中国科学: 地球科学 2025 年 第 55 卷 第 7 期: 2123 ~ 2140
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
评 述

孢粉类群自动化鉴定的研究进展与展望

杨暘, 李凯*, 许梦佳, 钟海璐, 刘立娜, 张潇, 廖梦娜, 倪健

浙江师范大学生命科学学院, 金华 321004
* 通讯作者, E-mail: likai@zjnu.edu.cn

收稿日期: 2024-07-27; 收修改稿日期: 2025-03-17; 接受日期: 2025-03-19; 网络版发表日期: 2025-06-13
国家自然科学基金项目(42177433)资助

摘要 孢粉是生态学、农业授粉学、环境学、第四纪古植被和古气候演变等研究的重要对象和指标, 而其准确鉴定是开展上述研究的基本前提. 传统的孢粉鉴定主要基于专家知识和经验, 依据孢粉外壁的物理形貌特征进行, 但存在物种分辨率低和鉴定耗时长等问题; 提高孢粉鉴定准确率和鉴定效率的内在需求推动了孢粉鉴定自动化的发展, 其实现途径主要有三种: 基于形态学、基于化学和光谱特征以及基于分子手段的自动鉴定. 目前, 基于形态学的孢粉自动鉴定技术较为主流, 其优势主要体现在孢粉图版丰富、支持鉴定结果的直接专家核验、结果可重复性高等; 基于化学和光谱特征的孢粉自动鉴定提供了更加精确的分类信息, 并可与显微镜鉴定联用, 提高了孢粉自动鉴定的效率; 基于孢粉DNA宏条形码的孢粉自动鉴定具有标准化的实验流程、不依赖分类专家、物种鉴定结果分辨率高等优势. 但这三种孢粉鉴定自动化实现方式均有一定不足, 比如孢粉图像识别技术存在误判和过度拟合等问题, 鉴定结果物种分辨率低, 且仍有待开发涵盖大地理范围、高精度孢粉鉴定的程序; 孢粉光谱鉴定则很难找到具有普适性的孢粉物种谱线, 光谱数据库的通用性低; 孢粉DNA宏条形码存在引物通用性不足、PCR特异性扩增、数据库不完备、丰度估计偏差等问题. 此外, 基于光谱和孢粉DNA的自动鉴定技术多应用于现代孢粉鉴定, 其在第四纪孢粉领域的应用还面临样品前处理和结果验证等问题. 如何联合不同技术手段, 推动实现第四纪孢粉鉴定自动化, 依然需要进一步探索. 由此我们建议: (1) 探索适用于光谱和DNA鉴定技术的第四纪孢粉样品的预处理方法; (2) 构建完备的本地孢粉形态、光谱和DNA条形码数据库, 提高孢粉自动鉴定的准确度; (3) 建立标准化分析流程, 准确评估孢粉自动鉴定效率以及实现不同研究结果对比. 通过梳理孢粉鉴定自动化研究的起源和发展及其优劣势, 以期能够推动孢粉学和相关学科的进一步发展.

关键词 孢粉鉴定, 自动化, 形态学, 光谱技术, 孢粉DNA宏条形码, 第四纪孢粉学

1 引言

孢粉是孢子植物的孢子和种子植物花粉的合称, 可以作为微体化石证据在地质记录中长期保存, 以孢粉为研究对象的第四纪孢粉学的发展, 对理解现代植被分布格局的形成过程、植被和环境长期变化、植物多样性以及人地关系演变等至关重要(Birks, 2019). 同时, 孢粉监测和研究也是维护环境和人类社会健康发展的重要方面, 对联合国可持续发展目标的实现具有重要贡献(Hornick等, 2022).

中文引用格式: 杨暘, 李凯, 许梦佳, 钟海璐, 刘立娜, 张潇, 廖梦娜, 倪健. 2025. 孢粉类群自动化鉴定的研究进展与展望. 中国科学: 地球科学, 55(7): 2123–2140, doi: 10.1360/SSTe-2024-0213
英文引用格式: Yang Y, Li K, Xu M, Zhong H, Liu L, Zhang X, Liao M, Ni J. 2025. The advance and prospect for automatic pollen classification (in Chinese). Scientia Sinica Terrae, 55(7): 2123–2140, doi: 10.1360/SSTe-2024-0213

© 2025 《中国科学》杂志社

准确的孢粉鉴定是开展孢粉学相关研究的基础和前提. 孢粉的直径通常以微米为单位, 很难通过肉眼直接观察, 直到显微镜发明后, 科学家才得以对孢粉的形貌特征进行观察描述, 孢粉分类学鉴定由此开始快速发展(Faegri等, 1989; Weber, 1998). 但是, 基于形态学的分类学鉴定严重依赖于专家的知识和经验, 存在鉴定耗时长、效率低、分类学精度低等问题(Stillman和Flenley, 1996). 许多孢粉类群的鉴定只能获得属或科一级的信息, 尤其是同科属或分类地位相近的孢粉颗粒在形态学上非常相近, 如禾本科、莎草科、大戟科等孢粉, 这不仅导致其在分类学鉴定中难以区分, 更限制了以孢粉为主要手段的学科发展和深入(许清海等, 2024). 例如, 现代生态学研究是以物种为单位, 而古生态学则以孢粉和化石作为主要研究载体, 孢粉形态学鉴定的低分辨率导致现代生态学和古生态学难以融合, 也是阻碍现代生态学理论和方法在古生态学中应用的主要因素. 因此, 提高孢粉鉴定效率和鉴定的物种分辨率, 是第四纪孢粉学等学科发展长期面临的挑战. 孢粉鉴定自动化的初衷是推动第四纪孢粉学等学科的发展(Stillman和Flenley, 1996), 其内在要求是在不依赖分类专家的前提下, 针对大量的混合孢粉样品实现标准化的、高分辨率的、可重复的、高可信度的快速分类学鉴定(Holt和Bennett, 2014). 在过去半个世纪, 基于形态学的孢粉自动鉴定取得了一定进步, 而随着光谱技术和DNA条形码(DNA metabarcoding)的发展, 基于孢粉化学组成和DNA信息的自动鉴定技术也得到了越来越多的重视和应用, 孢粉鉴定自动化如今正在步入快速发展时期. 目前, 孢粉鉴定自动化的实现途径主要包括3种, 分别是基于形态学的自动鉴定技术、基于化学和光谱特征的自动鉴定技术和基于分子手段的自动鉴定技术. 这些方法得到了不同程度的关注和应用, 也在不同的研究领域内展现出了各自的优势和不足. 本文尝试对此3种孢粉自动鉴定方法进行介绍, 以期推动孢粉鉴定自动化的进步, 助力孢粉学等相关学科的发展.

2 孢粉人工鉴定

孢粉学者基于形态学特征进行孢粉类群鉴定, 依然是目前最为主要的孢粉鉴定方式. Molpighi和Grev分别在1682年和1687年使用光学显微镜对孢粉颗粒进行了观察, 发现同种植物花粉的形态结构相似, 不同物种植物花粉之间存在着较大差异(王伏雄和喻诚鸿, 1954). Fischer(1890)深入研究了花粉壁的形态结构和化学性质, 将花粉从形态上细分为外壁和内壁. Mohl (1934)在德语专著*Über den Bau und die Formen der Pollenkörner*(孢粉粒的结构及其形状)中详细研究了植物花粉萌发孔、萌发沟的构造及排列方式, 并首次将孢粉按孢粉壁层次和孔沟分类, 开创了孢粉的简易分类系统, 由此孢粉形态学研究开始走上正轨. Erdtman设计了醋酸酐分解法以更好地观察花粉外壁的细微结构(Erdtman, 1960), 至今仍是使用率极高的孢粉处理方法, 也是光学显微镜下孢粉形态学观察的优先选择. 电子显微镜出现后, 孢粉学研究迈入了新的纪元, 科学家们逐渐揭示了常见植物的孢粉形态. Erdtman(1969)出版了*Handbook of Palynology*, 不仅详细介绍了各类孢粉的形态和分类依据, 还设计了沿用至今的制样方法和分类系统, 至今仍是孢粉学的权威参考书之一. 不同植物的孢粉在大小、形状和表面结构等形态表现上的差异, 为通过孢粉形态学进行植物类群鉴定提供了理论基础(Weber, 1998). 在具体的孢粉鉴定过程中, 孢粉学家主要基于孢粉的颗粒形态, 是否复合, 极性, 对称性, 萌发孔沟数量、分布和形态, 外壁纹饰形态和分布等对孢粉的科属进行分析和判别(王伏雄和喻诚鸿, 1954; Traverse, 2007).

与此同时, 世界各地大量的现代孢粉形态学研究为孢粉人工鉴定提供了有力支持. Erdtman最早开始进行系统化孢粉形态学研究, 其著作*Pollen Morphology and Plant Taxonomy—Angiosperms: An Introduction to Palynology*描述了327科2,400属约10,000种植物的孢粉形貌特征, 并含有261组细致的孢粉图版, 包含250科约600属植物孢粉的赤道观、极面观、表面细节等特征(Erdtman, 1952). 随后, 世界各地的孢粉形态学研究逐渐开展, 如欧洲(Punt等, 1976–2009)、非洲(Assemien等, 1974)、美洲(Colinvaux等, 1999)和亚洲(王伏雄等, 1995; 唐领余等, 2016)等, 这在许清海等(2024)的文章中有具体阐述. 中国孢粉学者也出版了各种孢粉图谱作为孢粉鉴定的主要参考, 主要有《中国蕨类植物孢子形态》(中国科学院北京植物研究所古植物研究室孢粉组, 1976)、《中国热带亚热带被子植物花粉形态》(中国科学院植物研究所古植物室孢粉组和中国科学院华南植物研究所形态研究室, 1982)、《中国

植物花粉形态》(第2版)(王伏雄等, 1995)、《中国常见栽培植物花粉形态: 地层中寻找人类痕迹之借鉴》(许清海, 2015)、《中国第四纪孢粉图鉴》(唐领余等, 2016)等. 最近, 为更加细致和准确地表征孢粉形貌特征, 李天庆等(2010)采集并发表了《中国木本植物花粉电镜扫描图志》, 许清海等(2022)发表了《中国北方山地常见植物花粉形态研究: 光学显微镜下精确鉴定方法探索》, 为孢粉的准确鉴定提供了更好支撑.

基于分类学专家的孢粉鉴定是第四纪孢粉学等研究的主要手段, 其优势主要是鉴定标准统一、鉴定结果可靠性高、通用性强等. 首先, 孢粉学家在独立进行孢粉鉴定之前都会经历专业的训练, 准确掌握孢粉分类系统和不同植物孢粉的形态特征, 这保证了孢粉鉴定结果的准确性. 其次, 广泛开展的孢粉形态学研究为孢粉学家提供了统一的参考标准, 保证了不同学者鉴定结果的一致性. 除此之外, 孢粉学家在鉴定过程中不需要昂贵的设备支持, 孢粉鉴定的经济成本低, 并且孢粉鉴定制片能够长期保存, 支持鉴定结果的验证. 最后, 由于鉴定参照标准一致, 不同学者的孢粉鉴定结果往往具有可比性, 这为全球和区域孢粉数据库的构建提供了支持, 例如全球晚第四纪孢粉数据集LegacyPollen 1.0(Herzschuh等, 2022)和中国地层和表土孢粉数据集(陈海燕等, 2021; 周伯睿等, 2023). 在过去百余年间, 孢粉学者通过光学显微镜鉴定的方式对全球表土、空气、冰芯、黄土、海洋和陆地沉积物等介质中保存的孢粉颗粒进行了大量的鉴定工作, 有力地推动了以孢粉为研究载体的第四纪植被演变、古生态学、考古学、系统地理学、植被模型、生态系统科学与服务、保护生物学、恢复生态学以及群落和种群动力学等学科的发展(Birks和Berglund, 2018; Birks, 2019).

但是, 基于专家进行孢粉形态学鉴定也存在一定的不足(Flenley, 1968; Holt等, 2011; 许清海等, 2024). 首先, 孢粉外貌特征较为复杂, 且孢粉鉴定必须掌握分类学知识, 因此需要进行专业的学习训练, 这一方面保证了鉴定结果的准确性, 但同时也为孢粉鉴定设置了较高的专业门槛. 其次, 部分鉴定结果存在主观性, 由于孢粉形态特征的复杂多变性, 仅靠孢粉学家的经验和记忆难以掌握全部的孢粉形态, 同时近缘植物的孢粉间存在较大的相似度, 导致依据形态特征进行鉴定过程中不可避免地出现一些主观判断. 再次, 专家鉴定过程耗时长, 并且在统计时容易出现数据错误等问题. 最后, 同属孢粉形态相似程度高, 难以区分, 尤其是对许多孢粉类群而言(如莎草科、禾本科), 高度的相似性导致了物种分辨率难以提高, 只能进行属或科一级的鉴定, 这就限制了从孢粉中获得的生物地理信息, 只能达到相对较粗的分类水平. 随着显微技术的发展, 扫描电子显微镜(Scanning Electron Microscope, SEM)的出现使得孢粉表面纹饰的观察更加精细, 显著提升了形态学鉴定的物种分辨率(李天庆等, 2010). 然而, SEM制样复杂、设备成本高昂, 且需逐粒观察, 难以适应大批量混合孢粉样本的快速鉴定需求.

3　孢粉自动鉴定

对快速和准确鉴定大批量孢粉的需求推动了孢粉鉴定自动化的出现和发展(Flenley, 1968; Stillman和Flenley, 1996). Mirkin和Bagdasaryan(1972)最早尝试根据全息图像总结不同孢粉大小和形状数据, 并据此进行孢粉鉴定. 但该研究存在两个主要问题: 一是孢粉壁的延展性和可变性远超研究中所用孢粉颗粒所能够涵盖的范围, 使其难以采用多变量统计分析孢粉大小和形状的方法来鉴定不同的孢粉(Witte, 1988); 二是在高放大倍率下焦点深度远小于孢粉粒直径, 难以分析孢粉纹理形态(Langford等, 1986). Haralick(1979)提出的表面纹理分析方法克服了孢粉的大小和形状可变性的问题, 并且在没有完美聚焦的光学图像上也能实现孢粉鉴定(Treloar, 1992), 自此通过机器学习进行自动孢粉鉴定开始了快速的发展, 并取得了较多成果, 这在早期的综述文献中有展示. 例如Stillman和Flenley(1996)从第四纪孢粉学角度, 强调了孢粉鉴定自动化的必要性, 并指出了孢粉自动鉴定存在的问题和未来前景. Holt和Bennett(2014)综述了孢粉鉴定自动化, 提出了孢粉自动鉴定的一般性原则和方法, 推动了该领域的进一步发展. 近年来, 利用光谱测定孢粉所含化学物质组成和含量的分类方法也多有报道(Zimmerman等, 2016; Kenđel和Zimmermann, 2020; Pereira等, 2021; Stiebing等, 2022), 基于孢粉图像或者化学特征的许多新的分析方法也与机器学习相结合, 大大提高了孢粉鉴定的速度和效率(Dunker等, 2021). 此外, 受益于分子生物学的迅速发展, 通过DNA宏条形码或浅层基因组测序等方法进行孢粉鉴定工作也在快速发展

(Keller等, 2015; Lang等, 2019; Peel等, 2019; Bell等, 2021, 2023).

3.1 基于形态学的孢粉自动鉴定

最初, 孢粉鉴定自动化是指采用计算机技术基于形态特征对孢粉颗粒进行自动分类学鉴定, 因此基于图像识别技术的孢粉自动鉴定是发展时间最久也是最为成熟的孢粉鉴定自动化方式. 为提高花粉鉴定效率, 减少鉴定过程中的主观性, 研究人员首先开发了基于模式特征分析的孢粉图像分析技术(Mirkin和Bagdasaryan, 1972; Langford等, 1990). 该技术基于人工选择孢粉模式特征设计孢粉识别系统, 并采用机器学习方法提取和分类识别孢粉特征(Stillman和Flenley, 1996), 通常分为如下几个步骤(图1): 模式孢粉数据采集、孢粉特征提取、特征选取和分类模型训练和自动鉴定的实现(Mander等, 2013; Holt和Bennett, 2014; Kubera等, 2021).

(1) 孢粉图像采集. 孢粉图像数据采集通常是将收集到的一定数量的孢粉颗粒进行醋解法或不做预处理直接制备孢粉玻片进行光学图像采集, 以及采用非制片方式进行孢粉样品图像采集, 其中需要选择性地对孢粉进行化学染色, 以方便后续的图像形态学分析. 为了便于成像, 孢粉颗粒必须在同一焦平面内可见, 并且不能被碎屑物质遮挡, 孢粉颗粒不应重叠或聚集, 同一样品至少需检测到50个花粉颗粒, 样品应染色充分, 以突出孢粉特征(von Allmen等, 2024). 图像采集应该尽量使用自动彩色图像获取设备, 如Leica Aperio CS2 slide scanner(Olsson等, 2021)、Zeiss AxioScan Z1 slide scanner(von Allmen等, 2024)、多光谱成像流式细胞仪(Dunker等, 2021)等. 另外, 在图像采集环节还涉及多图像采集融合过程(Matsuo等, 2022; Theuerkauf等, 2023), 以获取更加丰富的图像信息. 目前, 图像采集环节存在的主要问题是图像采集设备和方法不统一, 如采集过程中的放大倍率和图像分辨率的差异、样品是否染色、图像存储和预处理方式等. 虽然部分学者采用更高精度的显微镜进行孢粉图像采集, 如SEM、激光共聚焦显微镜等(Langford等, 1990; Treloar等, 2004), 这虽然可以获得更加准确的孢粉形态特征, 但也存在样品预处理难度高、图像采集效率低、需要手动操作等问题, 并不符合孢粉鉴定自动化的要求. 此外, 部分研究也采用孢粉图像数据库已有的图版进行方法学的验证(Gonçalves等, 2016; Astolfi等, 2020).

(2) 孢粉模式特征提取. 图像提取是形态学孢粉自动鉴定的核心, 主要包含2个关键环节: 一是准确快速地分割定位孢粉颗粒并提取孢粉颗粒图像; 二是对孢粉颗粒图像进行模式特征选择提取. 孢粉分割定位是指在一个显微镜视野或者图像视野内准确识别并定位孢粉颗粒, 排除非孢粉颗粒物质的干扰. 传统的孢粉颗粒识别包括使用模糊处理、形态和圆块检测算法、焦

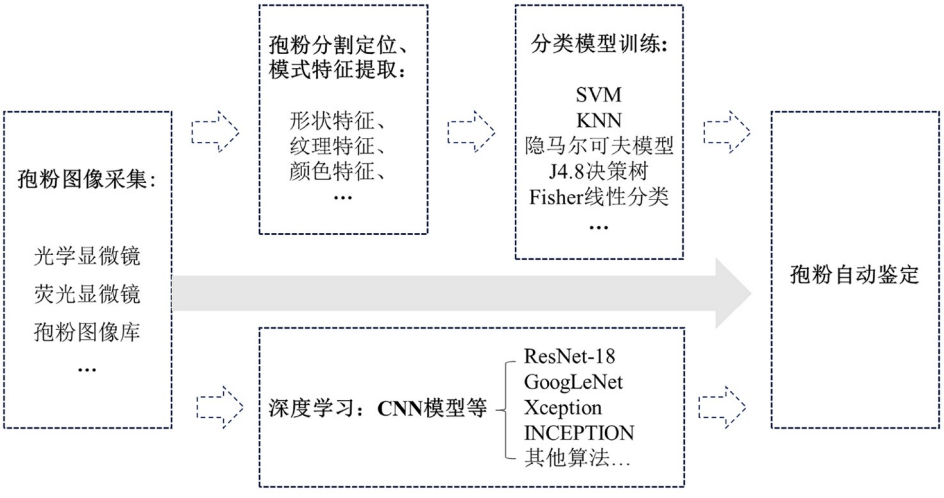

图 1　基于形态学的孢粉自动鉴定流程示意

SVM, 支持向量机; KNN, k-近邻算法; CNN, 卷积神经网络; ResNet-18、GoogLeNet、Xception、INCEPTION为不同的CNN深度学习架构模型

深过滤等步骤实现孢粉图像的定位和提取(Lozano-Vega等, 2014; Chudyk等, 2015; Yin等, 2021). 不同孢粉分割定位方法的识别准确率存在差异, 其原因可能在于图像中存在碎屑物质、孢粉颗粒聚集等(Theuerkauf等, 2023). 孢粉模式特征提取主要包括形状、纹理特征(Rodriguez-Damian等, 2006; Chudyk等, 2015)以及颜色特征(Gonçalves等, 2016; Dunker等, 2021), 常用的孢粉自动鉴定模式特征有周长、面积、圆度、极半径、赤道半径、轴差、萌发孔数量、孔饰特征、表面复杂度、颜色等(Li等, 2004; Treloar等, 2004; Chudyk等, 2015; Gonçalves等, 2016; Kubik-Komar等, 2018; Yin等, 2021). 孢粉模式特征提取方法较为多样, 如Gabor滤波、快速傅里叶变换、局部二元模式、定向梯度直方图、灰度共生矩阵、k曲率算法、视觉词袋模型等(Langford等, 1990; Chudyk等, 2015; Gonçalves等, 2016; Yin等, 2021; Chen等, 2022). 使用单一特征对后续分类器的效果存在一定的限制, 同时其可拓展性差, 抗干扰能力弱, 因此, 已有研究均使用多特征融合图像分类识别方法来提高后续分类的准确率, 如通过颜色、形状、纹理联合视觉词袋模型进行孢粉模式特征提取(Gonçalves等, 2016).

(3) 传统分类模型. 通过统计分析孢粉模式特征, 将已获取的孢粉颗粒图像进行分类, 确定其所属类型或种类. 首先复选上一阶段获取的孢粉特征, 以获得特异性较强的孢粉特征, 将复选后的孢粉特征输入分类器进行判断和分类(Kubik-Komar等, 2018). 基于模式特征的孢粉图像分析一般采用传统的机器学习分类方法, 通常包括支持向量机(Support Vector Machine, SVM)、k-近邻算法(K-Nearest Neighbors, KNN)、隐马尔可夫模型、主成分分析和线性判别分析(Principal Component Analysis-Linear Discriminant Analysis, PCA-LDA)、分层机器学习、J4.8决策树和人工神经网络模型(Artificial Neural Networks, ANN)等(Holt和Bennett, 2014; Punyasena等, 2022), 可在纹理特征突出的孢粉科属鉴定层级上达到较高的识别准确率. 例如, Kubik-Komar等(2018)采用J4.8决策树基于萌发孔数、最大轴、最小轴、轴差等孢粉模式特征进行了分类, 得到的鉴定准确率达94%. Gonçalves等(2016)通过使用顺序最小优化、C-SVM、J4.8决策树和KNN对孢粉模式特征进行分类, 交叉验证后达成了大于64%的鉴定准确率, 是多分类器中可以识别孢粉种类最多、标准化程度最高的自动识别系统之一. Ronneberger等(2002)采用支持向量机和径向基函数进行孢粉自动鉴定, 鉴定准确率在科属层级为92%; Silva等(2014)结合KNN和SVM对7种花粉图像进行识别的准确率为79%; Rodriguez-Damian等(2006)采用一种集合特征选择、多元感知机和径向基函数的方法, 能够89%地准确鉴定人工采集的荨麻科(Urticaceae)花粉样品, 不亚于专家鉴定结果; Marcos等(2015)基于15个花粉类别的纹理特征采集了1800多个图像样本, 使用Fisher线性分类器和KNN的孢粉图像识别模型, 鉴定准确率达到了95%. 虽然基于模式分析和简单机器学习的图像识别系统具有灵活性强、可用方法丰富等优点, 但其也存在图像清晰度要求高、特征提取步骤繁杂等缺点, 容易产生难以控制的精度误差, 影响最终的孢粉鉴定准确率(Marcos等, 2015).

(4) 深度学习模型. 随着深度学习理论在图像处理和分析领域的快速发展, 卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习模型在孢粉自动鉴定中的应用, 实现了比传统机器学习模型更高的孢粉鉴定准确率以及更高的鉴定效率(Kubera等, 2021). CNN模型通过多层卷积核自动提取孢粉图像的局部特征(如萌发孔形态、外壁纹饰等), 无需人工设计特征, 尤其擅长处理纹理复杂、对称性多变的孢粉图像, 已被许多学者运用到孢粉自动鉴定(Punyasena等, 2022), 开发了诸如DCNNs(Sevillano等, 2018; Olsson等, 2021)和Faster R-CNN(Theuerkauf等, 2023)等孢粉自动鉴定流程. CNN模型可根据孢粉形态特征进行孢粉图像自动识别处理(Punyasena等, 2022), 省去了复杂的特征选择和提取过程, 同时CNN模型对花粉图像清晰度要求较低, 相比传统的模式分析, 在孢粉鉴定方面更具优势(Sevillano等, 2018; Olsson等, 2021; Dunker等, 2021), 已经应用于第四纪孢粉的自动化鉴定(Bourel等, 2020; von Allmen等, 2024; Durand等, 2024). Sevillano等(2020)对经过乙酰预处理的46种形态存在较大差异的不同类别孢粉粒的图像进行分类, 通过将CNN与递归神经网络相结合来识别完整孢粉的z轴堆叠暗场图像, 最终鉴定准确率达到98%. Olsson等(2021)采集了来自17科83个物种的花粉并用品红染色, 输入ResNet-18、GoogLeNet和Xception模型进行学习, 使用数据分割法和留一法进行交叉验证实现模型训练, 并对采集的所有花粉样本都进行了物种和科

属鉴定, 但对物种层面的鉴定效果较差, 在科属层面的准确率(85%)和召回率却都达到了可接受的精度. 也有部分研究将明场图像与其他类型的光谱图像结合鉴定孢粉, 例如Dunker等(2021)通过自动多光谱成像流式细胞仪收集了17个科属的35个物种的现代花粉图像和每个花粉对应的荧光图像, 使用迁移学习(Transfer Learning)将图像输入在一个大规模数据集上, 并预训练具有12通道的CNN模型进行鉴定, 最佳通道的平均鉴定准确率可达96%. 但是, 当使用类似的多通道CNN模型对第四纪孢粉进行自动鉴定时, 其效果比现代孢粉的结果差(Durand等, 2024). 针对地层和化石孢粉, Bourel等(2020)建立了一个集成在决策树系统中的CNN模型, 为东非现代和化石花粉组合中常见的6种苋科花粉、1种禾本科花粉和1种莎草科花粉进行分类, 该系统的优越性在于能对部分受损和保存不佳的地层孢粉进行分类, 并可以灵活选择孢粉的分类级别.

(5) 优势和挑战. 基于图像识别技术的孢粉自动鉴定取得了长足的进步, 其优势主要体现在孢粉图版丰富、鉴定结果支持专家核验、结果可重复性高等, 在未来一定时间内依然会是孢粉鉴定自动化的主流趋势. 尤其是, 结合深度学习和成像流式细胞技术的孢粉自动鉴定方法, 能够在面对复杂的孢粉组合时提供更高的准确性和效率(Dunker等, 2021). 但是, 基于图像识别技术的孢粉自动鉴定同样存在一些不足之处和应用上的问题. 首先, 大部分的程序都只是对少量物种的孢粉进行鉴别和分类, 对大量的混合孢粉鉴定研究依然不足, 现有的孢粉图像数据库难以支持模型进行大规模的孢粉鉴定(Hornick等, 2022); 同时, 提高模型鉴定准确率需要更多特征参考图像, 但当参考图像数量过多时, 模型会出现过度拟合问题导致鉴定效果降低, 如何控制参考图像数量和鉴定效果之间的关系仍需要更深入的探索(Olsson等, 2021). 其次, 尽管CNN模型在分类识别任务中性能优越, 但其在训练阶段需要大量的含标志特征数据作为训练的样本, 罕见物种的孢粉图像数量过少会导致模型训练效果较差, 鉴定时容易出现误判(Dunker等, 2021). 第三, 孢粉自动鉴定在第四纪孢粉中的应用不足, 对大范围高精度的地层孢粉和化石孢粉的鉴定程序仍有待开发(Bourel等, 2020), 实现部分地层中变形和开裂孢粉的准确鉴定依然是一大挑战. 第四, 当前采用的模型精度判断方法过于理想化, 实际应用中模型的鉴定效果可能会出现较预测结果大幅降低的情况(Olsson等, 2021; Durand等, 2024). 最后, 基于图像识别技术的孢粉自动鉴定依然无法有效提高鉴定结果的物种分辨率.

3.2 基于光谱分析的孢粉自动鉴定

不同物种的孢粉具有不同的生物化学成分, 通过探究其光谱特征可以实现孢粉的分类学鉴定(Schulte等, 2008; Zimmermann, 2010, 2018; Muthreich等, 2020). 拉曼光谱(Raman Spectroscopy)和傅里叶变换红外光谱(Fourier Transform Infrared Spectroscopy, FTIR)是目前孢粉鉴定常用的两种光谱类型(Guedes等, 2014; Zimmermann等, 2015; Zimmermann, 2018; Kenđel和Zimmermann, 2020; Muthreich等, 2020), 此外还有荧光(朱江峰, 2017; Dunker等, 2021)、核磁共振光谱(Klimczak等, 2020)等也被用于孢粉自动鉴定. 在近期研究中, 研究人员将光谱数据集缩小到特定波段区域(Pereira等, 2021), 利用高通量红外光谱系统(Zimmermann等, 2017)以及自动化高通量拉曼光谱系统(Mondol等, 2019)构建了基于多种科属孢粉的综合光谱数据集, 有效推动了孢粉自动鉴定的发展. 基于光谱技术的孢粉自动鉴定一般流程如图2所示, 主要包括样品制备、光谱采集和孢粉自动鉴定.

(1) 孢粉样品制备. 不同研究在样品制备阶段采用的策略稍有差异, 但几乎所有的以光谱为手段的孢粉鉴定研究均使用新鲜花粉, 并且采用干燥后的花粉进行制片(Depciuch等, 2016; Diehn等, 2020; Kenđel和Zimmermann, 2020; Muthreich等, 2020), 以避免化学处理对孢粉内含物质成分的改变, 进而提高孢粉分类鉴定的准确性. 孢粉样品采集和保存流程主要包括采集花药、25℃干燥24h、过筛富集花粉、低温冻存等, 分析前取出样品并使其均匀散落在玻片上干燥(Pereira等, 2021; Stiebing等, 2022). 因孢粉内含物质保存完好, 光谱类型的选择也较为开放, 这种样品制备方式可能仅适用于现代孢粉(Zimmermann, 2010). 但是地层孢粉保存的完整性受多种因素的共同影响, 最终被提取的孢粉基本只保留了外壁, 这也对光谱类型的选择造成了一定限制. 一般而言, 拉曼光谱通常只能检测孢粉表层主要由孢粉素和纤维素组成的外壳(Pereira等, 2021), 使其有望应用于第四纪化石孢粉的自动鉴定(Zimmermann, 2010; Stiebing等, 2022), 但是其在物种层面分辨率较低(Stiebing等, 2022).

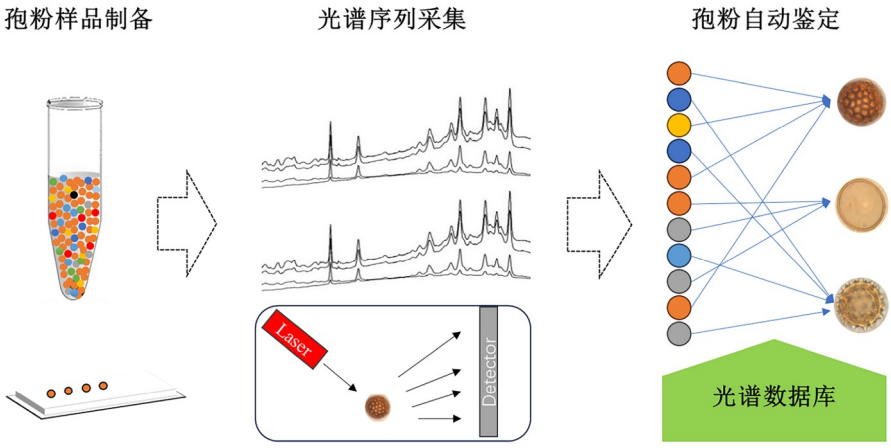

图 2　基于光谱技术的孢粉鉴定自动化流程示意

(2) 光谱序列采集. 孢粉的光谱采集和处理环节涉及光谱类型和激光器的选择、光谱范围的确定、光谱条件测试、孢粉光谱预处理和校验等. 在此环节需要首先确定光谱波段和采样频率, 连续多次重复采集光谱, 然后进行谱线预处理, 包括谱线平滑、信号校正、求一次(二次)导数等以去除光谱噪声等, 最后建立光谱峰值信号和样品信息之间的关系, 获取孢粉样品的物质组成、分子结构和状态等(Zimmermann, 2010; Guedes等, 2014; Kenđel和Zimmermann, 2020). 在已有的基于光谱进行孢粉分类学鉴定的工作中, 不同学者采用的光谱类型、激光强度和采集频率均有所不同(如Seifert等, 2016; Muthreich等, 2020; Diehn等, 2020). 拉曼光谱特征峰信号强, 峰位重叠较少见, 适合应用于定量分析和建立大范围的光谱数据库(Guedes等, 2014; Seifert等, 2016), 因此在孢粉自动鉴定中也较为常用, 如应用于孢粉自动鉴定的显微共聚焦拉曼光谱(Schulte等, 2008; Guedes等, 2014)、表面增强拉曼光谱(Sengupta等, 2005; Seifert等, 2016)、共振拉曼光谱(Schulte等, 2008)、傅里叶变换拉曼光谱(Zimmermann, 2010; Kenđel和Zimmermann, 2020)等. 作为一种吸收光谱, FTIR的信号比拉曼光谱更为强烈, FTIR与其他仪器或技术联用衍生出的红外检测方法包括透射法(Diehn等, 2020; Kenđel和Zimmermann, 2020)、衰减全反射法(Kasprzyk等, 2018; Zeghoud等, 2021; Swiatly-Blaszkiewicz等, 2021)、显微红外光谱法(Depciuch等, 2016)等也已经在孢粉自动鉴定中被使用. 但是不同学者所用的仪器和参数并不一致, 导致不同学者的结果间难以直接对比, 例如, Pereira等(2021)使用法国Horiba Scientific公司生产的XploRATM拉曼显微镜, 在波长为785nm的激发波长下采集, 光谱区域为1000~1800cm^{-1}, 分辨率约为1cm^{-1}; Stiebing等(2022)使用德国WITec产的alpha 300R激光共聚焦显微镜, 在波长785nm的激发波长下采集, 光谱区域为100~3300cm^{-1}, 分辨率约为6cm^{-1}; Diehn等(2020)使用美国Thermo产的Nicolet FTIR显微镜进行谱线采集, 光谱区域为800~1300nm和1300~1800nm, 光谱分辨率为4cm^{-1}.

(3) 孢粉光谱分类鉴定. 基于光谱技术的孢粉自动鉴定暂未有统一方法, 诸如主成分分析、偏最小二乘法、层次聚类分析、排序分析、SVM、ANN模型等均有报道(Zimmermann等, 2015; Seifert等, 2016; Diehn等, 2020; Pereira等, 2021; Swiatly-Blaszkiewicz等, 2021), 分类结果的验证多采用留一交叉验证(Leave One Out Cross Validation, LOOCV)等(Jardine等, 2019; Diehn等, 2020). 例如, Mondol等(2019)测量了来自20个科的37种孢粉的拉曼光谱, 使用层次聚类分析提供基于生活型的孢粉类型的初步估计, 并使用主成分分析-支持向量机对37种不同孢粉分类, 准确率超过90%. Seifert等(2016)将表面增强拉曼光谱技术与CNN结合, 提取14种市售鲜花花粉和在样地采集的10属71种花粉的水溶性成分进行拉曼光谱分析, 并根据成分对孢粉进行了分类. Zimmermann等(2015)和Diehn等(2020)均使用单粒孢粉的FTIR图像进行研究, 分别结合稀疏偏最小二乘回归模型和完全交叉验证中的判别分析模

型，获得了花粉粒的详细化学组成，验证了FTIR技术在花粉自动鉴定中的潜力. Muthreich等(2020)采集了葡萄牙境内栎属(*Quercus*)的5个种和1个亚种的花粉样本，采用偏最小二乘回归法对孢粉FTIR数据进行分类鉴定，研究结果表明FTIR在孢粉类群属级层面的鉴定准确率超过97%, 但是在种层级面临较大的挑战.

光谱技术进行孢粉自动鉴定的优势主要在于其提供了更加精确的分类信息，可以与显微镜联用提高孢粉鉴定自动化效率. 首先，光谱技术可以揭示不同种类孢粉的化学成分信息及其差异，可以在孢粉形态学鉴定的基础上进一步细化孢粉的分类. 如禾本科在形态学分类上很难达到种级别，而采用FTIR技术分别对12种和8种禾本科花粉进行化学分类学研究，采用KNN分类和LOOCV验证后得到的鉴定准确率均超过了80%(Julier等, 2016; Jardine等, 2019). 其次, 光谱技术的应用具有高度灵活性, 可以与显微技术结合, 组成显微共聚焦拉曼光谱、FTIR红外显微镜等，对孢粉样品进行聚焦检测, 减少焦点周围物质干扰, 同时支持与其他仪器或技术联用以提高孢粉自动鉴定的准确性和效率(Schulte等, 2008; Guedes等, 2014; Depciuch等, 2016; Mondol等, 2019; Diehn等, 2020). Mondol等(2019)采用显微共聚焦拉曼光谱技术结合自动图像定位分析技术建立了一种自动化高通量拉曼光谱平台, 实现了基于形态学和光谱技术的结合，为光谱技术在孢粉自动鉴定领域的发展迈出了重要一步. 第三，基于光谱的孢粉自动鉴定并不依赖于孢粉颗粒的完整性, 即使对于保存条件较差的部分变形、破碎孢粉, 理论上也可以采用光谱技术进行鉴定. 常用的Nicolet Nexus FTIR显微镜的光斑仅有15μm×15μm, 可以对大多数孢粉的碎片进行光谱采集，为孢粉鉴定提供额外的化学成分信息(Diehn等, 2020). 此外, 基于光谱的孢粉自动鉴定制样简单, 自动化程度高, 而且随着相关研究逐渐增多, 孢粉光谱数据也日益增加, 为后续基于光谱技术开展孢粉自动鉴定积累了数据，也为基于深度学习模型等算法进行快速孢粉鉴定提供了基础.

与此同时，基于光谱技术的孢粉自动鉴定依然面临诸多问题. 首先，拉曼或FTIR研究都缺乏成像方法, 光谱序列仅限于平均光谱或仅代表孢粉粒一小块区域的光谱信号. 焦点位置的不同、孢粉对称性的差异或孢粉上的球囊等特殊的结构均会引起光谱谱线的变化(Bağcıoğlu等, 2015), 单个物种内不同来源孢粉的化学组成也存在一定的变异性，这导致很难找到一条具有普适性的孢粉物种谱线(Zimmermann等, 2015). 虽然在实际应用中可以通过多次重复光谱采集以及光谱预处理的方式来拟合平均光谱曲线(如Kenđel和Zimmermann, 2020), 但这会无疑会损失部分信息以及降低孢粉鉴定的物种分辨率(Stiebing等, 2022). 其次，不同研究实验设置、仪器设备和光谱数据处理方法并不统一，造成研究结果间无法直接对比，例如，Mondol等(2019)和Kenđel和Zimmermann(2020)分别使用不同型号的显微镜和激光波长检测相同科属的花粉，最终结果中拉曼光谱的采样范围、分辨率和化学基团相关信号峰值并不一致. 这也使得已有的光谱数据库的通用性降低，可能是未来基于光谱技术开展大范围、快速孢粉自动鉴定的主要限制因素之一. 第三，孢粉所处环境的温度、湿度及其他环境因素都有可能对光谱结果造成一定干扰，这需要建立标准化的检测流程以保证光谱序列的准确性. 由于孢粉本身具有的强荧光特性(Zimmermann等, 2015), 采用可见光激发时，花粉吸收强光被加热，可能会出现光降解现象影响花粉的化学组成(Schulte等, 2008), 进而影响鉴定效果. 此外, 光谱技术目前多用于已知物种的少量孢粉甚至是单个花粉颗粒进行分类以达到物种鉴定的目的，属验证性研究(Diehn等, 2020; Pereira等, 2021; Stiebing等, 2022), 并且当前孢粉光谱库的数量和库内物种数量都较少(Guedes等, 2014; Zimmermann等, 2017; Mondol等, 2019), 在实时检测和模型算法上仍有很大的进步空间(Pereira等, 2021).

3.3 基于DNA组成的孢粉自动鉴定

孢粉遗传信息主要存在于细胞核、质体和线粒体中(Tanaka, 1997; Nagata等, 1999), 基于孢粉颗粒进行的研究均证实现代和部分化石孢粉可以提取足量的DNA用于分类学等相关研究(Petersen等, 1996; Suyama等, 1996; Parducci等, 2005; Bennett和Parducci, 2006). 孢粉DNA研究起步较早，但由于相关实验流程繁琐、污染风险大和成本高等问题(Parducci等, 2017), 采用分子手段进行孢粉分类学研究在21世纪初期发展缓慢(Bell等, 2016). 随着高通量测序技术的发展和DNA宏条形码技术的成熟，使用孢粉DNA开展分类学鉴定才重新焕发生机并展现出强劲的发展势头(Cornman等, 2015; Creer等, 2016; Bell等, 2023). 从古生态学

角度看, 孢粉DNA保存效果可能要优于植物组织DNA (Petersen等, 1996), 使其成为研究植物物种的迁移散布、物种进化和种群形成等关键科学问题较为理想的选择之一(Bennett和Parducci, 2006; Parducci等, 2015), 同时也为孢粉DNA自动鉴定提供了可能. DNA宏条形码技术是最为常用的孢粉DNA鉴定手段(Cornman等, 2015; Bell等, 2016; Smart等, 2017; Macgregor等, 2019; Leontidou等, 2021; Polling等, 2022; Bell等, 2023), 主要流程包括样品采集(纯化)、DNA提取、聚合酶链式反应(Polymerase Chain Reaction, PCR)扩增、高通量测序、生物信息学分析、数据库匹配和物种鉴定标记(图3).

(1) 样品制备. 孢粉样品的获取依研究目的的不同存在较大差异. 在现代生态学研究中多采用新鲜花粉、蜂蜜、捕获器孢粉等开展孢粉DNA研究(Hawkins等, 2015; Kraaijeveld等, 2015; Richardson等, 2015; de Vere等, 2017; Peel等, 2019; Lowe等, 2022), 而古生态学中则需要首先从地层中将孢粉分离纯化, 然后进行后续的实验工作(Parducci等, 2015). 从不同的介质中获得无污染的孢粉颗粒是基于分子手段进行孢粉自动鉴定的基础前提, 也是主要挑战之一. 由于空气中存在大量孢粉和孢粉碎屑(Wozniak等, 2018), 以及游离的植物DNA片段(Banchi等, 2020; Johnson等, 2021; Métris和Métris, 2023), 这可能意味着蜂蜜、大气颗粒物捕获器等获得的孢粉样品中含有非孢粉来源的DNA, 已有研究结果所报道的鉴定准确率可能偏高. 同时从第四纪沉积样品中分离孢粉样品同样面临较高的成本和污染风险(Parducci等, 2015).

(2) DNA提取. 孢粉DNA的提取需要将孢粉外壁打碎, 然后进行裂解孵化、DNA富集洗脱等步骤, 最后获得DNA样品. 孢粉外壁的破裂程度会影响孢粉DNA的提取效率, Swenson等(2021)的研究表明, 使用1~3mm的研磨珠在30Hz条件下震荡2~4min, 裂解孵化不超过1h的条件下, 可以获得最为理想的孢粉DNA产物. 孢粉DNA的提取多使用植物组织提取试剂盒, 但不同试剂盒的功效可能存在一定差异(Leontidou等, 2018; Bell等, 2019), 例如Leontidou等(2018)对比了DNeasy Plant mini kit (Qiagen)和Nucleomag kit (Macherey-Nagel), 后者提取的DNA产量要高于前者. 此外, 孢粉DNA提取环节也可使用自动DNA提取系统, 以降低工作强度以及人为操作带来的污染风险和实验结果偏差.

(3) PCR扩增. 常用的植物DNA条形码包括matK、trnL、ITS2、rbcL等(Hawkins等, 2015; Bell等, 2016; Leontidou等, 2018; Brennan等, 2019; Baksay等, 2020; Wasti等, 2024). 其中, ITS2、trnL内含子和非编码基因间隔区trnH-psbA在孢粉DNA分类学鉴定中更为常用(Hollingsworth等, 2011; Bell等, 2021), 尤其是nrITS2可以作为空气传播花粉监测的首选条形码(Polling等, 2022). 针对地层中的化石孢粉DNA, 高度保守、扩增体系稳定、在DNA高度降解的样本中仍可以成功扩增的trnL p6环(g-h)可能更为合适(Taberlet等, 2007). 值得注意的是, 不同条形码扩增引物的效率存在差异, 以及同种引物在不同研究中的功效也不

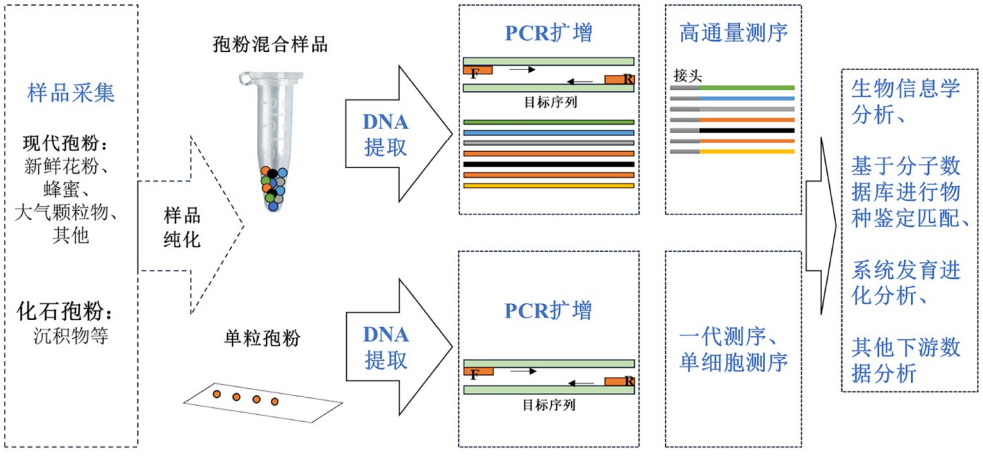

图 3　孢粉DNA宏条形码和孢粉基因分型研究流程示意

尽一致, 因此多引物共用可能是提高下游孢粉DNA分类学鉴定的选择之一(Taberlet等, 2018).

(4) 高通量测序. DNA宏条形码技术依赖于第二代测序技术, 其显著特征是高通量. 对高通量测序数据的主要分析流程包括测序数据质控和拆分、正反向数据匹配、污染排除、数据过滤、样品拆分、可操作分类单元(Operational Taxonomic Unit, OTU)聚类、物种匹配鉴定等. 相关数据处理可由专业生信分析程序实现, 例如Fastp(Chen, 2023)可以快速过滤测序环节出现的错误并评估测序数据质量, OBITools(Boyer等, 2016)支持数据过滤、样品拆分、OTU聚类和物种鉴定等, DAMe(Zepeda-Mendoza等, 2016)支持污染排除和准确的OUT聚类等, metabaR(Zinger等, 2021)通过识别污染和标签跳跃等产生的假阳性序列, 实现自动化数据质量评估.

(5) 物种鉴定. DNA条形码和Sanger测序技术获得的数据主要通过直接使用GenBank BLAST(Altschul等, 1990)进行比对鉴定. 随着测序深度和数据量的增加, 快速物种鉴定的生信处理方法也随之出现, 如OBITools(Boyer等, 2016)、PROTAX(Somervuo等, 2016)等. 孢粉DNA宏条形码鉴定的准确性依赖现代植物分子数据库的完善程度. 常用的全球分子数据库主要是美国国家生物技术信息中心的GenBank(www.ncbi.nlm.nih.gov/genbank/)、欧洲分子生物学实验室的ENA(https://www.ebi.ac.uk/ena)和生命条形码联盟构建的生命条形码数据库系统BOLD(https://www.boldsystems.org/)等. 但以全球物种条形码为参照会导致孢粉DNA序列鉴定出现高的假阳性结果(Taberlet等, 2018), 因此许多学者选择构建完备的本地植物分子数据库, 以提高物种鉴定准确性. 例如, Jones等(2021)使用Sanger测序的方法构建了英国全国性的DNA条形码数据库, 结合Genbank数据, 提供了英国所有的本土有花植物和针叶树的rbcL、matK和ITS2条形码, 并基于此开展了孢粉DNA宏条形码自动鉴定分类(Jones等, 2022). 针对北极地区现代植物物种构建了叶绿体trnL p6环序列数据库(Sønstebø等, 2010; Willerslev等, 2014), 在古生态学研究中使用较为广泛, 在现代孢粉DNA分类鉴定中也有使用(Milla等, 2021). 针对现代参比数据库不完备的情况, Somervuo等(2016)开发了DNA条形码的概率分类方法PROTAX, 可以提供在OTU中存在但没有参考序列的物种的可能物种分类及其概率, 以及计算出错误标记的参考序列, 即使在参比序列有限的前提下也可实现相对准确的分类学鉴定.

与专家鉴定相比, 孢粉DNA宏条形码技术的优势主要体现在: (1) 以高通量实现了高物种识别准确性, 相较于人工鉴定成本较低; (2) 具标准化的实验流程, 便于不同研究结果之间的对比和验证; (3) 不依赖分类专家, 经过短暂培训的科研人员也可以进行操作; (4) 物种鉴定结果分辨率高, 同时鉴定准确度也高于传统形态学分析; (5) 宏条形码直接针对孢粉中存留的DNA, 对孢粉保存的介质和孢粉粒的完整性要求低. 孢粉DNA宏条形码和孢粉人工鉴定的直接对比显示, 孢粉DNA宏条形码鉴定准确度不亚于或者优于形态学鉴定(Hawkins等, 2015; Keller等, 2015; Kraaijeveld等, 2015; Richardson等, 2015; Smart等, 2017), 例如, Keller等(2015)评估了16个混合花粉样本的鉴定结果, 孢粉DNA宏条形码分类结果中96%的分类群在属水平上可以正确识别, 70%的分类群在种水平上可以正确识别, 鉴定类群丰富度和分类学水平要优于基于形态学的专家鉴定.

然而, 基于DNA宏条形码技术开展的孢粉自动鉴定也存在一定的不足, 主要体现在实验过程中的污染风险大、引物的通用性不足、PCR特异性扩增、测序错误、数据库完备程度不高、丰度估计存在偏差等(Bell等, 2019; Swenson等, 2021; Polling等, 2022). 首先, 孢粉DNA实验从样品采集到测序的整个过程均存在很大的污染风险, 包括采样环节的交叉污染和实验室的污染等, 因此, 严格的采样控制、实验室污染预防规程、PCR实验前后实验室的物理隔离等是十分必要的(Pawlowski等, 2022; Capo等, 2023; Rieder等, 2024). 其次, 引物的通用性和DNA条码的鉴定功效存在差异(Taberlet等, 2018). 地层孢粉样品往往存在DNA降解的情况, 导致matK、rbcL等序列较长的片段在扩增时成功率较低, 而trnL p6环物种分辨率较低且无法区分近缘植物(Willerslev等, 2014). 针对现代孢粉样品, 可以采用多条形码共用以提高鉴定结果的物种分辨率, 但构建完善的本地DNA条形码参比数据库也面临一定挑战(Jones等, 2021). 此外, PCR扩增是DNA宏条形码的关键, 但其存在诸如扩增偏向性、错误扩增、扩增抑制等问题, 均会造成鉴定结果中存在假阳性和假阴性的现象. 孢粉DNA宏条形码鉴定结果中孢粉丰度

与通过专家鉴定的孢粉数量之间的相关性可能只达到中等水平(Bell等, 2019), 无法准确量化样本中每个物种的孢粉丰度. 虽然通过插入DNA内标的方法(SPIKEPIPE)可以得到相对准确的DNA序列丰度信息(Ji等, 2020), 但是SPIKEPIPE实验成本高且操作难度大, 目前在孢粉DNA宏条形码领域暂未使用. 基于鸟枪(Shotgun)测序进行的宏基因组学不需要对孢粉DNA进行扩增, 可作为DNA宏条形码的替代技术, 其提高了读取长基因的能力并提高了对混合孢粉的分类精度(Bell等, 2021; Peel等, 2019), 实现了接近100%准确的物种鉴定率、更加准确的丰度估计(Bell等, 2021), 并且结果具有很高的重复性(Lang等, 2019). 但是, 目前应用宏基因组技术对花粉鉴定的研究仍面临测序成本高、生信分析难度大, 以及现有的植物全基因组测序数据少而难以构建基因组参考文库等问题(Bell等, 2021). 近期, Peel等(2019)开发了一种RevMet(反向宏基因组学)方法, 通过比对混合孢粉的浅层基因组测序数据与参考物种的“基因组略读”直接进行孢粉鉴定, 在没有参考基因组的前提下实现了对混合孢粉样本的鉴定和半定量表征, 但如何将RevMet方法应用到第四纪孢粉鉴定中仍需进一步探索.

4　第四纪孢粉学中孢粉鉴定自动化的挑战和机遇

孢粉鉴定是第四纪孢粉学研究的基础和核心手段. 孢粉鉴定自动化的初衷是推动第四纪孢粉学的发展(Stillman和Flenley, 1996), 但是在第四纪孢粉研究中, 基于分类专家的人工孢粉鉴定依然是目前第四纪孢粉学研究的主要方式. 推动孢粉鉴定自动化、提高孢粉鉴定的物种分辨率长期以来都是第四纪孢粉学进一步发展的挑战和机遇之一(许清海等, 2024).

第一, 第四纪孢粉自动鉴定的实现面临诸多困难, 但其未来发展和实现是必然趋势. 狭义上的孢粉鉴定自动化就是采用计算机技术基于形态特征对孢粉颗粒进行分类学鉴定, 因此基于图像识别技术历来是孢粉自动鉴定的热门方向. 近期, 基于图像识别、CNN模型分类等自动鉴定方法在不同的第四纪孢粉记录中取得了一定的发展(Bourel等, 2020; Theuerkauf等, 2023; Durand等, 2024; von Allmen等, 2024). 但是目前基于图像识别的自动鉴定同样无法有效提高孢粉鉴定的物种分辨率(许清海等, 2022), 相对来讲, 基于孢粉化学成分、孢粉DNA进行的自动孢粉鉴定则有望提供更详细的物种鉴定结果(表1). 图像识别和多光谱自动成像流式细胞仪的联用在现代孢粉鉴定中得到的准确率超过90%(Dunker等, 2021), 因此, 笔者建议将传统的图像识别与光谱技术相结合, 采用机器学习和CNN模型等技术进行分类鉴定(Bourel等, 2020; Durand等, 2024), 有望进一步推进第四纪孢粉鉴定自动化、高分辨率的鉴定工作. 基于孢粉DNA条形码和宏基因组学开展的孢粉自动分类鉴定在第四纪孢粉学中鲜有报道, 与之相对的是沉积古DNA(sedaDNA)发展迅速(Parducci等, 2017). sedaDNA的来源可能包括了孢粉以及其他游离DNA片段, 可提供的植物分类学信号要多于孢粉(Parducci等, 2015, 2019; 吴铠等, 2024), 并且对sedaDNA进行多重引物扩增也可以获得某一种群的群体遗传信息(Lammers等, 2024). 但是考虑到sedaDNA存在诸如来源不明确、降解和保存问题复杂以及无法定量重建古植被等问题(吴铠等, 2021), 第四纪孢粉DNA自动鉴定在第四纪植被演变、植物种群形成和进化等研究中依然有一定的应用和发展前景.

第二, 探索适用于第四纪孢粉样品的预处理方法流程, 获取无污染且纯度高的地层孢粉样品, 推动自动化鉴定的实现. 第四纪孢粉散布于不同的沉积环境介质中, 对其进行自动分类鉴定的关键步骤之一是安全富集足量的孢粉颗粒. 第四纪孢粉样品的预处理方法主要是化学处理法和重液浮选法(Faegri等, 1989), 这种前处理或者孢粉纯化的方法显然适用于基于形态学的孢粉自动鉴定(von Allmen等, 2024), 尤其是可将此方法获得的孢粉样品直接制片或与流式细胞仪联用进行孢粉图像自动定位采集, 进而实现孢粉颗粒识别和自动鉴定. 但是, 化学预处理方法可能会引起孢粉内含物质损失、孢粉外壁局部化学成分的改变, 以及引起DNA的降解, 传统的第四纪孢粉预处理方法显然不适用于依靠孢粉光谱特征或者DNA信号的自动鉴定. 因此, 推动第四纪孢粉鉴定自动化需要发展适用于孢粉化学组成、DNA或者与图像识别技术联用的标准化孢粉分离富集流程.

第三, 构建完备的本地孢粉形态、光谱和DNA条形码数据库, 提高孢粉自动鉴定的准确度(Holt和Bennett, 2014; Durand等, 2024). 目前孢粉形态参比数据库的主要问题是缺少全面且通用的数据库, 已有数据

表 1　孢粉鉴定自动化实现途径的对比[a)]

参数	图像识别	化学组成	DNA宏条形码
鉴定分辨率	科、属	属、种、亚种	属、种、亚种
前处理要求	不破坏外壁结构	不改变化学性质	不引入分子污染
数据采集	光学、SEM、荧光等孢粉图像	拉曼、FTIR、荧光、核磁共振等光谱序列	孢粉DNA提取、PCR扩增和测序
代表性分类器/算法	CNN等	SVM、ANN等	BLAST
数据库完备程度	较完备	较少且通用性不足	通用且在快速发展
鉴定效率	对同环境下混合样本可以实现定量快速准确鉴定	对少量样本可实现定量准确鉴定	对大量混合样品可实现半定量快速准确鉴定
现代孢粉	适用	适用	适用
第四纪孢粉	有限适用	有限适用	有限适用
实验成本	低	中高(设备成本)	高
标准化流程	否	否	是
主要不足	分辨率较低、难以识别损坏或变形孢粉	无统一实验和分析标准、通用性较差	污染风险高、条码通用性不足、假阳性、定量估计偏差

a) SEM, 扫描电子显微镜; FTIR, 傅里叶红外光谱; PCR, 聚合酶链式反应; CNN, 卷积神经网络; SVM, 支持向量机; ANN, 人工神经网络

库中高质量的孢粉图像数量不足，部分参比图像无法提供清晰的纹饰细节，导致孢粉自动鉴定准确率和物种分辨率不高. 增加参考图像的数量可以提高模型鉴定准确率，但参考图像数量过多也会导致模型过度拟合问题(Olsson等, 2021). 因此，针对研究区域和具体科学问题，在全球或者区域孢粉图像数据库的基础上构建高质量的本地数据库，可能是未来推动基于图像识别技术进行孢粉自动鉴定应用的首选. 目前我国的植物物种信息系统收录的五万多幅孢粉图版和超过500万条植物DNA序列已经可以开放下载使用(https://www.iplant.cn/psis)，有望进一步推动基于图像识别和DNA条形码的孢粉自动鉴定在国内的进一步发展. 孢粉光谱数据库建设依然任重道远(Guedes等, 2014; Pereira等, 2021)，但考虑到光谱能够提供孢粉组成信息，基于构建的本地光谱数据库结合图像识别技术，可能是未来第四纪孢粉鉴定自动化值得发展的方向之一(Dunker等, 2021).

第四，建立标准化分析流程是评估孢粉自动鉴定效率以及实现不同研究结果对比的前提. 针对第四纪孢粉分析，Stillman和Flenley(1996)强调自动鉴定结果的可重复性，建议基于图像识别的自动鉴定，并采用永久玻片的形式对样品进行保存，以备后续其他方法的验证. 在本文中，我们归纳了包括图像识别技术在内的3种孢粉鉴定自动化实现途径，不同的孢粉自动鉴定实现途径在方法上存在较大差异，即使同样基于孢粉形态学的图像识别技术也没有标准的自动化流程，基于光谱技术和孢粉DNA的自动鉴定步骤更是纷繁多样, 导致对其鉴定效率无法进行对比评价. 基于图像识别的孢粉自动鉴定可能是未来第四纪孢粉学发展的主要手段，如何将多光谱流式细胞仪(Dunker等, 2021)等自动化图像采集设备、孢粉颗粒自动识别(Theuerkauf等, 2023)、机器学习或深度学习方法(Dunker等, 2021; Durand等, 2024)等相结合, 从形态学或形态学与光谱结合的途径获得更加准确的、物种分辨率更高的孢粉分类学结果，是现阶段第四纪孢粉自动鉴定的重要挑战和机遇. 与此同时，如何推动孢粉DNA在第四纪孢粉学中的应用也需要更多的努力.

5　结论

孢粉分析虽已历经百余年发展，但是孢粉鉴定依然是以形态学鉴定为主要手段，其优势主要是鉴定标准统一、鉴定结果可靠性高、通用性强等，有力地推动了以孢粉为研究载体的第四纪孢粉学的发展. 但是基于专家进行孢粉形态学鉴定在鉴定效率和物种分辨率等方面也存在一定的不足，这催生了孢粉鉴定自动化的需求. 目前，孢粉鉴定自动化的实现途径主要有3种，分别是基于孢粉形态学的自动鉴定、基于孢粉化

学和光谱特征的自动鉴定和基于孢粉DNA信息的自动鉴定.

基于图像识别技术的孢粉自动鉴定是发展时间最久也是最为成熟的孢粉鉴定自动化方式, 其优势主要体现在孢粉图版丰富、鉴定结果支持专家核验、结果可重复性高等, 在未来一定时间内依然会是孢粉鉴定自动化的主流趋势. 基于化学和光谱特征的孢粉自动鉴定在研究孢粉的化学组成方面有着显著优势, 不仅可以提供更加精确的分类信息, 而且可以与显微技术联用进一步提高孢粉鉴定自动化效率. 基于孢粉DNA宏条形码进行的孢粉鉴定主要优势在于具有标准化流程, 鉴定效率和物种分辨率高于专家鉴定, 并且能够对大量混合样品实现快速准确鉴定等. 然而, 上述3种孢粉鉴定自动化途径均存在一些不足之处和应用上的问题, 且在第四纪孢粉学中的应用有待进一步发展. 由于第四纪孢粉在地层中的保存伴随着内含有机质和遗传物质的损失, 这为基于光谱和分子手段的地层孢粉自动鉴定造成了较大的不确定性, 而地层中保存的孢粉外壁较完整, 使得基于形态学的孢粉自动鉴定技术更为合适, 同时基于形态学和光谱技术联用可能是未来第四纪孢粉鉴定自动化实现的重要方向. 最后我们建议探索适用于光谱和DNA鉴定技术的第四纪孢粉样品的预处理方法、构建和完善本地参比数据库以及建立标准分析流程等是目前孢粉鉴定自动化亟须解决的问题.

综上, 孢粉鉴定自动化虽面临诸多困难, 但其未来发展和实现乃是必然趋势. 随着技术的进步, 未来第四纪孢粉鉴定需要考虑纳入多种自动鉴定方法, 结合人工鉴定和自动鉴定的优势, 实现专家监督下的孢粉鉴定自动化, 以提高鉴定结果的准确性和效率以及鉴定的物种分辨率, 推动相关学科的发展.

参考文献

陈海燕, 徐德宇, 廖梦娜, 李凯, 倪健, 曹现勇, 程波, 郝秀东, 孔昭宸, 李升峰, 李小强, 刘光琇, 刘平妹, 刘兴起, 孙湘君, 唐领余, 魏海成, 许清海, 阎顺, 羊向东, 杨振京, 于革, 张芸, 张志勇, 赵克良, 郑卓, Herzchuh U. 2021. 中国现代花粉数据集. 植物生态学报, 45: 799–808

李天庆, 曹慧娟, 康木生, 张志翔, 赵楠, 张晖. 2010. 中国木本植物花粉电镜扫描图志. 北京: 科学出版社

唐领余, 毛礼米, 舒军武, 李春海, 沈才明, 周忠泽. 2016. 中国第四纪孢粉图鉴. 北京: 科学出版社

王伏雄, 钱南芬, 张玉龙, 杨惠秋. 1995. 中国植物花粉形态. 第2版. 北京: 科学出版社

王伏雄, 喻诚鸿. 1954. 花粉形态的研究1. 术语及研究方法. 植物学报, 3: 81–104

吴铠, 李凯, 贾伟瀚, 廖梦娜, 倪健. 2021. 湖泊沉积植物古DNA的现代过程. 植物生态学报, 46: 735–752

吴铠, 李凯, 贾伟瀚, Stoof-Leichsenring K R, Herzschuh U, 倪健, 廖梦娜, 田芳. 2024. 湖泊沉积植物DNA宏条形码技术在青藏高原植被调查中的应用. 中国科学: 地球科学, 54: 3629–3646

许清海, 黄小忠, 王涛. 2022. 中国北方山地常见植物花粉形态研究: 光学显微镜下精确鉴定方法探索. 北京: 科学出版社

许清海, 吕厚远, 郑卓. 2024. 第四纪孢粉学当前面临的主要挑战和机遇. 中国科学: 地球科学, 54: 2178–2192

许清海. 2015. 中国常见栽培植物花粉形态: 地层中寻找人类痕迹之借鉴. 北京: 科学出版社

中国科学院北京植物研究所古植物研究室孢粉组. 1976. 中国蕨类植物孢子形态. 北京: 科学出版社

中国科学院植物研究所古植物室孢粉组, 中国科学院华南植物研究所形态研究室. 1982. 中国热带亚热带被子植物花粉形态. 北京: 科学出版社

周伯睿, 廖梦娜, 李凯, 徐德宇, 陈海燕, 倪健, 曹现勇, 孔昭宸, 许清海, 张芸, Herzschuh U, 蔡永立, 陈碧珊, 陈敬安, 陈陵康, 程波, 高洋, 黄赐璇, 黄小忠, 李升峰, 李文漪, 廖淦标, 刘光琇, 刘平妹, 刘兴起, 马春梅, 宋长青, 孙湘君, 唐领余, 王曼华, 王永波, 夏玉梅, 徐家声, 阎顺, 羊向东, 姚轶锋, 叶传永, 张志勇, 赵增友, 郑卓, 朱诚. 2023. 中国地层花粉数据集. 植物生态学报, 47: 1453–1463

朱江峰. 2017. 基于双波长激发荧光光谱的花粉分类研究. 硕士学位论文. 杭州: 浙江大学

Altschul S F, Gish W, Miller W, Myers E W, Lipman D J. 1990. Basic local alignment search tool. J Mol Biol, 215: 403–410

Assemien P, Bonnefille R, Cambon-Bou G, Caratini Cl, Cerceau M, Dang C D, Fredoux A, Guers J, Guinet Ph, Hideux M, Hul-Thol S, Keddam-Malplanche M, Le Thomas A, Lobreau-Callen D, Lugardon B, Maley J, Mallea M, Masure E, Medus J, Nigaud M, Riollet G, Senesse S, Sivak J, Soler M, Stainier F, Thanikaimoni G, Thiam A. 1974. Pollen et Spores d'Afrique tropicale. Traveaux et documents de geographie tropical. In: Centre d'Etude de Geographie Tropical, Association des Palynogues de Langue Fran-çaise. Talence, France: Universite de Bordeaux

Astolfi G, Gonçalves A B, Menezes G V, Borges F S B, Astolfi A C M N, Matsubara E T, Alvarez M, Pistori H. 2020. POLLEN73S: An image dataset for pollen grains classification. Ecol Inf, 60: 101165

Bağcıoğlu M, Zimmermann B, Kohler A, Coles J A. 2015. A multiscale vibrational spectroscopic approach for identification and biochemical characterization of pollen. PLoS One, 10: e0137899

Baksay S, Pornon A, Burrus M, Mariette J, Andalo C, Escaravage N. 2020. Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. Sci Rep, 10: 4202

Banchi E, Pallavicini A, Muggia L. 2020. Relevance of plant and fungal DNA metabarcoding in aerobiology. Aerobiologia, 36: 9–23

Bell K L, Burgess K S, Botsch J C, Dobbs E K, Read T D, Brosi B J. 2019. Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. Mol Ecol, 28: 431–455

Bell K L, de Vere N, Keller A, Richardson R T, Gous A, Burgess K S, Brosi B J, Van der Bank M. 2016. Pollen DNA barcoding: Current applications and future prospects. Genome, 59: 629–640

Bell K L, Petit Iii R A, Cutler A, Dobbs E K, Macpherson J M, Read T D, Burgess K S, Brosi B J. 2021. Comparing whole-genome shotgun sequencing and DNA metabarcoding approaches for species identification and quantification of pollen species mixtures. Ecol Evol, 11: 16082–16098

Bell K L, Turo K J, Lowe A, Nota K, Keller A, Encinas-Viso F, Parducci L, Richardson R T, Leggett R M, Brosi B J, Burgess K S, Suyama Y, de Vere N. 2023. Plants, pollinators and their interactions under global ecological change: The role of pollen DNA metabarcoding. Mol Ecol, 32: 6345–6362

Bennett K D, Parducci L. 2006. DNA from pollen: Principles and potential. Holocene, 16: 1031–1034

Birks H J B. 2019. Contributions of Quaternary botany to modern ecology and biogeography. Plant Ecol Divers, 12: 189–385

Birks H J B, Berglund B E. 2018. One hundred years of Quaternary pollen analysis 1916–2016. Veget Hist Archaeobot, 27: 271–309

Bourel B, Marchant R, de Garidel-Thoron T, Tetard M, Barboni D, Gally Y, Beaufort L. 2020. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. Comput Geosci, 140: 104498

Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. 2016. OBITools: A Unix-inspired software package for DNA metabarcoding. Mol Ecol Resour, 16: 176–182

Brennan G L, Potter C, de Vere N, Griffith G W, Skjøth C A, Osborne N J, Wheeler B W, McInnes R N, Clewlow Y, Barber A, Hanlon H M, Hegarty M, Jones L, Kurganskiy A, Rowney F M, Armitage C, Adams-Groom B, Ford C R, Petch G M, Elliot A, Frisk C A, Neilson R, Potter S, Rafiq A M, Roy D B, Selby K, Steinberg N, Creer S. 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. Nat Ecol Evol, 3: 750–754

Capo E, Barouillet C, John P. 2023. Tracking Environmental Change Using Lake Sediments: Volume 6: Sedimentary DNA. Cham: Springer

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta, 2: e107

Chen H, Wang Z, An Y. 2022. Pollen recognition and classification method based on local binary pattern. In: Jiang D, Song H, eds. Simulation Tools and Techniques. LNICST, 424. Cham: Springer

Chudyk C, Castaneda H, Leger R, Yahiaoui I, Boochs F. 2015. Development of an automatic pollen classification system using shape, texture and aperture features. LWA 2015 Workshops

Colinvaux P, Oliveira P, Moreno J. 1999. Amazon Pollen Manual and Atlas. Amsteldijk: Hardwood Academic Publishers

Cornman R S, Otto C R V, Iwanowicz D, Pettis J S, Jan E. 2015. Taxonomic characterization of honey bee (*Apis mellifera*) pollen foraging based on non-overlapping paired-end sequencing of nuclear ribosomal loci. PLoS One, 10: e0145365

Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas W K, Potter C, Bik H M, Freckleton R. 2016. The ecologist's field guide to sequence-based identification of biodiversity. Methods Ecol Evol, 7: 1008–1018

de Vere N, Jones L E, Gilmore T, Moscrop J, Lowe A, Smith D, Hegarty M J, Creer S, Ford C R. 2017. Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. Sci Rep, 7: 42838

Depciuch J, Kasprzyk I, Roga E, Parlinska-Wojtan M. 2016. Analysis of morphological and molecular composition changes in allergenic *Artemisia vulgaris* L. pollen under traffic pollution using SEM and FTIR spectroscopy. Environ Sci Pollut Res, 23: 23203–23214

Diehn S, Zimmermann B, Tafintseva V, Bağcıoğlu M, Kohler A, Ohlson M, Fjellheim S, Kneipp J. 2020. Discrimination of grass pollen of different species by FTIR spectroscopy of individual pollen grains. Anal Bioanal Chem, 412: 6459–6474

Dunker S, Motivans E, Rakosy D, Boho D, Mäder P, Hornick T, Knight T M. 2021. Pollen analysis using multispectral imaging flow cytometry and deep learning. New Phytol, 229: 593–606

Durand M, Paillard J, Ménard M P, Suranyi T, Grondin P, Blarquez O, Sun X. 2024. Pollen identification through convolutional neural networks: First application on a full fossil pollen sequence. PLoS One, 19: e0302424

Erdtman G. 1952. Pollen Morphology and Plant Taxonomy. Angiosperms. An Introduction to Palynology. I. Stockholm: Almqvist and Wiksell

Erdtman G. 1960. The acetolysis method A revised description. Sven Bot Tidskr, 54: 561–564

Erdtman G. 1969. Handbook of Palynology. Copenhagen: Munksgaard

Faegri K, Kaland P, Krzywinski K. 1989. Textbook of Pollen Analysis. London: John Wiley & Sons Ltd

Fischer H. 1890. Beiträge zur vergleichenden Morphologie der

Pollenkörner. Thesis, Breslau (Wroclaw)
Flenley J. 1968. The problem of pollen recognition. In: Clowes M B, Penny J P, eds. Problems in Picture Interpretation. Canberrapp: CSIRO. 141–145
Lozano-Vega G, Benezeth Y, Marzani F, Boochs F. 2014. Analysis of relevant features for pollen classification. In: Iliadis L, Maglogiannis I, Papadopoulos H, eds. Artificial Intelligence Applications and Innovations. AIAI 2014. IFIP Advances in Information and Communication Technology, vol 436. Berlin, Heidelberg: Springer. 395–404
Gonçalves A B, Souza J S, Silva G G, Cereda M P, Pott A, Naka M H, Pistori H, Kestler H A. 2016. Feature extraction and machine learning for the classification of Brazilian savannah pollen grains. PLoS One, 11: e0157044
Guedes A, Ribeiro H, Fernández-González M, Aira M J, Abreu I. 2014. Pollen Raman spectra database: Application to the identification of airborne pollen. Talanta, 119: 473–478
Haralick R M. 1979. Statistical and structural approaches to texture. Proc IEEE, 67: 786–804
Hawkins J, de Vere N, Griffith A, Ford C R, Allainguillaume J, Hegarty M J, Baillie L, Adams-Groom B, Labra M. 2015. Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. PLoS One, 10: e0134735
Herzschuh U, Li C, Böhmer T, Postl A K, Heim B, Andreev A A, Cao X, Wieczorek M, Ni J. 2022. LegacyPollen 1.0: A taxonomically harmonized global late Quaternary pollen dataset of 2831 records with standardized chronologies. Earth Syst Sci Data, 14: 3213–3227
Hollingsworth P M, Graham S W, Little D P, Steinke D. 2011. Choosing and using a plant DNA barcode. PLoS One, 6: e19254
Holt K, Allen G, Hodgson R, Marsland S, Flenley J. 2011. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. Rev Palaeobot Palynol, 167: 175–183
Holt K A, Bennett K D. 2014. Principles and methods for automated palynology. New Phytol, 203: 735–742
Hornick T, Richter A, Harpole W S, Bastl M, Bohlmann S, Bonn A, Bumberger J, Dietrich P, Gemeinholzer B, Grote R, Heinold B, Keller A, Luttkus M L, Mäder P, Motivans Švara E, Passonneau S, Punyasena S W, Rakosy D, Richter R, Sickel W, Steffan-Dewenter I, Theodorou P, Treudler R, Werchan B, Werchan M, Wolke R, Dunker S. 2022. An integrative environmental pollen diversity assessment and its importance for the Sustainable Development Goals. Plants People Planet, 4: 110–121
Jardine P E, Gosling W D, Lomax B H, Julier A C M, Fraser W T. 2019. Chemotaxonomy of domesticated grasses: A pathway to understanding the origins of agriculture. J Micropalaeontol, 38: 83–95
Ji Y, Huotari T, Roslin T, Schmidt N M, Wang J, Yu D W, Ovaskainen O. 2020. SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. Mol Ecol Resour, 20: 256–267
Johnson M D, Fokar M, Cox R D, Barnes M A. 2021. Airborne environmental DNA metabarcoding detects more diversity, with less sampling effort, than a traditional plant community survey. BMC Ecol Evo, 21: 218
Jones L, Lowe A, Ford C R, Christie L, Creer S, de Vere N. 2022. Temporal patterns of honeybee foraging in a diverse floral landscape revealed using pollen DNA metabarcoding of honey. Integr Comp Biol, 62: 199–210
Jones L, Twyford A D, Ford C R, Rich T C G, Davies H, Forrest L L, Hart M L, McHaffie H, Brown M R, Hollingsworth P M, de Vere N. 2021. Barcode UK: A complete DNA barcoding resource for the flowering plants and conifers of the United Kingdom. Mol Ecol Resour, 21: 2050–2062
Julier A C M, Jardine P E, Coe A L, Gosling W D, Lomax B H, Fraser W T. 2016. Chemotaxonomy as a tool for interpreting the cryptic diversity of Poaceae pollen. Rev Palaeobot Palynol, 235: 140–147
Kasprzyk I, Depciuch J, Grabek-Lejko D, Parlinska-Wojtan M. 2018. FTIR-ATR spectroscopy of pollen and honey as a tool for unifloral honey authentication. The case study of rape honey. Food Control, 84: 33–40
Keller A, Danner N, Grimmer G, Ankenbrand M, von der Ohe K, von der Ohe W, Rost S, Härtel S, Steffan-Dewenter I, Mock H. 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol J, 17: 558–566
Kenđel A, Zimmermann B. 2020. Chemical analysis of pollen by FT-Raman and FTIR spectroscopies. Front Plant Sci, 11: 352
Klimczak L J, Ebner von Eschenbach C, Thompson P M, Buters J T M, Mueller G A. 2020. Mixture analyses of air-sampled pollen extracts can accurately differentiate pollen taxa. Atmos Environ, 243: 117746
Kraaijeveld K, de Weger L A, Ventayol García M, Buermans H, Frank J, Hiemstra P S, den Dunnen J T. 2015. Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour, 15: 8–16
Kubera E, Kubik-Komar A, Piotrowska-Weryszko K, Skrzypiec M. 2021. Deep learning methods for improving pollen monitoring. Sensors, 21: 3526
Kubik-Komar A, Kubera E, Piotrowska-Weryszko K. 2018. Selection of morphological features of pollen grains for chosen tree taxa. Biol

Open, 7: bio031237

Lammers Y, Taberlet P, Coissac E, Elliott L D, Merkel M F, Pitelkova I, Alsos I G. 2024. Multiplexing PCR allows the identification of within-species genetic diversity in ancient eDNA. Mol Ecol Resour, 24: e13926

Lang D, Tang M, Hu J, Zhou X. 2019. Genome-skimming provides accurate quantification for pollen mixtures. Mol Ecol Resour, 19: 1433–1446

Langford M, Taylor G, Flenley J. 1986. The application of texture analysis for automated pollen identification. In: Proceedings of the Conference on Identification and Pattern Recognition, Volume 2. Toulouse. 729–739

Langford M, Taylor G E, Flenley J R. 1990. Computerized identification of pollen grains by texture analysis. Rev Palaeobot Palynol, 64: 197–203

Leontidou K, Vernesi C, De Groeve J, Cristofolini F, Vokou D, Cristofori A. 2018. DNA metabarcoding of airborne pollen: New protocols for improved taxonomic identification of environmental samples. Aerobiologia, 34: 63–74

Leontidou K, Vokou D, Sandionigi A, Bruno A, Lazarina M, De Groeve J, Li M, Varotto C, Girardi M, Casiraghi M, Cristofori A. 2021. Plant biodiversity assessment through pollen DNA metabarcoding in Natura 2000 habitats (Italian Alps). Sci Rep, 11: 18226

Li P, Treloar W J, Flenley J R, Empson L. 2004. Towards automation of palynology 2: The use of texture measures and neural network analysis for automated identification of optical images of pollen grains. J Quat Sci, 19: 755–762

Lowe A, Jones L, Brennan G, Creer S, de Vere N. 2022. Seasonal progression and differences in major floral resource use by bees and hoverflies in a diverse horticultural and agricultural landscape revealed by DNA metabarcoding. J Appl Ecol, 59: 1484–1495

Macgregor C J, Kitson J J N, Fox R, Hahn C, Lunt D H, Pocock M J O, Evans D M. 2019. Construction, validation, and application of nocturnal pollen transport networks in an agro-ecosystem: A comparison using light microscopy and DNA metabarcoding. Ecol Entomol, 44: 17–29

Mander L, Li M, Mio W, Fowlkes C C, Punyasena S W. 2013. Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. Proc R Soc B, 280: 20131905

Marcos J V, Nava R, Cristóbal G, Redondo R, Escalante-Ramírez B, Bueno G, Déniz Ó, González-Porto A, Pardo C, Chung F, Rodríguez T. 2015. Automated pollen identification using microscopic imaging and texture analysis. Micron, 68: 36–46

Matsuo M, Kozuki N, Inomata Y, Kumagai Y, Shiba R, Hamaguchi K, Tanito M. 2022. Automated focal plane merging from a stack of gonioscopic photographs using a focus-stacking algorithm. Trans Vis Sci Tech, 11: 22

Métris K L, Métris J. 2023. Aircraft surveys for air eDNA: Probing biodiversity in the sky. PeerJ, 11: e15171

Milla L, Sniderman K, Lines R, Mousavi-Derazmahalleh M, Encinas-Viso F. 2021. Pollen DNA metabarcoding identifies regional provenance and high plant diversity in Australian honey. Ecol Evol, 11: 8683–8698

Mirkin G, Bagdasaryan L. 1972. The feasibility of identifying paleontological objects with the aid of optical analysing systems. Paleontol J, 6: 103–108

Mohl H. 1934. Beiträge zur Anatomie und Physiologie der Gewächse. Erstes Heft. Über den Bau und die Formen der Pollenkörner. Fischer und Co, Bern

Mondol A S, Patel M D, Rüger J, Stiebing C, Kleiber A, Henkel T, Popp J, Schie I W. 2019. Application of high-throughput screening raman spectroscopy (HTS-RS) for label-free identification and molecular characterization of pollen. Sensors, 19: 4428

Muthreich F, Zimmermann B, Birks H J B, Vila-Viçosa C M, Seddon A W R. 2020. Chemical variations in *Quercus* pollen as a tool for taxonomic identification: Implications for long-term ecological and biogeographical research. J Biogeogr, 47: 1298–1309

Nagata N, Saito C, Sakai A, Kuroiwa H, Kuroiwa T. 1999. The selective increase or decrease of organellar DNA in generative cells just after pollen mitosis one controls cytoplasmic inheritance. Planta, 209: 53–65

Olsson O, Karlsson M, Persson A S, Smith H G, Varadarajan V, Yourstone J, Stjernman M, Freckleton R. 2021. Efficient, automated and robust pollen analysis using deep learning. Methods Ecol Evol, 12: 850–862

Parducci L, Alsos I G, Unneberg P, Pedersen M W, Han L, Lammers Y, Salonen J S, Väliranta M M, Slotte T, Wohlfarth B. 2019. Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from Southern Sweden. Front Ecol Evol, 7: 189

Parducci L, Bennett K D, Ficetola G F, Alsos I G, Suyama Y, Wood J R, Pedersen M W. 2017. Ancient plant DNA in lake sediments. New Phytol, 214: 924–942

Parducci L, Suyama Y, Lascoux M, Bennett K D. 2005. Ancient DNA from pollen: A genetic record of population history in Scots pine. Mol Ecol, 14: 2873–2882

Parducci L, Väliranta M, Salonen J S, Ronkainen T, Matetovici I, Fontana S L, Eskola T, Sarala P, Suyama Y. 2015. Proxy comparison in ancient peat sediments: Pollen, macrofossil and plant DNA. Phil Trans R Soc B, 370: 20130382

Pawlowski J, Bruce K, Panksep K, Aguirre F I, Amalfitano S, Apothéloz-Perret-Gentil L, Baussant T, Bouchez A, Carugati L, Cermakova K, Cordier T, Corinaldesi C, Costa F O, Danovaro R,

Dell′Anno A, Duarte S, Eisendle U, Ferrari B J D, Frontalini F, Frühe L, Haegerbaeumer A, Kisand V, Krolicka A, Lanzén A, Leese F, Lejzerowicz F, Lyautey E, Maček I, Sagova-Marečková M, Pearman J K, Pochon X, Stoeck T, Vivien R, Weigand A, Fazi S. 2022. Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. Sci Total Environ, 818: 151783

Peel N, Dicks L V, Clark M D, Heavens D, Percival-Alwyn L, Cooper C, Davies R G, Leggett R M, Yu D W, Freckleton R. 2019. Semi-quantitative characterisation of mixed pollen samples using MinION sequencing and Reverse Metagenomics (RevMet). Methods Ecol Evol, 10: 1690–1701

Pereira S G, Guedes A, Abreu I, Ribeiro H. 2021. Testing the Raman parameters of pollen spectra in automatic identification. Aerobiologia, 37: 15–28

Petersen G, Johansen B, Seberg O. 1996. PCR and sequencing from a single pollen grain. Plant Mol Biol, 31: 189–191

Polling M, Sin M, de Weger L A, Speksnijder A G C L, Koenders M J F, de Boer H, Gravendeel B. 2022. DNA metabarcoding using nrITS2 provides highly qualitative and quantitative results for airborne pollen monitoring. Sci Total Environ, 806: 150468

Punt W, Blackmore S, Clarke G, et al. 1976–2009. The North-West European Pollen Flora, vol 1–9. Amsterdam: Elsevier

Punyasena S W, Haselhorst D S, Kong S, Fowlkes C C, Moreno J E. 2022. Automated identification of diverse Neotropical pollen samples using convolutional neural networks. Methods Ecol Evol, 13: 2049–2064

Richardson R T, Lin C, Sponsler D B, Quijia J O, Goodell K, Johnson R M. 2015. Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl Plant Sci, 3: 1400066

Rieder J, Jemmi E, Hunter M E, Adrian-Kalchhauser I. 2024. A guide to environmental DNA extractions for non-molecular trained biologists, ecologists, and conservation scientists. Environ DNA, 6: e70002

Rodriguez-Damian M, Cernadas E, Formella A, Fernandez-Delgado M, Pilar De Sa-Otero M. 2006. Automatic detection and classification of grains of pollen based on shape and texture. IEEE Trans Syst Man Cybern C, 36: 531–542

Ronneberger O, Schultz E, Burkhardt H. 2002. Automated pollen recognition using 3D volume images from fluorescence microscopy. Aerobiologia, 18: 107–115

Schulte F, Lingott J, Panne U, Kneipp J. 2008. Chemical characterization and classification of pollen. Anal Chem, 80: 9551–9556

Seifert S, Merk V, Kneipp J. 2016. Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods. J Biophoton, 9: 181–189

Sengupta A, Laucks M L, Davis E J. 2005. Surface-enhanced raman spectroscopy of bacteria and pollen. Appl Spectrosc, 59: 1016–1023

Sevillano V, Aznarte J L, Rutherford S. 2018. Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. PLoS One, 13: e0201807

Sevillano V, Holt K, Aznarte J L, Raza M. 2020. Precise automatic classification of 46 different pollen types with convolutional neural networks. PLoS One, 15: e0229751

Silva D, Nara P, Quinta B, Gonçalves A, Pistori H, Borth M. 2014. Application of wavelet transform in the classification of pollen grains. Afr J Agric Res, 9: 908–913

Smart M D, Cornman R S, Iwanowicz D D, McDermott-Kubeczko M, Pettis J S, Spivak M S, Otto C R V. 2017. A comparison of honey bee-collected pollen from working agricultural lands using light microscopy and ITS metabarcoding. Environ Entomol, 46: nvw159

Somervuo P, Koskela S, Pennanen J, Henrik Nilsson R, Ovaskainen O. 2016. Unbiased probabilistic taxonomic classification for DNA barcoding. Bioinformatics, 32: 2920–2927

Sønstebø J H, Gielly L, Brysting A K, Elven R, Edwards M, Haile J, Willerslev E, Coissac E, Rioux D, Sannier J, Taberlet P, Brochmann C. 2010. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. Mol Ecol Resour, 10: 1009–1018

Stiebing C, Post N, Schindler C, Göhrig B, Lux H, Popp J, Heutelbeck A, Schie I W. 2022. Revealing the chemical composition of birch pollen grains by Raman spectroscopic imaging. Int J Mol Sci, 23: 5112

Stillman E C, Flenley J R. 1996. The needs and prospects for automation in palynology. Quat Sci Rev, 15: 1–5

Suyama Y, Kawamuro K, Kinoshita I, Yoshimura K, Tsumura Y, Takahara H. 1996. DNA sequence from a fossil pollen of *Abies* spp. from Pleistocene peat. Genes Genet Syst, 71: 145–149

Swenson S J, Gemeinholzer B, Heinze B. 2021. Testing the effect of pollen exine rupture on metabarcoding with Illumina sequencing. PLoS One, 16: e0245611

Swiatly-Blaszkiewicz A, Pietkiewicz D, Matysiak J, Czech-Szczapa B, Cichocka K, Kupcewicz B. 2021. Rapid and accurate approach for honeybee pollen analysis using ED-XRF and FTIR spectroscopy. Molecules, 26: 6024

Taberlet P, Bonin A, Zinger L, Coissac É. 2018. Environmental DNA: For Biodiversity Research and Monitoring. Oxford: Oxford University Press

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E. 2007. Power and

limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res, 35: e14

Tanaka I. 1997. Differentiation of generative and vegetative cells in angiosperm pollen. Sexual Plant Reprod, 10: 1–7

Theuerkauf M, Siradze N, Gillert A. 2023. A trainable object finder, selector and identifier for pollen, spores and other things: A step towards automated pollen recognition in lake sediments. Holocene, 34: 297–305

Traverse A. 2007. Paleopalynology. 2nd ed. Cham: Springer

Treloar W J, Taylor G E, Flenley J R. 2004. Towards automation of palynology 1: Analysis of pollen shape and ornamentation using simple geometric measures, derived from scanning electron microscope images. J Quat Sci, 19: 745–754

Treloar W. 1992. Digital image processing techniques and their application to the automation of palynology. Dissertation for Doctoral Degree. Hull: University of Hull

von Allmen R, Brugger S O, Schleicher K D, Rey F, Gobet E, Courtney Mustaphi C J, Tinner W, Heiri O. 2024. Method development and application of object detection and classification to Quaternary fossil pollen sequences. Quat Sci Rev, 327: 108521

Wasti Q Z, Sabar M F, Farooq A, Khan M U. 2024. Stepping towards pollen DNA metabarcoding: A breakthrough in forensic sciences. Forensic Sci Med Pathol, 20: 1464–1474

Weber R W. 1998. Pollen identification. Ann Allergy Asthma Immunol, 80: 141–148

Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards M E, Lorenzen E D, Vestergård 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, Sønstebø 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

Witte H. 1988. Preliminary research into possibilities of automated pollen counting. Pollen et Spores, 30: 111–124

Wozniak M C, Solmon F, Steiner A L. 2018. Pollen rupture and its impact on precipitation in clean continental conditions. Geophys Res Lett, 45: 7156–7164

Yin H, Chen Y, Xiong J, Xia R, Xie J, Yang K. 2021. An improved local binary pattern method for pollen image classification and recognition. Comput Electrical Eng, 90: 106983

Zeghoud S, Rebiai A, Hemmami H, Ben Seghir B, Elboughdiri N, Ghareba S, Ghernaout D, Abbas N. 2021. ATR-FTIR spectroscopy, HPLC chromatography, and multivariate analysis for controlling bee pollen quality in some algerian regions. ACS Omega, 6: 4878–4887

Zepeda-Mendoza M L, Bohmann K, Carmona Baez A, Gilbert M T P. 2016. DAMe: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. BMC Res Notes, 9: 255

Zimmerman B, Tafintseva V, Bağcıoğlu M, Høegh Berdahl M, Kohler A. 2016. Analysis of allergenic pollen by FTIR microspectroscopy. Anal Chem, 88: 803–811

Zimmermann B. 2010. Characterization of pollen by vibrational spectroscopy. Appl Spectrosc, 64: 1364–1373

Zimmermann B. 2018. Chemical characterization and identification of Pinaceae pollen by infrared microspectroscopy. Planta, 247: 171–180

Zimmermann B, Bağcıoğlu M, Sandt C, Kohler A. 2015. Vibrational microspectroscopy enables chemical characterization of single pollen grains as well as comparative analysis of plant species based on pollen ultrastructure. Planta, 242: 1237–1250

Zimmermann B, Bağcıoğlu M, Tafinstseva V, Kohler A, Ohlson M, Fjellheim S. 2017. A high-throughput FTIR spectroscopy approach to assess adaptive variation in the chemical composition of pollen. Ecol Evol, 7: 10839–10849

Zinger L, Lionnet C, Benoiston A, Donald J, Mercier C, Boyer F, Kembel S. 2021. metabaR: An R package for the evaluation and improvement of DNA metabarcoding data quality. Methods Ecol Evol, 12: 586–592

(责任编委: 赵艳)