



基于 ITS2 测序的中国四个城市 灰尘样本地域推断研究*

张文君^{1,2)} 冯耀森²⁾ 彭加金^{2,3)} 冯凯⁴⁾ 邓晔⁴⁾ 康克莱^{2)**} 王乐^{1,2)**}

(¹⁾ 山西医科大学法医学院, 晋中 030600; (²⁾ 公安部鉴定中心, 法医遗传学公安部重点实验室, 北京 100038;

(³⁾ 昆明医科大学法医学院, 昆明 650500; (⁴⁾ 中国科学院生态环境研究中心, 中国科学院环境生物技术重点实验室, 北京 100085)

摘要 目的 灰尘作为重要的微量物证检材, 在法庭科学领域有极大的应用潜力, 随着 DNA 测序技术的发展, 越来越多的研究者开始关注灰尘中的微生物群落分布。不同地区灰尘中的微生物群落分布存在差异, 利用灰尘中的微生物群落信息进行生物地理位置推断在法庭科学领域有重要意义。通过分析不同地区灰尘样本的微生物 DNA, 可以揭示该地区的微生物组成, 从而推断该地区的生物地理位置。本研究旨在探索中国 4 个城市的真菌群落组成结构差异, 评估基于真菌的内转录间隔区 (internal transcribed spacer, ITS) 2 区测序对灰尘样本进行生物地理位置推断的准确率, 为公安实战基于环境 DNA 进行地域推断奠定基础。方法 使用通用引物 5.8S-Fun/ITS4-Fun 扩增真菌基因组的 ITS2 区, 在 Illumina MiSeq FGx 平台进行测序。采用非度量多维标度分析 (non-metric multidimensional scaling, NMDS) 可视化样本之间的差异。使用相似性分析 (analysis of similarities, ANOSIM) 和多元方差分析 (permutational multivariate analysis of variance, PERMANOVA) 两种不同的非参数分析方法衡量样本间的群落差异程度。使用 LDA Effect Size (LefSe) 分析筛选出不同城市间有显著差异的物种。使用微生物来源分析软件 SourceTracker 对灰尘样本进行生物地理位置来源推断。结果 北京、福州、昆明、乌鲁木齐 4 个城市的灰尘样本中, 北京的物种丰富度最高, 物种注释结果显示, 4 个城市的真菌群落物种组成结构和相对丰度有明显差异。NMDS 结果显示, 灰尘样本在多维空间中基于生物地理来源进行分组, 同一城市的灰尘样本有明显的聚类现象, 不同城市的灰尘样本在第一轴明显分开。ANOSIM 和 PERMANOVA 结果也表明, 4 个城市间的真菌群落组成均具有显著差异, 其中以福州和乌鲁木齐两个城市之间的差异最为明显。所有已知来源的灰尘样本的生物地理位置均被成功预测。结论 北京、福州、昆明、乌鲁木齐 4 个城市的真菌群落物种组成结构和相对丰度具有显著差异, 基于真菌的 ITS2 区测序对这 4 个城市的灰尘样本进行生物地理位置推断具有较高的准确率, 在法庭科学领域利用灰尘中的真菌群落信息进行生物地理位置推断有较高的可行性。

关键词 法庭科学, 真菌, 灰尘 DNA, 地域推断, 高通量测序

中图分类号 R89, Q938, DF795.2

DOI: 10.16476/j.pibb.2024.0300

CSTR: 32369.14.pibb.20240300

灰生物证在法庭科学领域的应用由来已久, 早在 20 世纪初, 法国犯罪学家埃德蒙·洛卡德 (Edmond Locard)^[1] 就提出, 灰尘、土壤等微量检材在法庭科学领域具有重要的检验价值。灰生物证作为微量检材具有量小体微的特点, 在案发现场极易借助衣物、鞋子、武器等载体附着到罪犯或受害者身上; 通过检测案件中收集到的灰尘样本, 可以将犯罪嫌疑人或受害者与犯罪现场联系起来, 为侦破案件提供线索和侦查方向。在法庭科学的漫长历史中, 针对灰尘的研究通常是使用显微镜或放大镜

进行鉴定, 基于形态学的灰尘分析被用于生物地理位置推断的案例已被报道^[2-3], 这些真实的案例有力地证明了灰生物证是法医案件工作中的理想工具。然而基于形态学的鉴定往往是主观的, 依赖于对广泛生物类群有丰富经验和知识的专家。随着

* 公安部鉴定中心基本科研业务费 (2022JB022) 和公安部科技强警基础工作专项 (2023JC14) 资助。

** 通讯联系人。

康克莱 Tel: 010-83752707, E-mail: kang_kl@yeah.net

王乐 Tel: 010-83752707, Email: wangle_02@163.com

收稿日期: 2024-07-07, 接受日期: 2024-11-21

DNA 测序技术和环境 DNA 分析的出现, 针对灰尘的分析发生了革命性变化。动物、植物、微生物、人体细胞或其他生物类群在灰尘中普遍存在, 新的 DNA 技术的出现, 意味着可以直接基于 DNA 序列信息对它们进行物种鉴定^[4]。已有研究表明, DNA 测序方法比显微镜分析发现的物种数量更多, 能够检测到显微镜下未检测到的物种^[5-6]。

生态学研究很好地揭示了灰尘中微生物群落的空间异质性分布特征^[7-8], 基于此特征, 灰尘物证可以为案件侦查中的生物地理位置推断提供有价值的信息。然而, 自然界中大量的微生物无法利用纯培养手段在实验室条件下进行培养^[9], DNA 测序技术的出现使得对灰尘中微生物群落的研究摆脱了传统培养手段的限制, 目前 DNA 测序技术已成为检测环境样本中微生物群落的关键工具^[10]。核糖体 RNA (rRNA) 基因广泛存在于所有细胞生物体中, 具有一系列进化保守性较高和较低的区域, 常被用作多样性调查的标记基因^[11-12]。根据研究目标的不同应用于灰尘分析的遗传标记主要包括 16S rRNA 基因^[13-15]、18S rRNA 基因^[13] 和内转录间隔区 (internal transcribed spacer, ITS)^[16-17] 等, 基于 DNA 测序技术对这些位点的某个可变区序列进行检测, 可以实现对灰尘中的细菌、真菌等微生物群落的表征^[18]。

在过去的十几年中, 针对灰尘中微生物组的研究主要集中在细菌^[19-20], 也有少部分研究关注于真菌^[21]。已有研究表明, 细菌群落结构的稳定性易受各种因素的影响, 相比之下, 真菌对某些环境变量相对不敏感, 如降水、温度、保存条件等^[22-23]。Barberán 等^[24] 为了研究微生物群落不同地理区域之间如何变化, 采集了美国大陆各地的室外灰尘样本, 发现真菌的地理分布比细菌更窄, 并且随气候的变化更具有可预测性。真菌的物种多样性是非常丰富的, 已描述的真菌种类近 10 万种, 更多的真菌种类仍未被描述^[25], 灰尘样品中含有大量的真菌分类单元, 对灰尘中的真菌进行表征在法医学领域可以为犯罪侦查、案件审理和法医学鉴定提供新的技术手段和科学依据。

灰尘中的真菌分布已被证明与地理位置有关, Allwood 等^[26] 以灰尘中的真菌类群为特征开展了全球范围内的探索地理定位工作, 来源国家与训练模型相同的单国家拭子样本预测准确率达 82.8%。Grantham 等^[8] 使用从灰尘中提取的真菌 DNA 预测来自美国大陆的近 1 000 个灰尘样品的地理定位,

预测误差的中位数为 230 km。针对灰尘真菌的相关研究增多, 这意味着广泛地理来源的灰尘真菌测序数据可以在公共数据库中找到, 这些数据可以被整合后创建参考数据库, 用于地理位置推断。然而, 以环境 DNA 为目标的法医学研究还处于可行性评估阶段, 更深入地了解灰尘中微生物群落的自然分布模式在法庭科学领域有重要意义。本研究基于二代测序技术对不同城市灰尘中的真菌 DNA 进行检测, 旨在探索北京、福州、昆明和乌鲁木齐 4 个城市的真菌群落自然分布模式, 评估基于真菌的 ITS2 区测序对灰尘样本进行生物地理位置推断的准确率, 为环境 DNA 在法庭科学领域的应用奠定基础。

1 材料与方法

1.1 样本采集

本研究采集了中国北京 (115°20′-117°30′E, 39°28′-41°05′N)、福州 (118°08′-120°31′E, 25°15′-26°39′N)、昆明 (102°10′-103°40′E, 24°23′-26°33′N) 和乌鲁木齐 (86°37′-88°58′E, 42°45′-44°08′N) 4 个城市的住宅小区灰尘样本。采样点主要分布在反映城市生活重要特征的主城区 (表 1), 相邻采样点间相距 5~10 km。北京采样点的经度范围为 116°11′-116°36′E, 纬度范围为 39°50′-40°01′N, 分布于海淀区、朝阳区、东城区、西城区、丰台区和石景山区; 福州采样点的经度范围为 119°15′-119°32′E, 纬度范围为 25°54′-26°08′N, 分布于马尾区、闽侯县、仓山区、晋安区、鼓楼区和台江区; 昆明采样点的经度范围为 102°37′-102°52′E, 纬度范围为 24°50′-25°59′N, 分布于呈贡区、官渡区、西山区、五华区 and 盘龙区; 乌鲁木齐采样点的经度范围为 87°24′-87°41′E, 纬度范围为 43°44′-43°59′N, 分布于水磨沟区、天山区、沙依巴克区、头屯河区、米东区、新市区。门框上缘是不经常被打扫的位置, 人源影响因素较少, 因此选择门框上缘作为灰尘采集的载体。使用 Copan Liquid Amies Elution 拭子 (ESwab, Copan Diagnostics 公司, 意大利) 擦拭小区单元门门框上缘的灰尘, 擦拭完毕将拭子头放入样本管后放入干冰箱内保存。本研究共采集灰尘样本 91 份, 包括北京样本 22 份、福州样本 23 份、昆明样本 23 份、乌鲁木齐样本 23 份, 所有样本均保存于 -80℃ 冰箱内待进一步处理。

Table 1 The information of sampling sites

City	Community	Longitude/°	Dimensional/°	City	Community	Longitude/°	Dimensional/°
Beijing	BJ01-A	116.286 111	40.009 444	Kunming	KM02-A	102.823 889	24.850 556
	BJ02-C	116.252 500	39.970 833		KM03-A	102.845 556	24.843 611
	BJ03-A	116.287 222	39.920 278		KM05-B	102.782 500	25.010 833
	BJ05-C	116.502 500	39.855 833		KM06-C	102.822 500	24.981 111
	BJ06-A	116.513 333	39.891 111		KM09-B	102.753 333	25.030 278
	BJ09-A	116.328 611	39.897 500		KM10-A	102.624 444	25.010 278
	BJ10-B	116.410 833	39.893 611		KM11-A	102.637 500	25.042 778
	BJ11-A	116.406 944	39.920 833		KM12-A	102.653 333	25.054 722
	BJ12-B	116.394 167	39.948 889		KM13-D	102.718 333	25.110 833
	BJ13-C	116.501 389	39.959 167		KM14-C	102.666 389	25.096 111
	BJ14-A	116.412 500	39.978 333		KM15-A	102.822 222	25.006 111
	BJ15-B	116.428 056	39.850 833		KM16-C	102.874 167	25.026 111
	BJ17-A	116.304 444	39.841 389		KM17-C	102.795 278	25.053 333
	BJ18-B	116.295 278	39.888 611		KM18-A	102.741 944	25.046 389
	BJ19-A	116.185 556	39.907 222		KM19-A	102.727 778	25.085 556
	BJ20-B	116.203 056	39.941 944		KM20-B	102.650 556	25.041 667
	BJ21-A	116.373 333	39.879 722		KM21-B	102.697 222	25.075 278
	BJ22-A	116.360 278	39.913 611		KM22-A	102.745 000	25.118 889
	BJ23-B	116.360 278	39.943 611		KM23-C	102.779 444	25.953 889
	BJ24-C	116.601 389	39.919 444		KM24-A	102.816 944	25.188 333
	BJ27-C	116.275 000	39.956 667		KM25-B	102.781 111	25.003 056
	BJ28-A	116.282 500	40.032 222		KM26-B	102.782 778	25.998 056
Fuzhou	FZ01-A	119.542 778	25.953 333	Urumqi	KM27-C	102.861 111	25.255 556
	FZ02-C	119.491 111	25.941 389		WQ01-B	87.634 722	43.812 222
	FZ03-A	119.456 389	25.997 778		WQ02-A	87.610 833	43.795 278
	FZ04-B	119.394 722	26.026 389		WQ03-A	87.608 333	43.813 611
	FZ05-A	119.252 500	25.983 889		WQ04-A	87.559 722	43.744 722
	FZ06-D	119.370 556	25.903 889		WQ05-C	87.647 500	43.765 000
	FZ07-C	119.371 111	25.990 000		WQ06-A	87.542 500	43.815 278
	FZ08-B	119.346 389	26.006 111		WQ07-B	87.524 722	43.798 889
	FZ09-D	119.321 389	26.018 056		WQ08-A	87.553 056	43.792 500
	FZ10-B	119.365 833	26.048 333		WQ09-C	87.656 944	43.790 278
	FZ11-A	119.331 667	26.070 556		WQ10-C	87.678 056	43.830 278
	FZ12-A	119.282 222	26.088 333		WQ11-C	87.401 389	43.842 778
	FZ13-D	119.279 444	26.065 000		WQ12-C	87.426 667	43.868 611
	FZ14-A	119.264 167	26.047 222		WQ13-C	87.496 944	43.840 833
	FZ15-A	119.278 333	26.011 389		WQ14-C	87.651 111	43.883 611
	FZ16-B	119.318 611	26.048 333		WQ16-A	87.620 000	43.818 611
	FZ17-B	119.343 889	26.053 056		WQ17-B	87.420 556	43.985 000
	FZ18-C	119.301 389	26.061 389		WQ19-C	87.660 556	43.965 833
	FZ19-C	119.335 833	26.090 000		WQ20-B	87.690 278	43.953 611
	FZ20-A	119.333 611	26.146 944		WQ21-B	87.650 000	43.908 333
	FZ21-A	119.311 667	26.147 778		WQ22-C	87.493 056	43.890 833
	FZ22-D	119.283 889	26.133 333		WQ23-A	87.575 556	43.906 944
	FZ23-D	119.288 333	26.106 667		WQ24-B	87.597 778	43.863 056
					WQ25-A	87.540 000	43.857 500

“BJ” stands for Beijing, “FZ” stands for Fuzhou, “KM” stands for Kunming, “WQ” stands for Urumqi.

1.2 DNA提取、文库制备及测序

使用DNeasy PowerSoil Pro试剂盒(QIAGEN公司, 德国)进行DNA提取, 提取方法遵循已报道的优化方案^[27]。使用Qubit® dsDNA HS Assay试剂盒(Thermo Fisher公司, 美国)和Qubit® 3.0 Fluorometer定量仪器(Thermo Fisher公司)进行DNA定量。使用通用引物5.8S-Fun (5'-AACTTTYRRC AAYGGATCWCT-3')和ITS4-Fun (5'-AGCCTCCGCTTATTGATATGCTTAART-3')扩增真菌基因组的ITS2区。PCR反应体系为: 10 µl Platinum™ II热启动PCR预混液(Thermo Fisher公司), 0.2 µmol/L正向引物, 0.2 µmol/L反向引物, 10 ng模板DNA, 无核酸酶水补齐至25 µl。反应程序为: 94℃预变性3 min; 94℃变性45 s, 58℃退火60 s, 72℃延伸90 s, 35个循环; 72℃延伸10 min。磁珠(Beckman公司, 美国)纯化后使用Qubit® dsDNA HS Assay试剂盒和Qubit® 3.0 Fluorometer定量仪器进行文库DNA定量。使用2100生物分析仪(Agilent公司, 美国)和Agilent DNA 1000试剂盒(Agilent公司)进行文库质检。使用MiSeq® Reagent Kit V3试剂盒(Illumina公司, 美国)在Illumina MiSeq FGx平台(Illumina公司)进行测序。

1.3 测序数据处理

测序得到的原始数据使用FastQC^[28](版本0.12.1)进行初步质量评估, 使用Cutadapt^[29](版本4.7)去除引物序列, 使用FLASH^[30](版本1.2.11)将至少有30 bp重叠的配对读段(Reads)连接成长读段。使用Fastp^[31](版本0.23.2)对拼接后的数据进行数据过滤, 窗口大小设置为20, 去除碱基平均质量分数小于20的不合格Reads, 去除片段长度小于200 bp的Reads。使用USEARCH^[32](版本11.0.667_i86linux64)对数据进行去冗余、去嵌合体。基于UNOISE3算法^[33]进行去噪, 丰度阈值设为10, 鉴定扩增子序列变异体(amplicon sequence variant, ASV)。对所有灰尘样本的总Reads数进行标准化, 重新抽取至27 571(样本最小Reads数)条。通过将序列与UNITE数据库^[34]进行比对, 注释ASV的分类信息, 注释相似度阈值设为97%。

1.4 统计分析

所有统计分析均使用R语言(版本4.3.2; <https://www.r-project.org/>)完成。基于USEARCH

计算4个城市真菌群落的 α 多样性指数。使用vegan包对真菌群落间的相异性距离进行非度量多维标度分析(non-metric multidimensional scaling, NMDS)以可视化样本之间的差异。同时使用veganEx包和pairwiseAdonis包对不同城市的灰尘样本进行相似性分析(analysis of similarities, ANOSIM)和多元方差分析(permutational multivariate analysis of variance, PERMANOVA), 置换检验次数设为999, 通过Bonferroni方法进行多重比较 P 值校正, 衡量样本间的真菌群落差异程度。使用LEfSe^[35]分析筛选出不同城市间有显著差异的物种。使用微生物来源分析软件SourceTracker, 以4个城市的所有ASV为特征估算灰尘样本的来源地组成成分, 根据灰尘样本的主要来源占比推断其生物地理位置^[36]。

2 结 果

2.1 数据特征

4个城市的91份灰尘样本测序后共得到11 438 662条原始Reads, 经过质量过滤、降噪和去嵌合体, 最终用于后续数据分析的真菌数据集包括7 126 627条Reads。长度分布显示, 文库的片段长度为303~442 bp, 平均片段长度为334 bp。序列聚类后共生成2 473个ASV, 其中178个ASV未能在门的水平上成功进行物种注释(表2)。

2.2 真菌群落多样性分析

为探究4个城市真菌群落的物种丰富度和均匀度, 本研究计算了不同的 α 多样性指数, 分别为丰富度(Richness)、皮洛指数(Pielou's Evenness Index)、香农指数(Shannon Index)。结果显示: 4个城市中北京的真菌物种丰富度最高(图1a), 乌鲁木齐真菌物种均匀度最高(图1b); 当综合考虑物种均匀度和丰富度时, 乌鲁木齐高于其他城市(图1c); 福州的物种丰富度和均匀度均最低(图1a~c)。为了比较4个城市间真菌群落的组成差异, 本文基于非加权的Unifrac距离在ASV水平上, 利用NMDS对真菌群落进行了 β 多样性分析。NMDS结果显示, 灰尘样本在多维空间中基于生物地理位置进行分组, 同一城市的灰尘样本有明显的聚类现象, 不同城市来源的灰尘样本在第一轴明显分开(图2)。进一步对数据进行了相异性统计检验, ANOSIM和PERMANOVA结果表明, 4个城市间

的真菌群落组成具有显著差异 (P 值均小于 0.001), 其中以福州和乌鲁木齐两个城市之间的差异最为明显 (ANOSIM, $R=0.96$; PERMANOVA, $R^2=0.91$) (表3)。

Table 2 Features of the sequencing data

	Raw reads ¹⁾		Clean reads ²⁾			ASVs ³⁾		Unclassified reads/% ⁴⁾
	Total	Minimum	Merged reads	Filtered reads	Reads after clustering	Total	Unclassified ASVs/%	
Beijing	2 496 317	62 673	2 361 261	2 157 431	1 531 372	1 771	7.06	4.47
Fuzhou	3 311 357	100 890	3 161 081	2 900 132	2 139 431	1 166	5.14	0.34
Kunming	3 002 400	81 706	2 846 640	2 606 190	1 862 208	1 830	5.68	5.68
Urumqi	2 628 588	47 250	2 460 791	2 244 910	1 593 616	1 361	7.93	5.45

^{1,2)} The number of reads before or after USEARCH based on 91 dust samples. ³⁾ The number of ASVs , and the percentage of ASVs that could not be assigned at the phylum level with UNITE database. ⁴⁾ The percentage of reads that could not be classified at the phylum level with UNITE database.

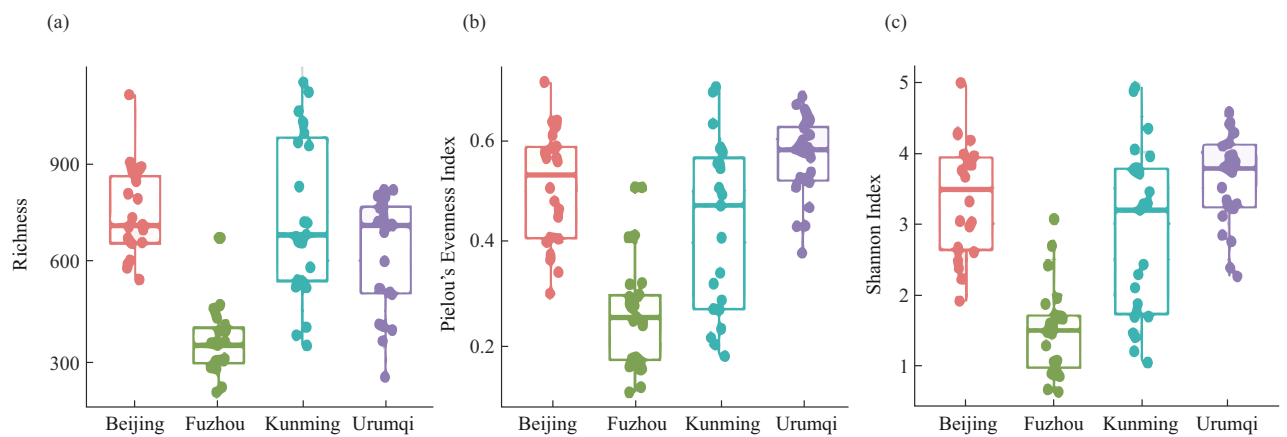


Fig. 1 Alpha diversity of the fungal communities in four cities

(a) Species richness of fungal communities in four cities. (b) Species Pielou's Evenness Index of fungal community in four cities. (c) Species Shannon Index of fungal communities in four cities.

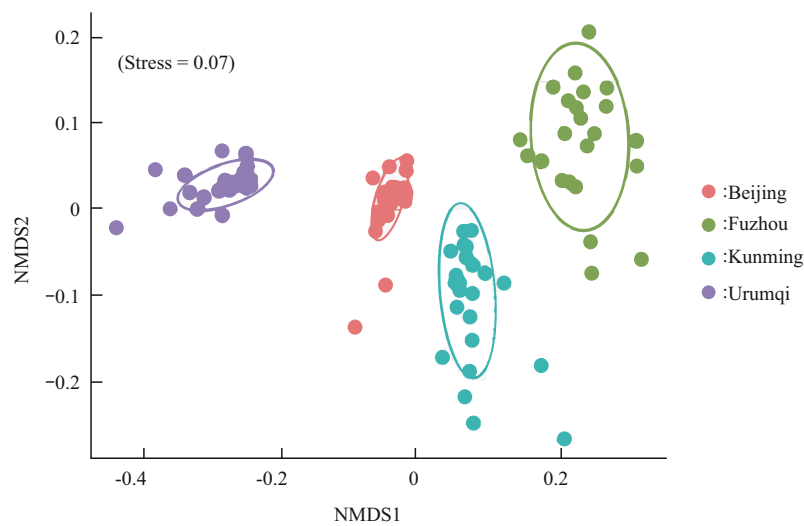


Fig. 2 Non-metric multidimensional scaling (NMDS) analysis in four cities

Table 3 Dissimilarity test by ANOSIM and PERMANOVA of the dust fungal community in four cities

Groups	ANOSIM		PERMANOVA		
	R	P	F.Model	R ²	P
Beijing versus Fuzhou	0.87	0.001***	330.23	0.88	0.001***
Beijing versus Kunming	0.39	0.001***	76.24	0.64	0.001***
Beijing versus Urumqi	0.66	0.001***	380.11	0.90	0.001***
Fuzhou versus Kunming	0.56	0.001***	85.86	0.66	0.001***
Fuzhou versus Urumqi	0.96	0.001***	425.31	0.91	0.001***
Kunming versus Urumqi	0.82	0.001***	281.12	0.86	0.001***

***P<0.001.

2.3 真菌群落组成

在韦恩图中显示了所有灰尘样本中独有和共有的科 (family)、属 (genus) 和种 (species)。4个城市共有 112 个科 (图 3a)、190 个属 (图 3b)、278 个种 (图 3c)。随着分类水平的细化, 4个城市共有物种占总物种的比例有降低的趋势, 而城市的独有物种占比逐渐升高 (图 3)。基于 UNITE 数据库对检测到的读段进行物种注释, 结果显示, 在门 (phylum) 的水平上, 子囊菌门 (Ascomycota) 在 4 个城市中都占主导地位, 担子菌门

(Basidiomycota) 排第二位, 壶菌门 (Chytridiomycota) 和接合菌门 (Zygomycota) 也被检测到, 但丰度较低 (图 4a)。在目 (order) 的水平, 4个城市检测到的相对丰度大于 5% 的优势菌目包括煤炱目 (Capnodiales)、格孢腔菌目 (Pleosporales)、座囊菌目 (Dothideales)、散囊菌目 (Eurotiales)、银耳目 (Tremellales)、肉座菌目 (Hypocreales)、伞菌目 (Agaricales) 和锁掷酵母目 (Sporidiobolales), 其中 Capnodiales 是丰度最高的真菌目, 在福州的丰度远高于另外 3 个城市, 在乌鲁木齐的丰度较低。Pleosporales 在乌鲁木齐的丰度最高, 在福州的丰度最低 (图 4b)。在属水平上, 相对丰度大于 2% 的优势菌属中枝孢霉属 (*Cladosporium*) 丰度最高, 链格孢属 (*Alternaria*) 排第二位 (图 4c)。*Cladosporium*、*Aureobasidium* 和 *Pleospora* 在昆明的丰度明显高于其他 3 个城市。*Alternaria*、*Cryptococcus*、*Phoma*、*Aspergillus*、*Botryotinia* 和 *Chalastospora* 在乌鲁木齐的丰度较高, *Didymella* 在北京有较高丰度, *Penidiella* 在福州的丰度远远高于其他 3 个城市。

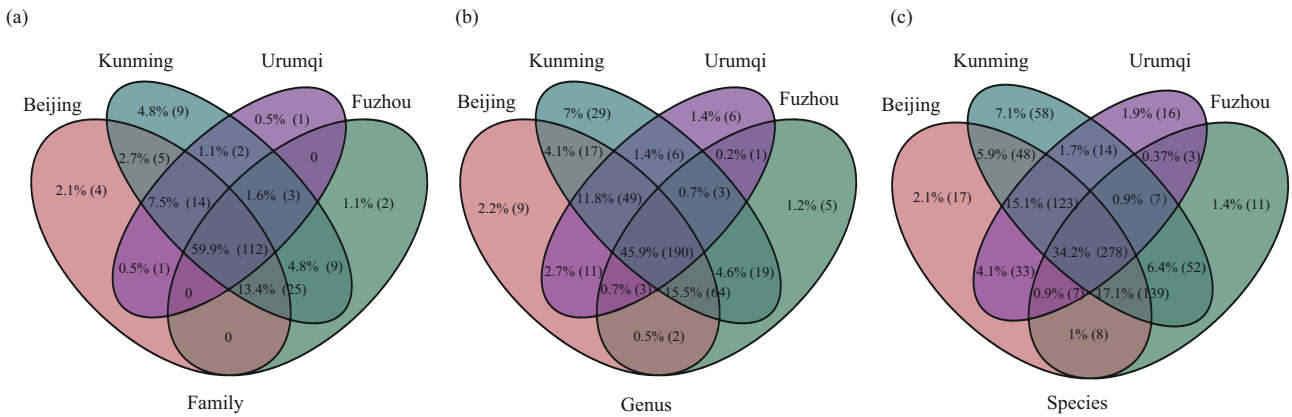


Fig. 3 The proportion of common and unique species at different taxonomic levels (family, genus or species) among the four cities

(a) The proportion of common and unique species in the four cities at the family level. (b) The proportion of common and unique species in the four cities at the genus level. (c) The proportion of common and unique species in the four cities at the species level.

2.4 灰尘样本生物地理位置推断

此外, 基于 LEfSe 分析在属的水平上筛选了 4 个城市间有显著差异的物种 (图 5)。线性判别分析得分 (linear discriminant analysis, LDA) 大于 4 时, 在属的水平上 *Didymella*、*Penidiella*、*Cladosporium*、*Aureobasidium*、*Pleospora*、

Capnobotryella、*Alternaria*、*Cryptococcus*、*Phoma*、*Botryotinia*、*Chalastospora* 和 *Aspergillus* 被筛选为 4 个城市间的显著差异物种。*Didymella* 被分析为北京的指示物种; *Penidiella* 被分析为福州的指示物种; *Cladosporium*、*Aureobasidium*、*Pleospora* 和 *Capnobotryella* 可作为 4 个城市中昆明

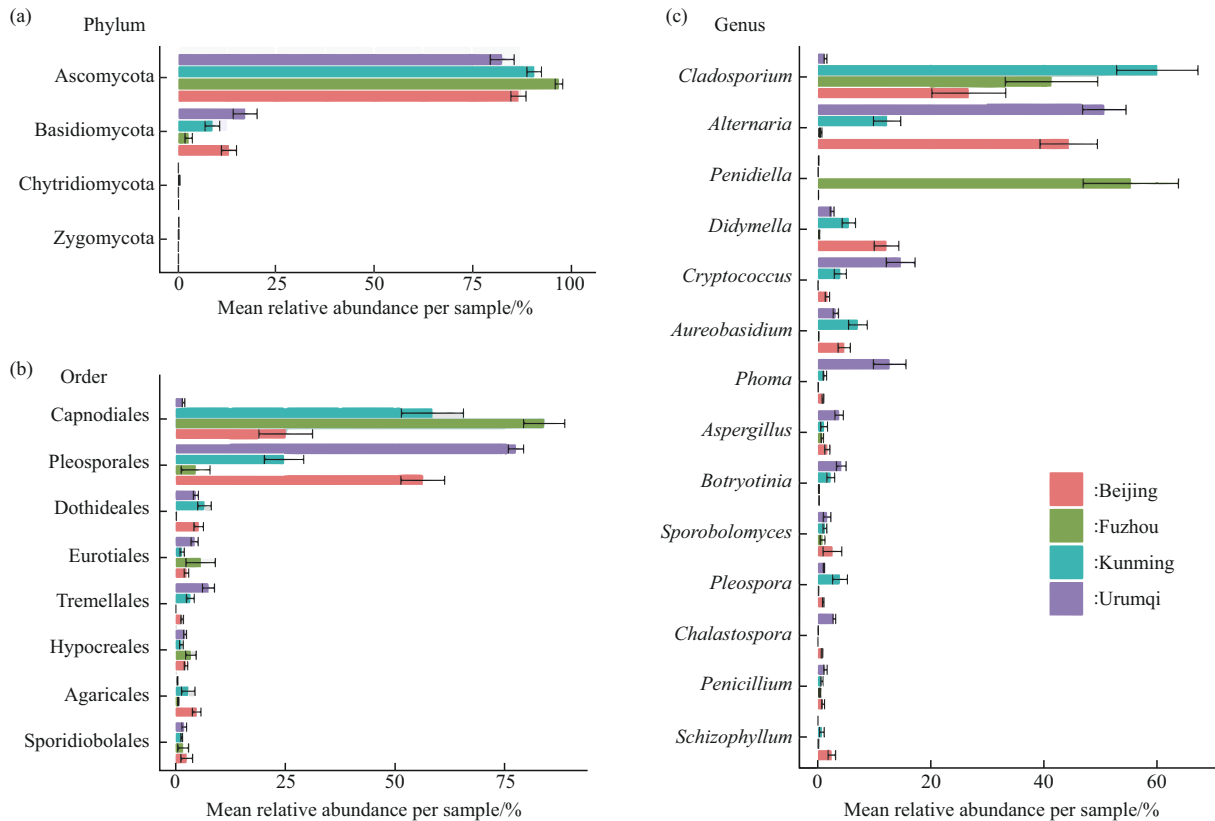


Fig. 4 Taxonomic distribution of the dust fungal community at three taxonomic levels between the four cities

(a) All classified species at the phylum level in four cities. (b) Distribution of species with relative abundance greater than 5% at the order level in four cities. (c) Distribution of species with relative abundance greater than 2% at the genus level in four cities.

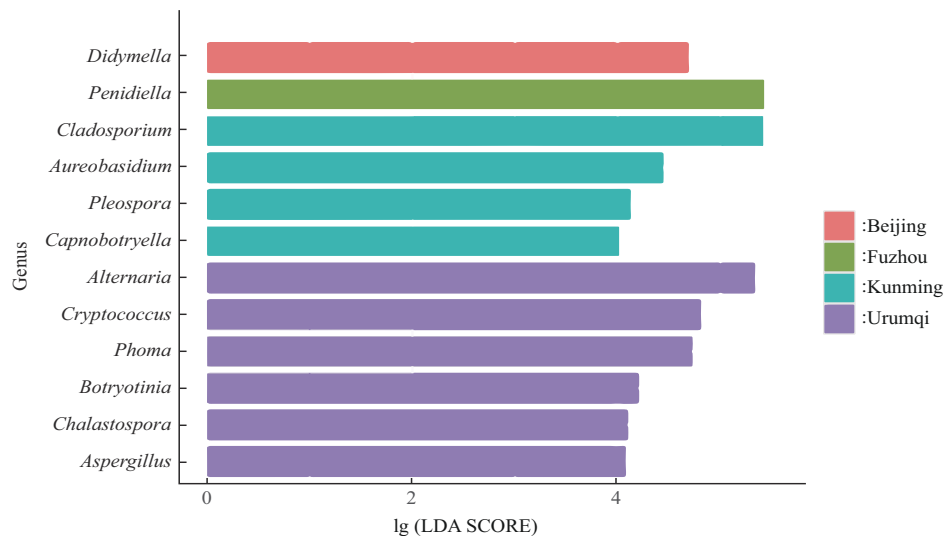


Fig. 5 The Linear discriminant analysis Effect Size (LEfSe) was used to select species with significant differences between the four cities

的指示物种；*Alternaria*、*Cryptococcus*、*Phoma*、*Botryotinia*、*Chalastospora* 和 *Aspergillus* 被分析为

乌鲁木齐的指示物种。每个城市随机选择 3 个样本 ($n=12$) 作为测试集，剩余样本 ($n=79$) 作为训练

集,使用微生物来源分析软件SourceTracker对4个城市的灰尘样本进行生物地理位置来源推断。SourceTracker结果表明(图6),所有已知来源的灰尘样本的生物地理位置均被正确预测。其中北京的灰尘样本被正确预测为其来源地的百分比为

86.4%~92.4%,福州的灰尘样本被正确预测为其来源地的百分比为97.4%~99%,昆明样本被正确预测为其来源地的百分比为71.2%~95.1%,乌鲁木齐样本被正确预测为其来源地的百分比为93.3%~98%。

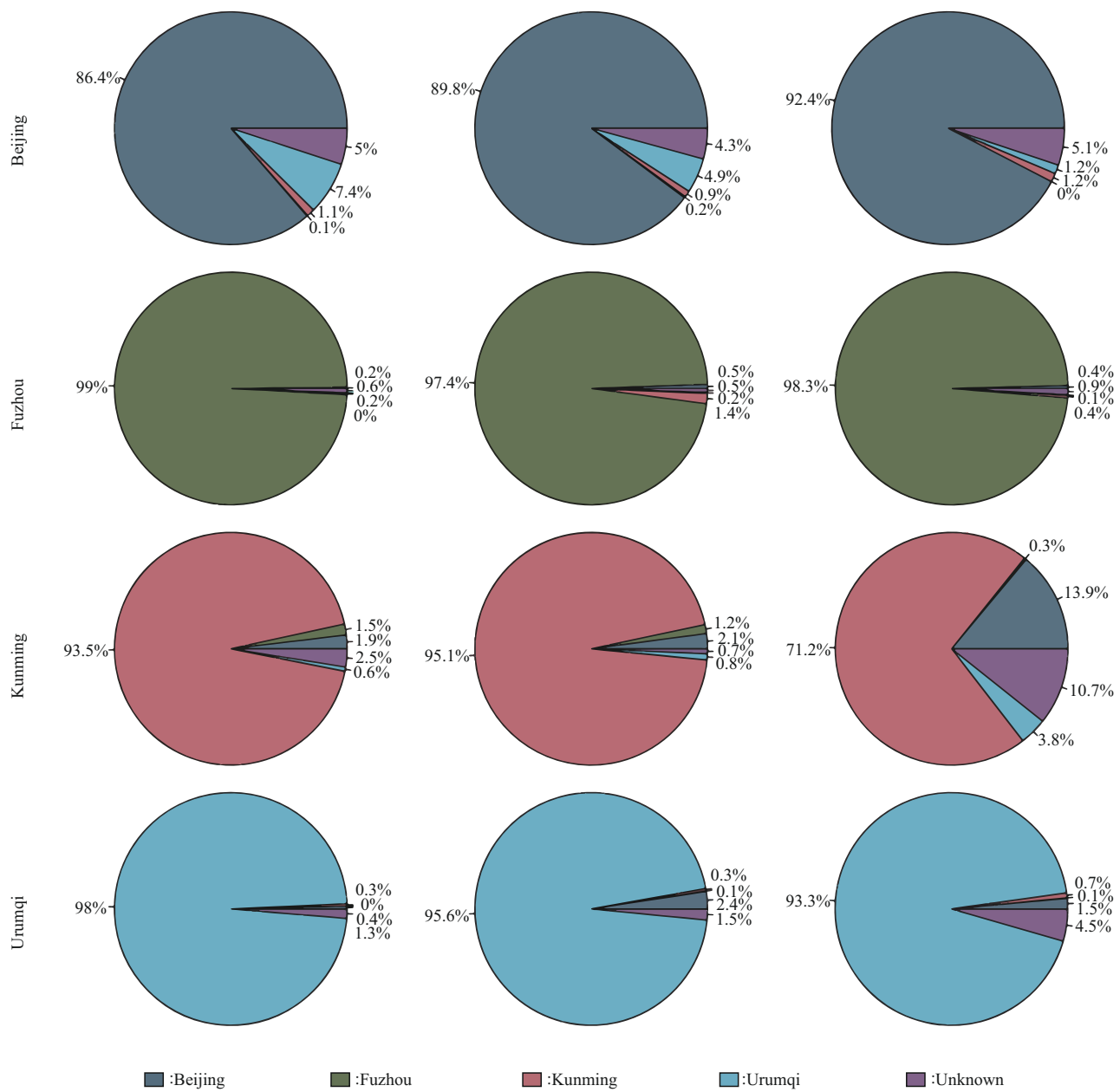


Fig. 6 Geographical inference of dust samples from four Chinese cities using SourceTracker

3 讨 论

当前在信息爆炸时代,科技的飞速发展各个领域带来了深远的影响,罪犯的反侦查意识和能力不断提高,犯罪现场遗留下的人源样本逐渐减少,破坏证据、伪造现场以及不留证据等现象日益突

出,使得传统的侦查手段面临严峻挑战,迫切需要新的技术和方法来提升案件侦破的效率和准确性。随着生物技术的发展,法庭科学领域针对微生物的研究扩展至与案件相关的死亡时间推断、死亡原因推断、生物地理位置推断、个体识别等方面。灰尘中包含丰富的环境微生物成分,具有广泛性、异质

性和易迁移性等特点,通过比较分析灰尘中的微生物群落组成,可以实现样本间的比对或样品来源地推断,在一些疑难案件中越来越显现出其良好的物证属性和证据价值^[37]。本研究基于二代测序技术,探索北京、福州、昆明和乌鲁木齐4个城市的灰尘真菌群落自然分布模式,评估基于真菌的ITS2区测序对这4个城市的灰尘样本进行生物地理位置推断的准确率。

已有研究报道显示,真菌群落组成受多种因素的影响^[38-39],这些因素包括地理位置、气候条件、土壤类型和人类活动等。本研究选择中国北京、福州、昆明、乌鲁木齐4个具有明显地理和气候差异的城市,深入探索4个城市的灰尘真菌群落组成。研究结果表明,4个城市间的灰尘真菌组成存在显著差异,其中以福州和乌鲁木齐两个城市之间的差异最为明显(ANOSIM, $R=0.96$; PERMANOVA, $R^2=0.91$, 表3)。4个城市中北京的真菌物种丰富度最高,其次是乌鲁木齐(图1a),这一结果与已有的研究报道相符,支持了真菌物种丰富度随纬度增加而升高这一观点^[39-40]。这一现象的出现可能是因为相较于热带及亚热带地区,温带地区的降水量、温度变化等因素为真菌的生长和繁殖提供了更为有利的环境。在4个城市的灰尘样本中,子囊菌门(Ascomycota)和担子菌门(Basidiomycota)被确定为主要的优势菌群(图4a),这与以往针对环境真菌群落的研究结果高度一致^[16, 41],子囊菌门作为真菌门中最大的亚门,具有极强的抗逆性,能够在多样的环境条件下存活,在自然界中扮演着重要的生态角色。

目前在法庭科学领域基于环境DNA进行地域推断的研究多针对细菌,关于灰尘真菌的研究较少,已有研究显示,相较于细菌,真菌具有更高的地理区分能力^[42]。真菌基因组的ITS区有较高的分类覆盖率,常被用作真菌群落的通用遗传条形码^[43]。然而,整个ITS区域太长,无法基于现有的二代测序技术进行完全测序,因此目前关注于真菌的研究主要针对ITS1或ITS2区域进行。使用不同的标记对真菌群落进行表征会带来结果的偏差,已有研究表明^[44],相较于真菌的ITS1区,ITS2标记能提供更高的真菌丰富度和分辨率。本研究在大空间尺度下,探索中国4个城市的灰尘真菌群落自然分布模式,基于二代测序技术检测真菌基因组的ITS2区,所有已知来源的灰尘样本的生物地理位置均被成功预测,这与已报道的基于灰尘真菌群落

信息进行生物地理推断的研究结果^[8]相比获得了更高的预测准确率。但本研究仅采集了4个城市的灰尘样本,样本数有限,且4个城市间的地理距离较远,在小空间尺度下利用灰尘真菌进行生物地理位置推断的可行性尚待研究,后续可以考虑收集更多城市的灰尘样本,进一步评估基于真菌的ITS2区测序对灰尘样本进行生物地理位置推断的准确率。

4 结 论

本研究基于二代测序技术检测了北京、福州、昆明和乌鲁木齐4个城市的灰尘真菌DNA,以此探索这4个城市的灰尘真菌群落自然分布模式,为公安实战基于环境DNA进行地域推断奠定基础。研究结果表明,4个城市真菌群落组成具有显著差异,随着分类水平的细化,4个城市的物种组成结构和相对丰度的差异越来越明显。所有已知来源的灰尘样本的生物地理位置均被成功预测,基于真菌的ITS2区测序对这4个城市的灰尘样本进行生物地理位置推断具有较高的准确率。

参 考 文 献

- [1] Locard E. The analysis of dust traces. Part I. Am J Police Sci, 1930, 1(3): 276-293
- [2] Hawksworth D L, Wiltshire P. Forensic mycology: current perspectives. Res Rep Forensic Med Sci, 2015, 5: 75-83
- [3] Wiltshire P E J, Hawksworth D L, Webb J A, *et al.* Two sources and two kinds of trace evidence: enhancing the links between clothing, footwear and crime scene. Forensic Sci Int, 2015, 254: 231-242
- [4] Foster N R, Martin B, Hoogewerff J, *et al.* The utility of dust for forensic intelligence: exploring collection methods and detection limits for environmental DNA, elemental and mineralogical analyses of dust samples. Forensic Sci Int, 2023, 344: 111599
- [5] Kraaijeveld K, de Weger L A, Ventayol García M, *et al.* Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol Ecol Resour, 2015, 15(1): 8-16
- [6] Richardson R T, Lin C H, Sponsler D B, *et al.* Application of ITS2 metabarcoding to determine the provenance of pollen collected by honey bees in an agroecosystem. Appl Plant Sci, 2015, 3(1): apps.1400066
- [7] De Deckker P, Munday C I, Brocks J, *et al.* Characterisation of the major dust storm that traversed over eastern Australia in September 2009; a multidisciplinary approach. Aeolian Res, 2014, 15: 133-149
- [8] Grantham N S, Reich B J, Pacifici K, *et al.* Fungi identify the geographic origin of dust samples. PLoS One, 2015, 10(4): e0122605
- [9] Amann R I, Ludwig W, Schleifer K H. Phylogenetic identification and *in situ* detection of individual microbial cells without

- cultivation. *Microbiol Rev*, 1995, **59**(1): 143-169
- [10] Lindahl B D, Nilsson R H, Tedersoo L, *et al.* Fungal community analysis by high-throughput sequencing of amplified markers—a user's guide. *New Phytol*, 2013, **199**(1): 288-299
- [11] Amann R I. Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology. *Mol Ecol*, 1995, **4**(5): 543-554
- [12] 吴悦妮, 冯凯, 厉舒祺, 等. 16S/18S/ITS扩增子高通量测序引物的生物信息学评估和改进. *微生物学通报*, 2020, **47**(9): 2897-2912
Wu Y N, Feng K, Li S Z, *et al.* *Microbiol China*, 2020, **47**(9): 2897-2912
- [13] Gou H, Lu J, Li S, *et al.* Assessment of microbial communities in PM₁ and PM₁₀ of Urumqi during winter. *Environ Pollut*, 2016, **214**: 202-210
- [14] Nishimura Y, Kenzaka T, Sueyoshi A, *et al.* Similarity of bacterial community structure between Asian dust and its sources determined by rRNA gene-targeted approaches. *Microbes Environ*, 2010, **25**(1): 22-27
- [15] Yamaguchi N, Ichijo T, Sakotani A, *et al.* Global dispersion of bacterial cells on Asian dust. *Sci Rep*, 2012, **2**: 525
- [16] Martin-Sanchez P M, Estensmo E L F, Morgado L N, *et al.* Analysing indoor mycobiomes through a large-scale citizen science study in Norway. *Mol Ecol*, 2021, **30**(11): 2689-2705
- [17] Niemeier-Walsh C, Ryan P H, Meller J, *et al.* The mycobiomes and bacteriomes of sputum, saliva, and home dust. *Indoor Air*, 2021, **31**(2): 357-368
- [18] 杨琪, 康克莱, 梅宏成, 等. 环境DNA检验及地域来源推断研究进展. *中国法医学杂志*, 2024, **39**(3): 349-356
Yang Q, Kang K L, Mei H C, *et al.* *Chin J Forensic Med*, 2024, **39**(3): 349-356
- [19] Adams R I, Bateman A C, Bik H M, *et al.* Microbiota of the indoor environment: a meta-analysis. *Microbiome*, 2015, **3**: 49
- [20] Park J, Li P F, Ichijo T, *et al.* Effects of Asian dust events on atmospheric bacterial communities at different distances downwind of the source region. *J Environ Sci*, 2018, **72**: 133-139
- [21] Täubel M, Jalanka J, Kirjavainen P V, *et al.* Fungi in early-life house dust samples and the development of asthma: a birth cohort study. *Ann Am Thorac Soc*, 2023, **20**(10): 1456-1464
- [22] MacDonald L M, Singh B K, Thomas N, *et al.* Microbial DNA profiling by multiplex terminal restriction fragment length polymorphism for forensic comparison of soil and the influence of sample condition. *J Appl Microbiol*, 2008, **105**(3): 813-821
- [23] Young J M, Weyrich L S, Breen J, *et al.* Predicting the origin of soil evidence: high throughput eukaryote sequencing and MIR spectroscopy applied to a crime scene scenario. *Forensic Sci Int*, 2015, **251**: 22-31
- [24] Barberán A, Ladau J, Leff J W, *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci USA*, 2015, **112**(18): 5756-5761
- [25] Blackwell M. The fungi: 1, 2, 3.. 5.1 million species?. *Am J Bot*, 2011, **98**(3): 426-438
- [26] Allwood J S, Fierer N, Dunn R R, *et al.* Use of standardized bioinformatics for the analysis of fungal DNA signatures applied to sample provenance. *Forensic Sci Int*, 2020, **310**: 110250
- [27] 杨琪, 彭加金, 王乐, 等. 物体表面灰尘采集及DNA提取方法优化研究. *生物化学与生物物理进展*, 2024, **51**(12): 3310-3320
Yang Q, Peng J J, Wang L, *et al.* *Prog. Biochem. Biophys*, 2024, **51**(12): 3310-3320
- [28] Wingett S W, Andrews S. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res*, 2018, **7**: 1338
- [29] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*, 2011, **17**(1): 10
- [30] Magoč T, Salzberg S L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 2011, **27**(21): 2957-2963
- [31] Chen S, Zhou Y, Chen Y, *et al.* Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 2018, **34**(17): i884-i890
- [32] Edgar R C, Haas B J, Clemente J C, *et al.* UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 2011, **27**(16): 2194-2200
- [33] Edgar R C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, 2016. DOI: 10.1101/081257
- [34] Kõljalg U, Larsson K H, Abarenkov K, *et al.* UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol*, 2005, **166**(3): 1063-1068
- [35] Chang F, He S, Dang C. Assisted selection of biomarkers by linear discriminant analysis effect size (LEfSe) in microbiome data. *JoVE*, 2022(183): e61715
- [36] Knights D, Kuczynski J, Charlson E S, *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat Methods*, 2011, **8**(9): 761-763
- [37] 许敏敏, 时玉, 刘洋, 等. 土壤微生物在法庭科学领域中的应用研究进展. *微生物学通报*, 2021, **48**(10): 3823-3834
Xu M M, Shi Y, Liu Y, *et al.* *Microbiol China*, 2021, **48**(10): 3823-3834
- [38] Talbot J M, Bruns T D, Taylor J W, *et al.* Endemism and functional convergence across the North American soil mycobiome. *Proc Natl Acad Sci USA*, 2014, **111**(17): 6341-6346
- [39] Tedersoo L, Bahram M, Põlme S, *et al.* Fungal biogeography. Global diversity and geography of soil fungi. *Science*, 2014, **346**(6213): 1256688
- [40] Amend A S, Seifert K A, Samson R, *et al.* Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc Natl Acad Sci USA*, 2010, **107**(31): 13748-13753
- [41] Li P D, Jeewon R, Aruna B, *et al.* Metabarcoding reveals differences in fungal communities between unflooded versus tidal flat soil in coastal saline ecosystem. *Sci Total Environ*, 2019, **690**: 911-922
- [42] Young J M, Weyrich L S, Cooper A. Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers. *Forensic Sci Int Genet*, 2014, **13**: 176-184
- [43] Schoch C L, Seifert K A, Huhndorf S, *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci USA*, 2012, **109**(16): 6241-6246
- [44] Li S, Deng Y, Wang Z, *et al.* Exploring the accuracy of amplicon-based internal transcribed spacer markers for a fungal community. *Mol Ecol Resour*, 2020, **20**(1): 170-184

Geographical Inference Study of Dust Samples From Four Cities in China Based on ITS2 Sequencing*

ZHANG Wen-Jun^{1,2)}, FENG Yao-Sen²⁾, PENG Jia-Jin^{2,3)}, FENG Kai⁴⁾, Deng Ye⁴⁾,
KANG Ke-Lai²⁾**, WANG Le^{1,2)}**

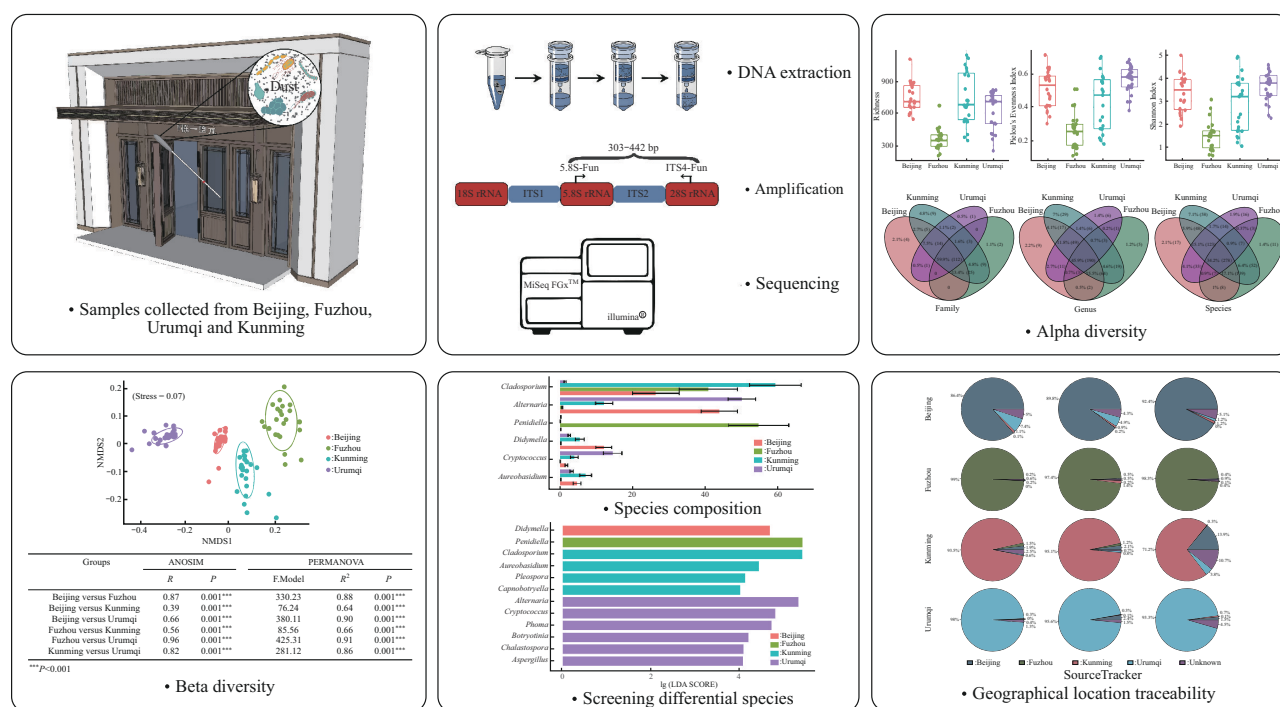
⁽¹⁾School of Forensic Medicine, Shanxi Medical University, Jinzhong 030600, China;

⁽²⁾Key Laboratory of Forensic Genetics of Ministry of Public Security, Institute of Forensic Science, Ministry of Public Security, Beijing 100038, China;

⁽³⁾School of Forensic Medicine, Kunming Medical University, Kunming 650500, China;

⁽⁴⁾CAS Key Laboratory of Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China)

Graphical abstract



Abstract Objective In the realm of forensic science, dust is a valuable type of trace evidence with immense potential for intricate investigations. With the development of DNA sequencing technologies, there is a heightened interest among researchers in unraveling the complex tapestry of microbial communities found within

* This work was supported by grants from the Institute of Forensic Science, Ministry of Public Security of China (2022JB022) and Ministry of Public Security of China (2023JC14).

** Corresponding author.

KANG Ke-Lai. Tel: 86-10-83752707, E-mail: kang_kl@yeah.net

WANG Le. Tel: 86-10-83752707, E-mail: wangle_02@163.com

Received: July 7, 2024 Accepted: November 21, 2024

dust samples. Furthermore, striking disparities in the microbial community composition have been noted among dust samples from diverse geographical regions, heralding new possibilities for geographical inference based on microbial DNA analysis. The pivotal role of microbial community data from dust in geographical inference is significant, underscoring its critical importance within the field of forensic science. This study aims to delve deeply into the nuances of fungal community composition across the urban landscapes of Beijing, Fuzhou, Kunming, and Urumqi in China. It evaluates the accuracy of biogeographic inference facilitated by the internal transcribed spacer 2 (ITS2) fungal sequencing while concurrently laying a robust foundation for the operational integration of environmental DNA into geographical inference mechanisms. **Methods** ITS2 region of the fungal genomes was amplified using universal primers known as 5.8S-Fun/ITS4-Fun, and the resulting DNA fragments were sequenced on the Illumina MiSeq FGx platform. Non-metric multidimensional scaling analysis (NMDS) was employed to visually represent the differences between samples, while analysis of similarities (ANOSIM) and permutational multivariate analysis of variance (PERMANOVA) were utilized to statistically evaluate the dissimilarities in community composition across samples. Furthermore, using Linear Discriminant Analysis Effect Size (LEfSe) analysis to identify and filter out species that exhibit significant differences between various cities. In addition, we leveraged SourceTracker to predict the geographic origins of the dust samples. **Results** Among the four cities of Beijing, Fuzhou, Kunming and Urumqi, Beijing has the highest species richness. The results of species annotation showed that there were significant differences in the species composition and relative abundance of fungal communities in the four cities. NMDS analysis revealed distinct clustering patterns of samples based on their biogeographic origins in multidimensional space. Samples from the same city exhibited clear clustering, while samples from different cities showed separation along the first axis. The results from ANOSIM and PERMANOVA confirmed the significant differences in fungal community composition between the four cities, with the most pronounced distinctions observed between Fuzhou and Urumqi. Notably, the biogeographic origins of all known dust samples were successfully predicted. **Conclusion** Significant differences are observed in the fungal species composition and relative abundance among the cities of Beijing, Fuzhou, Kunming, and Urumqi. Employing fungal ITS2 sequencing on dust samples from these urban areas enables accurate inference of biogeographical locations. The high feasibility of utilizing fungal community data in dust for biogeographical inferences holds particular promise in the field of forensic science.

Key words forensic sciences, fungus, dust DNA, regional inference, high-throughput sequencing

DOI: 10.16476/j.pibb.2024.0300

CSTR: 32369.14.pibb.20240300