

# 博物馆基因组学为解析蜜蜂及其肠道菌的协同演化开辟新路径

李子聪<sup>1,2†</sup>, 郭楚蓉<sup>3†</sup>, 邱立飞<sup>1</sup>, 胡家玮<sup>3</sup>, 姜春燕<sup>4</sup>, 朱朝东<sup>1,5</sup>, 周欣<sup>3</sup>, 陈军<sup>1,4</sup>, 张雪<sup>3\*</sup>, 刘山林<sup>1,5\*</sup>

1. 中国科学院动物研究所, 动物多样性保护与有害动物防控全国重点实验室, 北京 100101

2. 中国科学院大学生命科学学院, 北京 100049

3. 中国农业大学植物保护学院昆虫学系, 北京 100193

4. 中国科学院动物研究所, 国家动物标本馆资源库, 北京 100101

5. 中国科学院动物研究所, 中法生物多样性研究网络, 北京 100101

† 同等贡献

\* 联系人, E-mail: [zhangxue05@cau.edu.cn](mailto:zhangxue05@cau.edu.cn); [shanlin.liu@ioz.ac.cn](mailto:shanlin.liu@ioz.ac.cn)

2026-01-20 收稿, 2026-03-20 修回, 2026-05-20 接受, 2026-05-22 网络版发表

国家自然科学基金(32470455)、中国科学院动物研究所自主项目(202410Z0108, 202310Z0104)和动物多样性保护与有害动物防控全国重点实验室自主部署项目(SKLA2501)资助

**摘要** 博物馆标本作为重要的生物资源宝库, 蕴藏着历史节点的遗传数据, 已被广泛应用于宿主遗传进化分析. 蜜蜂作为重要的传粉昆虫, 其肠道菌群在宿主世代间社会性传播, 与宿主蜂产生紧密的互作关系并长期协同演化, 对从全基因组(hologenome)角度追溯物种演化历程、理解生物对环境变化的适应机制不可或缺. 然而, 如何从历史标本中同时获取宿主蜂和肠道菌群的遗传信息, 其方法和可行性却尚未有报道. 本研究通过对中国科学院动物研究所国家动物标本资源库保藏的16只采自1955~1988年的东方蜜蜂(*Apis cerana*)历史标本进行DNA提取, 系统比较了后足提取与腹部肠道组织提取的DNA总量、降解程度、DNA损伤程度、内源(宿主蜂和肠道核心菌)序列比对率等指标. 结果显示, 两类组织中均可获得可靠的蜜蜂基因组序列, 所获序列的读长集中分布在40 bp左右, 且序列两端均显示出典型的古DNA损伤特征; 与后足相比, 腹部肠道组织额外获取了大量来源于蜜蜂肠道核心菌的序列, 可用于揭示其肠道菌与宿主蜂的协同演化. 同时, 本研究所用的DNA提取方法, 最大限度减少了对标本的损伤. 虽仅能获得痕量DNA, 但通过对同一组织样本进行反复提取, 可以有效提高数据产出, 获得更为充足的有效数据. 综上, 本研究证实蜜蜂腹部肠道组织是利用历史标本同步开展宿主遗传与肠道菌协同演化研究的可靠材料, 此发现既为全基因组水平解析蜜蜂进化提供了新支撑, 也进一步提升了馆藏昆虫标本的科研价值.

**关键词** 东方蜜蜂, 历史标本, 博物馆基因组学, 肠道菌群, 协同演化

人类活动引发的全球变化, 如气候变暖、栖息地变化、环境污染和大量农药施用等, 正严重威胁着地球物种的生物多样性<sup>[1~4]</sup>. 准确量化生物多样性变化并解析其驱动机制已成为当下研究的热点与难点<sup>[5~7]</sup>. 目前, 大量研究依赖于现生样本与短期监测数据, 但这类

数据往往缺乏足够的时间分辨率与历史连续性, 因而难以完整重构生物多样性的长期动态及响应过程. 即便是备受关注的旗舰物种, 能够反映生物多样性动态的连续性样本与信息目前依然非常有限<sup>[8,9]</sup>. 随着古遗传学与高通量测序技术的发展, 各类自然历史博物馆

**引用格式:** 李子聪, 郭楚蓉, 邱立飞, 等. 博物馆基因组学为解析蜜蜂及其肠道菌的协同演化开辟新路径. 科学通报, 2026, 71: 6806~6818

Li Z, Guo C, Qiu L, et al. Museum-omics opens a new avenue for investigating honey bee host-microbiome coevolution (in Chinese). Chin Sci Bull, 2026, 71: 6806~6818, doi: [10.1360/CSB-2026-0099](https://doi.org/10.1360/CSB-2026-0099)

中保藏的不同历史时期的动植物标本,为不同历史时间节点的遗传信息获取和环境影响下生物进化及多样性变化的驱动机制研究提供了宝贵的原始数据来源<sup>[10]</sup>。

自18世纪以来,自然历史的发展跨越了两个多世纪的宏伟历程。这不仅是人类对地球生命多样性认知的缩影,更是自然科学探索的文明结晶。万千博物学家踏遍荒原、横跨重洋,从人迹罕至的原始雨林到波涛汹涌的远洋深海,将海量动植物的形态与生命密码以生物标本的形式尽可能如实保留下来。这些跨越时空的“生命档案”如今被妥善保存在全球各地的科研机构中,如伦敦自然历史博物馆(<https://www.nhm.ac.uk>)、史密森尼国家自然历史博物馆(<https://naturalhistory.si.edu>),以及我国的国家动物标本资源库(<http://museum.ioz.ac.cn>,依托中国科学院动物研究所)和国家植物标本资源库(<http://pe.ibcas.ac.cn>,依托中国科学院植物研究所)。这些标本不仅是物种多样性研究的物质基础,更是研究进化生物学、生物地理学及全球气候变迁的宝库。每一件标本,都记录了特定历史时期的生态信息(包括表型、基因型及共生微生物等多层次信息),为研究人类活动与气候变化对物种的影响、预测物种未来命运提供了关键的时间序列数据<sup>[11~14]</sup>。尽管由于年代久远,标本中的DNA通常高度碎片化,并伴随碱基损伤,如胞嘧啶脱氨生成尿嘧啶<sup>[15,16]</sup>,得益于高通量测序及古DNA提取技术的发展,从博物馆标本中获取可靠的遗传信息已具备充分可行性<sup>[17,18]</sup>,而且历史DNA的碎片化和碱基损伤特征也为标本内源性序列与外源污染的区别提供了关键依据。

昆虫作为植食者、捕食者、寄生者、传粉者、分解者和食腐者,广泛参与地球生态网络的构建与运行,是驱动全球变化的关键生物类群。同时,它们生命周期短、世代更替快,能够迅速响应生态环境中生物与非生物因子的变化,因此也成为监测生物多样性动态的重要指示物种<sup>[19~21]</sup>。蜜蜂是陆地生态系统中最重要的传粉者类群之一,其种群动态与遗传多样性对维持生态系统功能和农业生产具有至关重要的作用<sup>[22]</sup>。近年来,气候变化、农药暴露以及外来蜂群引入等多种人类活动,已对全球蜂类多样性和种群结构施加了显著压力<sup>[23~25]</sup>。基于博物馆历史标本的蜜蜂类群研究揭示了蜜蜂种群的遗传多样性变化及其基因组的演化特征,例如Liu等人<sup>[23]</sup>利用法国历史博物馆距今超过100年历史的东方蜜蜂(*Apis cerana*)标本揭示农药施用是其种

群遗传多样性下降的核心驱动因素。Mikheyev等人<sup>[24]</sup>通过西方蜜蜂(*Apis mellifera*)博物馆标本揭示了瓦螨(*Varroa destructor*)入侵后西方蜜蜂种群发生的快速基因组演化并出现了与抗寄生虫相关的遗传变异。蜜蜂具有保守的肠道菌群,由5~9种特异的肠道菌属组成,对宿主的营养代谢、致病菌抵抗、生长发育及行为调控具有重要作用<sup>[26]</sup>,这些菌属在其他环境中则不存在<sup>[27]</sup>。稳定的肠道菌群组成在携粉足蜜蜂物种分化前已获得,与蜜蜂物种具有超过8千万年的共进化历史<sup>[28]</sup>。在与宿主协同进化的过程中,其保守的肠道菌群在促进宿主物种分化和环境适应中扮演着重要角色<sup>[26,29,30]</sup>。但是,当前基于博物馆标本的时空比较研究多聚焦于分析物种自身遗传信息<sup>[23,31]</sup>,而忽略了共生菌群与宿主蜂在全基因组层面的协同演化。因此,同时获取宿主蜂和肠道共生菌群的基因组,对揭示蜜蜂在全球生态变化下的快速适应机制和驱动因子,深入解析其肠道菌群与宿主的协同演化规律具有重要意义<sup>[23,32]</sup>。

然而,在单个历史标本中同时获取宿主蜂与其肠道菌群的遗传信息的方法和可行性尚缺乏系统性评估。一方面,不同组织来源、保存方式与标本年代等因素,可能导致宿主DNA与肠道菌DNA在含量、完整性、损伤模式和污染水平上存在差异<sup>[33]</sup>,因而难以从同一标本组织同步获取高质量的宿主蜂与肠道菌遗传信息。另一方面,博物馆长期保存过程中易引入环境来源的细菌或真菌污染<sup>[34]</sup>,因此在分析肠道菌DNA时,必须谨慎鉴别其来源,并依据序列特征与损伤模式进行系统评估,以确保分析结果的可靠性<sup>[35]</sup>。

基于上述背景,本文针对20世纪50~60年代的东方蜜蜂历史标本,系统比较了后足与腹部(肠道)两种组织来源DNA的特性,并验证肠道组织在同步获取宿主蜂与共生肠道菌群基因组信息方面的可行性。本研究将为在博物馆蜜蜂历史标本中研究宿主与肠道菌群协同演化提供方法参考。

## 1 材料与方法

### 1.1 标本DNA提取与建库测序

研究选取保存在中国科学院动物研究所国家动物标本资源库的东方蜜蜂(*A. cerana*)干制针插标本。国家动物标本资源库为标本提供统一温湿度恒定(温度18~23℃,平均22℃;相对湿度30%~50%,平均35%)的

馆藏环境,并杜绝皮蠹及霉菌侵害.实验共计包括16头蜜蜂标本:采集地包括北京、福建、广西与云南等,采集时间主要集中于两个时期(1955~1963与1972~1988年;表1).为避免因保存时长引入偏倚,实验设计按采集时间与产地将标本尽量均匀分配为两组,分别选取后足和腹部以进行无损DNA提取.后足和腹部组织的DNA提取分别参照Campos等人<sup>[36]</sup>、Cavill等人<sup>[37]</sup>、Wibowo等人<sup>[37]</sup>和Dabny等人<sup>[38]</sup>,简述操作如下:首先,将分离的组织置于含1 mL无菌水的1.5 mL离心管中,通过反复颠倒清洗去除表面污染物(为避免消毒液可能导致的样本漂白,本步骤仅使用无菌水),随后转移至新的1.5 mL离心管中,重复洗涤3次.随后针对不同组织类型进行差异化降解孵育:对于腹部样本,使用无菌镊子剖取内部肠道组织,加入1 mL CTAB缓冲液和10  $\mu$ L溶菌酶,于37℃振荡孵育1 h;对于后足样本,则加入1 mL消化缓冲液<sup>[36]</sup>,于56℃振荡孵育3 h.首次孵育后,进行如下分步处理:(1)将残余组织转移至等量新鲜裂解缓冲液中,于相同条件下进行第二次孵育以充分提取DNA.(2)向首次孵育上清液中加入12.5  $\mu$ L

蛋白酶K,并于56℃继续孵育3 h以彻底消化蛋白质.待蛋白质消化完成后,将裂解液与13 ml PB Buffer充分混匀,采用二氧化硅吸附法(MinElute PCR Purification Kit (50), Qiagen, 德国)进行DNA结合、洗涤与洗脱.二次提取则将二次孵育的组织裂解液按上述方法再次进行DNA纯化,最终获得同一组织来源的两份独立DNA样品.为保持标本形态完整性,提取过程中涉及的几丁质外壳及表皮附属结构均在自然条件下风干后归置于原标本保存(图1).使用Qubit(Qubit 1X dsDNA HS, Thermo Fisher Scientific, 美国)测定DNA浓度,并采用Wilcoxon秩和检验评估组织类型及提取批次对提取DNA浓度的影响.初提与复提的DNA产物参考SCR protocol<sup>[39]</sup>单独构建DNA测序文库.DNA提取与测序文库构建均在超净实验室完成,全过程严格执行标准化操作规范,确保实验数据的真实与可靠.DNA文库利用核酸分析仪(Agilent 2100)定量后均量混合,测序工作由北京优吉科技有限公司(中国)完成,采用因美纳X-plus平台双末端150 bp(PE150)测序策略.原始测序数据已上传至Science Data Bank(<https://www.scidb.cn>)并公开,

表 1 样本信息表

Table 1 Sample information in this study

| 样本名称 | 馆藏编号   | 采集地   | 采集年份 | 组织类型 | DNA浓度(ng/ $\mu$ L) <sup>a)</sup> |                 | 原始序列数 <sup>b)</sup> |           | 宿主序列比例 <sup>c)</sup> |        | 重复序列比例 <sup>d)</sup> |        |
|------|--------|-------|------|------|----------------------------------|-----------------|---------------------|-----------|----------------------|--------|----------------------|--------|
|      |        |       |      |      | F <sup>e)</sup>                  | R <sup>e)</sup> | F                   | R         | F                    | R      | F                    | R      |
| F02  | 647912 | 北京海淀  | 1959 | 腹部   | 0.166                            | <0.05           | 90244887            | 87445374  | 48.02%               | 43.83% | 14.07%               | 12.10% |
| F04  | 649055 | 福建武夷山 | 1960 | 腹部   | 1.04                             | 0.063           | 315906012           | 155949918 | 0.85%                | 3.19%  | 0.11%                | 1.70%  |
| F07  | 649021 | 广西桂林  | 1974 | 腹部   | <0.05                            | <0.05           | 111419525           | 304380702 | 1.04%                | 0.36%  | 0.37%                | 0.20%  |
| F10  | 648837 | 云南芒市  | 1980 | 腹部   | 0.168                            | 0.106           | 113864221           | 104984158 | 30.05%               | 6.71%  | 12.82%               | 2.36%  |
| F15  | 649280 | 广西龙胜  | 1963 | 腹部   | 0.168                            | <0.05           | 63911266            | 120011131 | 13.07%               | 2.70%  | 2.87%                | 0.53%  |
| F16  | 678635 | 四川乡城  | 1982 | 腹部   | 1.15                             | <0.05           | 407354177           | 114621786 | 60.82%               | 37.08% | 13.63%               | 10.06% |
| F18  | 678281 | 西藏察雅  | 1976 | 腹部   | 0.895                            | <0.05           | 453282077           | 82468425  | 58.25%               | 35.04% | 9.80%                | 5.19%  |
| F23  | 619490 | 广西隆林  | 1959 | 腹部   | 0.348                            | <0.05           | 91538696            | 145553441 | 5.17%                | 0.88%  | 1.60%                | 0.33%  |
| T06  | 649424 | 云南丽江  | 1962 | 后足   | <0.05                            | <0.05           | 98104531            | 120693439 | 29.10%               | 1.23%  | 6.47%                | 0.32%  |
| T08  | 648815 | 福建厦门  | 1982 | 后足   | 0.065                            | <0.05           | 240818096           | 135111050 | 16.56%               | 39.79% | 9.02%                | 14.54% |
| T10  | 649336 | 北京延庆  | 1972 | 后足   | 0.254                            | 0.412           | 239222817           | 278557763 | 0.23%                | 0.30%  | 0.12%                | 0.12%  |
| T12  | 619947 | 上海    | 1955 | 后足   | 0.163                            | <0.05           | 98482098            | 127129186 | 8.41%                | 3.37%  | 2.02%                | 0.83%  |
| T13  | 647591 | 湖南沅陵  | 1959 | 后足   | <0.05                            | <0.05           | 79028035            | 71229838  | 13.44%               | 2.01%  | 3.85%                | 0.40%  |
| T15  | 648772 | 四川马尔康 | 1961 | 后足   | <0.05                            | <0.05           | 102757094           | 159286591 | 15.13%               | 9.12%  | 2.59%                | 1.90%  |
| T21  | 649011 | 贵州江口  | 1988 | 后足   | 0.125                            | <0.05           | 94648131            | 132423859 | 41.18%               | 36.33% | 6.36%                | 7.06%  |
| T22  | 649337 | 北京延庆  | 1975 | 后足   | 0.063                            | <0.05           | 101857606           | 126949111 | 11.62%               | 12.61% | 3.14%                | 4.48%  |

a) DNA浓度< 0.05 ng/ $\mu$ L表示样本浓度低于检出限; b) 原始序列指双端测序数据中正向测序的序列数; c) 宿主内源序列比例指原始序列中成功比对至参考基因组的序列比例(含重复序列); d) 重复序列比例指比对至参考基因组的长度一致,且在基因组上起始位置完全重合序列的占比; e) F和R分别代表首次组织裂解和二次裂解(深色背景)提取DNA对应结果

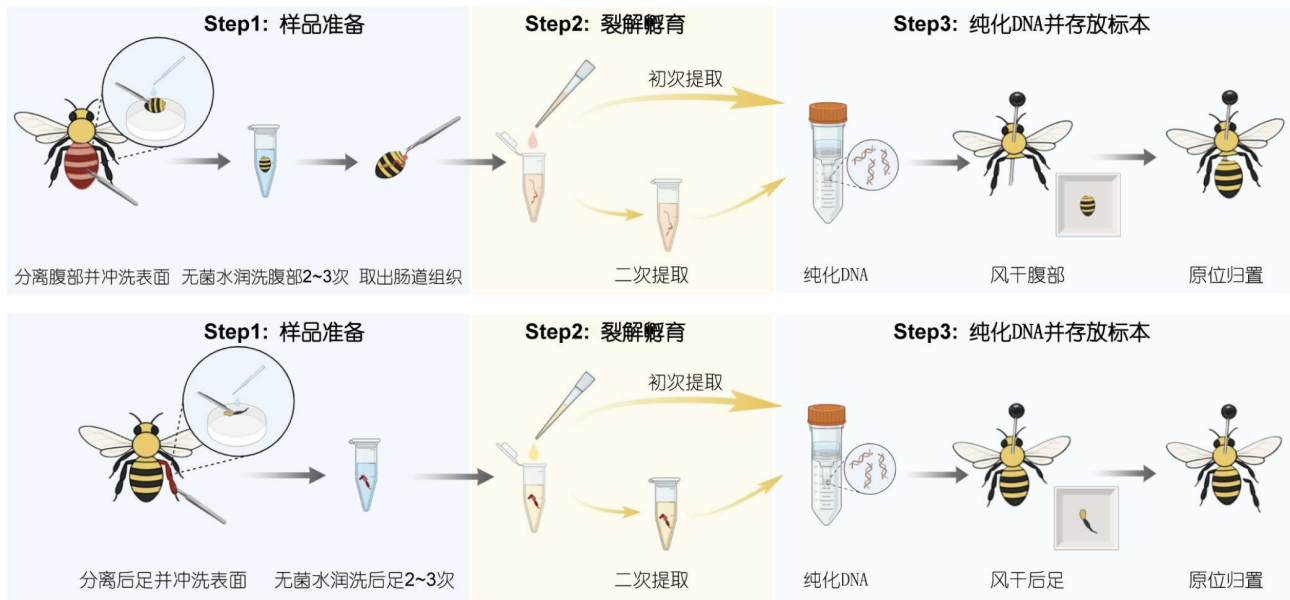


**图 1** (网络版彩色)蜜蜂博物馆标本取样与DNA提取流程图. 首先, 从博物馆馆藏的蜜蜂标本中解剖取出腹部肠道或后足组织经清洗后进行差异化孵育. 孵育结束后采用硅柱法提取并纯化DNA, 将提取得到的DNA  $-20^{\circ}\text{C}$  保存以备后续测序; 剩余标本组织经晾干并与对应的原始标本一并放入库归档, 作为溯源凭证. 此图使用BioRender.com生成

**Figure 1** (Color online) Workflow for sampling and DNA extraction from museum bee specimens. Abdominal gut tissue and a metathoracic leg were dissected from the bee specimen, cleaned, and subjected to differential incubation strategies. Following incubation, DNA was extracted and purified from the lysates using a silica column-based method. The extracted DNA was stored at  $-20^{\circ}\text{C}$  for subsequent sequencing. The remaining tissue was air-dried and archived together with the corresponding original specimen as a traceability voucher. This figure was created with BioRender.com

数据DOI为10.57760/sciencedb.zoology.00012.

## 1.2 测序质量评估

我们首先使用Seqkit v2.9.0<sup>[40]</sup>统计原始测序序列数目, 随后采用AdapterRemoval v2.3.1<sup>[41]</sup>去除测序接头并合并双端序列. 为进一步提升数据质量, 使用Fastp v0.24.0<sup>[42]</sup>去除序列末端的技术性来源多聚核苷酸序列(poly序列), 最终仅保留长度 $\geq 30$  bp且碱基质量值 $\geq 15$ 的序列, 并统计质控后保留数据的序列长度分布特征. 使用BWA-ALN v0.7.17<sup>[43]</sup>将质控数据比对到*A. cerana cerana*参考基因组(未发表, 基因组相关参数见表S1). 为适应历史DNA片段化严重、末端易损伤的特点, 比对中设置参数 $-l\ 1024$ 以关闭seed延伸, 从而提高对短序列、末端损伤与测序错误的容忍度. 在此基础上, 分别使用Samtools v1.21<sup>[44]</sup>的flagstats统计序列比对率, Samtools sort对比对结果排序, Picard v3.2.0(<https://github.com/broadinstitute/picard>)标记PCR重复、Perl脚本(reE-valueRdp.pl, <https://github.com/liushanlin/rhinoceros-comparative-genome>)<sup>[18]</sup>对Picard可能存在的错误标记进行矫正. 最后通过Samtools coverage计算基因组的平

均覆盖度与测序深度, 利用Preseq<sup>[45]</sup>估算两次提取产物理论可达到的测序深度, 并采用Mapdamage v2.2.2<sup>[46]</sup>检测DNA序列的末端损伤信号.

## 1.3 腹部组织肠道菌序列评估

我们将比对到参考基因组的宿主序列去除后, 使用Samtools将剩余非宿主蜂序列转换为FASTA格式. 随后, 构建了一个包含600株已知蜜蜂肠道核心菌基因组的本地参考数据库(详见表S2), 并采用Kraken v2.1.3<sup>[47]</sup>(K-mer=29)进行两轮分类注释: 首轮注释中, 使用该本地数据库对数据进行初步筛选, 以富集候选菌群序列并排除大量非目标类群, 从而提升后续分析效率; 第二轮注释则基于Kraken公共数据库(包括古菌、细菌和病毒, 2025年11月30日下载)对首轮获得的候选序列进行严格注释(置信度阈值confidence=0.9)以进一步降低假阳性. 具体判定规则如下: 第二轮注释中直接被归类为已知蜜蜂肠道菌的序列予以保留, 而明确注释为其他非蜜蜂肠道微生物的序列则直接去除. 对于被注释到科、目等上级分类单元或未注释的序列, 我们依据其K-mer注释分布进行如下判定: 1) 上级分类

单元注释的序列: 若该序列的科、目不属于已知蜜蜂肠道菌类群对应的科目, 则直接剔除, 若属于, 则进一步检查其K-mer注释中归属于蜜蜂肠道菌类群的比例. 该比例 $\geq 50\%$ 时保留, 否则剔除, 以避免假阳性; 2) 未注释序列: 若序列中未被注释的K-mer比例 $\geq 50\%$ , 则判定为无法分类并予以排除, 否则, 进一步统计已注释K-mer中属于蜜蜂肠道菌(种/属水平)的比例, 若该比例 $\geq 50\%$ , 则将序列归入肠道菌, 否则同样予以排除(图2). 随后, 针对各样本获得的肠道核心菌序列进行统计, 筛选出每个腹部样本中相对丰度最高的肠道菌物种, 并将其序列比对至相应参考基因组, 进而分析比对序列的长度特征及损伤模式. 最后, 在属级水平上分析肠道核心菌群多样性.

## 2 结果

### 2.1 标本DNA提取浓度与序列特征

我们首先比较了蜜蜂历史标本不同部位及不同样

本间DNA提取量的差异. DNA定量结果显示, 除少数腹部肠道标本可提取获得 $\sim 50$  ng的DNA总量(浓度 $\sim 1$  ng/ $\mu$ L), 几乎所有样品都只能提取到痕量DNA. 大部分样品的单次获取DNA总量在5 ng左右(浓度 $\sim 0.1$  ng/ $\mu$ L), 很多样品获取的DNA浓度甚至低于检测标线, 无法准确定量(表1). 腹部肠道组织( $0.49 \pm 0.42$ , mean $\pm$ SD,  $n=8$ )能够获得显著高于后足( $0.10 \pm 0.07$ , mean $\pm$ SD,  $n=8$ )的DNA总量(图3(a),  $W=54.50$ ,  $P=0.02$ ). 该结果符合两者对应的可裂解肌肉、脂肪组织量.

进一步比较同一组织的首次提取与二次提取的DNA总量. 结果显示, 几乎所有样本二次提取获取的DNA总量均低于首次提取(图3(b),  $W=206.50$ ,  $P<0.01$ ). 统计能够定量(浓度 $>0.05$  ng/ $\mu$ L)的DNA产物样品, 首次提取和二次提取的DNA总量分别占88.80%和11.20%, 表明首次组织裂解提取能够获得样品的大部分DNA. 然而, 考虑到博物馆标本的稀缺性, 通过组织进行二次裂解以获取更多痕量DNA仍具有重要的尝试价值.

我们对所有样品两次提取的DNA(包括可准确定

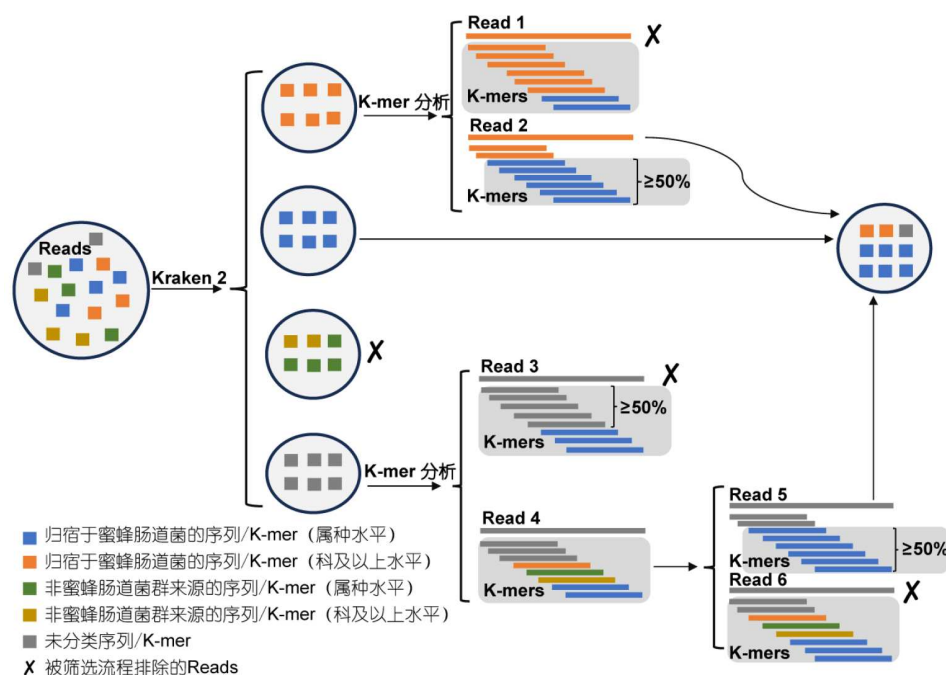


图2 (网络版彩色) 基于K-mer分析的序列分类判定流程. 该流程对初步分类结果进行验证与筛选: 对直接注释到蜜蜂肠道菌群属或种水平的序列予以保留; 注释为其他微生物属或种的序列则直接排除. 对仅能注释到更高分类阶元(如科、目等)或未能注释的序列, 进一步通过K-mer分析进行判定, 仅保留其中蜜蜂肠道菌群来源K-mer比例 $\geq 50\%$ 的序列

Figure 2 (Color online) Sequence classification based on K-mer analysis. It validates and filters preliminary classification results as follows: sequences directly annotated to the genus or species level of honey bee gut microbiota are retained, whereas those assigned to other microbes (genus or species level) are excluded. For sequences that can only be annotated to higher taxonomic ranks (e.g., family, order) or remain unannotated, a K-mer-based analysis is further applied, retaining only those sequences in which  $\geq 50\%$  of the K-mers originate from honey bee gut microbiota

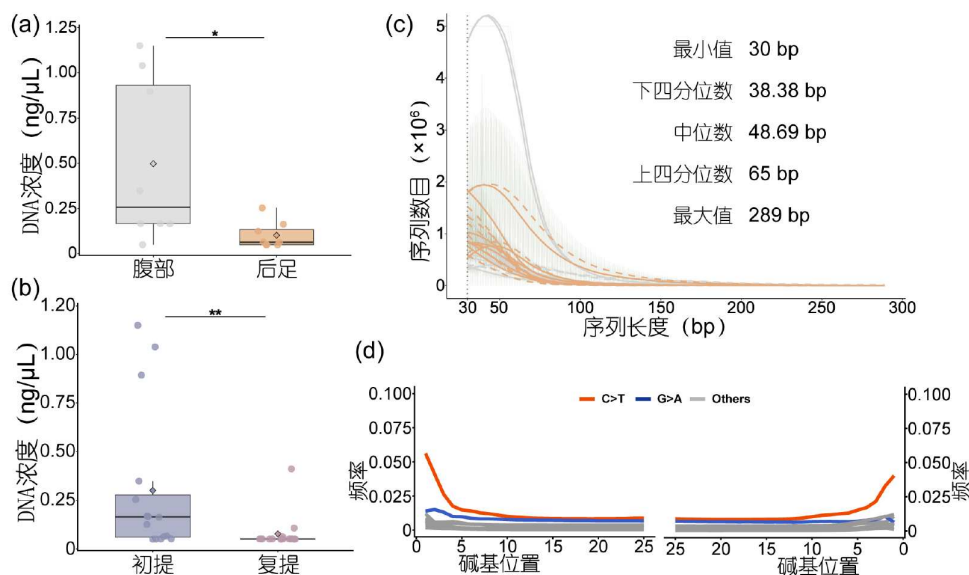


**图 3** (网络版彩色)DNA产量、片段长度分布及典型末端损伤特征。(a) 腹部与后足组织初次DNA提取浓度的分布(仅展示初提)。\*表示组间差异显著( $P<0.05$ )。(b) 初次提取与复提DNA浓度的比较。\*\*表示组间差异极显著( $P<0.01$ )。(c) 所有样本测序片段长度分布, 豆绿色折线为序列长度真实数目分布, 其他为趋势曲线(灰色: 腹部样本, 棕色: 后足样本; 实线: 初提, 虚线: 复提)。(d) 片段末端替换频率(F02样本, 左侧5'端, 右侧3'端)红色曲线表示C→T替换率, 蓝色曲线表示G→A替换率, 灰色曲线表示其他替换类型, 如A→G(因比对采用双端合并序列, 因此5'和3'端均为C→T的DNA损伤模式)

**Figure 3** (Color online) DNA yield, fragment length distributions, and characteristic terminal damage patterns. (a) DNA concentration distributions from initial extractions of abdominal and hind leg tissues (only initial extractions are shown). \*, Significant difference at  $P<0.05$ . (b) Comparison of DNA concentrations between the initial and re-extraction steps. \*\*, Highly significant difference at  $P<0.01$ . (c) Sequence length distribution of all samples. The dark green line shows the actual distribution of sequence lengths, and the other lines indicate trend curves (gray: abdominal tissues; brown: hind leg tissues; solid line: first extraction; dashed line: re-extraction). (d) Terminal nucleotide substitution rates at fragment ends from the F02 sample (5' ends on the left, 3' ends on the right of the fragments). Red curves represent C-to-T substitution rates, blue curves represent G-to-A substitution rates, and gray curves represent all other substitution patterns, for example, A-to-G (since paired-end reads were overlapped and merged before being used for sequence alignment, C-to-T DNA damage patterns were observed at both the 5' and 3' termini)

量的15个样本, 浓度 $>0.05$  ng/μL, 以及无法定量的17个样本, 浓度 $<0.05$  ng/μL)分别构建了测序文库并进行高通量测序。共获得4969235041条原始序列。通过数据拆分将序列划归到其对应的样品, 发现即便是无法准确定量的极低浓度样本, 通过单链建库技术仍能获得有效的测序数据, 总测序量中2118613118条(占比42.63%)来源于DNA浓度低于检出限的样本提取产物。但该部分测序数据的序列质量、内源序列比例等还需要进一步验证(见2.2节)。

我们对原始数据进行质控, 并分析序列的长度分布特征和损伤模式。原始序列经过去除接头、合并正反反向测序读段、修剪技术性来源末端poly序列、并去除长度小于30 bp的序列后, 共获得2643769927条序列(占原始序列的53.20%)。这些序列的长度集中分布于39~49 bp(图3(c)), 表明产出的测序数据以短片段为主。该结果符合博物馆标本DNA的基本特征, 即长期室温干燥保存条件导致DNA片段发生严重降解并断裂, 从

而形成短片段<sup>[16]</sup>。另外, 历史DNA序列两端断裂处易发生胞嘧啶(C)自发脱氨基生成尿嘧啶(U), 从而在测序扩增中产生链特异性损伤模式: 正向链5'端的C→T, 及其互补链对应的3'端的G→A的损伤突变模式。因分析采用双端合并序列进行比对, 因此两端均应呈现末端C→T的突变, 分析结果显示测序序列两端具有C→T突变的梯度富集, 表明相关序列来自历史标本而非现生样本的环境污染(图3(d), 图S1)。综上, 序列的长度分布与末端碱基损伤模式共同支持所获序列确实来源于历史馆藏标本, 为后续分析提供了可靠的数据基础。

## 2.2 宿主蜂内源性序列含量与基因组覆盖度特征

我们对后足和腹部肠道组织的测序数据进行质控后, 统计了序列中来源于东方蜜蜂宿主的内源性DNA比例。其中, 将可比对至蜜蜂参考基因组的序列定义为宿主内源序列(简称宿主蜂内源序列或宿主蜂内源性DNA), 以区别于核心菌群的内源序列(简称核心菌内源

序列或核心菌内源性DNA)。结果显示, 宿主蜂内源DNA比例在样品间差异较大, 最高为60.82%, 最低仅0.23%。整体而言, 除少数样品(F04、T08和T22)外, 二次裂解释放的DNA中内源性DNA比例普遍低于首次裂解(表1; 图4(a))。腹部肠道与后足间的宿主蜂内源性DNA比例无显著差异( $W=37$ ,  $P=0.65$ ; 图4(b))。后足样本来源的DNA中, 除T10样品宿主蜂内源性DNA占比较低( $<1\%$ )外, 其余样本的宿主蜂内源性DNA比例差异较小( $SD=0.13$ )。相比而言, 尽管宿主蜂内源性DNA比例最高的3个样品均来自腹部肠道组织, 但该组织整体表现出更高的异质性( $SD=0.25$ ), 其中三个样品的宿主蜂内源性DNA比例明显低于或接近5%, 这些非宿主蜂内源序列是否来源于蜜蜂肠道核心菌, 需进一步分析(见2.3节)。

我们对同一样本两次提取获得的宿主蜂内源序列中重复序列的比例进行了比较, 以评估其对测序结果的影响。总体上, 多数腹部肠道样本在二次提取中重复序列占比呈下降趋势, 而后足样本则未呈现一致的变化模式(表1)。配对 $t$ 检验结果显示, 两种组织中两次提取间的重复序列占比差异均未达到统计学显著水平(腹部:  $P=0.06$ , 后足:  $P=0.70$ ; 图S2)。进一步选择在两次提取中均获得较高测序深度的样本(F02与T21)进行文库复杂度分析。结果表明, 除F02的二次提取样品外, 其余样品的理论可测得深度在约25亿条序列处趋于平缓(图4(c))。整体而言, 基于首次提取所构建文库的可测得深度比二次提取高出约2 $\times$ 。

我们以基因组覆盖度与测序深度为指标, 评估了博物馆历史标本用于获取有效宿主基因组信息的效率。结果显示, 相较于现生样本, 历史标本通常显著更高的测序深度才能达到相当的基因组覆盖水平。具体而言, 在约5 $\times$ 测序深度下, 多数历史样本的实际覆盖度明显偏离基于泊松分布的理论预期值(图4(d)), 反映出历史样本在文库复杂度和覆盖均匀性方面存在固有局限性。

## 2.3 肠道菌序列含量和特征

为评估从历史蜜蜂标本腹部肠道组织中同步获取宿主及其共生肠道菌群基因组信息的可行性, 我们通过比对实验室自建的蜜蜂肠道核心菌数据库<sup>[48]</sup>, 对蜜蜂肠道核心菌来源序列的占比和序列特征进行了分析。通过Kraken2注释筛选, 腹部样本均检出较高比例的蜜蜂肠道核心菌来源序列(图5(a); 表S3), 涵盖了东方蜜蜂的主要核心肠道菌属, 包括*Lactobacillus*、*Bifidobac-*

*terium*、*Gilliamella*、*Apibacter*、*Snodgrassella*, 以及蜂巢环境的核心菌属*Apilactobacillus*。而后足样品中肠道核心菌的注释比例极低(占比均低于0.1%; 表S4), 支持了腹部肠道组织样品中肠道微生物来源序列的可靠性。在宿主蜂内源性DNA占比较低的F04、F07和F23三个样本中(表1), 注释到肠道菌的序列比例同样较低, 但不具有显著性。这一结果在排除F15极端值后, 宿主蜂内源性DNA占比与肠道菌序列占比呈显著正相关(Pearson correlation:  $r=0.92$ ,  $P<0.01$ ; 图5(b)), 表明宿主蜂内源性DNA比例下降并非由肠道菌DNA的过度富集造成, 而更可能与样本个体保存质量差异有关。以上结果说明, 从历史标本腹部肠道组织中可同步获取宿主蜂与肠道微生物DNA, 两者在数据量上不存在此消彼长的竞争关系。

此外, 我们注意到, 虽然样本F04和F23的DNA浓度和产量明显偏高, 后足样本T10也呈现类似趋势, 但这三例样本的宿主蜂与肠道核心菌内源性序列占比均较低, 提示其较高DNA总量可能来源于其他外源DNA污染。进一步分析这三个样本的非内源测序数据(排除比对到宿主蜂和肠道核心菌的序列), 其序列长度也主要集中于30~50 bp(图S3), 与其他样本的长度特征一致, 提示历史性的外源污染可能是导致内源宿主及肠道菌DNA比例偏低的主要原因。

为进一步评估腹部组织中肠道核心菌的DNA序列源自历史标本, 而非现生菌的DNA污染, 我们从各腹部样本中选取注释丰度最高的代表肠道菌, 将其序列分别比对至对应参考基因组(*Apilactobacillus kunkeei*: NZ\_CP080568, *Lactobacillus panisapium*: NZ\_CP048268, *Bifidobacterium asteroides*: NZ\_CP017696)。长度分析显示, 比对至核心菌的序列的长度分布呈现明显的短片段富集特征, 在35~50 bp区间高度集中(图5(c)), 与宿主内源性DNA长度分布一致, 符合历史DNA的片段降解特征。以F02样本中注释到的*Lactobacillus panisapium*菌种为代表肠道菌, 评估其DNA片段的末端损伤程度, 结果鉴定到典型的末端碱基替换模式(图5(d)), 证实该序列为历史样本内源DNA, 而非现生污染。

最后, 我们对腹部肠道组织DNA中鉴定到的蜜蜂肠道核心菌进行了丰度与组成分析。结果显示, 所有样本中均可检测到肠道核心菌, 且总体由少数优势菌属主导: *Lactobacillus*(乳杆菌属)、*Bifidobacterium*(双歧杆菌属)和*Gilliamella*的累计相对丰度最高, 而*Snod-*

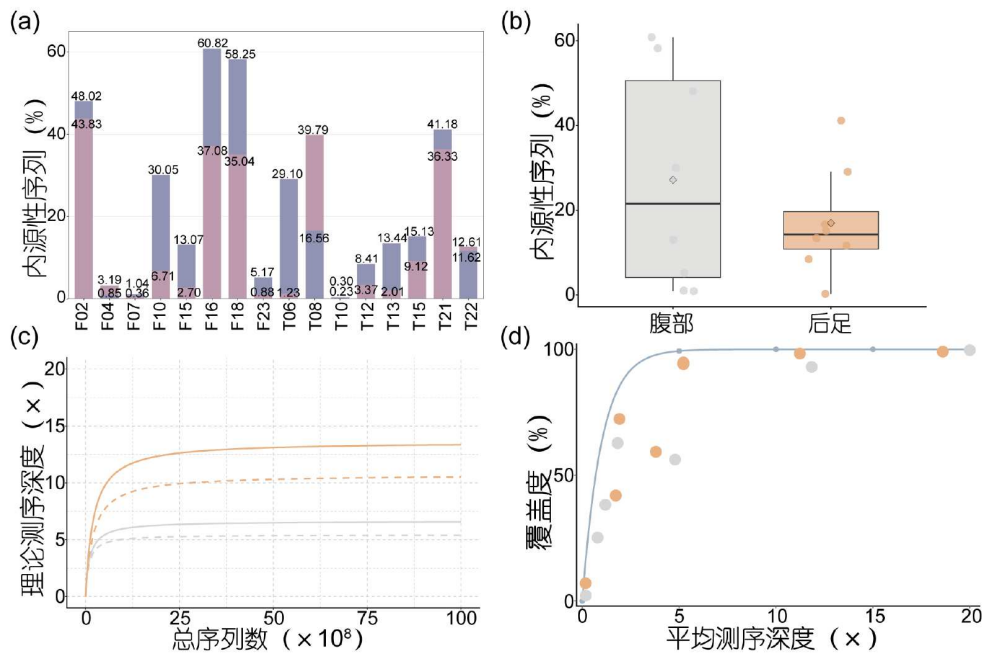


**图 4** (网络版彩色)宿主蜂内源性序列含量与基因组覆盖度特征。(a) 各样本初次提取(蓝色)与复提(紫色)的内源性序列含量。(b) 样本(初提+复提)内源性序列含量( $W=37$ ,  $P=0.65$ )。(c) 基于现有数据对理论测序深度随总测序量增加的饱和曲线估计(棕色: T21, 灰色: F02; 实线: 初提, 虚线: 复提)。理论测序深度=(预期独特序列数 $\times$ 平均序列长度/基因组大小) $\times$ 内源性序列比例(平均序列长度=40 bp, 基因组=217.69 Mb, 去重后内源性序列比例(初体/复提): F02=0.34/0.32, T21=0.35/0.29)。(d) 覆盖度与平均测序深度(深度 $>20\times$ 的样本统一设定为 $20\times$ 以便于显示; 棕色: 后足, 灰色: 腹部; 平滑曲线为理论饱和曲线)

**Figure 4** (Color online) Endogenous sequence content and genome-coverage characteristics. (a) Endogenous sequence content for each sample in the initial extraction (blue) and the re-extraction (purple). (b) Endogenous sequence content per sample after combining data from both extractions (initial +re-extraction;  $W=37$ ,  $P=0.65$ ). (c) Estimated saturation curves showing theoretical sequencing depth as total sequencing output increases, based on available data (brown: T21; gray: F02; solid lines: initial extraction; dashed lines: re-extraction). Theoretical sequencing depth was calculated as (expected number of unique reads  $\times$  mean read length/genome size)  $\times$  endogenous proportion (mean read length=40 bp; genome size=217.69 Mb). Deduplicated endogenous proportions (initial/re-extraction) were: F02=0.34/0.32 and T21=0.35/0.29. (d) Relationship between genome coverage and mean sequencing depth (depth  $>20\times$  was capped at  $20\times$ ; brown: hind leg; gray: abdomen). The smooth curve shows the theoretical saturation trend

*grassella*的丰度则相对较低(图5(e)). 同时, 不同个体间的核心菌属种组成与相对丰度存在明显差异(图5(f)), 提示群落结构在个体层面具有较高的异质性。然而由于样本DNA提取质量、内源DNA比例及测序深度等因素的影响, 观察到的异质性是否反映真实的菌群多样性变化, 仍需后续通过更多针插标本及增加测序深度等手段进一步解析。

### 3 讨论

本研究以博物馆馆藏昆虫历史干制针插标本为研究对象, 将腹部和后足分别进行DNA提取与高通量测序, 比较了两种组织在DNA回收效率、测序数据产出、内源DNA比例等方面的差异。结果表明, 腹部组织通常能回收更多DNA, 但是, 我们也注意到腹部肠道组织DNA总量最高的几个样品, 其目标DNA(包括蜜蜂自身

以及肠道微生物)的比例却极低。相反, 多个DNA浓度和总量极低的样品, 其最终的目标DNA反而达到了平均水平。因此, 本研究结果提示DNA得率无法作为标本是否适合进行后续文库构建与测序的有效指标。昆虫后足组织主要由厚实的角质层和韧带等构成, 其中的DNA被角蛋白和多糖牢固包裹, 导致降解不充分、提取效率较低<sup>[49,50]</sup>。而标本腹部获取的肠道组织能与裂解液充分接触, 从而能够促使更充分的DNA释放和获取。同时, 我们关注到测序结果中非宿主蜂和肠道微生物来源的序列具有类似的短片段特征, 其更可能来自历史性的外源污染, 因此推测腹部组织出现大量非目标DNA的原因可能与标本制备过程有关: 昆虫腹部结构复杂, 含有多种腺体, 形成了独特的微环境。在组织降解时, 这一环境可能更利于非肠道核心菌或机会性微生物的定植与扩增。然而, 这一解释目前尚属推

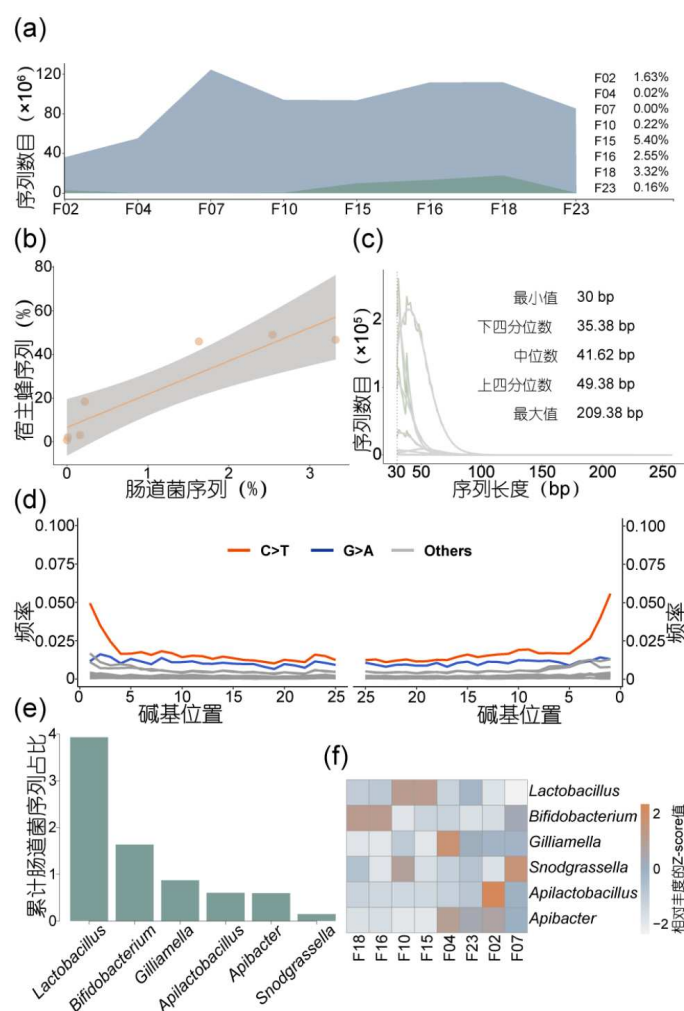


**图 5** (网络版彩色)历史蜜蜂标本中肠道微生物序列特征与群落组成特征. (a) 腹部样本的非中蜂序列与肠道菌序列数(浅灰蓝色: 非中蜂序列; 灰绿色: 肠道菌序列). (b) 样本宿主蜂序列占比与肠道菌序列占比的相关性(Pearson correlation:  $r=0.92$ ,  $P=0.0037$ ). (c) 比对至肠道菌参考基因组的序列片段长度分布(豆绿色折线为序列长度的真实计数分布, 灰色曲线为对应的趋势拟合曲线). (d) F02乳酸杆菌(*Lactobacillus panisapium*) DNA损伤模式分析片段末端替换频率(F02样本, 左侧5'端, 右侧3'端). 红色曲线表示C→T替换率, 蓝色曲线表示G→A替换率, 灰色曲线表示其他替换类型, 如A→G(因比对采用双端合并序列, 因此5'和3'端均为C→T的DNA损伤模式). (e) 在八个腹部肠道样本中累计肠道核心菌序列占比前6的属以及其累计序列占比. (f) 热图显示在至少一个样本中相对丰度>5%的6个细菌属的相对丰度分布. 行(属)按累计丰度从高到低排序, 高丰度属位于顶部. 数据经过行标准化(Z-score转换), 颜色表示每个高丰度属在不同样本中的相对变化: 红色表示该属在对应样本中的丰度高于其平均值, 蓝色表示低于平均值

**Figure 5** (Color online) Characteristics of gut microbial sequences and community composition in historical bee specimens. (a) Number of non-*Apis cerana* sequences and gut microbial sequences in abdominal samples (dusty blue: non-*A. cerana* sequences; sage green: gut microbial sequences). (b) Correlation between honey bee-derived sequence content and gut microbial sequence content across samples (Pearson correlation:  $r=0.92$ ,  $P=0.0037$ ). (c) Length distribution of sequencing reads mapped to the gut microbial reference genomes (the dark green line represents the actual count distribution of sequence lengths, and the gray curve indicates the corresponding fitted trend). (d) Terminal substitution frequencies in DNA damage pattern analysis for *Lactobacillus panisapium* F02. Red curves represent C-to-T substitution rates, blue curves represent G-to-A substitution rates, and gray curves represent all other substitution types, for example, A-to-G (since paired-end merged reads were used for sequence alignment, C-to-T DNA damage patterns were observed at both the 5' and 3' termini.). (e) Cumulative abundance of the top six genera across eight specimens. (f) Heatmap showing the relative abundance distribution of six bacterial genera with relative abundance >5% in at least one specimen. Rows (genera) are sorted by cumulative abundance in descending order, with high-abundance genera at the top. Data are row-scaled (Z-score transformed), with colors representing the relative variation of each high-abundance genus across specimens: red indicates abundance above the mean for that genus, and blue indicates abundance below the mean

测, 未来需要更系统的对照实验来阐明其背后机制.

目前已有多项方法学研究为提升历史标本DNA回

收率提供了技术支撑, 如Patzold等人<sup>[51]</sup>开发的低损伤提取方案, 使得保存历史达241年的昆虫标本的DNA提

取量平均提升17.3倍; Holmquist等人<sup>[52]</sup>提出了一种基于SPRI磁珠的高通量低成本DNA提取流程,该方法在显著降低单位耗材成本的同时,仍能实现与常用商业试剂盒可比的DNA回收率与产量; Korlević等人<sup>[53]</sup>则通过标本再水化、温和蛋白酶K裂解及关键点干燥处理等组合策略,实现了DNA最大回收与标本损伤最小化平衡。此外,本研究与已有文献<sup>[49]</sup>的结果均证实,对同一样本进行重复DNA提取时,即便单次获得的DNA浓度较低,多次提取的产物仍能贡献大量有效测序序列,因此多轮提取是标本DNA研究中的有效技术策略。

蜜蜂后足为高度特化的采集与清洁工具,在活动中频繁与体表接触,易沾染体表残留的排泄物或肠道内容物,即便对样品进行表面消毒,少量的粪便或肠道残留仍可能被带入DNA提取流程。相比之下,腹部作为肠道的载体,通常保留丰富的肠道菌群DNA,因此在后足样本中检出的肠道细菌序列更可能源于蜜蜂的生物学特性<sup>[27,54]</sup>。需要注意的是,不同博物馆标本的保存条件存在显著差异,包括如保存前处理方式、储存环境温度波动及标本采集年代等,这些因素对肠道微生物群落检测结果的影响呈现复杂性<sup>[55]</sup>。一项涵盖鱼类、两栖类、爬行类及哺乳类的跨类群研究表明,福尔马林或乙醇会改变检出的宿主肠道微生物的组成,并导致 $\alpha$ 多样性降低,但宿主物种间固有的微生物群落结构差异仍能被有效识别<sup>[51]</sup>。此外,博物馆标本中普遍存在保存环境或实验室来源的外源微生物DNA污染<sup>[56,57]</sup>,因此在下游数据分析过程中,必须通过严格的筛选和判断流程剔除这类非目标序列。蜜蜂肠道菌群在属水平上具有高度保守性,不受地理差异影响<sup>[58]</sup>,但在种及菌株水平存在差异<sup>[59,60]</sup>。然而,受样本个体保存条件等

因素干扰,尽管所有样本均能检测到全部核心菌属,但部分核心菌在不同个体中的丰度极低,严重偏离现生蜜蜂肠道菌群组成。因此,基于现有博物馆DNA样本的分析结果,尚无法解析地理、环境等差异对蜜蜂肠道菌群组成的影响。这一瓶颈有待后续通过扩大样本量、提升测序深度等策略加以突破。综上,通过优化提取体系、建立严格的外源污染排除等技术手段,博物馆昆虫标本可为基因组学和微生物组学研究提供具有重要价值的遗传信息<sup>[61]</sup>。

综合本研究结果可知,利用昆虫干制针插标本的腹部肠道组织,不仅能够获取物种自身的高深度基因组数据,还能同步解析肠道菌群的遗传信息,因此在昆虫的博物馆基因组学研究中具有显著优势。但需要注意,不同于昆虫具有成对的六足,其腹部内的肠道组织仅有一份。虽然取样过程几乎不影响标本的形态特征,该部分所含的遗传信息具有唯一性,一旦消耗无法重复获取。同时已有研究证实,仅取单条后足即可获得高质量的基因组数据<sup>[23,49,62]</sup>,因此博物馆昆虫样本研究中最优取材部位需紧密结合具体研究目标。若研究核心目标为获取高深度宿主基因组数据的同时解析肠道微生物组特征,建议优先选择腹部组织,并保留重复提取流程,以最大化挖掘历史样本的基因组及微生物组信息;若研究仅以宿主基因组数据获取为核心,且需将标本损伤控制在最低水平,则后足组织是更适宜的选择。本研究仍存在一定局限性,样本量及物种代表性相对有限,且标本保存历史对结果的潜在影响干扰尚未实现全控制。这些问题需在后续研究中通过扩大样本覆盖范围、细化保存条件分组设计等方式进一步验证与优化,以期为博物馆昆虫历史DNA研究提供更具普适性的技术参考。

**致谢** 感谢中国科学院动物研究所国家动物标本资源库在标本使用方面提供的支持。

## 参考文献

- 1 Keck F, Peller T, Alther R, et al. The global human impact on biodiversity. *Nature*, 2025, 641: 395–400
- 2 Diaz S, Settele J, Brondizio E S, et al. Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, 2019, 366: eaax3100
- 3 Verissimo D, Blake K, Flint H B, et al. Changing human behavior to conserve biodiversity. *Annu Rev Environ Resour*, 2024, 49: 419–448
- 4 Ye F Y, Zhang X, Zheng H. The fate of bugs: decline under the looming threats of climate, habitat loss, and pesticides. *Zool Res-Diversity Conservation*, 2024, 1: 201–203
- 5 Yao Y G, Jiang X L, Li S Q. Forging ahead towards a brighter future in biodiversity conservation. *Zool Res-Diversity Conservation*, 2025, 2: 1–2
- 6 Zhang W, Sheldon B C, Grenyer R, et al. Habitat change and biased sampling influence estimation of diversity trends. *Curr Biol*, 2021, 31: 3656–

- 3662.e3
- 7 Boënnec M, Dakos V, Devictor V. Sources of confusion in global biodiversity trends. *Oikos*, 2024, 2024(6): e10732
- 8 Turkalo A K, Wrege P H, Wittemyer G, et al. Long-term monitoring of Dzanga Bai forest elephants: forest clearing use patterns. *PLoS One*, 2013, 8: e85154
- 9 Dornelas M, Magurran A E, Buckland S T, et al. Quantifying temporal change in biodiversity: challenges and opportunities. *Proc R Soc B*, 2013, 280: 20121931
- 10 Card D C, Shapiro B, Giribet G, et al. Museum genomics. *Annu Rev Genet*, 2021, 55: 633–659
- 11 Schmitt C J, Cook J A, Zamudio K R, et al. Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene. *Phil Trans R Soc B*, 2018, 374: 20170387
- 12 Zeuss D, Brandl R, Brändle M, et al. Global warming favours light-coloured insects in Europe. *Nat Commun*, 2014, 5: 3874
- 13 Nakahama N. Museum specimens: an overlooked and valuable material for conservation genetics. *Ecol Res*, 2020, 36: 13–23
- 14 Benham P M, Bowie R C K, Zamudio K. Natural history collections as a resource for conservation genomics: understanding the past to preserve the future. *J Heredity*, 2022, 114: 367–384
- 15 Hofreiter M. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res*, 2001, 29: 4793–4799
- 16 Dabney J, Meyer M, Paabo S. Ancient DNA damage. *Cold Spring Harb Perspect Biol*, 2013, 5: a012567
- 17 Huerta-Sánchez E, Jin X, Asan X, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 2014, 512: 194–197
- 18 Liu S, Westbury M V, Dussex N, et al. Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell*, 2021, 184: 4874–4885.e16
- 19 Stanbrook-Buyer R, Jones M, Wheeler C P. Biotic and abiotic drivers of dung beetle assemblage structure in Kenyan Afromontane forest: trees and elephants matter. *J Trop Ecol*, 2025, 41: e11
- 20 Lister B C, Garcia A. Climate-driven declines in arthropod abundance restructure a rainforest food web. *Proc Natl Acad Sci USA*, 2018, 115: E10397
- 21 Salcido D M, Forister M L, Garcia Lopez H, et al. Loss of dominant caterpillar genera in a protected tropical forest. *Sci Rep*, 2020, 10: 422
- 22 Abrol D P. Asiatic honeybee *Apis cerana*: biodiversity conservation and agricultural production. Dordrecht, New York: Springer, 2013
- 23 Liu S, Qiu L, Liang D, et al. Temporal genomics reveal rapid adaptation to pesticide exposure in Eastern honeybees. *Natl Sci Rev*, 2025, 12: nwaf438
- 24 Mikheyev A S, Tin M M Y, Arora J, et al. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nat Commun*, 2015, 6: 7991
- 25 Qiu L, Dong J, Li X, et al. Defining honeybee subspecies in an evolutionary context warrants strategized conservation. *Zool Res*, 2023, 44: 483–493
- 26 Motta E V S, Moran N A. The honeybee microbiota and its impact on health and disease. *Nat Rev Microbiol*, 2023, 22: 122–137
- 27 Kwong W K, Moran N A. Gut microbial communities of social bees. *Nat Rev Microbiol*, 2016, 14: 374–384
- 28 Kwong W K, Medina L A, Koch H, et al. Dynamic microbiome evolution in social bees. *Sci Adv*, 2017, 3: e1600513
- 29 Guo L, Tang J, Tang M, et al. Reactive oxygen species are regulated by immune deficiency and Toll pathways in determining the host specificity of honeybee gut bacteria. *Proc Natl Acad Sci USA*, 2023, 120: e2219634120
- 30 Han B, Hu J, Yang C, et al. *Lactobacillus* firm-5-derived succinate prevents honeybees from having diabetes-like symptoms. *Proc Natl Acad Sci USA*, 2024, 121: e2405410121
- 31 Gauthier J, Pajkovic M, Neuenschwander S, et al. Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Mol Ecol Resour*, 2020, 20: 1191–1205
- 32 Ghisbain G, Thiery W, Massonnet F, et al. Projected decline in European bumblebee populations in the twenty-first century. *Nature*, 2023, 628: 337–341
- 33 Heintzman P D, Elias S A, Moore K, et al. Characterizing DNA preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera). *Mol Ecol Resour*, 2013, 14: 606–615
- 34 Barnes I, Thomas M G. Evaluating bacterial pathogen DNA preservation in museum osteological collections. *Proc R Soc B*, 2005, 273: 645–653
- 35 Willerslev E, Cooper A. Pathogenic microbial ancient DNA: a problem or an opportunity? *Proc R Soc B*, 2005, 273: 643
- 36 Campos P F, Gilbert T M P. DNA Extraction from Keratin and Chitin. New York: Humana Press, 2012. 43–49
- 37 Wibowo M C, Yang Z, Borry M, et al. Reconstruction of ancient microbial genomes from the human gut. *Nature*, 2021, 594: 234–239
- 38 Dabney J, Knapp M, Glocke I, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA*, 2013, 110: 15758–15763

- 39 Kapp J D, Green R E, Shapiro B, et al. A Fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *J Heredity*, 2021, 112: 241–249
- 40 Shen W, Sipos B, Zhao L. SeqKit2: a Swiss army knife for sequence and alignment processing. *iMeta*, 2024, 3: e191
- 41 Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes*, 2012, 5: 337
- 42 Chen S, Zhou Y, Chen Y, et al. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 2018, 34: i884–i890
- 43 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009, 25: 1754–1760
- 44 Danecek P, Bonfield J K, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience*, 2021, 10: giab008
- 45 Daley T, Smith A D. Predicting the molecular complexity of sequencing libraries. *Nat Methods*, 2013, 10: 325–327
- 46 Ginolhac A, Rasmussen M, Gilbert M T P, et al. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*, 2011, 27: 2153–2155
- 47 Wood D E, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*, 2019, 20: 257
- 48 Yang C, Hu J, Su Q, et al. A review on recent taxonomic updates of gut bacteria associated with social bees, with a curated genomic reference database. *Insect Sci*, 2025, 32: 2–23
- 49 Cavill E L, Liu S, Zhou X, et al. To bee, or not to bee? One leg is the question. *Mol Ecol Resour*, 2022, 22: 1868–1874
- 50 Mullin V E, Stephen W, Arce A N, et al. First large-scale quantification study of DNA preservation in insects from natural history collections using genome-wide sequencing. *Methods Ecol Evol*, 2023, 14: 360–371
- 51 Patzold F, Zilli A, Hundsdoerfer A K, et al. Advantages of an easy-to-use DNA extraction method for minimal-destructive analysis of collection specimens. *PLoS One*, 2020, 15: e0235222
- 52 Holmquist A, Tavris H, Kim G, et al. Towards large-scale museomics projects: a cost-effective and high-throughput extraction method for obtaining historical DNA from museum insect specimens. *Mol Ecol Resour*, 2025, 25: e14117
- 53 Korlević P, McAlister E, Mayho M, et al. A minimally morphologically destructive approach for DNA retrieval and whole-genome shotgun sequencing of pinned historic dipteran vector species. *Genome Biol Evol*, 2021, 13: evab226
- 54 Hong S, Sun Y, Sun D, et al. Microbiome assembly on *Drosophila* body surfaces benefits the flies to combat fungal infections. *iScience*, 2022, 25: 104408
- 55 De Cock M, Virgilio M, Vandamme P, et al. Impact of sample preservation and manipulation on insect gut microbiome profiling. A test case with fruit flies (Diptera, Tephritidae). *Front Microbiol*, 2019, 10: 2833
- 56 Staats M, Erkens R H J, van de Vossen B, et al. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One*, 2013, 8: e69189
- 57 Salter S J, Cox M J, Turek E M, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*, 2014, 12: 87
- 58 Luo S, Zhang X, Zhou X. Temporospatial dynamics and host specificity of honeybee gut bacteria. *Cell Rep*, 2024, 43: 114408
- 59 Moran N A, Hansen A K, Powell J E, et al. Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS One*, 2012, 7: e36393
- 60 Su Q, Tang M, Hu J, et al. Significant compositional and functional variation reveals the patterns of gut microbiota evolution among the widespread Asian honeybee populations. *Front Microbiol*, 2022, 13: 934459
- 61 Raxworthy C J, Smith B T. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol Evol*, 2021, 36: 1049–1060
- 62 Rayo E, Ulrich G F, Zemp N, et al. Minimally destructive hDNA extraction method for retrospective genetics of pinned historical Lepidoptera specimens. *Sci Rep*, 2024, 14: 12875

## 补充材料

- 图S1 片段末端的替换频率
- 图S2 不同采样批次间重复序列含量的相关性
- 图S3 外源DNA序列片段的长度分布
- 表S1 *Apis cerana*参考基因组信息
- 表S2 蜜蜂肠道菌自建库物种信息表
- 表S3 腹部样本蜜蜂肠道菌序列统计表
- 表S4 后足样本蜜蜂肠道菌序列统计表

本文以上补充材料见网络版csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “博物馆基因组学为解析蜜蜂及其肠道菌的协同演化开辟新路径”

# Museum-omics opens a new avenue for investigating honey bee host-microbiome coevolution

Zicong Li<sup>1,2†</sup>, Churong Guo<sup>3†</sup>, Lifei Qiu<sup>1</sup>, Jiawei Hu<sup>3</sup>, Chunyan Jiang<sup>4</sup>, Chaodong Zhu<sup>1,5</sup>, Xin Zhou<sup>3</sup>, Jun Chen<sup>1,4</sup>, Xue Zhang<sup>3\*</sup> & Shanlin Liu<sup>1,5\*</sup>

<sup>1</sup> State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> College of Biological Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

<sup>3</sup> Department of Entomology, College of Plant Protection, China Agricultural University 100193, China

<sup>4</sup> National Animal Collection Resource Center of China, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

<sup>5</sup> CAS-CNRS International Research Network on Biodiversity, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

† Equally contributed to this work

\* Corresponding authors, E-mail: zhangxue05@cau.edu.cn; shanlin.liu@ioz.ac.cn

Museum specimens are irreplaceable archives of historical genetic diversity, yet their potential to yield both host and symbiotic microbial DNA has remained largely unexplored. Here, we developed a coordinated host-microbiome DNA extraction and analysis pipeline using 16 historical dry-pinned specimens of *Apis cerana* collected between 1955 and 1988. We compared traditional hind leg tissue with abdominal gut tissue in terms of DNA yield, degradation patterns, endogenous mapping rates for both the honeybee host and core gut bacteria, and the feasibility of repeated extraction from the same specimen.

Our results show that both tissue types yielded reliable genomic sequences with typical ancient DNA damage signals, including short fragment lengths (39–49 bp) and elevated C→T substitutions at the 5' ends, confirming their historical origin. Abdominal gut tissue yielded significantly more DNA than hind legs ( $0.49 \pm 0.42$  ng total DNA, mean  $\pm$  SD vs.  $0.10 \pm 0.07$  ng; Wilcoxon test,  $P=0.02$ ). However, endogenous honeybee DNA content varied widely among samples (0.23%–60.8%), and the highest DNA concentrations often corresponded to very low endogenous content because of historical exogenous contamination. Thus, DNA yield alone is not a reliable indicator of sequencing success. Re-extraction from the same tissue consistently generated additional usable data, with the second extraction contributing 11.2% of the total DNA even when the first extraction yielded only trace amounts, making repeated extraction a valuable strategy for precious museum material.

Crucially, abdominal gut tissues harbored significantly higher proportions of sequences assigned to core honeybee gut bacterial genera, including *Lactobacillus*, *Bifidobacterium*, *Gilliamella*, *Apibacter*, *Snodgrassella*, and *Apilactobacillus*, whereas hind leg samples contained negligible microbial reads (<0.1%). The gut-derived microbial sequences also showed the expected short-fragment size distributions and clear deamination damage patterns, indicating that they are endogenous to the historical specimens rather than modern contaminants.

Our extraction protocol minimizes physical damage to the specimen: after processing, the abdomen and hind leg can be air-dried and reattached beneath the original specimen using triangular mounts. Although only trace amounts of DNA are obtained initially, repeated extractions from the same abdominal gut tissue substantially increase total data output and provide sufficient valid reads for both host genomic and microbial community analyses. Based on these findings, we recommend prioritizing abdominal gut tissue when the goal is to recover high-depth host genomic data together with the associated gut microbiome. For studies focused solely on the host genome while minimizing specimen damage, a single hind leg remains a suitable choice.

In conclusion, this study demonstrates for the first time that dry-pinned insect museum specimens, especially abdominal gut tissue, can reliably yield both host and gut microbial genetic information. Our work provides a robust framework for studying honeybee coevolution and for leveraging museum collections through a hologenomic lens, with broad applicability to other insects in which gut symbionts play important ecological roles.

*Apis cerana*, historical specimens, museum-omics, gut microbiota, coevolution

doi: 10.1360/CSB-2026-0099