



研究报告

16S/18S/ITS 扩增子高通量测序引物的生物信息学评估和改进

吴悦妮^{1,2} 冯凯^{1,2} 厉舒祯^{1,3} 王朱琚^{1,2} 张照婧^{3,4} 邓晔^{*1,2,4}

1 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

2 中国科学院大学资源与环境学院 北京 100049

3 大连理工大学环境学院 工业生态与环境工程教育部重点实验室 辽宁 大连 116024

4 山东大学海洋研究院 山东 青岛 266237

摘要:【背景】在过去的十几年里,基于核糖体 RNA 基因的扩增子测序技术被广泛用于各种生态系统中微生物群落的多样性检测。扩增子测序的使用极大地促进了土壤、水体、空气等环境中微生物生态的相关研究。【目的】随着高通量测序技术的不断发展和参考数据库的不断更新,针对不同的环境样本的引物选择和改进仍然需要更深入的校验。【方法】本文收集了目前在微生物群落研究中被广泛采用的标记基因扩增通用引物,包括 16S rRNA 基因扩增常用的 8 对通用引物和 2 对古菌引物、9 对真菌转录间隔区(internal transcribed spacer, ITS)基因扩增引物,以及 18S rRNA 基因扩增的 4 对真核微生物通用引物和 1 对真菌特异性引物。这些引物中包括了地球微生物组计划(Earth Microbiome Project, EMP)推荐的 2 对 16S rRNA 基因扩增引物、1 对 ITS1 基因扩增引物和 1 对 18S rRNA 基因扩增引物。采用最近更新的标准数据库对这些引物进行了覆盖度和特异性评价。【结果】EMP 推荐的引物依然具有较高的覆盖度,而其他引物在覆盖度及对特定环境或类群的特异性上也各有特点。此外,最近有研究对这些通用引物进行了一些改进,而我们也发现,一个碱基的变化都可能会导致评价结果或扩增产物发生明显变化,简并碱基的引入既可以覆盖更多的物种,但同时也会在一定程度上降低关注物种的特异性。【结论】研究结果为微生态研究中标记基因的引物选择提供了一个广泛的指导,但在关注具体科学问题时,引物的选择仍需数据指导与实验尝试。

关键词: 扩增子测序, 16S rRNA 基因, ITS 基因, 18S rRNA 基因, PCR 扩增, 引物, 引物覆盖度, 引物特异性, 扩增产物片段长度

Foundation items: National Natural Science Foundation of China (91851106, U1906223, 31861133002); Key Research Program of Frontier Science of Chinese Academy of Sciences (QYZDB-SSW-DQC026)

***Corresponding author:** Tel: 86-10-62840082; E-mail: yedeng@rcees.ac.cn

Received: 17-01-2020; **Accepted:** 12-04-2020; **Published online:** 09-05-2020

基金项目: 国家自然科学基金(91851106, U1906223, 31861133002); 中国科学院前沿科学重点研究项目(QYZDB-SSW-DQC026)

***通信作者:** Tel: 010-62840082; E-mail: yedeng@rcees.ac.cn

收稿日期: 2020-01-17; **接受日期:** 2020-04-12; **网络首发日期:** 2020-05-09

***In-silico* evaluation and improvement on 16S/18S/ITS primers for amplicon high-throughput sequencing**

WU Yue-Ni^{1,2} FENG Kai^{1,2} LI Shu-Zhen^{1,3} WANG Zhu-Jun^{1,2} ZHANG Zhao-Jing^{3,4}
DENG Ye^{*1,2,4}

1 Key Laboratory for Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

2 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

3 Key Laboratory of Industrial Ecology and Environmental Engineering, Ministry of Education; School of Environmental Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China

4 Institute of Marine Science and Technology, Shandong University, Qingdao, Shandong 266237, China

Abstract: [Background] In the past decade, high-throughput sequencing technology on ribosomal RNA amplicons has been widely used to explore the diversity and composition of microbial communities in a variety of ecosystems. [Objective] With the development of sequencing technology and the expanding of reference databases, commonly adopted primers still need to be cautiously verified for different environmental samples. [Methods] We collected several marker gene amplification primers for microbial communities (*i.e.* bacteria, archaea, fungi and other eukaryotic microorganisms), including 8 universal primer sets and 2 archaea primer sets of 16S rRNA amplification, 9 universal primer sets for internal transcribed spacer (ITS) amplification, 4 universal primer sets and 1 fungal universal primer set for 18S rRNA amplification, which contained two 16S primer sets, one ITS1 primer set and one 18S primer set recommended by the Earth Microbiome Project (EMP). These primers were evaluated for the coverage and specificity using a recently updated standard reference databases. [Results] The EMP recommended primers all still have high coverage in current databases, while other primers have their own advantages in coverage and specificity to specific environments or taxonomic groups. In addition, recent studies have made some improvements on these universal primers. Changes in one single base may lead to changes in evaluation results or amplification products. The single degenerate base added to primers may cover more species, but it may also reduce the species specificity to some extent. [Conclusion] For different environmental samples, the selection of primers and experimental process still need more detailed verification. Our results could be a guide for the selection and improvement of amplified primers in microbial ecological studies.

Keywords: Amplicon sequencing, 16S rRNA gene, Internal transcribed spacer (ITS) gene, 18S rRNA gene, PCR amplification, Primer, Primer coverage, Primer specificity, Amplification product fragment length

扩增子测序是一种高靶向性地分析特定基因组区域中基因变异的方法,是发现单核苷酸多态性(single nucleotide polymorphisms, SNPs)的理想方法。其利用PCR的引物来扩增基因组的特定区域,靶向地捕获目标区域的DNA,达到目的DNA片段的富集目标;最后针对扩增产物(也被称为扩增子)进行高通量测序,分析序列中的遗传变异等信息。一般认为,核糖体RNA(rRNA)基因广泛存在于所有细胞生物体中,由于很少发生大规模的横向基因迁移,因此具有一系列由非常保守到高变的区域,

适合于微生物分类信息的确定。由于核糖体操纵子具有足够的保守性,因此常被用作多样性调查的标记基因^[1],主要包括16S rRNA、18S rRNA和转录间隔区(internal transcribed spacer, ITS)等。用下一代测序技术(next generation sequencing, NGS)测定16S/18S/ITS某个可变区的序列,可以反映环境样品在细菌、真菌、古菌等组成的微生物群落之间的差异,对研究海洋、土壤、宿主等不同环境中的微生物群落组成及动态变化有重要的指导作用,同时也是系统发育和分类学研究中一种广泛使用的

方法^[2]。

原核生物(细菌、古菌) rRNA 按核酸的沉降系数分为 3 种, 分别为 5S、16S 和 23S。16S rRNA 是核糖体 RNA 的一个亚基, 16S rRNA 基因就是编码该亚基的基因。其存在于所有细菌染色体基因组中, 参与生物蛋白质的合成过程, 并在生物进化的漫长历程中保持一定的遗传保守性, 是细菌系统分类研究中最有用和最常用的分子标记^[3]。16S rRNA 基因分子大小适中(长度约为 1 540 bp), 在结构与功能上又具有高度的保守性, 检测方便并且与原核生物的系统发育关联密切, 因此逐渐成为最常用的原核生物标记基因。16S rRNA 编码基因序列共有 9 个保守区和 9 个高可变区(图 1), 其中 V4–V6 区数据库信息全、特异性好, 是细菌多样性分析的常用目标区域^[4]。

在真核微生物中, 核糖体 DNA 是由核糖体基因及与之相邻的间隔区组成, 其基因组序列从 5'到 3'依次为: 外部转录间隔区(external transcribed spacer, ETS)、18S rRNA 基因、内部转录间隔区 1 (ITS1)、5.8S rRNA 基因、内部转录间隔区 2 (ITS2)、28S rRNA 基因和基因间隔序列(intergenic spacer, IGS)(图 2)。其中, 18S rRNA 基因是编码真核生物核糖体小亚基的 DNA 序列, 其中既有保守区, 也有可变区(V1–V9)。与 16S rRNA 基因相似, 保守区

反映了生物物种间的亲缘关系, 而可变区则能体现物种间的差异, 适用于作为物种分类单元及系统发育的分子标记。在真菌中, 相比于 rRNA 中的 18S、5.8S 和 28S 的基因组序列, 内转录间隔区 ITS1 和 ITS2 作为非编码区具有更高的丰富度和分辨率, 承受的选择压力较小, 相对变化较大, 并且能够提供详尽的系统学分析所需要的可遗传性状^[5]。同时, ITS 的保守型表现为种内相对一致, 种间差异较明显, 能够反映出种属间甚至菌株间的差异。此外, ITS 序列片段较小(ITS1 和 ITS2 长度分别为 350 bp 和 400 bp 左右, 但不同物种之间存在一定差异), 易于分析。也就是说, ITS 具有更高的扩增和测序成功率以及分类学分辨率, 目前已被广泛应用于真菌不同种属的系统发育分析。

在过去的 10 年中, NGS 技术被广泛用于探索各种生态系统中微生物群落的多样性和组成结构^[6], 诸如全球气候变化及微生物响应^[7–9]、生物地理分布^[10–11]、海洋生态系统^[12]、农业生态系统^[13]、生物降解及修复^[14]、野生动物^[15]及人体共生微生物^[16]等各个领域。随着高通量测序技术的不断发展和参考数据库的不断更新, 利用核糖体操纵子作为 DNA 标记可以揭示不同生态系统中的微生物多样性和组成^[17]。然而, 针对不同环境样本的引物选择和实验过程仍然需要更细致的

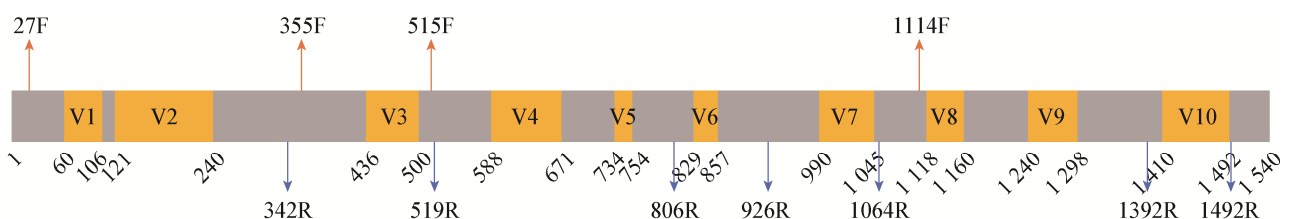


图 1 8 对原核生物 16S rRNA 基因扩增引物及其相对位置

Figure 1 Eight primers and their positions for 16S rRNA gene amplification in prokaryotes



图 2 真核微生物核糖体 DNA 结构示意图

Figure 2 Structural diagram of eukaryotic microbial ribosomal DNA

验证^[18–20]。本文收集了微生物群落中针对细菌、古菌、真菌所常用的标记基因扩增引物, 包括 16S rRNA 基因(细菌和古菌)、ITS 基因(真菌)和 18S rRNA 基因(真核微生物), 并用不同数据库对其进行了生物信息学的评价, 以期对不同环境样

本和研究对象的扩增引物选择能起到一定的指导意义。

1 材料与方法

1.1 16S rRNA 基因扩增引物与数据库

我们收集了 16S rRNA 基因扩增常用的 8 对通用引物(图 1),其中包括 4 条正向引物: 27F^[21]、355F^[22]、515F^[23]、1114F^[24]; 以及 7 条反向引物: 342R^[25]、519R^[26]、806R^[27]、926R^[28]、1064R^[22]、1392R^[29]、1492R^[30]。针对不同的扩增区域,关注了扩增 V1–V2 区的 27F/342R 引物对,扩增 V3 区的成对引物 355F/519R 引物对,扩增 V3–V5 区的成对引物 355F/806R 引物对,扩增 V4–V5 区的成对引物 515F/806R 引物对,扩增 V4–V6 区的成对引物 515F/926R 引物对,扩增 V4–V7 区的成对引物 515F/1064R 引物对,扩增 V9 区的成对引物 1114F/1392R 引物对,以及扩增 16S rRNA 基因全长(V1–V10 区)的成对引物 27F/1492R 引物对(表 1)。

选用 SILVA 数据库(SILVA123_QIIME_release, 2015)对以上 16S rRNA 基因通用引物进行覆盖度的评价。SILVA 是一个 rRNA 基因序列的综合数据库,

收录了原核和真核微生物的小亚基 rRNA 基因序列(简称 SSU, 即 16S 和 18S rRNA)和大亚基 rRNA 基因序列(简称 LSU, 即 23S 和 28S rRNA)。细菌和真菌的都有,能够提供定期的更新和全面优质的检查。选用 SILVA 的 16S rRNA 基因数据库来评价 16S 通用引物,数据库包含原核生物 16S rRNA 基因序列 226 267 条。为了只保留数据库中 16S 全长序列,以确保每个可变区扩增引物的覆盖度评价结果更具有可比性,删除了数据库中序列长度小于 1 500 bp 的序列,保留了 19 084 条长度在 1 500 bp 以上的序列,构建为“SILVA 16S rRNA 基因全长序列数据库”(表 2)。

此外还收集了用来扩增古菌 16S rRNA 基因的正向引物 A519F^[31]及其改进版 A519F*^[32],反向引物 A806R^[33]和 A908R^[34]。其中,A519F 和 A806R 的扩增区域为 V4–V5 区,A519F*和 A908R 的扩增区域为 V4–V6 区(表 3)。为了评价古菌 16S rRNA 基因特异引物,从 SILVA 的 16S rRNA 基因数据库中分离出古菌序列共 9 825 条,构建了“SILVA 16S rRNA 基因古菌序列数据库”用来评价古菌 16S rRNA

表 1 16S rRNA 基因通用引物基本信息

Table 1 Basic information of 16S rRNA gene universal primers

No.	Primers name	Sequences (5'→3')	Amplified region	Amplified fragment length (bp)
1	27F	AGRGTTYGATYMTGGCTCAG	V1–V2	315
	342R	CTGCTGCSYCCCGTAG		
2	355F	ACWCCTACGGGWGGCWGC	V3	164
	519R	GWATTACCGCGGCKGCTG		
3	355F	ACWCCTACGGGWGGCWGC	V3–V5	451
	806R	GGACTACNVGGGTWTCTAAT		
4	515F	GTGYCAGCMGCCGCGGTAA	V4–V5	291
	806R	GGACTACNVGGGTWTCTAAT		
5	515F	GTGYCAGCMGCCGCGGTAA	V4–V6	411
	926R	CCGYCAATTYMTTTRAGTTT		
6	515F	GTGYCAGCMGCCGCGGTAA	V4–V7	550
	1064R	AYCTCACGRACGAGCTGAC		
7	1114F	GCAACGAGCGCAACCC	V9	278
	1392R	ACGGGCGGTGTGTRC		
8	27F	AGRGTTYGATYMTGGCTCAG	V1–V10	1 465
	1492R	TACGGYTACCTTGTTACGACTT		

表 2 用于本文引物评价的数据库

Table 2 Databases for primer evaluation in this paper

Gene name	Database	Number of sequences
16S rRNA gene	SILVA 16S rRNA 基因全长序列数据库	19 084
	SILVA 16S rRNA gene full-length sequence database	
	SILVA 16S rRNA 基因古菌序列数据库	9 825
	SILVA 16S rRNA gene archaeal sequence database	
	SILVA 16S rRNA 基因非古菌序列数据库	216 442
	SILVA 16S rRNA gene non-archaeal sequence database	
ITS gene	Unite ITS1 基因序列数据库	3 970
	Unite ITS1 gene sequence database	
	Unite ITS2 基因序列数据库	11 367
	Unite ITS2 gene sequence database	
	Unite ITS 基因全长序列数据库	1 871
	Unite ITS gene full-length sequence database	
	海洋真菌 ITS1 基因序列数据库	683
	Marine ITS1 gene sequence database	
	海洋真菌 ITS2 基因序列数据库	1 281
	Marine ITS2 gene sequence database	
18S rRNA gene	海洋真菌 ITS 基因全长序列数据库	113
	Marine ITS gene full-length sequence database	
	SILVA 18S rRNA 基因真核微生物数据库	25 497
	SILVA 18S rRNA gene database for eukaryotic microbes	
	SILVA 18S rRNA 基因真菌序列数据库	3 321
	SILVA 18S rRNA gene fungal sequence database	
	SILVA 18S rRNA 基因非真菌序列数据库	22 176
	SILVA 18S rRNA gene non-fungal sequence database	
	海洋真菌 18S rRNA 基因序列数据库	1 116
	Marine 18S rRNA gene sequence database	

基因扩增引物的覆盖度。同时，用去除古菌序列之后 SILVA 16S rRNA 数据库中剩余的 216 442 条序列构成了“SILVA 16S rRNA 基因非古菌序列数据库”，用该数据库的引物覆盖度来评价古菌引物对于古菌序列的特异性。由于这两对引物的扩增区域均位于 16S rRNA 基因中部，因此未对数据库序列长度进行筛选。扩增 16S rRNA 基因 V4–V5 区的通用引物对 515F/806R 和扩增 V4–V6 区的通用引物对

515F/926R 也一起用来评价古菌引物的相对覆盖度与特异性。

1.2 ITS 引物与数据库

收集了 9 对用于扩增 ITS 基因的通用引物(表 4), 包括 7 条正向引物: ITS1^[35]、ITS1F^[36]、gITS7^[37]、gITS7ngs^[38]、5.8S-Fun^[39]、ITS3_KYO2^[40]、ITS9MUNng^[41]; 以及 4 条反向引物: ITS2^[35]、ITS4R^[35]、ITS4ngs^[42]、ITS4-Fun^[39]。针对不同的扩

表 3 古菌 16S rRNA 基因扩增引物基本信息

Table 3 Basic information of primers for archaea 16S rRNA gene amplification

Primers name	Sequences (5'→3')	Amplified region	Amplified fragment length (bp)
A519F	CAGCCGCCGCGGTAA	V4–V5	287
A806R	GGACTACNSGGGTMTCTAAT		
A519F*	CAGCMGCCGCGGTAA	V4–V6	389
A908R	CCCGCCAATTCCTTTAAGTT		

表 4 ITS 基因通用引物基本信息

Table 4 Basic information of ITS gene universal primers

No.	Primers name	Sequence (5'→3')	Position	Amplified region	Amplified fragment length (bp)
1	ITS1	TCCGTAGGTGAACCTGCGG	18S	ITS1	220
	ITS2	GCTGCGTTCTTCATCGATGC	5.8S		
2	ITS1F	CTTGGTCATTTAGAGGAAGTAA	18S		190
	ITS2	GCTGCGTTCTTCATCGATGC	5.8S		
3	gITS7	GTGARTCATCGARTCTTTG	5.8S	ITS2	315
	ITS4R	TCCTCCGCTTATTGATATGC	28S		
4	gITS7ngs	GTGARTCATCRARTYTTTG	5.8S		315
	ITS4ngs	TCCTSCGCTTATTGATATGC	28S		
5	5.8S-Fun	AACTTTYRRC AAYGGATCWCT	5.8S		325
	ITS4-Fun	(AG)CCTCCGCTTATTGATATGCTTAART	28S		
6	ITS3_KYO2	GATGAAGAACGYAGYRAA	5.8S		360
	ITS4R	TCCTCCGCTTATTGATATGC	28S		
7	ITS1	TCCGTAGGTGAACCTGCGG	18S	ITS1+5.8S+ITS2	510
	ITS4R	TCCTCCGCTTATTGATATGC	28S		
8	ITS1F	CTTGGTCATTTAGAGGAAGTAA	18S		540
	ITS4R	TCCTCCGCTTATTGATATGC	28S		
9	ITS9MUNgs	TTGTACACACCGCCGTCG	18S		650
	ITS4ngs	TCCTSCGCTTATTGATATGC	28S		

增区域关注了 9 对引物对，其中包括扩增 ITS1 区的 2 对引物对 ITS1/ITS2 和 ITS1F/ITS2，扩增 ITS2 区的 4 对引物对 gITS7/ITS4R、gITS7ngs/ITS4ngs、5.8S-Fun/ITS4-Fun 和 ITS3_KYO2/ITS4R，扩增 ITS 全长序列的 3 对引物对 ITS1/ITS4R、ITS1F/ITS4R 和 ITS9MUNgs/ITS4ngs。

由于真核微生物的核糖体 DNA 是由核糖体基因及与之相邻的间隔区(ITS)组成(图 2)，使扩增 ITS1 的正向引物落在 18S 亚基上，扩增 ITS1 的反向引物和扩增 ITS2 的正向引物落在 5.8S，而扩增 ITS2 的反向引物则是落在 28S 亚基上。然而目前并没有一个较为准确完整的数据库能够提供从 18S 至 28S 全长序列。因此，我们用 ITSx Extractor 从专门针对真菌 ITS 序列的数据库 Unite 中分别筛选出的 ITS1、ITS2 和 ITS 全长序列^[43]，并构建了 3 个新的数据库。其中，ITS1 数据库的筛选条件为：包含 ITS1 基因，且 ITS1 基因前、IT1 基因后序列长度>100 bp；ITS2 数据库的筛选条件为：包含 ITS2 基因，且 ITS2 基因前、ITS2 基因后序列长度>100 bp；ITS 基因全

长序列数据库的筛选条件为：序列包含 ITS1 基因，ITS2 基因、且 ITS1 基因前、ITS1 基因与 ITS2 基因间隔、ITS2 基因后序列长度>100 bp。基于此，筛选出 ITS1 序列 3 970 条、ITS2 序列 11 367 条、ITS 全长序列 1 871 条，并以此分别构建了“Unite ITS1 基因序列数据库”“Unite ITS2 基因序列数据库”和“Unite ITS 基因全长序列数据库”，用于评价 ITS 通用引物的覆盖度(表 2)。此外，还从 GenBank 中收集了海洋真菌的 ITS 序列 15 468 条(关键词：Marine、Ocean、Costal、Estuary)，并用上述方法筛选出 ITS1 序列 683 条、ITS2 序列 1 281 条以及 ITS 全长序列 113 条，基于此构建了“海洋真菌 ITS1 基因序列数据库”“海洋真菌 ITS2 基因序列数据库”和“海洋真菌 ITS 基因全长序列数据库”，用于判断通用引物对海洋环境的适用性(表 2)。

1.3 18S rRNA 基因扩增引物与数据库

对于 18S rRNA 基因，收集了 5 对引物(表 5)，包括 4 条正向引物：SSU_F_22^[44]、TAReuk454FWD1^[45]、nu-SSU-0817^[46]、Euk1391f^[47]；以及 5 条反向引物：

SSU_R_09^[44]、SSU_R_13^[44]、R1200^[48]、nu-SSU-1196^[46]、EukBr^[47]。针对不同的扩增区域关注了扩增 V3 区的引物对 SSU_F_22/SSU_R_09、扩增 V3-V7 区的引物对 SSU_F_22/SSU_R_13、扩增 V4-V5 区的引物对 TAREuk454FWD1/R1200、扩增 V5 区的引物对 nu-SSU-0817/nu-SSU-1196 以及扩增 V9 区的引物对 Euk1391f/EukBr。其中，nu-SSU-0817/nu-SSU-1196 为真菌特异性引物对，其他 4 对引物为真核微生物通用引物对。由于 Euk1391f/EukBr 的扩增片段是 18S rRNA 基因最末端的一个可变区 V9 区，反向引物 EukBr 落在了 ITS1 基因上，其他引物则均位于 18S rRNA 基因的保守区(表 5)。

我们选择 SILVA 18S rRNA 基因真核微生物数据库来评价以上通用引物，数据库共包含 25 497 条真核微生物 18S rRNA 基因序列，且序列长度均>900 bp。其中包括真菌序列 3 321 条，基于此我们构建了“SILVA 18S rRNA 基因真菌序列数据库”，剩余序列则构建为“SILVA 18S rRNA 基因非真菌序列数据库”，包括非真菌序列 22 176 条(表 2)。由于扩增 V9 区的反向引物 EukBr 落在 ITS1 基因上，而并没有一个较为完整的数据库可以包含 18S+ITS1 全长序列，因此我们用 Unite ITS1 基因序列数据库来另外评价反向引物 EukBr 的覆盖度(表 2)。跟 ITS 基因一样，我们从 GenBank 收集了海洋真菌的 18S rRNA 基因序列 13 459 条(关键词: Marine、Ocean、

表 5 18S rRNA 基因通用引物基本信息

Table 5 Basic information of 18S rRNA gene universal primers

No.	Primers name	Sequences (5'→3')	Amplified fragment length (bp)	Amplified region
1	SSU_F_22	TCCAAGGAAGGCAGCAGGC	137	V3
	SSU_R_09	AGCTGGAATTACCGCGGCTG		
2	SSU_F_22	TCCAAGGAAGGCAGCAGGC	990	V3-V7
	SSU_R_13	GGGCATCACAGACCTGTTA		
3	TAREuk454FWD1	CCAGCASCYGCGGTAATTCC	600	V4-V5
	R1200	CCCGTGTTGAGTCAAATTAAGC		
4	nu-SSU-0817	TTAGCATGGAATAATRRATAGGA	385	V5
	nu-SSU-1196	TCTGGACCTGGTGAGTTCC		
5	Euk1391f	GTACACACCGCCCGTC	130	V9
	EukBr*	TGATCCTTCYGCAGGTTACCTAC		

Costal、Estuary)。由于本研究关注的引物中最长的扩增长度为 1 200 bp，因此从该数据库中筛选出长度>1 300 bp 的序列 1 116 条，基于此构建了“海洋真菌 18S rRNA 基因序列数据库”，用于判断通用引物对海洋环境的适用性。海洋真菌 ITS1 基因序列数据库也用来另外评价反向引物 EukBr 对于海洋样本的覆盖度(表 2)。

1.4 覆盖度与特异性计算

引物覆盖度的评价方法选用了 BLASTn 序列比对。引物与数据库序列进行比对，并通过比对结果来计算每条引物和成对引物在数据库序列的覆盖度。计算方法如下：

用集合 S 表示一个有 N 条序列的数据库，那么 $S=\{S_1,S_2,...,S_N\}$ ， $N=|S|$ ，其中 $|S|$ 是集合 S 的模，表示集合中序列的个数；假定简并引物 P 有 m 种可能性，分别记为 $P_i(i=1,2,...,m)$ ， P_i 与数据库中的序列进行比对后，覆盖到的序列可以构成一个集合，记为 $S(P_i)$ ；

那么，上述 $S(P_i)$ 的并集 $\bigcup_{i=1}^m S(P_i)$ 就是引物 P 在数据库 S 中覆盖序列的集合，引物 P 的覆盖度可以表示为

Coverage(P)= $\frac{|\bigcup_{i=1}^m S(P_i)|}{|S|}$

记 P 的反向引物为 P' ，相似的， P' 的覆盖度为

Coverage(P')= $\frac{|\bigcup_{i=1}^{m'} S(P'_i)|}{|S|}$

记 $P-P'$ 为 P 和 P' 这一对引物, 那么, $P-P'$ 在数据库 S 中覆盖序列的集合, 就是正向引物 P 和反向引物 P' 在数据库 S 中覆盖序列的交集, 即:

$(\bigcup_{i=1}^m S(P_i)) \cap (\bigcup_{i=1}^{m'} S(P'_i))$, 那么 $P-P'$ 的覆盖度为:

$$\text{Coverage}(P-P') = \frac{|\bigcup_{i=1}^m S(P_i) \cap \bigcup_{i=1}^{m'} S(P'_i)|}{|S|}.$$

2 结果与分析

2.1 16S rRNA 基因通用引物评价结果

本研究收集和整理了 SILVA 数据库中 19 084 条长度在 1 500 bp 以上的序列, 构建为 16S 全长序列数据库。该数据库共包含 69 个细菌和古菌门(phylum), 2 202 个可鉴定的属(genus), 5 105 个可鉴定的种(species) (图 3)。基于此, 我们系统地评价了在 16S rRNA 基因的高通量测序中最常使用的 8 对通用引物, 其中包括 4 条正向引物和 7 条反向引物(表 1)。在不允许碱基错配和仅允许一个碱基错配的情况下, 分别计算了这些引物对于 SILVA 16S rRNA 全长序列数据库的覆盖度。

在不允许错配的情况下, 这 11 条引物中除了 27F、1114F 和 1492R 三条引物之外, 其他引物的覆盖度均在 80% 以上。其中 355F、515F、926R 和 1392R 四条引物的覆盖度达到了 90% 以上, 806R

和 1064R 的覆盖度也达到了 85% 以上。从引物对来看, V4-V6 区 515F/926R 引物对的评价结果最好, 覆盖度能达到 87.05%, V3-V5 区 355F/806R 引物对、V4-V5 区 515F/806R 引物对、V4-V7 区 515F/1064R 引物对的覆盖度也均达到 80% 以上(图 4)。在允许一个碱基错配的情况下, 引物的覆盖度几乎都得到了较大的提升。除了 27F、342R、1114F 和 1492R 四条引物之外, 其他引物的覆盖度均在 90% 以上, 515F、519R 和 1392R 三条引物的覆盖度达到了 95% 以上。从引物对来看, V3 区引物对 355F/519R 和 V4-V6 区引物对 515F/926R 的评价结果最好, 覆盖度能达到 91% 以上。V3-V5 区引物对 355F/806R、V4-V5 区引物对 515F/806R、V4-V7 区引物对 515F/1064R 的覆盖度也均达到 87% 以上(图 4)。

采用构建的 SILVA 16S rRNA 基因古菌序列数据库和 SILVA 16S rRNA 基因非古菌序列数据库(表 2), 分别评价了古菌 16S rRNA 基因扩增引物对于古菌序列的覆盖度和特异性。图 5 结果表明, A519F 作为扩增古菌 16S rRNA 基因的最常用正向引物, 有着较好的覆盖度和特异性, 该引物可以覆盖到 SILVA 16S rRNA 基因数据库中高达 90.46% 的古菌序列和仅仅匹配 0.03% 的其他细菌序列。之后的研

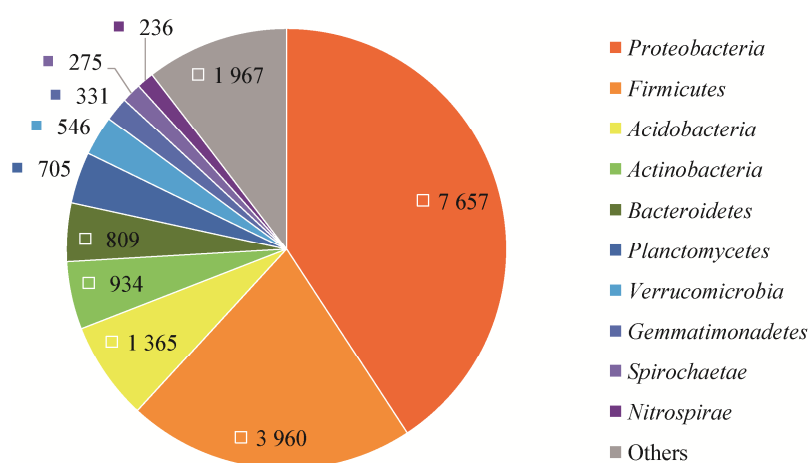


图 3 SILVA 16S rRNA 基因全长序列数据库物种组成(门分类单元)

Figure 3 The composition of SILVA database for 16S rRNA gene full-length sequences in phylum level

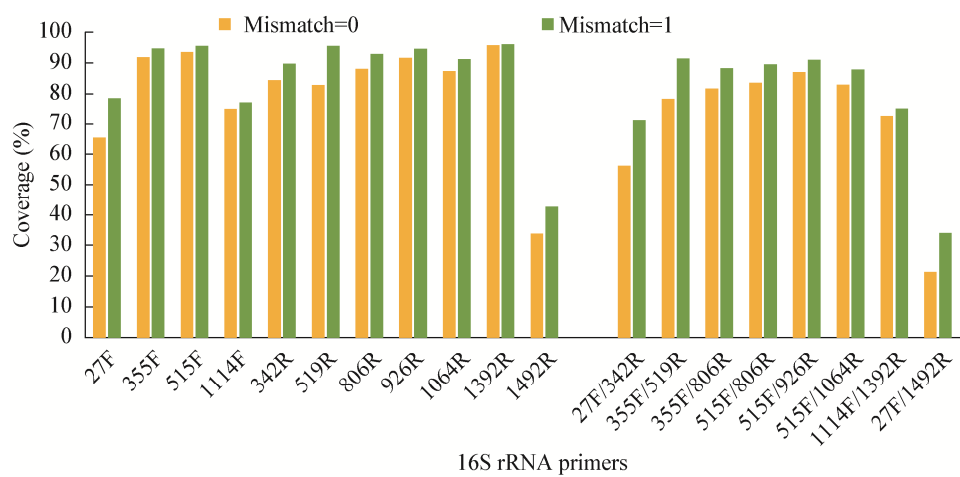


图 4 16S rRNA 基因扩增引物在 SILVA 16S rRNA 基因全长序列数据库中的覆盖度
Figure 4 Coverage of 16S rRNA gene amplification primers in SILVA 16S rRNA gene full-length sequence database

究中, 研究者对 A519F 引物进行了改进, 即 A519F* 引物, 原本的第 5 个碱基 C 改为简并碱基 M, 加入了 A 的可能性^[32]。然而从 A519F 和 A519F* 引物的对比可以看出, 简并碱基的加入虽然略微提高了覆盖度(92.52%), 但是却大大降低了针对古菌的特异性, 而对细菌的覆盖度也大大提高了, 对细菌序列的覆盖度高达 93.32%。另外, 反向引物 A806R 和 A908R

均表现出较高的覆盖度和较好的特异性。在不允许碱基错配和允许 1 个碱基错配的情况下, A806R 对古菌序列的覆盖度分别可以达到 87.60%和 94.63%, 特异性为 11.57%。A908R 的覆盖度相对低一点, 在不允许碱基错配和允许 1 个碱基错配的情况下, 分别可以达到 82.78%和 90.44%, 但特异性评价结果更好, 仅能覆盖数据库中不到 0.01%的非古菌序列。

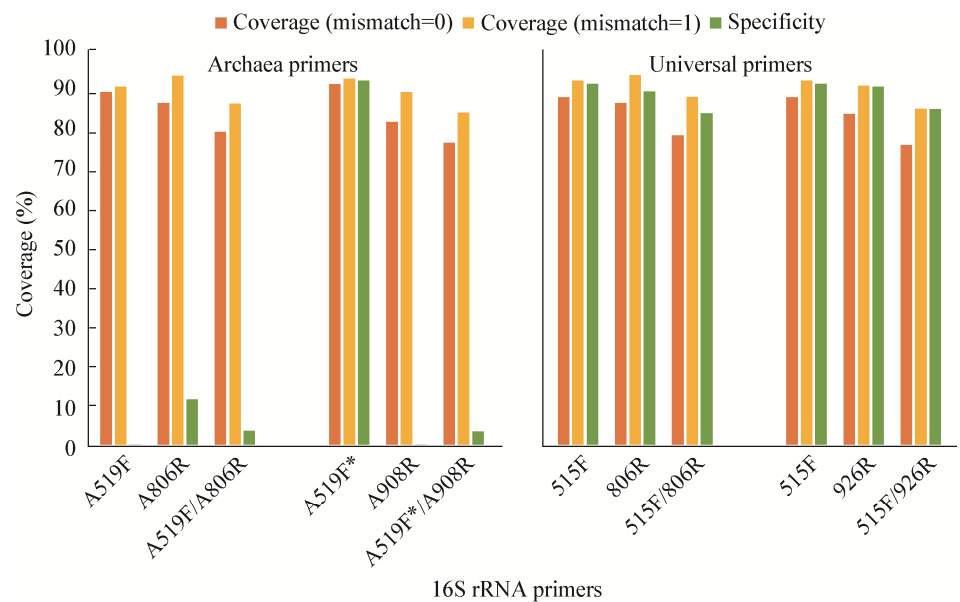


图 5 16S rRNA 基因古菌扩增引物在 SILVA 16S rRNA 基因古菌序列数据库中的覆盖度及其对古菌的特异性
Figure 5 Coverage of 16S rRNA gene archaea amplification primers in SILVA 16S rRNA gene archaea database and their specificity to archaea

为了评价古菌引物的相对覆盖度与特异性,扩增 16S rRNA 基因 V4-V5 区的通用引物对 515F/806R 和扩增 V4-V6 区的通用引物对 515F/926R 也分别用 SILVA 16S rRNA 基因古菌序列数据库和 SILVA 16S rRNA 基因非古菌序列数据库进行了覆盖度以及对古菌序列特异性的评价。结果显示,通用引物对于古菌序列的覆盖度也都很高(图 5)。在不允许碱基错配的情况下,古菌引物对于古菌序列的覆盖度均略高于通用引物;但在允许 1 个碱基错配的情况下,通用引物的覆盖度反而更高。也就是说,这两对通用引物都能较好地反映群落细菌与古菌组成。

2.2 ITS 基因通用引物评价结果

通过收集和整理 Unite 数据库中的真菌 ITS 基因序列,重新构建了 Unite ITS1 基因序列数据库、Unite ITS2 基因序列数据库和 Unite ITS 基因全长序

列数据库(表 2),并以此为基础系统地评价了在 ITS 基因的高通量测序中常使用的 9 对通用引物,包括扩增 ITS1 区的 2 对引物、扩增 ITS2 区的 4 对引物和扩增 ITS 全长序列的 3 对引物(表 4)。其中,3 条正向引物 ITS1、ITS1F 和 ITS9MUNngs 位于 18S 亚基上;1 条反向引物 ITS2 以及 4 条正向引物 gITS7、gITS7ngs、5.8S-Fun 和 ITS3_KYO2 位于 5.8S 亚基上;3 条反向引物 ITS4R、ITS4ngs 和 ITS4-Fun 位于 28S 亚基上。在不允许碱基错配和仅允许 1 个碱基错配的情况下,分别计算了这些引物对于 Unite 3 个 ITS 序列数据库的覆盖度。此外,分别基于海洋真菌 ITS1 基因序列数据库、海洋真菌 ITS2 基因序列数据库和海洋真菌 ITS 基因全长序列数据库,系统评价了 ITS 通用引物对海洋真菌序列的普适性。

对于 ITS1 的扩增主要关注了 2 条正向引物 ITS1 和 ITS1F 以及 1 条反向引物 ITS2 (图 6A)。正

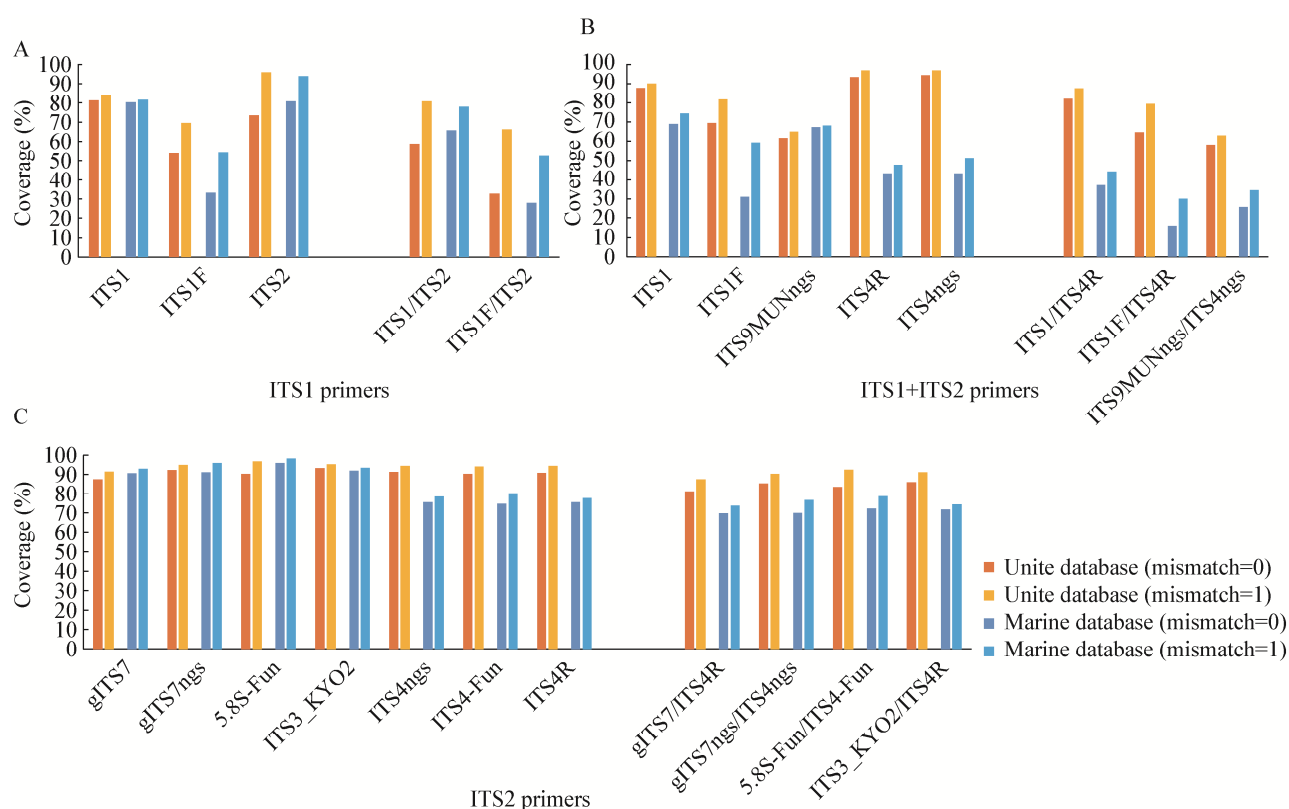


图 6 ITS 扩增引物对于 Unite ITS 3 个数据库和海洋真菌 ITS 3 个数据库的覆盖度

Figure 6 The coverage of ITS amplification primers to 3 Unite databases and 3 marine ITS fungal databases

向引物 ITS1 对于 Unite 数据库和海洋数据库都表现出较高的覆盖度, 可以达到 80% 以上, 而 ITS1F 相对来说则无法覆盖到足够多的序列。对于反向 ITS2 来说, 允许 1 个碱基错配的情况下覆盖度会得到大大提升, 在 Unite 数据库和海洋数据库均能覆盖到 93% 以上的序列。

对于 ITS 全长序列的扩增, 收集了 3 条正向引物 ITS1、ITS1F 和 ITS9MUNgs, 以及 2 条反向引物 ITS4R 和 ITS4ngs。正向引物中, ITS1 的覆盖度最高, 在允许 1 个碱基错配的情况下, 可以覆盖 Unite ITS 全长数据库中 90% 以上的序列, 这条正向引物对海洋 ITS 全长序列数据库也是具有最高的适用性(图 6B)。相对来说, ITS1F 和 ITS9MUNgs 的覆盖度则不够好, 当允许 1 个碱基错配时, ITS1F 对于 Unite 数据库的覆盖度会有一个较大的提升。从引物对来看, ITS1/ITS4R 引物对的评价结果最好, 能够覆盖到 Unite 数据库中 80% 以上的全长序列。然而 3 对引物对于海洋真菌 ITS 数据库来说覆盖度都不理想, 即使允许了 1 个碱基错配, 覆盖度仍均未达到 50% 以上。也就是说, 这 3 对引物并不适用于海洋真菌的 ITS 基因扩增。

针对 ITS2 区的扩增, 收集了 4 条正向引物 gITS7、gITS7ngs、5.8S-Fun 和 ITS3_KYO2, 其中 gITS7ngs 是在 gITS7 的基础上增加了 2 个简并碱基而来(表 4)。4 条正向引物对于 Unite ITS2 数据库的覆盖度都很高, 不允许碱基错配的情况下均可达到 85% 以上, 允许一个碱基错配的情况下覆盖度则可以达到 90% 以上。对于海洋 ITS2 数据库, 这些正向引物的评价结果更好, 覆盖度均能达到 90% 以上。此外还收集了 3 条反向引物 ITS4R、ITS4ngs 和 ITS4-Fun, 其中 ITS4ngs 是在 ITS4R 的基础上增加了一个简并碱基, ITS4-Fun 则是将 ITS4R 前端去除第 1 个碱基, 后端延长了 6 个碱基。从评价结果来看, 简并碱基的引入可以在一定程度上提高 ITS4R 的覆盖度, 而引物的增长则会使信息学的评价结果变低, 但评价结果差别并不大。从引物对来看, 这 4 对引物的评价结果相差不大, 不允许错配

的情况下均能覆盖 80% 以上 Unite 数据库中的 ITS2 序列, 在允许 1 个碱基错配的情况下, 覆盖度则还能再提升 5% 左右。此外, 这 4 对引物均能覆盖 70%–80% 的海洋 ITS2 序列(图 6C)。

2.3 18S rRNA 基因通用引物评价结果

通过收集和整理 SILVA 数据库中 18S rRNA 基因序列(25 497 条), 系统地评价了在 18S rRNA 基因的高通量测序中针对不同区域常用的 5 对引物, 包括 4 条正向引物和 5 条反向引物(表 5), 其中 4 对引物是真核微生物通用引物, 还有 1 对引物是针对真菌扩增的特异性引物。因此额外构建了 SILVA 18S rRNA 基因真菌数据库和非真菌数据库用于评价这些引物对于真菌序列的覆盖度和特异性。此外, 基于海洋真菌 18S rRNA 基因数据库海洋和真菌 ITS1 序列数据库(表 2)系统评价了 18S rRNA 基因通用引物对海洋真菌序列的普适性。

由于反向引物 EukBr 对于真菌序列的覆盖度太低(不足 1%), 且错配大比例发生在第 10 个碱基, 因此根据 BLASTn 的比对结果, 将第 10 个碱基 T 加入碱基 C 的简并, 将原先的引物 EukBr (5'-TGAT CCTTCTGCAGGTTTCACCTAC-3')改良为 EukBr* (5'-TGATCCTTCYGCAGGTTTCACCTAC-3')。在不允许错配的情况下, 改良后的引物覆盖度可以从原先的不到 1% 提升至 75% 以上(图 7)。本文后续对于引物 EukBr* 的评价均为改良后的结果。此外, 由于这条反向引物落在 ITS1 基因上, 因此 Unite ITS1 基因序列数据库用来评价其覆盖度。

在不允许错配的情况下, 7 条通用引物中, 除了 SSU_F_22 仅能覆盖 55.95% 的真核微生物序列, 其余 6 条引物的覆盖度均在 70% 以上(图 8)。在允许 1 个碱基错配的情况下, 所有引物的覆盖度则均达到了 80% 以上。由于用来评价反向引物 EukBr* 数据库与正向引物数据库不同, 无法得到扩增 V9 区的引物对 Euk1391f/EukBr* 的覆盖度。对于其他 4 对引物的评价结果来看, 扩增 V4–V5 的引物对 TAReuk454FWD1/R1200 的覆盖度最高, 不允许碱基错配的覆盖度可以达到 69.58%, 允许 1 个碱基

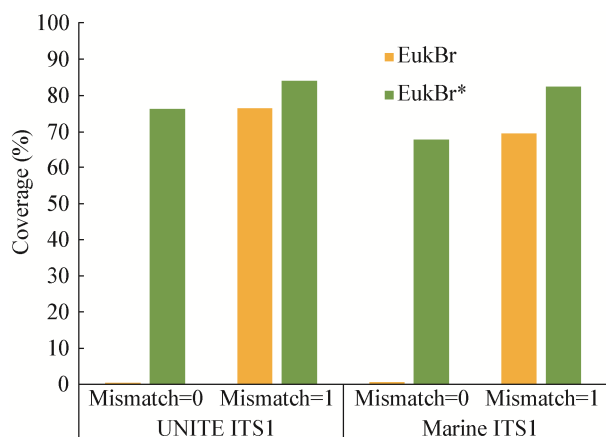


图 7 18S rRNA 基因扩增引物 EukBr 改进前后对于 ITS1 数据库的覆盖度

Figure 7 18S rRNA gene primer EukBr's coverage of ITS1 databases before and after the improvement

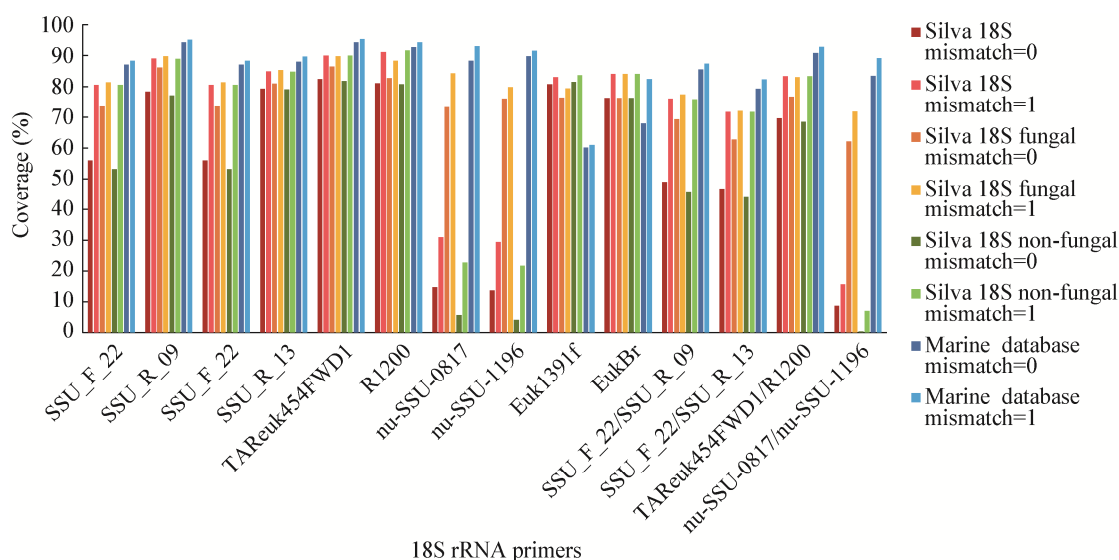


图 8 18S rRNA 基因扩增引物对于对应数据库的覆盖度

Figure 8 Coverage of the 18S rRNA gene amplification primer to their corresponding database

3 讨论与结论

下一代测序技术的迅速发展大大降低了微生物多样性分析的检测成本和复杂程度。因此,选择引物扩增标记基因就成为确定序列长度和覆盖范围的关键。2010年,地球微生物组计划(Earth Microbiome Project, EMP)成立,慢慢建立起来自世界各地生态系统的微生物多样性目录^[4],以创建微生物组样本数据库。基于此,微生物生态学家可以

错配的覆盖度则能达到 83.32%。此外,这 4 对通用引物对真菌序列的适配性也较好,在不允许碱基错配的情况下均能覆盖到 60% 以上的真菌序列,允许 1 个碱基错配的覆盖度则都能达到 70% 以上。用于扩增真菌 18S rRNA 基因 V5 区的特异引物对 nu-SSU-0817/nu-SSU-1196 对于真菌序列的覆盖度反而略低于通用引物。然而,nu-SSU-0817/nu-SSU-1196 对于真菌序列的特异性非常好,在不允许碱基错配的情况下仅能覆盖到 0.39% 的非真菌序列。对于海洋真菌 18S rRNA 序列,这 5 对其他引物的覆盖度都比较高,TAReuk454FWD1/R1200 甚至能覆盖到 90% 以上的海洋真菌序列,SSU_F_22/SSU_F_9 的覆盖度也能达到 85% 以上(图 8)。

控制分析方法的一致性,以促进全球微生物组的比较。因此,EMP 提出了标准引物和实验方法,以保证样本间的多样性比较的准确性。对于 16S rRNA 基因,EMP 推荐的标准引物是扩增 V4-V5 区的 515F/806R 引物对和扩增 V4-V6 区的 515F/926R 引物对。16S rRNA 基因的 V4-V6 区特异性好、数据库信息全,是细菌多样性分析注释的最佳选择。从我们的评价结果来看,在不允许碱基错配的情况

下, 515F/806R 和 515F/926R 这 2 对引物覆盖度也是最高(图 3)。其中, 引物对 515F/926R 的覆盖度略高于 515F/806R, 但 515F/806R 的扩增产物长度为 291 bp, 相对 515F/926R 的扩增产物长度 411 bp 更易测序与分析。对于真核生物 18S rRNA 基因, EMP 则推荐了扩增 V9 区的引物对 Euk1391f/EukBr。由于引物 EukBr 的覆盖度较低, 我们对其进行了一个碱基的改良, 改良后的引物 EukBr* 的覆盖度得到了大幅度的提升(图 7)。

对于 ITS 基因扩增, 由于整个 ITS 区域太长, 目前的第二代测序无法对其进行完全测序。因此, 我们捕捉的目的片段目前只针对 ITS1 或 ITS2 区域, EMP 则是推荐了扩增 ITS1 基因的引物对 ITS1F/ITS2。由于这对引物是在真菌物种中只有一小部分分子变异已知的情况下设计的, 因此在我们的结果中其覆盖度并不高。在过去的研究中, 这对引物在担子菌等部分真菌类群的扩增中有错配, 甚至不匹配的比例很高^[38,49], 因此当我们允许一个碱基的错配时, 这对引物对 Unite ITS1 数据库的覆盖度可以提高一倍。而这对引物之所以仍然被广泛使用, 是因为其对覆盖真菌序列的特异性很好, 尤其能避免植物序列的扩增^[50]。对于 ITS2 区域的扩增, EMP 并未推荐引物, 事实上, ITS1 或 ITS2 究竟哪个片段能更好地反映真菌群落组成仍存有争议。ITS1 通常比 ITS2 短, 对物种分辨率低一些, 但是扩增和测序错误也低一些; ITS2 分辨率高, 但是测序的误差会增加 OTU 的数量。研究表明, 在不同的生态环境和群落复杂性下, ITS1 和 ITS2 测序结果反映的群落组成既存在一致性, 也存在差异性。例如, 在部分土壤和外生菌根真菌群落研究中, ITS1 和 ITS2 呈现出了相似的结果; 而在部分土壤和人类肠道真菌群落研究, ITS1 和 ITS2 的结果则存在差异^[19]。我们的结果表明, 针对 ITS2 片段的扩增引物覆盖度均高于针对 ITS1 片段的扩增引物。总之, 这些结果表明, 真菌群落结构研究所关注的标记基因片段和引物选择尚未标准化。

为了关注真菌的两个标记基因 ITS 和 18S rRNA 基因通用引物对于海洋样本的覆盖度, 我们还收集了海洋真菌 ITS 与 18S rRNA 基因序列构建了海洋数据库。结果表明, 覆盖度高的通用引物并不一定能覆盖最多的海洋序列。比如 EMP 推荐的真核微生物 18S rRNA 通用引物 Euk1391f 和 EukBr* 在不允许错配的情况下分别仅能覆盖 60% 多的海洋真菌序列, 相比其他引物的覆盖度要低一些。对于 ITS 通用引物来说, 扩增 ITS2 的 5.8S-Fun/ITS4-Fun 不管对综合数据库还是海洋数据库都表现出较高的覆盖度。我们的结果表明, 由于真菌群落结构研究所关注的标记基因片段和引物选择尚未标准化, 针对不同的环境样本, 标记基因与引物的选择也要另外考量。此外, 独特的物种关注偏好可能也需要不同的引物选择、组合或重新设计。

随着我们对未知群落的认识越来越深, 信息技术也在不断提升, 一些之前设计的引物可以进行改进, 以提高对微生物群落认识的准确性。本文扩增 ITS2 的正向引物 gITS7ngs 是由 gITS7 改动了两个碱基, 对 Unite 数据库中真菌 ITS 序列的覆盖度提升了 4.66%, 反向引物 ITS4ngs 则是由 ITS4R 改动了 1 个碱基, 覆盖度也有所提升, gITS7ngs/ITS4ngs 的覆盖度相对于 gITS7/ITS4R 也提升了 5.09%。此外, 我们的结果中, 许多引物允许 1 个碱基错配的覆盖度评价结果会比不允许碱基错配时有很大的提升, 如 ITS2、SSU_F_22 等。本文改进了 EukBr 的第 10 个碱基, 而一些其他引物也还有改进的空间。然而对于引物来说, 一个碱基的变化都可能会导致扩增产物发生变化^[28]。例如本文中, 同一位置的古菌 16S rRNA 基因引物 A806R 与通用引物 806R 相比只有两个碱基的差别, 但 A806R 无法扩增细菌序列。简并碱基的引入也许可以覆盖更多的物种, 但也会在一定程度上降低关注物种的特异性。例如, 古菌 16S rRNA 基因通用引物 A519F 由于一个简并碱基 M 替换了 C, 使改良后的 A519F* 覆盖度

提高了 2%, 但相对地, 对细菌序列的覆盖度提升了 90%多(图 5)。也就是说, A519F*不再是针对古菌序列具有特异性的古菌引物了。因此, 当我们在对引物进行改进时, 既要确保能够识别微生物群落及其生存的生态系统之间的联系, 还要确保不会引入新的偏差。这除了需要用已知数据库进行覆盖度与特异性的计算, 也需要一定的 PCR 实验与测序数据结果分析。如 16S rRNA 基因通用引物对 515F/806R 的简并碱基(Y 和 N)发生改动时, 针对不同环境样本的大量实验用于验证其优势与偏差, 结果也表明这些改动适用于海洋、土壤等环境^[51]。

在过去的十几年中, 下一代测序(NGS)技术被广泛用于探索各种生态系统中微生物群落的多样性和组成结构, 对研究海洋、土壤、宿主等环境中的微生物构成有重要的指导作用, 同时也是系统发

育和分类学研究中一种广泛使用的方法, 特别是应用于不同的环境宏基因组学样本中^[6]。随着高通量测序技术的不断发展和参考数据库的不断更新, 利用核糖体操纵子作为 DNA 标记可以揭示不同生态系统中的微生物多样性和组成。采用第二代高通量测序平台测定的 16S/18S/ITS 某个高变区域的序列, 可以反映环境样品在细菌、真菌、古菌等分类方面物种之间的差异。然而, 针对不同的环境样本, 引物的选择和实验过程仍然需要更细致的验证。根据本研究的结果与过去的研究, 综合考虑覆盖度、特异性、产物长度的合理性等因素, 我们为环境样本的测序引物做了推荐(表 6), 为微生态研究中标记基因的引物选择提供了一个广泛的指导, 但在关注具体科学问题时, 引物的选择仍需数据指导与实验尝试。

表 6 环境样本测序引物推荐

Table 6 The final recommended primers for amplicon sequencing of environmental samples

Primers name	EMP proposal	Proposal of this study	Proposal for marine samples
16S rRNA gene universal primers	515F/806R, 515F/926R	515F/806R	—
16S rRNA gene archaeal primers	—	A519F/A806R	—
ITS1 primers	ITS1F/ITS2	ITS1F/ITS2, ITS1/ITS2	ITS1/ITS2
ITS2 primers	—	5.8S-Fun/ITS4-Fun	5.8S-Fun/ITS4-Fun
18S rRNA gene primers	Euk1391f/EukBr*	Euk1391f/EukBr*	Euk1391f/EukBr*

REFERENCES

- [1] Amann RI. Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology[J]. *Molecular Ecology*, 1995, 4(5): 543-554
- [2] Wei ZY, Jin DC, Deng Y. Bioinformatics tools and applications in the study of environmental microbial metagenomics[J]. *Microbiology China*, 2015, 42(5): 890-901 (in Chinese)
魏子艳, 金德才, 邓晔. 环境微生物宏基因组学研究中的生物信息学方法[J]. *微生物学通报*, 2015, 42(5): 890-901
- [3] Zhang JY, Zhu BC, Xu C, et al. Strategy of selecting 16S rRNA hypervariable regions for matagenome-phylogenetic marker genes based analysis[J]. *Chinese Journal of Applied Ecology*, 2015, 26(11): 3545-3553 (in Chinese)
张军毅, 朱冰川, 徐超, 等. 基于分子标记的宏基因组 16S rRNA 基因高变区选择策略[J]. *应用生态学报*, 2015, 26(11): 3545-3553
- [4] Gilbert JA, Meyer F, Jansson J, et al. The Earth Microbiome Project: Meeting report of the “1st EMP meeting on sample selection and acquisition” at Argonne National Laboratory October 6th 2010[J]. *Standards in Genomic Sciences*, 2010, 3(3): 249-253
- [5] Tedersoo L, Anslan S, Bahram M, et al. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi[J]. *Mycology*, 2015, 10: 1-43
- [6] Gohl DM, Vangay P, Garbe J, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies[J]. *Nature Biotechnology*, 2016, 34(9): 942-949
- [7] Zhou JZ, Xue K, Xie JP, et al. Microbial mediation of carbon-cycle feedbacks to climate warming[J]. *Nature Climate Change*, 2012, 2(2): 106-110
- [8] Du XF, Yu H, Wang S, et al. Metagenomics reveal responses of soil microbial community in grassland to global changes[J]. *Chinese Journal of Ecology*, 2019, 38(11): 3516-3526 (in Chinese)
杜雄峰, 于皓, 王尚, 等. 宏基因组方法揭示草地土壤微生物群落响应全球变化[J]. *生态学杂志*, 2019, 38(11):

- 3516-3526
- [9] Wang ZJ, Lu GX, Yuan MT, et al. Elevated temperature overrides the effects of N amendment in Tibetan grassland on soil microbiome[J]. *Soil Biology and Biochemistry*, 2019, 136: 107532
 - [10] Wu LW, Ning DL, Zhang B, et al. Global diversity and biogeography of bacterial communities in wastewater treatment plants[J]. *Nature Microbiology*, 2019, 4(7): 1183-1195
 - [11] Maron PA, Mougél C, Ranjard L. Soil microbial diversity: methodological strategy, spatial overview and functional interest[J]. *Comptes Rendus Biologies*, 2011, 334(5/6): 403-411
 - [12] Simon HM, Smith MW, Herfort L. Metagenomic insights into particles and their associated microbiota in a coastal margin ecosystem[J]. *Frontiers in Microbiology*, 2014, 5: 466
 - [13] Wlasiuk G, Vercelli D. The farm effect, or: when, what and how a farming environment protects from asthma and allergic disease[J]. *Current Opinion in Allergy and Clinical Immunology*, 2012, 12(5): 461-466
 - [14] Li SZ, Deng Y, Zhang ZJ, et al. Advances in molecular detection on aromatic bioremediation[J]. *China Environmental Science*, 2019, 39(6): 2577-2587 (in Chinese) 厉舒祯, 邓晔, 张照婧, 等. 生物降解芳香族化合物的分子检测技术研究进展[J]. *中国环境科学*, 2019, 39(6): 2577-2587
 - [15] Wu YN, Yang YZ, Cao L, et al. Habitat environments impacted the gut microbiome of long-distance migratory swan geese but central species conserved[J]. *Scientific Reports*, 2018, 8(1): 13314
 - [16] Reyes A, Semenkovich NP, Whiteson K, et al. Going viral: next-generation sequencing applied to phage populations in the human gut[J]. *Nature Reviews Microbiology*, 2012, 10(9): 607-617
 - [17] Simon C, Daniel R. Metagenomic analyses: past and future trends[J]. *Applied and Environmental Microbiology*, 2011, 77(4): 1153-1161
 - [18] Frank JA, Reich CI, Sharma S, et al. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes[J]. *Applied and Environmental Microbiology*, 2008, 74(8): 2461-2470
 - [19] Li S, Deng Y, Wang Z, et al. Exploring the accuracy of amplicon-based internal transcribed spacer markers for a fungal community[J]. *Molecular Ecology Resources*, 2020, 20(1): 170-184
 - [20] Bradley IM, Pinto AJ, Guest JS. Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities[J]. *Applied and Environmental Microbiology*, 2016, 82(19): 5878-5891
 - [21] Lane DJ. 16S/23S rRNA sequencing[A]//Stackebrandt E, Goodfellow M. *Nucleic Acid Techniques in Bacterial Systematics*[M]. New York: John Wiley and Sons, 1991: 115-175
 - [22] Winsley T, van Dorst JM, Brown MV, et al. Capturing greater 16S rRNA gene sequence diversity within the domain bacteria[J]. *Applied and Environmental Microbiology*, 2012, 78(16): 5938-5941
 - [23] Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(Supplement 1): 4516-4522
 - [24] Brosius J, Palmer ML, Kennedy PJ, et al. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1978, 75(10): 4801-4805
 - [25] Vescio PA, Nierzwicki-Bauer SA. Extraction and purification of PCR amplifiable DNA from lacustrine subsurface sediments[J]. *Journal of Microbiological Methods*, 1995, 21(3): 225-233
 - [26] Turner S, Pryer KM, Miao VPW, et al. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis[J]. *The Journal of Eukaryotic Microbiology*, 1999, 46(4): 327-338
 - [27] Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms[J]. *The ISME Journal*, 2012, 6(8): 1621-1624
 - [28] Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples[J]. *Environmental Microbiology*, 2016, 18(5): 1403-1414
 - [29] Lane DJ, Pace B, Olsen GJ, et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1985, 82(20): 6955-6959
 - [30] Macrae A. The use of 16S rDNA methods in soil microbial ecology[J]. *Brazilian Journal of Microbiology*, 2000, 31(2): 77-82
 - [31] Stahl D, Amann R. Development and application of nucleic acid probes[A]//Stackebrandt E, Goodfellow M. *Nucleic Acid Techniques in Bacterial Systematics*[M]. Chichester: John Wiley & Sons Ltd., 1991: 205-248
 - [32] Ovreås L, Forney L, Daae FL, et al. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA[J]. *Applied and Environmental Microbiology*, 1997, 63(9): 3367-3373
 - [33] Wang N, Chang ZZ, Xue XM, et al. Biochar decreases nitrogen oxide and enhances methane emissions via altering

- microbial community composition of anaerobic paddy soil[J]. *Science of the Total Environment*, 2017, 581-582: 689-696
- [34] Wang YZ, Feng XY, Natarajan VP, et al. Diverse anaerobic methane-and multi-carbon alkane-metabolizing archaea coexist and show activity in Guaymas Basin hydrothermal sediment[J]. *Environmental Microbiology*, 2019, 21(4): 1344-1355
- [35] White TJ, Bruns TD, Lee SB, et al. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics[A]/PCR Protocols: A Guide to Methods and Applications[M]. San Diego: Academic Press, 1990: 315-322
- [36] Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts[J]. *Molecular Ecology*, 1993, 2(2): 113-118
- [37] Ihrmark K, Bödeker ITM, Cruz-Martinez K, et al. New primers to amplify the fungal ITS2 region-evaluation by 454-sequencing of artificial and natural communities[J]. *FEMS Microbiology Ecology*, 2012, 82(3): 666-677
- [38] Tedersoo L, Lindahl B. Fungal identification biases in microbiome projects[J]. *Environmental Microbiology Reports*, 2016, 8(5): 774-779
- [39] Taylor DL, Walters WA, Lennon NJ, et al. Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing[J]. *Applied and Environmental Microbiology*, 2016, 82(24): 7217-7226
- [40] Toju H, Tanabe AS, Yamamoto S, et al. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples[J]. *PLoS One*, 2012, 7(7): e40863
- [41] Tedersoo L, Anslan S. Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences[J]. *Environmental Microbiology Reports*, 2019, 11(5): 659-668
- [42] Tedersoo L, Bahram M, Pölme S, et al. Global diversity and geography of soil fungi[J]. *Science*, 2014, 346(6213): 1256688
- [43] Bengtsson-Palme J, Ryberg M, Hartmann M, et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data[J]. *Methods in Ecology and Evolution*, 2013, 4(10): 914-919
- [44] Blaxter ML, de Ley P, Garey JR, et al. A molecular evolutionary framework for the phylum Nematoda[J]. *Nature*, 1998, 392(6671): 71-75
- [45] Stoeck T, Bass D, Nebel M, et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water[J]. *Molecular Ecology*, 2010, 19(2): 21-31
- [46] Borneman J, Hartin RJ. PCR primers that amplify fungal rRNA genes from environmental samples[J]. *Applied and Environmental Microbiology*, 2000, 66(10): 4356-4360
- [47] Medlin L, Elwood HJ, Stickel S, et al. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions[J]. *Gene*, 1988, 71(2): 491-499
- [48] Hadziavdic K, Lekang K, Lanzen A, et al. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers[J]. *PLoS One*, 2014, 9(2): e87624
- [49] Bellemain E, Carlsen T, Brochmann C, et al. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases[J]. *BMC Microbiology*, 2010, 10(1): 189
- [50] Leff JW, Bardgett RD, Wilkinson A, et al. Predicting the structure of soil communities from plant community taxonomy, phylogeny, and traits[J]. *The ISME Journal*, 2018, 12: 1794-1805
- [51] Walters W, Hyde ER, Berg-Lyons D, et al. Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys[J]. *mSystems*, 2016, 1(1): e00009-15