Shen CC, et al. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau[J]. *Nature*, 2019, 569(7756): 409-412
- [25] Welker F, Ramos-Madriral J, Gutenbrunner P, et al. The dental proteome of *Homo antecessor*[J]. *Nature*, 2020, 580: 235-238
- [26] Cappellini E, Collins MJ, Gilbert MTP. Unlocking ancient protein palimpsests[J]. *Science*, 2014, 343(6177): 1320-1322
- [27] Tomiak P, Penkman K, Hendy E, et al. Testing the limitations of artificial protein degradation kinetics using known-age massive *Porites* coral skeletons[J]. *Quaternary Geochronology*, 2013, 16: 87-109
- [28] Allentoft ME, Collins M, Harker D, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils[J]. *Proceedings of the Royal Society of London B: Biological Sciences*, 2012, 279(1748): 4724-4733
- [29] Cappellini E, Jensen LJ, Szklarczyk D, et al. Proteomic analysis of a Pleistocene mammoth femur reveals more than one hundred ancient bone proteins[J]. *Journal of Proteome Research*, 2011, 11(2): 917-926
- [30] Abelson PH. *Paleobiochemistry: Organic constituents of fossils*[M]. Washington, DC: Carnegie Institution of Washington, Yearbook, No 53, 1954: 97-101
- [31] Abelson PH. Amino acids in fossils[J]. *Science*, 1954, 119(3096): 576
- [32] Demarchi B, Collins M. Amino acid racemization dating[J]. *Encyclopedia of Earth Sciences*, 2015: 13-26
- [33] Penkman KE, Preece RC, Bridgland DR, et al. A chronological framework for the British Quaternary based on *Bithynia* opercula[J]. *Nature*, 2011, 476(7361): 446-449
- [34] Craig OE, Collins MJ. An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture[J]. *Journal of Immunological Methods*, 2000, 236(1): 89-97
- [35] Högberg A, Puseman K, Yost C. Integration of use-wear with protein residue analysis – a study of tool use and function in the south Scandinavian Early Neolithic[J]. *Journal of Archaeological Science*, 2009, 36(8): 1725-1737
- [36] Seaman MF, Nilsson NE, Summers GL, et al. Evaluating protein residues on Gainey phase Paleoindian stone tools[J]. *Journal of Archaeological Science*, 2008, 35(10): 2742-2750
- [37] Cartechini L, Vagnini M, Palmieri M, et al. Immunodetection of proteins in ancient paint media[J]. *Accounts of Chemical Research*, 2010, 43(6): 867-876
- [38] Scott DA, Warmlander S, Mazurek J, et al. Examination of some pigments, grounds and media from Egyptian cartonnage fragments in the Petrie Museum, University College London[J]. *Journal of Archaeological Science*, 2009, 36(3): 923-932
- [39] Zheng HL, Yang HL, Zhang W, et al. Insight of silk relics of mineralized preservation in Maoling Mausoleum using two enzyme-linked immunological methods[J]. *Journal of Archaeological Science*, 2020, 115: 105089
- [40] Schweitzer MH, Suo Z, Avci R, et al. Analyses of soft tissue from *Tyrannosaurus rex* suggest the presence of protein[J]. *Science*, 2007, 316(5822): 277-280
- [41] Bailleul AM, Zheng W, Horner JR, et al. Evidence of proteins, chromosomes and chemical markers of DNA in exceptionally preserved dinosaur cartilage[J]. *National Science Review*, 2020, 7(4): 815-822
- [42] Pan YH, Zheng WX, Sawyer RH, et al. The molecular evolution of feathers with direct evidence from fossils[J]. *Proceedings of the National Academy of Sciences*, 2019, 116(8): 3018-3023
- [43] Pan YH, Zheng WX, Moyer AE, et al. Molecular evidence of keratin and melanosomes in feathers of the Early Cretaceous bird *Eoconfuciusornis*[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(49): E7900-E7907
- [44] Ostrom PH, Schall M, Gandhi H, et al. New strategies for characterizing ancient proteins using matrix-assisted laser desorption/ionization mass spectrometry[J]. *Geochimica et Cosmochimica Acta*, 2000, 64(6): 1043-1050
- [45] Buckley M, Kansa SW, Howard S, et al. Distinguishing between archaeological sheep and goat bones using a single collagen peptide[J]. *Journal of Archaeological Science*, 2010, 37(1): 13-20
- [46] Buckley M, Collins M, Thomas-Oates J, et al. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry[J]. *Rapid Communications in Mass Spectrometry*, 2009, 23(23): 3843-3854
- [47] Welker F, Soressi M, Rendu W, et al. Using ZooMS to identify fragmentary bone from the late Middle/Early Upper Palaeolithic sequence of Les Cottes, France[J]. *Journal of Archaeological Science*, 2015, 54: 279-286
- [48] Biard V, Gol'din P, Gladilina E, et al. Genomic and proteomic identification of Late Holocene remains: Setting baselines for Black Sea odontocetes[J]. *Journal of Archaeological Science: Reports*, 2017, 15: 262-271
- [49] Presslee S, Wilson J, Woolley J, et al. The identification of archaeological eggshell using peptide markers[J]. *STAR: Science & Technology of Archaeological Research*, 2018, 4(1): 13-23
- [50] Brown S, Higham T, Slon V, et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis[J]. *Scientific Reports*, 2016, 6: 23559
- [51] Slon V, Mafessoni F, Vernot B, et al. The genome of the offspring of a Neanderthal mother and a Denisovan father[J]. *Nature*, 2018, 561(7721): 113-116

- [52] Sinet-Mathiot V, Smith GM, Romandini M, et al. Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy)[J]. *Scientific Reports*, 2019, 9(1): 12350
- [53] Dekker J, Sinet-Mathiot V, Spithoven M, et al. Human and cervid osseous materials used for barbed point manufacture in Mesolithic Doggerland[J]. *Journal of Archaeological Science: Reports*, 2021, 35: 102678
- [54] Kirby D, Buckley M, Promise E, et al. Identification of collagen-based materials in cultural heritage[J]. *Analyst*, 2013, 138(17): 4849-4858
- [55] Martisius NL, Welker F, Dogandžić T, et al. Non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals[J]. *Scientific Reports*, 2020, 10(1): 7746
- [56] Hendy J, Colonese AC, Franz I, et al. Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers[J]. *Nature Communications*, 2018, 9(1): 4064
- [57] Bleasdale M, Richter KK, Janzen A, et al. Ancient proteins provide evidence of dairy consumption in eastern Africa[J]. *Nature Communications*, 2021, 12(1): 632
- [58] Maixner F, Turaev D, Cazenave-Gassiot A, et al. The Iceman's last meal consisted of fat, wild meat, and cereals[J]. *Current Biology*, 2018, 28(14): 2348-2355.e9
- [59] Brandt LØ, Schmidt AL, Mannering U, et al. Species identification of archaeological skin objects from Danish Bogs: Comparison between mass spectrometry-based peptide sequencing and microscopy-based methods[J]. *PLoS ONE*, 2014, 9(9): e106875
- [60] Li L, Gong YX, Yin H, et al. Different types of peptide detected by mass spectrometry among fresh silk and archaeological silk remains for distinguishing modern contamination[J]. *PLoS ONE*, 2015, 10(7): e0132827
- [61] Solazzo C, Heald S, Ballard MW, et al. Proteomics and Coast Salish blankets: A tale of shaggy dogs?[J]. *Antiquity*, 2011, 85(330): 1418-1432
- [62] Dallongeville S, Richter M, Schäfer S, et al. Proteomics applied to the authentication of fish glue: Application to a 17th century artwork sample[J]. *Analyst*, 2013, 138(18): 5357-5364
- [63] Rao HY, Li B, Yang YM, et al. Proteomic identification of organic additives in the mortars of ancient Chinese wooden buildings[J]. *Analytical Methods*, 2015, 7: 143-149
- [64] Rao HY, Yang YM, Abuduresule I, et al. Proteomic identification of adhesive on a bone sculpture-inlaid wooden artifact from the Xiaohe Cemetery, Xinjiang, China[J]. *Journal of Archaeological Science*, 2015, 53: 148-155
- [65] Rebay-Salisbury K, Janker L, Pany-Kucera D, et al. Child murder in the Early Bronze Age: Proteomic sex identification of a cold case from Schleinbach, Austria[J]. *Archaeological and Anthropological Sciences*, 2020, 12: 265
- [66] Stewart NA, Gerlach RF, Gowland RL, et al. Sex determination of human remains from peptides in tooth enamel[J]. *Proceedings of the National Academy of Sciences*, 2017, 114(52): 13649-13654
- [67] Warinner C, Rodrigues JFM, Vyas R, et al. Pathogens and host immunity in the ancient human oral cavity[J]. *Nature Genetics*, 2014, 46(4): 336-344
- [68] Hendy J, Collins M, Teoh KY, et al. The challenge of identifying tuberculosis proteins in archaeological tissues[J]. *Journal of Archaeological Science*, 2016, 66: 146-153
- [69] Buckley M. A molecular phylogeny of *Plesiorcterus* reassigns the extinct mammalian order 'Bibymalagasiasia'[J]. *PLoS ONE*, 2013, 8(3): e59614
- [70] Welker F, Collins MJ, Thomas JA, et al. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates[J]. *Nature*, 2015, 522(7554): 81-84
- [71] Welker F, Smith GM, Hutson JM, et al. Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae[J]. *PeerJ*, 2017, 5: e3033
- [72] Buckley M, Lawless C, Rybczynski N. Collagen sequence analysis of fossil camels, *Camelops* and c.f. *Paracamelus*, from the Arctic and sub-Arctic of Plio-Pleistocene North America[J]. *Journal of Proteomics*, 2019, 194: 218-225
- [73] Welker F. Palaeoproteomics for human evolution studies[J]. *Quaternary Science Reviews*, 2018, 190: 137-147
- [74] Warren M. Move over, DNA: Ancient proteins are starting to reveal humanity's history[J]. *Nature*, 2019, 570(7762): 433-436
- [75] Qu T, Tomar V. Understanding straining induced changes in thermal properties of tropocollagen-hydroxyapatite interfacial configurations[J]. *International Journal of Experimental and Computational Biomechanics*, 2015, 3(1): 62-81
- [76] Smith CEL, A. PJ, Agne A, et al. Amelogenesis imperfecta; Genes, proteins, and pathways[J]. *Frontiers in Physiology*, 2017, 8: 435
- [77] Schweitzer MH, Schroeter ER, Cleland TP, et al. Paleoproteomics of mesozoic dinosaurs and other Mesozoic fossils[J]. *Paleoproteomics*, 2019, 19(16): 1800251
- [78] Saitta ET, Liang R, Lau MCY, et al. Cretaceous dinosaur bone contains recent organic material and provides an environment conducive to microbial communities[J]. *eLife*, 2019, 8: e46205
- [79] Asara JM, Schweitzer MH, Freemark LM, et al. Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry[J]. *Science*, 2007, 316(5822): 280-285
- [80] Pevzner PA, Kim S, Ng J. Comment on "Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry"[J]. *Science*, 2008, 321(5892): 1040

- [81] Buckley M, Walker A, Ho SY, et al. Comment on “Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry”[J]. *Science*, 2008, 319(5859): 33c
- [82] Schroeter ER, DeHart CJ, Cleland TP, et al. Expansion for the *Brachylophosaurus canadensis* collagen I sequence and additional evidence of the preservation of Cretaceous protein[J]. *Journal of Proteome Research*, 2017, 16(2): 920-932
- [83] Matthiesen H, Høier Eriksen AM, Hollesen J, et al. Bone degradation at five Arctic archaeological sites: Quantifying the importance of burial environment and bone characteristics[J]. *Journal of Archaeological Science*, 2021, 125: 105296
- [84] Kimura S, Ohno Y. Fish type I collagen: Tissue-specific existence of two molecular forms, $(\alpha 1)2\alpha 2$ and $\alpha 1\alpha 2\alpha 3$, in Alaska pollack[J]. *Comparative Biochemistry Physiology Part B: Comparative Biochemistry*, 1987, 88(2): 409-413
- [85] Buckley M. Zooarchaeology by mass spectrometry (ZooMS) collagen fingerprinting for the species identification of archaeological bone fragments[M]. *Zooarchaeology in Practice*. Springer. 2018: 227-247
- [86] Hajdinjak M, Fu Q, Hübner A, et al. Reconstructing the genetic history of late Neanderthals[J]. *Nature*, 2018, 555(7698): 652-656
- [87] Prüfer K, de Filippo C, Grote S, et al. A high-coverage Neanderthal genome from Vindija Cave in Croatia[J]. *Science*, 2017, 358(6363): 655-658
- [88] Meyer M, Arsuaga JL, de Filippo C, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins[J]. *Nature*, 2016, 531(7595): 504-507
- [89] Prüfer K, Racimo F, Patterson N, et al. The complete genome sequence of a Neanderthal from the Altai mountains[J]. *Nature*, 2014, 505(7481): 43-49
- [90] Castellano S, Parra G, Sánchez-Quinto F, et al. Patterns of coding variation in the complete exomes of three Neanderthals[J]. *Proceedings of the National Academy of Sciences*, 2014, 111(18): 6666-6671
- [91] Meyer M, Kircher M, Gansauge MT, et al. A high-coverage genome sequence from an archaic Denisovan individual[J]. *Science*, 2012, 338(6104): 222-226
- [92] Hendy J, Welker F, Demarchi B, et al. A guide to ancient protein studies[J]. *Nature Ecology & Evolution*, 2018, 2(5): 791-799
- [93] Hollund H, Ariele F, Fernandes R, et al. Testing an alternative high - throughput tool for investigating bone diagenesis: FTIR in Attenuated Total Reflection (ATR) mode[J]. *Archaeometry*, 2013, 55(3): 507-532
- [94] Pal Chowdhury M, Wogelius R, Manning PL, et al. Collagen deamidation in archaeological bone as an assessment for relative decay rates[J]. *Archaeometry*, 2019, 61(6): 1382-1398
- [95] Sponheimer M, Ryder CM, Fewlass H, et al. Saving old bones: A non-destructive method for bone collagen prescreening[J]. *Scientific Reports*, 2019, 9(1): 13928
- [96] Madden O, Chan DMW, Dundon M, et al. Quantifying collagen quality in archaeological bone: Improving data accuracy with benchtop and handheld Raman spectrometers[J]. *Journal of Archaeological Science: Reports*, 2018, 18: 596-605
- [97] Rao HY, Yang YM, Liu JY, et al. Palaeoproteomic analysis of Pleistocene cave hyenas from east Asia[J]. *Scientific Reports*, 2020, 10: 16674
- [98] Wilson J, van Doorn NL, Collins MJ. Assessing the extent of bone degradation using glutamine deamidation in collagen[J]. *Analytical Chemistry*, 2012, 84(21): 9041-9048
- [99] Ramsøe A, van Heekeren V, Ponce P, et al. DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes[J]. *Journal of Archaeological Science*, 2020, 115: 105080
- [100] Simpson JP, Fascione M, Bergström E, et al. Ionisation bias undermines the use of matrix-assisted laser desorption/ionisation for estimating peptide deamidation: Synthetic peptide studies demonstrate electrospray ionisation gives more reliable response ratios[J]. *Rapid Communications in Mass Spectrometry*, 2019, 33(12): 1049-1057
- [101] Simpson JP, Penkman KEH, Demarchi B, et al. The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone[J]. *Journal of Archaeological Science*, 2016, 69: 29-38
- [102] Schroeter ER, Cleland TP. Glutamine deamidation: An indicator of antiquity, or preservational quality?[J]. *Rapid Communications in Mass Spectrometry*, 2016, 30(2): 251-255
- [103] Pal Chowdhury M, Buckley M. Trends in deamidation across archaeological bones, ceramics and dental calculus[J]. *Methods*, 2022, 200: 67-79
- [104] Robinson NE, Robinson ZW, Robinson BR, et al. Structure-dependent nonenzymatic deamidation of glutaminy and asparaginy pentapeptides[J]. *The Journal of Peptide Research*, 2004, 63(5): 426-436
- [105] Van Doorn NL, Wilson J, Hollund H, et al. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration[J]. *Rapid Communications in Mass Spectrometry*, 2012, 26(19): 2319-2327
- [106] Welker F. Elucidation of cross-species proteomic effects in human and hominin bone proteome identification through a bioinformatics experiment[J]. *BMC Evolutionary Biology*, 2018, 18(1): 23

- [107]Nielsen-Marsh CM, Stegemann C, Hoffmann R, et al. Extraction and sequencing of human and Neanderthal mature enamel proteins using MALDI-TOF/TOF MS[J]. *Journal of Archaeological Science*, 2009, 36(2009): 1758-1763
- [108]Nielsen-Marsh CM, Richards MP, Hauschka PV, et al. Osteocalcin protein sequences of Neanderthals and modern primates[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(12): 4409-4413
- [109]Fiddymment S, Holsinger B, Ruzzier C, et al. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(49): 15066-15071
- [110]McGrath K, Rowsell K, Gates St-Pierre C, et al. Identifying archaeological bone via non-destructive ZooMS and the materiality of symbolic expression: Examples from Iroquoian bone points[J]. *Scientific Reports*, 2019, 9(1): 11027
- [111]Ntasi G, Kirby D, Stanzione I, et al. A versatile and user-friendly approach for the analysis of proteins in ancient and historical objects[J]. *Journal of Proteomics*, 2021, 231: 104039
- [112]Zilberstein G, Zilberstein R, Zilberstein S, et al. Proteomics and metabolomics composition of the ink of a letter in a fragment of a Dead Sea scroll from Cave 11 (P1032-Fr0)[J]. *Journal of Proteomics*, 2021, 249: 104370
- [113]Righetti PG, Zilberstein G, Zilberstein S. New baits for fishing in cultural heritage's Mare Magnum[J]. *Journal of Proteomics*, 2021, 235: 104113
- [114]Wang NH, Brown S, Ditchfield P, et al. Testing the efficacy and comparability of ZooMS protocols on archaeological bone[J]. *Journal of Proteomics*, 2021, 233: 104078
- [115]Lanigan LT, Mackie M, Feine S, et al. Multi-protease analysis of Pleistocene bone proteomes[J]. *Journal of Proteomics*, 2020, 228: 103889
- [116]Gu MX, Buckley M. Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting[J]. *BMC Bioinformatics*, 2018, 19(1): 241
- [117]Hickinbotham S, Fiddymment S, Stinson TL, et al. How to get your goat: Automated identification of species from MALDI-ToF spectra[J]. *Bioinformatics*, 2020, 36(12): 3719-3725
- [118]Tran NH, Zhang X, Xin L, et al. De novo peptide sequencing by deep learning[J]. *Proceedings of the National Academy of Sciences*, 2017, 114(31): 8247-8252
- [119]Tiwary S, Levy R, Gutenbrunner P, et al. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis[J]. *Nature Methods*, 2019, 16(6): 519-525
- [120]Zohora FT, Rahman MZ, Tran NH, et al. DeepIso: A deep learning model for peptide feature detection from LC-MS map[J]. *Scientific Reports*, 2019, 9(1): 17168
- [121]Qiao R, Tran NH, Xin L, et al. Computationally instrument-resolution-independent de novo peptide sequencing for high-resolution devices[J]. *Nature Machine Intelligence*, 2021, 3(5): 420-425
- [122]Zhou WJ, Wei ZH, He SM, et al. pValid 2: A deep learning based validation method for peptide identification in shotgun proteomics with increased discriminating power[J]. *Journal of Proteomics*, 2022, 251: 104414
- [123]Parker GJ, Yip JM, Eerkens JW, et al. Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel[J]. *Journal of Archaeological Science*, 2019, 101: 169-180
- [124]Sawafuji R, Cappellini E, Nagaoka T, et al. Proteomic profiling of archaeological human bone[J]. *Royal Society Open Science*, 2017, 4(6): 161004
- [125]Runge AKW, Hendy J, Richter KK, et al. Palaeoproteomic analyses of dog palaeofaeces reveal a preserved dietary and host digestive proteome[J]. *Proceedings of the Royal Society B*, 2021, 288(1954): 20210020
- [126]高锐, 张晓鹏, 崔永镇, 等. 蛋白质组学研究进展 [J]. *畜牧兽医科技信息*, 2007(12): 13-16
- [127]Yang YM, Shevchenko A, Knaust A, et al. Proteomics evidence for kefir dairy in Early Bronze Age China[J]. *Journal of Archaeological Science*, 2014, 45: 178-186
- [128]Ni XJ, Ji Q, Wu WS, et al. Massive cranium from Harbin in northeastern China establishes a new Middle Pleistocene human lineage[J]. *The Innovation*, 2021, 2(3): 100130
- [129]Gokhman D, Mishol N, de Manuel M, et al. Reconstructing Denisovan anatomy using DNA methylation maps[J]. *Cell*, 2019, 179(1): 180-192.e110