

DOI: 10.19663/j.issn2095-9869.20200519001

http://www.yykxjz.cn/

牟铭, 李昂, 赵新宁, 柳淑芳, 庄志猛. 人工模拟条件下环境 DNA 宏条形码技术的定量分析初探. 渔业科学进展, 2021, 42(5): 24–30

MU M, LI A, ZHAO X N, LIU S F, ZHUANG Z M. Preliminary study on the quantitative analysis of environmental DNA metabarcoding under ideal conditions. Progress in Fishery Sciences, 2021, 42(5): 24–30

人工模拟条件下环境 DNA 宏条形码 技术的定量分析初探*

牟 铭^{1,2} 李 昂² 赵新宁^{1,2} 柳淑芳^{2①} 庄志猛²

(1. 上海海洋大学水产与生命学院 上海 201306; 2. 中国水产科学研究院黄海水产研究所 青岛海洋科学与技术试点国家实验室海洋渔业科学与食物产出过程功能实验室 山东 青岛 266071)

摘要 近年来,环境 DNA 宏条形码技术(eDNA metabarcoding)在水生生态系统生物多样性评估等相关领域中得到广泛应用,因其具有快速测算群落中物种丰度的潜能,eDNA 宏条形码技术成为资源保护和管理中颇具应用前景的调查工具。虽然大量证据表明 eDNA 高通量测序获得的 reads 数与自然环境中生物相对数量具有相关性,但一直不能得到明确的量化关系结果。eDNA 的富集、扩增过程中的偏倚等诸多不确定因素,制约了该技术在生物资源调查领域的推广应用。假定水体中的 eDNA 全部回收,且 PCR 扩增时不存在引物偏倚性,这种理想状态下的水体中 eDNA 组成与其高通量测序 reads 数是否存在线性关系?为此,本研究在实验室可控条件下,选择 2 个同属近缘的凡纳滨对虾(*Penaeus vannamei*)和墨吉对虾(*Penaeus merguensis*),对其 DNA 样品进行不同比例混合,模拟从自然水体中富集到的 eDNA 复合样品,既保证了样品的回收率,又降低了引物偏倚的干扰。以此为模板,探究 eDNA 宏条形码技术检测种群相对数量的准确性。结果显示,当 2 个物种 DNA 模板浓度比例为 1:1 时,高通量测序结果注释得到的 2 个物种 reads 数比值为 13/24(墨吉对虾/凡纳滨对虾),可见,即使是同属近缘种间依然存在轻微的引物偏倚现象,引物偏移率为 1.5%。同时,根据 7 个实验组获得的高通量测序结果注释得到的 2 个物种 reads 数比值与对应模板中 2 个物种 DNA 浓度比值之间的线性回归分析表明,水体中 eDNA 组成与其高通量测序 reads 数间呈明显线性关系,即 $y=0.0716x+0.7043$ ($r^2=0.9824$)。综上所述,本研究为验证 eDNA 宏条形码技术监测水生生物资源量的可行性提供了直接证据,也为后续 DNA 宏条形码技术的定量研究提供了思路。

关键词 环境 DNA 宏条形码; PCR 引物偏好; 高通量测序; 线性函数关系; 生物学调查
中图分类号 S932.2 **文献标识码** A **文章编号** 2095-9869(2021)05-0024-07

环境 DNA 宏条形码技术(environmental DNA metabarcoding)可以通过高通量测序对环境 DNA (eDNA)进行多物种分类鉴定。随着 eDNA 宏条形码技术的应用和发展,该技术对群落中单一或组合物种

* 山东省支持青岛海洋科学与技术试点国家实验室重大科技专项(2018SDKJ0501)、国家自然科学基金项目(31502168)和山东省泰山学者建设工程专项经费项目共同资助 [This work was supported by the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (2018SDKJ0501), National Natural Science Foundation of China (31502168), and Special Funds for Taishan Scholar Project of Shandong Province]. 牟 铭, E-mail: 1065028922@163.com

① 通讯作者: 柳淑芳, 研究员, E-mail: liusf@ysfri.ac.cn

收稿日期: 2020-05-19, 收修改稿日期: 2020-05-22

的定量评估逐渐显现出优势。无论是水族箱或中观实验等小型生态系统(Minamoto *et al*; Moyer *et al*, 2014; Pilliod *et al*, 2013; Thomsen *et al*, 2012; Doi *et al*, 2017), 还是江河湖泊等自然淡水系统(Doi *et al*, 2017; Lacoursiereroussel *et al*, 2016), 甚至是开放式的海洋系统(Jo *et al*, 2017), 研究人员都尝试过使用 eDNA 宏条形码技术测算种群相对数量。大量研究表明, 特定环境中物种的个体数量和基于 eDNA 宏条形码技术获得的高通量测序 reads 数之间具有明显的相关性。Evans 等(2016)设计了一个包含鱼类和两栖类的中观实验, 证明水体中物种个体数或生物量相对丰度与 eDNA 宏条形码技术获得的高通量测序 reads 数呈正相关; Hanfling 等(2016)对比较了一个天然湖泊的长期监测鱼类丰度结果与 eDNA 宏条形码技术得到的高通量测序 reads 数之间的相关性, 结果表明二者呈正相关; Thomsen 等(2016)研究发现, 当将鱼类的 eDNA 高通量测序 reads 数聚类到“科”的分类等级时, eDNA 宏条形码技术测得的深海生境中的生物量与传统拖网捕捞数据比较吻合。可以预见, eDNA 宏条形码技术具有快速测算群落中物种丰度的潜能, 将成为资源保护和管理中具有应用前景的调查工具。

然而, 通过对已发表的相关文献进行梳理和归纳, 发现当前大多数利用 eDNA 高通量测序结果的 reads 数来评估自然环境中生物相对数量的研究并不能得到明确的量化关系结果(Lim *et al*, 2016)。究其原因, 是因为从生物到 eDNA 以及从 eDNA 再到高通量测序结果的 2 个过程中的数学关系难以理清。其中, 实验室操作过程中的主要问题是: (1)自然水体中, 不同浓度 eDNA 富集效率不同(Kelly *et al*, 2016); (2)PCR 扩增过程中, 通用引物对不同物种 DNA 存在偏倚性(Elbrecht *et al*, 2015; Pinol *et al*, 2015)。假定水体中的 eDNA 全部回收, 且 PCR 扩增时不存在引物偏倚性, 这种理想状态下的水体中 eDNA 组成与其高通量测序 reads 数是否完全呈线性关系? 为了探究这个问题, 本研究在实验室可控条件下, 选择 2 个同属近缘种, 对其 DNA 样品进行不同比例混合, 模拟从自然水体中富集到的 eDNA 复合样品, 既保证了样品的回收率, 又降低了引物偏倚的干扰。以此为模板, 探究 eDNA 宏条形码技术检测种群相对数量的准确性, 旨在为验证 eDNA 宏条形码技术监测水生生物资源量的可行性提供直接证据。

1 材料与方法

1.1 实验方案设计

1.1.1 实验样品的选择

鉴于通用引物对不同物

种的扩增效率有所不同, 本研究实验样品选取 2 个同属近缘物种用于制备模拟 eDNA 样品, 以确保将 PCR 过程中因引物对物种 DNA 的偏移引起的实验误差降低到最小程度。

实验样品选择了 2 种常见对虾: 凡纳滨对虾(*Penaeus vannamei*)和墨吉对虾(*Penaeus merguensis*), 它们同为节肢动物门(Arthropoda)、软甲纲(Malacostraca)、十足目(Decapoda)、对虾科(Penaeidae)、对虾属(*Penaeus*), 是较为常见的同属近缘种。

1.1.2 模板 DNA 的提取 采用传统的酚-氯仿-异戊醇方法分别提取 2 种对虾肌肉组织 DNA, 经纯化后使用超微量核酸蛋白分析仪(GE, 美国)检测浓度, 稀释 DNA 浓度至 20 ng/ μ L ($\pm 1\%$), -20°C 保存备用。

1.1.3 模拟 eDNA 的制备 人工模拟 eDNA 复合样品, 共设 7 个实验组: 将凡纳滨对虾和墨吉对虾的 DNA 模板分别按照 1:1、1:2、1:4、1:8、1:10、1:50 和 1:100 的比例均匀混合, 且混合后 DNA 终浓度一致。设阳性对照 2 组: 分别选用凡纳滨对虾和墨吉对虾 DNA 样品。设阴性对照 1 组: 用纯水代替 DNA 样品。每组样品设 3 个重复。

将上述制备的 eDNA 复合样品作为 eDNA 宏条形码检测的模板 DNA。

1.2 通用引物的选取

考虑到后续高通量测序及 DNA 条形码数据库信息资源的丰富度, 本研究选取了 Leray 等(2013)推荐的长度为 313 bp 的线粒体 CO I 基因序列作为待测 eDNA 目标片段, 该片段长度适中、数据资源丰富、物种鉴定特异性好, 且引物通用性强。通用引物序列见表 1, 由华大基因有限公司合成。

表 1 CO I 基因 PCR 扩增引物信息

Tab.1 Primer sets of CO I used for PCR amplification

引物 Primer	序列 Sequence (5'~3')	退火温度 $T_m/^{\circ}\text{C}$
mICO I intF	GGWACWGGWTGAACWGTWTA YCCYCC	62
jgHCO2198	TAIACYTCIGGRTGICRAARAAY CA	61

1.3 PCR 扩增

PCR 扩增体系: 2 \times Rapid *Taq* Master Mix (Vazyme) 25 μ L, 上下游引物(mICO I intF/jgHCO2198) (10 mmol/L)各 2 μ L, DNA 模板 5 μ L, 无菌水补齐总体积至 50 μ L。

为了增强扩增特异性, 使用 Touch down PCR 反

应程序: 95℃预变性 10 min; 95℃变性 30 s, 62℃退火 30 s, 72℃延伸 30 s, 16 个循环; 95℃变性 30 s, 50℃退火 30 s, 72℃延伸 30 s, 25 个循环; 72℃延伸 3 min; 4℃保存。

用 1%琼脂糖凝胶电泳检测 PCR 产物质量, 选用 DL2000 DNA marker, 目标条带约为 300 bp。挑选电泳条带单一且明亮的 PCR 产物用于高通量测序。

1.4 测序及结果注释

将所得 PCR 产物送至青岛欧易生物科技有限公司, 进行高通量测序。采用 Illumina Miseq 平台测序。

按照 97%相似性对非重复序列(不含单序列)进行 OTU(operational taxonomic units)聚类, 在聚类过程中去除嵌合体(chimera, 融合了 2 个不同序列信息的序列)、异源双链序列等干扰序列, 得到 OTU 聚类结果, 然后对不同 OTU 的代表序列进行结果注释。在结果注释时, 参照 NCBI 公共数据库(<https://www.ncbi.nlm.nih.gov/>)来进行分类学信息的标注, 同时使用中国重要渔业生物 DNA 条形码信息平台(<http://www.fisherybarcode.cn>)对注释结果进行辅助校正。

1.5 数据处理

统计每个样品高通量测序的原始 reads 数, 以所有样本中 reads 数最低值为基准, 进行抽平, 使每个样品测序结果的 reads 数相同, 用于后续数据分析。

在推导数学关系时, 设: eDNA 样品来自 2 个物种, 分别为 A、B; 实际参与 PCR 反应的底物模板 DNA 的量为 n ; 参与 PCR 反应的底物模板 DNA 理论量为 x ; 每个循环中, 参与扩增的 2 个物种底物模板 DNA 比率为 λ ; 扩增的循环数为 c ; PCR 产物的高通量测序 reads 数为 y 。假定理想情况为: 在 PCR 扩增的不同循环中, 参与扩增的 2 个物种底物模板 DNA 比率固定; PCR 扩增全程都保持底数为 2 的指数增长型(反应物足够), 则高通量测序 reads 数可以反应 eDNA 样品中 2 个物种的相对量, 即推得公式(1):

$$\frac{y_A}{y_B} = \frac{2^c \times n_A}{2^c \times n_B} = \lambda^c \times \frac{x_A}{x_B} \quad (1)$$

经变形得公式(2):

$$y = \lambda^c \cdot x \quad (2)$$

2 结果与分析

2.1 PCR 扩增结果

以人工模拟的 7 组 eDNA 复合样品及对照组样品为模板, 使用 eDNA 宏条形码通用引物进行 PCR 扩

增, 经 1%琼脂糖凝胶电泳检测, 结果显示, 阳性对照和 7 个实验组均成功扩增出 300 bp 左右的目标片段, 阴性对照无扩增产物(图 1)。

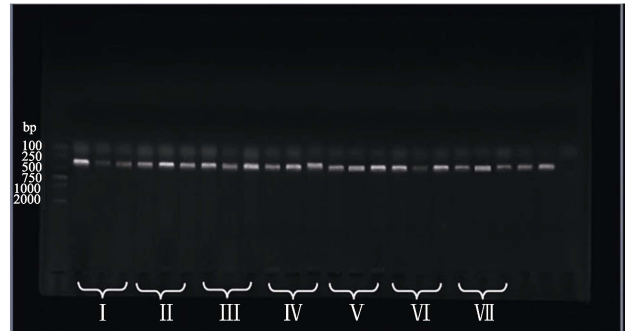


图 1 琼脂糖凝胶电泳检测结果
Fig.1 Results of agarose gel electrophoresis

Marker: DNA marker DL2000; I~VII分别为实验组 I~VII 的 PCR 产物, 每组各 3 个重复; 右起第 1 泳道为阴性对照, 第 2~3 泳道分别为 2 个阳性对照

Marker: DNA marker DL2000; Lanes I~VII: PCR products of the 7 samples, with 3 parallel groups; Lane 1(From right): Negative control; Lane 2~3(From right): Positive control

2.2 测序数据

2.2.1 高通量测序结果 本研究对 7 个实验组共 21 个 PCR 产物进行了高通量测序。统计高通量测序获得的序列长度分布, 显示 99.80%以上序列的长度分布于 301~350 bp 之间(图 2)。抽平后共计获得有效序列 674,205 条, 序列平均长度为 312.41 bp。

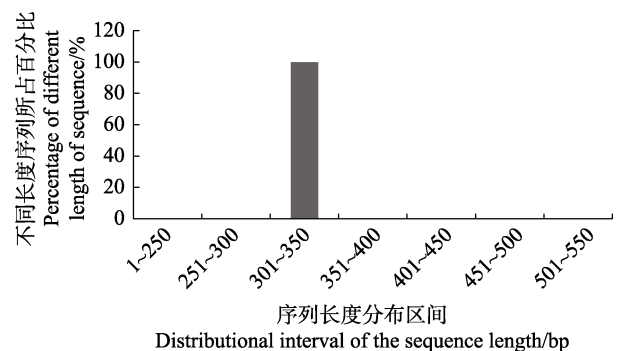


图 2 测序数据长度分布统计

Fig.2 Result of distributive interval of the sequence length

2.2.2 分类学信息注释及 reads 数统计结果 将每个样品 reads 总数抽平后, 对 674,205 条序列进行分类学信息注释, 保留有效序列(序列长度在 301~350 之间, 且分类学信息注释结果是墨吉对虾和凡纳滨对虾)的 reads。各实验组高通量测序 reads 数及其分类注释结果详见表 2。实验组 I 的 3 组平行样共获得 94,774 条序列, 每个样品注释出的墨吉对虾和凡纳滨

对虾平均 reads 数分别为 11,097 和 20,494, 二者的比值为 13/24; 实验组 II 的 3 组平行样共获得 95,208 条序列, 二者 reads 数比值为 18/23; 实验组 III 的 3 组平行样共获得 95,572 条序列, 二者 reads 数比值为 77/73; 实验组 IV 的 3 组平行样共获得 95,250 条序列,

二者 reads 数比值为 67/44; 实验组 V 的 3 组平行样共获得 94,408 条序列, 二者 reads 数比值为 43/24; 实验组 VI 的 3 组平行样共获得 94,640 条序列, 二者 reads 数比值为 127/35; 实验组 VII 的 3 组平行样共获得 95,526 条序列, 二者 reads 数比值为 187/23。

表 2 各实验组获得的序列数及分类学信息注释
Tab.2 Results of taxonomic information and numbers of reads

项目 Items	组别 Groups							分类注释结果 Taxonomic results
模板 DNA 浓度比(墨吉对虾/凡纳滨对虾) Ratio of DNA template (<i>P. merguensis</i> / <i>P. vannamei</i>)	I	II	III	IV	V	VI	VII	
模板 DNA 浓度比(墨吉对虾/凡纳滨对虾) Ratio of DNA template (<i>P. merguensis</i> / <i>P. vannamei</i>)	1/1	2/1	4/1	8/1	10/1	50/1	100/1	—
3 组平行样的 Reads 数 Reads of three duplicates	11,097	13,928	16,353	19,165	20,197	24,731	28,355	墨吉对虾 <i>P. merguensis</i>
	20,494	17,808	15,505	12,585	11,272	6816	3487	凡纳滨对虾 <i>P. vannamei</i>
Reads 平均数比值(墨吉对虾/凡纳滨对虾) Ratio of average reads (<i>P. merguensis</i> / <i>P. vannamei</i>)	13/24	18/23	77/73	67/44	43/24	127/35	187/23	—

2.2.3 混合模板 PCR 扩增的引物偏倚率 当模板中 2 个物种 DNA 浓度比例为 1 : 1 (墨吉对虾/凡纳滨对虾)时, 2 个物种的模板 DNA 含量相同, 扩增过程中主要是引物对不同底物的偏好引起的扩增产物差别, 此时主要是引物偏倚引起的 reads 数差距。根据公式(2)变形得:

$$\lambda = \sqrt{\frac{y}{x}} \quad (3)$$

代入扩增循环数, 由 reads 平均数比值, 计算得到每个循环中参与反应的 2 个物种模板 DNA 的平均比率为 66/67, 换言之, PCR 扩增的每个循环中, 平均偏倚率为 1/67, 约为 1.5%, 即该引物在同一 PCR 体系中, 每次退火过程对墨吉对虾 DNA 模板的结合率比凡纳滨对虾 DNA 模板大约少 1.5%。

2.2.4 eDNA 组成与高通量测序 reads 数的函数关系

以各实验组设置的底物中墨吉对虾 DNA 与凡纳滨对虾 DNA 的比值为横坐标, 以各实验组样品高通量测序结果注释的墨吉对虾与凡纳滨对虾 reads 平均数的比值为纵坐标, 绘制 eDNA 组成与高通量测序 reads 数的函数关系图。在公式(2)的前提假设条件下, 在高通量结果中, 2 种对虾的 reads 数之比和初始 PCR 模板 DNA 中 2 种对虾的 DNA 比例应呈现 $y=kx$ 线性关系, 对表 2 中的 7 个实验组的 reads 平均数比值和 PCR 模板 DNA 比值的 2 组数据进行一元线性回归分析, 最终得出 eDNA 组成与高通量测序 reads 数的函

数关系: $y=0.0716x+0.7043$ ($r^2=0.9824$)(图 3)。

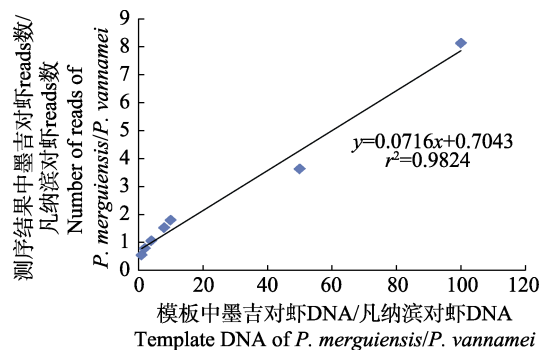


图 3 2 个物种高通量测序结果的序列数比值与模板比例的关系

Fig.3 The curve of ratio of two species' reads and ratio of template DNA

3 讨论

利用 eDNA 宏条形码技术来评估水体中不同物种的相对数量, 是该技术最具潜力的应用前景之一。与传统调查方法相比, eDNA 宏条形码技术具有操作简便、省时高效且对自然资源破坏小等特点。如何使用该技术来监测水体中的生物量, 一直是学界关注的焦点, 但水体中的 eDNA 难以完全富集, 且富集到的 eDNA 成分复杂; PCR 扩增时, 引物对不同种类 DNA 模板的偏好还会导致扩增效率不同, 这些问题亦是

eDNA 宏条形码技术推广应用的最大障碍。为运用 eDNA 宏条形码技术检测环境生物相对数量而设计的中观实验,大都是在尝试推导生物量与高通量测序结果 reads 数的数学关系,主要包括环境中生物量与 eDNA 量、环境中 eDNA 量与提取的 eDNA 量和提取的 eDNA 模板与 PCR 产物的高通量测序结果等 3 个对应关系,其中任何一个关系不明确,都将影响 eDNA 宏条形码技术的检测结果,所以,必须逐步理清这 3 种对应关系的函数关系,进而得到“高鲁棒性”(模式体系固定且成熟,效果准确且稳定)的 eDNA 宏条形码方法应用体系。

分解困难是解决问题的有效方法之一,剖析 DNA 宏条形码技术全过程,当只关注富集到的 eDNA 扩增及测序时,如果能创造一种 eDNA 成分简单、无引物偏好影响的理想条件,进而探索 eDNA 高通量测序 reads 数和底物 eDNA 含量之间的数学关系,便可以收获“柳暗花明”的结果。

3.1 混合模板 PCR 扩增的引物偏倚性

影响多物种 eDNA 混合模板 PCR 扩增效率的主要原因之一,便是 PCR 过程中引物对不同模板 DNA 碱基组成存在偏好性(Kanagawa, 2003)。理论上,引物的每个脱氧核苷酸和模板 DNA 链上同源区段的每个脱氧核苷酸可以实现一一对应的关系,但实践过程中通常难以实现,任何一个错配位点的出现都会导致相应的扩增效率降低,从而减少产物。为此,通用引物中加入了简并碱基以降低错配的影响。然而,在进行复杂模板的 PCR 扩增时,即使使用含简并碱基的通用引物,引物结合能力也会因为不同模板 DNA 引物结合区段序列的 GC 碱基含量高低而不同(Acinas *et al.*, 2005), GC 含量高的引物结合力较强(Fonseca, 2018)。为了提高模板 DNA 引物结合区段碱基组成的一致性,从而降低引物偏倚的影响,本研究选择较为常见的同属近缘种凡纳滨对虾和墨吉对虾作为实验对象。同时,基于基因的进化速率及其数据库信息的丰富程度等因素,决定以 mtDNA 中的 CO I 基因片段作为检测的目的基因。

通过控制上述引物偏好的条件,最终高通量测序结果中引物的偏倚程度如何。研究分析发现,当 eDNA 模板中凡纳滨对虾和墨吉对虾 DNA 组成比例为 1:1 时,高通量测序获得对应物种 reads 数比例为 13/24。根据扩增循环数和公式(2),推导此时参与扩增的 2 种 DNA 比约为 66/67,即 PCR 扩增的平均偏倚率为 1/67,约 1.5%。可见选择凡纳滨对虾和墨吉对虾制备简单 eDNA 混合模板产生的引物偏倚程度

影响较小,符合最初的实验设计预期。此时,2 个物种的模板浓度相同,高通量测序结果的偏倚主要是由于引物对不同模板的“偏好”程度不同产生的。在后续 eDNA 宏条形码技术的数量性研究探索中,设计实验时可选用近源种作为实验对象,从而降低扩增过程中引物对不同模板偏好性不同而导致的“偏倚”现象。

3.2 eDNA 组成与高通量测序 reads 数的相关性

当面对实际环境样本时,如果直接应用 eDNA 宏条形码技术分析高通量测序 reads 数与待测样本生物量的关系,从生物到 eDNA 以及从 eDNA 再到高通量测序结果的两个过程中的数学关系则难以理清。鉴于此,本研究采取“化繁为简”的策略,主要关注 eDNA 与高通量测序 reads 数之间的关系。同时,考虑到复合底物的 PCR 过程比较复杂(Harper *et al.*, 2019),最终注释到每个物种的测序 reads 数都受其反应体系的影响,故直接比较样本间的 reads 数会产生一定偏差。充分考虑这些影响因素,本研究采用了样本内的 2 个物种 reads 数比值,用于确定其与模板 eDNA 浓度的相关性。

统计分析实验组的 reads 平均数比值与 PCR 模板 eDNA 比值 2 组数据,进行一元线性回归分析,最终得出 eDNA 组成与高通量测序 reads 数的函数关系: $y = 0.0716x + 0.7043$ ($r^2=0.9824$)。等式中的斜率可反应出 PCR 扩增过程中的偏倚程度(主要是引物偏倚),当扩增循环数相同,斜率越小,扩增过程中偏倚程度越大。等式中的截距主要是实验过程中的误差或者污染导致的。该等式可描述为:当环境水样中只存在墨吉对虾和凡纳滨对虾的 DNA 时,可通过高通量测序的 reads 数推算出 2 个物种的相对生物量。

在实际操作中,可能由于模板 eDNA 的回收率存在差异,使之不能完全符合预设的比例;另外,PCR 扩增时并不是全程以底数为 2 的指数式增长,且存在其他不可避免的系统误差,这些因素都会导致上述线性关系式出现一定偏离。但实验结果已充分表明,高通量测序 reads 数与初始模板 eDNA 含量呈线性正相关。

参 考 文 献

- ACINAS S G, SARMA-RUPAVTARM R, KLEPAC-CERAJ V, *et al.* PCR- induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, 2005, 71(12): 8966–8969
- DOI H, INUI R, AKAMATSU Y, *et al.* Environmental DNA analysis for estimating the abundance and biomass of stream

- fish. *Freshwater Biology*, 2017, 62(1): 30–39
- ELBRECHT V, LEESE F. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass: Sequence relationships with an innovative metabarcoding protocol. *PLoS One*, 2015, 10(7): e0130324
- EVANS N T, OLDS B P, RENSHAW M A, *et al.* Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 2016, 16(1): 29–41
- FONSECA V G. Pitfalls in relative abundance estimation using eDNA metabarcoding. *Molecular Ecology Resources*, 2018, 18(5): 923–926
- HANFLING B, HANDLEY L L, READ D S, *et al.* Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, 2016, 25(13): 3101–3119
- HARPER L R, BUXTON A S, REES H C, *et al.* Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, 2019, 826(1): 25–41
- JO T, MURAKAMI H, MASUDA R, *et al.* Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*, 2017, 17(6): e25–e33
- KANAGAWA T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering*, 2003, 96(4): 317–323
- KELLY R P, PORT J A, YAMAHARA K M, *et al.* Harnessing DNA to improve environmental management. *Science*, 2014, 344(6191): 1455–1456
- LACOURSIEREROUSSSEL A, COTE G, LECLERC V, *et al.* Quantifying relative fish abundance with eDNA: A promising tool for fisheries management. *Journal of Applied Ecology*, 2016, 53(4): 1148–1157
- LERAY M, YANG J Y, MEYER C P, *et al.* A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 2013, 10(1): 1–14
- LIM N K M, TAY Y C, SRIVATHSAN A, *et al.* Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science*, 2016, 3(11): 160635
- MINAMOTO T, YAMANAKA H, TAKAHARA T, *et al.* Surveillance of fish species composition using environmental DNA. *Limnology*, 2012, 13(2): 193–197
- MOYER G R, DIAZFERGUSON E E, HILL J E, *et al.* Assessing environmental DNA detection in controlled lentic systems. *PLoS One*, 2014, 9(7): e103767
- PILLIOD D S, GOLDBERG C S, ARKLE R S, *et al.* Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 2013, 70(8): 1123–1130
- PINOL J, MIR G, GOMEZPOLO P, *et al.* Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, 2015, 15(4): 819–830
- THOMSEN P F, KIELGAST J, IVERSEN L, *et al.* Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 2012, 21(11): 2565–2573
- WILLERSLEV E. Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS One*, 2016, 11(11): e0165252

(编辑 冯小花)

Preliminary Study on the Quantitative Analysis of Environmental DNA Metabarcoding under Ideal Conditions

MU Ming^{1,2}, LI Ang², ZHAO Xinning^{1,2}, LIU Shufang^{2①}, ZHUANG Zhimeng²

(1. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China; 2. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences; Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory of Marine Science and Technology (Qingdao), Qingdao, Shandong 266071, China)

Abstract In recent years, environmental DNA (eDNA) metabarcoding has been widely used in the field of biological investigation, which includes species detection and biodiversity assessment. eDNA metabarcoding has the potential to rapidly assess species abundances in communities, making it a promising investigation tool in resource conservation and management. Although previous reports concerning eDNA metabarcoding found that the high-throughput sequencing (HTS) reads were related to the biomass in the natural environmental sample, a clear quantitative relationship was not found in this study. In the field and laboratory, the enrichment efficiency of eDNA is difficult to evaluate. Meanwhile, primer bias inevitably occurs in the eDNA amplification process, resulting in uncertainty in the eDNA HTS results, which restricts the application of eDNA metabarcoding for biological resource investigations. Assuming that the eDNA is completely recovered, and there is no primer bias during PCR amplification, an ideal state for deciphering whether there is a linear relationship between the eDNA and HTS reads is created. In this study, under controlled conditions in the laboratory, a sister group was selected (*Penaeus vannamei* and *Penaeus merguensis*) and their DNA samples were mixed in different proportions to simulate eDNA samples enriched from natural waters. In this way, the recovery of the sample and primer bias was at optimal levels. Then, this eDNA template was used to explore the accuracy of eDNA metabarcoding in detecting species biomass. The results showed that when the concentration ratio of the DNA templates of two species was 1:1, the HTS ratio of the two species was 13/24 (*P. merguensis*/*P. vannamei*). Therefore, even between the closest relatives there is still a slight primer bias (primer migration rate: 1.5%). At the same time, the HTS results from the seven test groups showed an obvious linear relationship between the composition of eDNA in the water and the number of high-throughput sequencing reads, that is, $y=0.0716x+0.7043$ ($r^2=0.9824$). In summary, this study provides direct evidence to verify the feasibility of eDNA metabarcoding in monitoring aquatic biological resources, and also provides ideas for the subsequent quantitative study of DNA metabarcoding.

Key words eDNA metabarcoding; PCR primer bias; High-throughput sequencing; Linear relationship; Biological investigation

① Corresponding author: LIU Shufang, E-mail: liusf@ysfri.ac.cn