

对虾幼体培育全阶段细菌群落 16S rRNA 基因扩增子测序数据集

王艳婷¹, 王可¹, 高婧¹, 刘小爽¹, 於俊琦², 李明¹, 董鹏生^{1*}

1 河南农业大学 动物科技学院, 河南 郑州

2 浙江省海洋水产养殖研究所, 全省近岸生物种质资源保护与利用重点实验室, 浙江 温州

王艳婷, 王可, 高婧, 刘小爽, 於俊琦, 李明, 董鹏生. 对虾幼体培育全阶段细菌群落 16S rRNA 基因扩增子测序数据集[J]. 微生物学报, 2026, 66(8): 4276-4288.

WANG Yanting, WANG Ke, GAO Jing, LIU Xiaoshuang, YU Junqi, LI Ming, DONG Pengsheng. A 16S rRNA gene amplicon sequencing dataset of the bacterial community across all larviculture stages of shrimp[J]. *Acta Microbiologica Sinica*, 2026, 66(8): 4276-4288.

摘要: 凡纳滨对虾作为全球水产养殖产量和经济价值最高的甲壳类经济动物之一, 其健康与养殖系统微生物群落演替和微生态平衡密切相关。苗种的稳定供给是对虾养殖业高质量发展的基础与核心环节之一, 解析对虾早期发育阶段细菌群落演替规律是实现育苗微生态调控的重要前提。【目的】建立凡纳滨对虾幼体培育全阶段的标准化细菌群落数据集, 为系统开展对虾幼体发育过程中微生物群落研究提供数据支撑。【方法】使用规范的采样和实验室处理流程, 全程跟踪苗种培育系统中对虾幼体的发育过程, 获取对虾幼体与水体样本, 采用 16S rRNA 基因扩增子高通量测序技术, 结合 Dix-seq 扩增子数据分析流程, 构建细菌群落数据集。【结果】该数据集涵盖对虾幼体发育全程 102 个样品(48 个幼体和 54 个水体)的双端测序原始序列, 共 204 个 fastq.gz 文件和 4 709 988 条原始序列, 质控后获得 4 255 994 条有效序列及 202 505 个零半径操作分类单元(zero-radius operational taxonomic units, ZOTUs)。数据产生严格遵循统一的标准和规范, 通过多点混合采样保障微生物样本代表性, 采用标准化的样本采集与核酸提取流程, 并利用 Dix-seq 工作流的参数卡机制公开完整的数据分析参数, 确保数据分析的可重复性与可靠性。【结论】本数据集样本覆盖幼体发育整个过程, 实现了对虾苗种培育全程养殖系统细菌群落的同步追踪, 不仅为探究对虾早期发育阶段细菌群落演替及“宿主-环境”互作机制提供了基础数据, 助力构建基于微生态调控的绿色苗种培育技术体系; 同时也为同类别高通量测序数据的规范化管理与分析提供了范例, 对提升农业生态系统中微生物组学研究的规范性和可操作性具有重要意义。

关键词: 凡纳滨对虾; 苗种培育; 细菌群落; 群落演替; “宿主-环境”互作; 数据集

资助项目: 国家自然科学基金(30900978); 河南省科技攻关项目(252102111033); 浙江省农业(水产新品种选育)新品种选育重大科技专项(2021C02069)

This work was supported by the National Natural Science Foundation of China (30900978), the Science and Technology Research Project of Henan Province (252102111033), and the Zhejiang Provincial Major Special Project on Agricultural (Aquatic) New Variety Breeding (2021C02069).

*Corresponding author. E-mail: dpsh@henau.edu.cn

Received: 2026-02-14; Accepted: 2026-04-21; Published online: 2026-05-11

数据集基本信息简介

数据集名称	对虾幼体培育全阶段细菌群落 16S rRNA 基因扩增子测序数据集
数据通信作者	董鹏生(dpsh@henau.edu.cn)
数据作者	王艳婷,王可,高婧,刘小爽,於俊琦,李明,董鹏生
数据产生时间	2025年5月
地理区域	浙江省海洋水产养殖研究所(28.28°N, 121.11°E)
数据量	553 MB
数据格式	*.fastq.gz、.xlsx、.txt
数据服务系统网址	https://doi.org/10.57760/sciencedb.j00231.00047
基金项目	国家自然科学基金(30900978);河南省科技攻关项目(252102111033);浙江省农业(水产新品种选育)新品种选育重大科技专项(2021C02069)
数据集组成	数据集包含1个文件夹,1个Excel数据表,1个TXT文件:“原始测序数据(rawdata)”文件夹包括对虾幼体和苗种培育水体的16S rRNA 基因扩增子测序原始序列,共204个fastq.gz文件;“数据质量控制统计(summary of illumina sequence characteristics)”数据表记录各样本序列数目及质量评估;“metadata.txt”文件记录了对测序原始序列进行序列拼接、引物截除、去噪、物种注释等分析的参数

A 16S rRNA gene amplicon sequencing dataset of the bacterial community across all larviculture stages of shrimp

WANG Yanting¹, WANG Ke¹, GAO Jing¹, LIU Xiaoshuang¹, YU Junqi², LI Ming¹, DONG Pengsheng^{1*}

1 College of Animal Science and Technology, Henan Agricultural University, Zhengzhou, Henan, China
2 Zhejiang Key Laboratory of Coastal Biological Germplasm Resources Conservation and Utilization, Zhejiang Mariculture Research Institute, Wenzhou, Zhejiang, China

Abstract: *Litopenaeus vannamei* is one of the crustacean species with the highest production and economic value in the global aquaculture industry. Its health status is closely linked to the community succession and balance of the microbiota in the aquaculture system, and the stable supply of seeds is one of the fundamental and core components of the high-quality development of the shrimp aquaculture industry. Characterizing the pattern of bacterial community succession during the early developmental stages of shrimp is an essential prerequisite for achieving microbiome-based regulation in larviculture. **[Objective]** To establish a standardized bacterial community dataset covering the entire cycle of *L. vannamei* larviculture, thereby providing data support for systematic investigations on the microbial communities during shrimp larvae development. **[Methods]** Focusing on the complete developmental stages of *L. vannamei* larvae, larval shrimps and rearing water samples were collected from the larviculture system of the Zhejiang Mariculture Research Institute following standardized sampling and laboratory processing protocols. Bacterial community datasets were constructed by high-throughput 16S rRNA gene amplicon sequencing combined with the standard Dix-seq amplicon analysis pipeline. **[Results]**

This dataset encompassed the raw paired-end sequencing reads from 102 samples (48 larval samples and 54 rearing water samples) across the entire cycle of shrimp larviculture, comprising a total of 204 fastq.gz files and 4 709 988 raw paired-end reads. After quality control, 4 255 994 effective sequences and 202 505 zero-radius operational taxonomic units (ZOTUs) were obtained. Data generation strictly followed standardized protocols, with multi-point composite sampling ensuring sample representativeness of microbial samples, and unified sample collection and nucleic acid extraction procedures were employed. All analytical parameters were publicly disclosed through the parameter card mechanism of the Dix-seq pipeline, guaranteeing the reproducibility and reliability of the data analyses. **[Conclusion]** The samples in this dataset cover the key developmental nodes of *L. vannamei* larvae, achieving synchronous monitoring of the bacterial communities throughout the entire larviculture period. It provides fundamental data for exploring bacterial community succession and host-environment interaction mechanisms during the early developmental stages of shrimp, thereby facilitating the development of green larviculture technologies based on microbiome regulation. Additionally, it serves as a benchmark for the standardized management and analysis of similar high-throughput sequencing data, which is of great significance for enhancing the standardization and operability of microbiome research in agricultural ecosystems.

Keywords: *Litopenaeus vannamei*; larviculture; bacterial community; community succession; host-environment interaction; dataset

Dataset profile

Title	A 16S rRNA gene amplicon sequencing dataset of the bacterial community across all larviculture stages of <i>Litopenaeus vannamei</i>
Data corresponding author	DONG Pengsheng (dpsh@henau.edu.cn)
Data authors	WANG Yanting, WANG Ke, GAO Jing, LIU Xiaoshuang, YU Junqi, LI Ming, DONG Pengsheng
Data generation time	May 2025
Geographical scope	Zhejiang Mariculture Research Institute (28.28°N, 121.11°E)
Data volume	553 MB
Data format	*.fastq.gz, .xlsx, .txt
Data service system	https://doi.org/10.57760/sciencedb.j00231.00047
Sources of funding	National Natural Science Foundation of China (30900978); Science and Technology Research Project of Henan Province (252102111033); Zhejiang Provincial Major Special Project on Agricultural (Aquatic) New Variety Breeding (2021C02069)
Dataset composition	The dataset was composed of one folder, one Excel data table, and one TXT file. The folder named “Raw sequencing data (Rawdata)” contained the raw 16S rRNA gene amplicon sequences from 48 shrimp larval samples and 54 aquaculture water samples, comprising a total of 204 fastq.gz files. The Excel table “summary of illumina sequence characteristics” documented the sequence quantity and quality assessment for each sample. The “metadata.txt” file documented the parameters employed for the analyses of raw sequencing reads, including sequence assembly, primer trimming, denoising, and taxonomic annotation

水产养殖是全球粮食安全保障体系的重要组成部分,近年来实现规模化快速发展,其中对虾养殖凭借显著的经济效益占据重要地位。我国是对虾养殖大国,凡纳滨对虾(*Litopenaeus vannamei*)作为主导养殖品种,已形成集苗种培育、成虾养殖、加工和销售于一体的完整产业链^[1]。当前,对虾养殖正经历从传统劳动密集型向集约化、工厂化及智慧化的深刻变革,形成多元化的养殖模式格局。然而,其产业发展仍面临严峻挑战:一方面,对虾种质资源退化问题日益凸显;另一方面,在高密度养殖模式下,对虾养殖系统微生态平衡相对脆弱,对养殖环境生态调控与精细化管理技术提出了更高的要求^[2-3]。上述问题导致对虾养殖疾病频发、养殖系统不稳定,已成为制约产业绿色高质量发展的关键瓶颈之一^[2-4]。对虾健康状况直接决定养殖产量与产业效益,而苗种培育作为对虾养殖的初始关键环节,其环境微生态稳定是维持幼体存活和生长发育的重要基础^[5-6]。微生物群落被喻为宿主“看不见的器官”,在对虾养殖生态系统中占据核心地位,不仅参与水体物质循环与污染物降解,还能通过构建共生屏障抵御病原入侵;因此,养殖系统微生物群落组成和演替与对虾幼体健康以及苗种质量直接相关^[7-9],甚至显著影响对虾的生长与抗逆性能、营养品质和疾病发展进程^[10-14]。水体微生物作为对虾肠道微生物最重要的来源之一,可通过迁移与定殖直接影响肠道菌群结构和组成,尤其对早期肠道菌群的构建具有决定性作用^[15-17]。因此,解析苗种培育系统细菌群落特征及演替机制,不仅能为优化育苗工艺提供精准的数据支持,而且可以为以微生物调控为核心的对虾绿色养殖技术提供科学依据与原理支撑。

尽管微生物组学已成为对虾养殖与幼体发育领域的研究热点,但目前部分研究存在数据表型信息缺失、数据分析参数和流程不清晰、结果难以复现和微生物群落机制解析不深入等问题,制约了该方向研究的纵深突破与技术转

化。这些问题主要体现在3个方面。(1)数据公开性与规范性不足:微生物组原始数据未(完整)公开,或公开数据的元信息(如采样时间、养殖环境参数、样本处理细节等)缺失、格式混乱,导致数据整合分析困难,不利于形成系统性认知。(2)数据分析核心参数公开不全面:许多研究仅注重分析结果的呈现,对数据分析工具使用的参数描述并不详细,尤其是微生物测序数据的质控、拼接、去噪等过程中的参数,以及物种注释相似度阈值、数据库选择等关键信息公开得并不全面,导致研究结果难以复现或重复验证。例如,针对相同的16S rRNA基因扩增子原始数据,采用不同的去噪算法或聚类策略可能导致物种多样性指数及群落组成分析结果存在差异,进而影响结论的可靠性与结果的可重复性^[18]。(3)技术流程缺乏统一标准:在样本采集、核酸提取及组学数据分析等环节,不同研究的采样方式、DNA提取试剂盒选择、测序区域偏好以及生物信息学分析流程等关键技术流程并不统一,不仅降低了跨研究样本的可比性,还可能因处理流程差异产生批次效应,进而带来显著偏差的结果^[19-23]。

针对当前对虾微生态研究中数据标准化程度不足、元信息缺失,以及“宿主-环境”互作机制研究视角局限等问题,本研究提出以全流程标准化数据产出结合多维度数据集构建为核心的解决思路。在实验设计上,本数据集以对虾养殖环境本底为基线,结合“宿主-环境”动态同步的监测体系,不仅可以确立养殖水源的本底微生物群落结构和组成,而且通过对苗种培育关键节点的对虾幼体与环境水体严格配对采集,实现解析“水体-幼体”群落演替规律及互作机制的目的。在数据规范上,本数据集不仅公开了所有的微生物序列信息,而且同步公开了样品收集、核酸提取等实验的标准化流程及数据分析方法和参数。这种原始数据搭配规范化流程与分析方法的数据集生产模式可降低方法差异导致的结果偏差,提升数据可重复性与二次挖

掘潜力。本数据集的建立不仅系统性地补充了对虾苗种培育全过程中养殖系统的细菌群落数据,而且为解决当前研究中技术流程不规范、数据传播与重复利用困难等共性问题提供了数据资源与技术范例。本数据集对于阐明“宿主-微生物”互作机制、研发幼体培育阶段微生态调控技术、推动对虾健康养殖领域研究范式向标准化转型具有重要的学术价值与实践意义,同时也为其他水产微生物组学的规范化数据研究提供参考。

1 数据采集和处理方法

2025年5月在浙江省海洋水产养殖研究所(28.28°N, 121.11°E)随机选取12个规格一致的苗种培育池(6 m×5 m×1.3 m)进行对虾幼体发育的全程监测。为确保实验条件的均一性,所有苗种培育系统执行严格一致的管理措施,包括放苗前砂滤海水的输入与肥水处理、受精卵来源、培育管理方式,以及饵料类型、投喂频次和时间。实验周期覆盖对虾幼体培育全程:从受精卵孵化池转移至育苗池开始,对虾历经无节幼体、蚤状幼体、糠虾幼体至仔虾期共4个发育阶段,共设置5个采样时间点(图1),分别采集砂滤海水、对虾幼体及苗种养殖环境样本,共获得48个幼体样本和54个水体样本。

本研究中所有动物实验均获得河南农业大学科学伦理委员会批准,编号为HNND2023031420。

使用QIAamp® DNA Stool Mini Kit (Qiagen公司)提取对虾幼体总DNA,采用PowerSoil® DNA Kit (MOBIO公司)提取水体样本对应滤膜上的总DNA。用NanoDrop ND-2000型核酸测定仪(Thermo公司)测定DNA的浓度和纯度。用带双Barcode的引物515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3')和806R-B (5'-GGACTACN VGGGTWTCTAAT-3')对细菌16S rRNA基因V4区域进行PCR扩增,在Illumina NovaSeq 6000平台以双端250 bp模式完成测序。

利用Dix-seq扩增子标准分析流程(v1.0.0)^[18],调用Trimmomatic (v0.38)^[24]、USEARCH (v11.0667)^[25]等软件对原始序列进行序列拼接、引物截除、去噪等分析,获得零半径操作分类单元(zero-radius operational taxonomic unit, ZOTU)序列和丰度信息,并以SILVA数据库(v138.2)^[26]为参考,用sintax算法^[27]对ZOTU序列进行物种分类注释。本研究数据分析的所有参数均储存在参数卡文件“metadata.txt”中,并在科学数据银行(Science Data Bank, ScienceDB)数据库同步公开。共获得4 709 988条原始序列,质控后得到4 255 994条有效序列(平均每个样品41 725.4条有效序列),去噪后获得202 505个

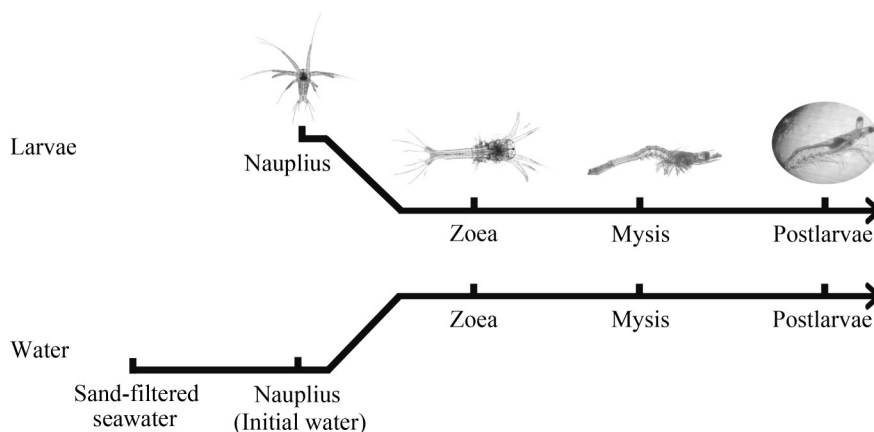


图1 实验设计和采样时间轴

Figure 1 Experimental design and sampling schedules.

ZOTUs (平均每个样品 1 985.4 个 ZOTUs)序列并构建特征序列表。

2 数据样例描述

2.1 数据集命名规则及数据量

本数据集为凡纳滨对虾苗种培育全过程细菌群落 16S rRNA 基因扩增子测序数据, 数据集总大小为 553 MB。文件名为“对虾幼体培育全阶段细菌群落 16S rRNA 基因扩增子测序数据集”, 含有 1 个文件夹、1 个 Excel 数据表和 1 个 txt 文件, 分别命名为“原始测序数据(rawdata)”“数据质量控制统计(summary of illumina sequence characteristics)”“metadata”。

“原始测序数据(rawdata)”文件夹内含 204 个 fastq.gz 文件, fastq.gz 文件为每个样本的原始双端测序文件。fastq.gz 文件命名规则为“分组+处理编码”, 分组以字母 L 结尾的是幼体样本, 如文件“ZL1_1.fastq.gz”和“ZL1_2.fastq.gz”是蚤状幼体时期幼体样本 1 的双端测序文件; 分组以字母 W 结尾的是水体样本, 如文件“ZW1_1.fastq.gz”和“ZW1_2.fastq.gz”是蚤状幼体时期培育水体样本 1 的双端测序文件。

“数据质量控制统计(summary of illumina sequence characteristics)”数据表是数据质量控制统计结果, 记录各样本序列数目及质量评估。

“metadata.txt”是参数卡, 记录对测序原始序

列进行去噪、拼接、引物匹配等分析的参数信息。

2.2 数据样本描述

样本原始测序数据以 FASTQ 格式存储, 采用 ASCII 字符记录生物序列及其对应质量信息, 是当前高通量测序领域通用的标准文件格式。FASTQ 文件中每条序列由 4 行构成: 第 1 行为序列标识行, 以“@”起始, 包含唯一的序列编号及测序相关描述信息, 二者以空格分隔; 第 2 行为核酸序列, 由碱基 A、T、C、G 组成; 第 3 行以“+”起始, 无额外描述信息; 第 4 行为碱基质量值编码, 字符长度与核酸序列完全一致, 可通过质量值换算规则得到各碱基的测序质量分数, 用于表征测序过程中单个碱基的识别准确度和错误概率。“ZL1_1.fastq.gz”文件中部分序列信息如图 2 所示。

数据表“数据质量控制统计”的具体内容、表头示例、字段含义、数据类型、数据项含义等信息见表 1。

“metadata.txt”参数卡中参数示例及参数含义见表 2。

3 数据质量控制

本数据集主要从样本采集、样品保存、核酸提取及测序数据分析 4 个方面进行数据质量控制。

```
@Z1CKG1_26015 1:N:0:CGGAAT
TCCACCCTCTATGTGTCAGCCGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATTACTGGGCGTAAA
GCGCACGTAGGCGGTCGGATAAGTCGGATGTGAAAGCCCTGGGCTTAACCTGGGAAGTGCATTGATACTGT
CTGGCTAGAGTATGGGGGAGGGAAGCGGACTTTCCGGTGTAGCGGTGAAATGCGTAGATATCGGAAGGAAC
ATCAATGGCGAAGGCAGCTTCTGGCCCAATACTG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
F:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFF
F:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFF
F:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFFFFF:FFF
```

图2 “ZL1_1.fastq.gz”文件中部分序列信息

Figure 2 Part of the sequence information in the “ZL1_1.fastq.gz” file.

表1 本数据集Excel表内容及字段含义

Table 1 Data content and descriptions in excel table

分类名称 Category name	字段名称 Field name	数据类型 Data type	量纲 Unit	数据项说明 Description
Sample information	Sample ID	字符型	无	样本名称
	Sample description	字符型	无	样本类型说明
Raw reads	Raw PE reads	数字型	个	原始双端序列数
Quality Control	Reads	数字型	个	去除低质量碱基后的双端序列数
	Base	数字型	个	去除低质量碱基后的序列数中总碱基数目
	Q20	数字型	%	质量值大于等于20的碱基占总碱基数的百分比
	Q30	数字型	%	质量值大于等于30的碱基占总碱基数的百分比
Merged paired sequence	Merged reads	数字型	个	完成双端序列合并的序列数
	Merged rate	数字型	%	有效合并的序列数占质控后总序列的百分比
Matched primer sequence	Primer_match reads	数字型	个	引物匹配的序列数
	Primer_match rate	数字型	%	引物匹配的序列数占质控后总序列的百分比
ZOTU	Tags number	数字型	个	该样本中的有效序列总数
	ZOTUs number	数字型	个	ZOTU特征序列数

3.1 样本采集质量控制

样本采集严格按照董鹏生等^[28]的方法和步骤进行。

水样采集：使用有机玻璃采水器在苗种培育池的每条边的中点分别采集 1 000 mL 水体(共设置 4 个采样点)，装入 4 L 无菌采样瓶，混合均匀。

对虾幼体微生物样品采集：以同一育苗池为独立生物学重复，对虾幼体在同一池中培育至仔虾期；根据幼体发育时序，在无节幼体、蚤状幼体、糠虾幼体、仔虾期 4 个关键发育节点分别从池中取样。使用 160 目(约 95 μm 孔径)灭菌筛绢过滤水样，富集对应发育时期的对虾幼体，将截留的幼体用灭菌海水冲洗 3 次，以去除体表附着的水体微生物及杂质，并转入 2 mL 无菌离心管，4 $^{\circ}\text{C}$ 、100 $\times g$ 离心 1 min，沉淀物即为该育苗池的对虾幼体微生物样本。为确保各样本生物量一致，每个育苗池收集的幼体湿重约为 0.8 g。

苗种培育水体微生物样品收集：水体微生物的富集采用真空抽滤法。利用真空泵将

160 目灭菌筛绢预过滤的水体微生物富集至 0.2 μm 聚碳酸酯膜上。每张滤膜抽滤 200–500 mL 水样(视水体浊度调整，但同一批次样本保持体积一致)。为获得足量的微生物 DNA，每次水样采集过滤 3 张滤膜，合并作为该时间点的单一水体微生物样本。

通过多点混合采样确保样品的代表性；幼体采集严格遵循无菌操作，并使用灭菌海水冲洗以去除体表杂质及杂菌，从而获得幼体样本；在水体微生物富集前，对过滤装置进行预处理，选用 0.2 μm 滤膜并严格控制过滤水量，高效富集水体微生物并保证样品的平行性。上述操作旨在获取无污染且具代表性的幼体及水体微生物样本，为后续实验及微生物群落分析的准确性奠定基础。

3.2 样品保存质量控制

样品采集后迅速置于液氮中速冻，运回实验室后转入 -80 $^{\circ}\text{C}$ 超低温冰箱保存。该处理可有效减少运输和储存过程中因温度波动导致的微生物群落结构改变，降低 DNA 降解风险，保障后续 16S rRNA 基因扩增及高通量测序的

表2 本数据集“metadata.txt”参数卡内容及含义

Table 2 The content and descriptions in “metadata.txt” file

参数名称	参数项说明
Parameter name	Description
project_home	工作目录绝对路径
project_id	分析结果文件名
raw_data	原始数据储存绝对路径
mapping_file	样本名与分组信息表(mapping_file)文件绝对路径
cpus	项目分配的总CPU/线程数
parallel	并行处理的样本数
threads	单样本最多使用的线程数
trim_param	剪切接头和低质量碱基的参数
mergepairs_param	合并双端序列 usearch 参数
search_pcr_param	引物切除参数
maxee	过滤掉错误率之和大于指定值(默认 1)的序列
unoise3 param	usearch 去噪参数(必须指定, 建议 minisize 为 8)
norm	抽平参数, 默认进行抽平操作
db	数据库文件夹绝对路径
classifier	指定分类学方法(推荐/默认 syntax 算法)
syntax_cutoff	指定物种分类的预测准确性阈值
filter	指定需要过滤的物种分类名(默认参数不过滤)
tree	指定代表序列的发育树构建算法
rarefaction	指定稀释曲线参数
rarefaction_text	指定稀释曲线上进行文字描述
text	指定描述β多样性分析中数据点标注样本名称
pathway	设置是否对“picrust2”软件在 pathway 层进行预测
vegdist	计算β多样性采用的算法
DESeq2 qval	指定 DESeq2 鉴定差异类群的参数
DESeq2 foldchange	设置 DESeq2 的 log ₂ fold change 阈值
volcano_text	设置可视化部分是否显示样本名称
quant_trim	指定阈值, 用于过滤 ZOTU 表中丰度低于阈值的 ZOTU
kwt_qval	指定 Kruskal-Wallis 检验的 <i>q</i> 值, 默认参数为 0.05
kwt_pval	逻辑型参数; 若选择“T”, Kruskal-Wallis 检验使用 <i>P</i> 值, “F” (默认) 则为 <i>q</i> 值
wilcoxon_qval	指定 wilcoxon 检验的 <i>q</i> 值, 默认参数为 0.05
wilcoxon_pval	逻辑型参数; 若选择“T”, wilcoxon 检验使用 <i>P</i> 值, “F” (默认) 则为 <i>q</i> 值
lefse_para	设置线性判别分析(linear discriminant analysis effect size, LEfSe)参数阈值

DNA 质量, 确保微生物群落分析结果的准确性与可靠性。

3.3 核酸提取质量控制

对虾幼体微生物组 DNA 提取: 在灭菌后的研钵中加入液氮, 将幼体样品充分研磨成粉状。

称取 0.2 g 研磨后样品, 使用 QIAamp® DNA Stool Mini Kit 提取微生物组 DNA。

水体微生物组 DNA 提取: 在无菌环境下, 用镊子将滤膜置于培养皿中, 用剪刀将截留微生物的滤膜剪碎, 转移至-4 °C预冷的 2 mL 无

菌离心管中, 并使用 PowerSoil® DNA Kit 提取微生物组 DNA。

液氮研磨可实现组织与微生物细胞的低温破碎, 同时抑制核酸酶活性以减少 DNA 降解; 剪碎滤膜能增大微生物与提取试剂的接触面积, 预冷无菌离心管可维持核酸稳定性; 定量取样结合针对性试剂盒可高效富集目标微生物 DNA, 减少干扰。上述操作旨在获得高纯度、高完整性的 DNA 模板, 保障后续扩增、测序及微生物群落分析结果的准确性。

3.4 高通量测序数据质量控制

利用参数卡和 Dix-seq 扩增子标准分析流程^[18]对原始序列进行处理, 包括质控、去噪、拼接、引物匹配、物种注释、群落多样性分析、功能预测和数量生态学等分析。规范记录并公开质控和去噪参数(如低质量序列阈值)、物种注释准确度及统计分析量, 确保数据处理的规范与参数透明, 减少因处理流程、方法和参数差异导致结果偏差, 为后续研究结果的准确解析和规律总结奠定基础。

4 数据使用方法和建议

本数据集的核心为凡纳滨对虾苗种培育全阶段幼体及水体微生物 16S rRNA 基因扩增子测序数据, 含原始测序数据、数据分析参数卡及样本元数据, 具体建议如下:

数据分析规范: 建议使用前查阅数据集附带的质控报告, 并采用本研究提供的参数卡和 Dix-seq 扩增子标准分析流程完成原始序列去噪、拼接与引物匹配等处理。同时, 应统一过滤与去噪参数、注释数据库版本号及物种分类阈值, 以确保数据分析流程的规范性, 减少因分析流程差异引入偏差。

Dix-seq 扩增子数据分析流程使用指南: 课题组已将 Dix-seq 分析流程的安装指南、使用教程及参数卡文件配置示例上传至多个主流代码与文档共享平台。

快速入门版(CSDN): 提供了 Dix-seq 软件参数卡设置和数据分析的简洁教程, 适合初学者快速了解数据结构与基础分析流程, 网址为 https://blog.csdn.net/dpsh_6/article/details/146340577。

详细完整版(Gitee & GitHub): 提供了详细的 Dix-seq 工具安装说明、参数卡配置详解及分析教程, 网址为 <https://gitee.com/dpsh/dix-seq/wikis>(Gitee) 和 <https://github.com/jameslz/dix-seq/wiki>(GitHub)。

metadata.txt 完整示例文件: <https://gitee.com/dpsh/dix-seq/tree/master/example>。

核心应用方向: (1) 群落特征分析: 利用 Dix-seq 工作流调用 USEARCH 软件、R 语言 vegan 包等工具, 基于 ZOTU 特征表解析不同苗种培育阶段微生物多样性与物种组成差异; (2) 关联机制探究: 通过相关性分析、共线网络分析以及群落构建机制解析等方法, 挖掘幼体与水体微生物群落的互作关系, 或初始微生物群落与苗种培育成效的潜在关联; (3) 支持跨场景拓展应用: 如在研究不同对虾品系或养殖系统间微生物群落对比时需严格对齐本数据集的实验设计和分析标准, 以提升结果的可比性。

关键注意事项: 本数据集源于特定苗种培育条件(浙江省某苗种培育池), 结果解读需结合具体育苗环境与技术背景。数据二次开发应遵循 ScienceDB 数据库使用协议, 限于非商业科研用途。引用时需标注数据集存储地址及相关论文来源, 确保数据溯源规范。

5 数据集价值

对虾苗种培育系统是涵盖幼体宿主、共生/寄生微生物及自由态环境微生物的复杂生态系统, 微生物群落作为环境与幼体健康的纽带, 直接影响虾苗质量及后续养殖成效。本数据集系统收录了凡纳滨对虾苗种培育全阶段的幼体及水体微生物群落数据, 具有完整的元数据信息、系统的采样设计及“宿主-环境”互作的创新

视角, 为该领域研究提供了高质量且参考性强的基础资源。

本数据集不仅公开了原始测序数据, 还提供了完整的元数据信息(含样本表型信息)、记录所有分析参数的参数卡文件及数据分析教程, 确保分析流程的规范性和参数公开透明, 有效减少了因处理方法和参数差异导致的偏差, 提升了本数据集分析结果的可重复性及二次数据挖掘潜力。本数据集基于严格的实验设计与标准化的样品采集及实验室处理流程, 测序数据采用统一分析标准, 覆盖了对虾从无节幼体到仔虾期的全部关键发育节点, 适用于系统研究对虾幼体发育阶段微生物群落的演替规律。此

外, 本数据集同步采集并测序了养殖原水(砂滤海水)、初始培育用水及养殖水体等环境微生物样本, 这种配对采样的设计有利于深入探究对虾育苗系统中“水体-幼体”微生物的迁徙机制及互作网络。基于本数据集, 可通过 α 多样性分析揭示对虾幼体从无节幼体到仔虾期发育过程中细菌群落多样性的变化规律, 并比较环境样本与对虾样本间菌群多样性的差异(图 3); 细菌群落组成分析可进一步揭示优势菌群在不同发育阶段的演替特征、特定菌群与对虾发育阶段的潜在关联, 以及幼体与水体样本间类群组成的异同(图 4)。

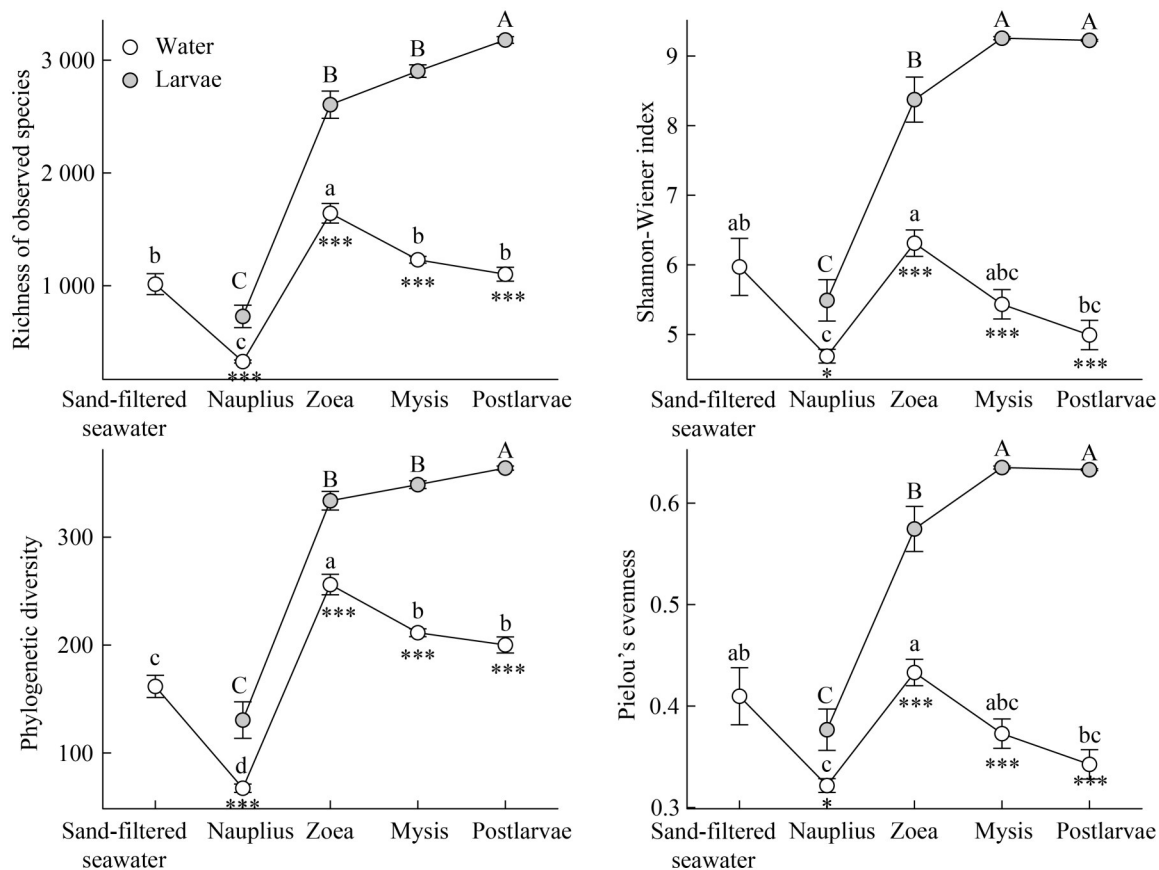


图3 对虾幼体和水体样品的细菌 α 多样性和均匀度指数

Figure 3 Bacterial α diversity and evenness indices of shrimp larvae and water samples. Data present means $\pm$ standard errors. The different letters indicate significant differences between stages ($P<0.05$) (uppercase for larvae; lowercase for water samples). Significance of differences between larvae and water at each stage was tested using Mann-Whitney U test (*: $P<0.05$; ***: $P<0.001$).

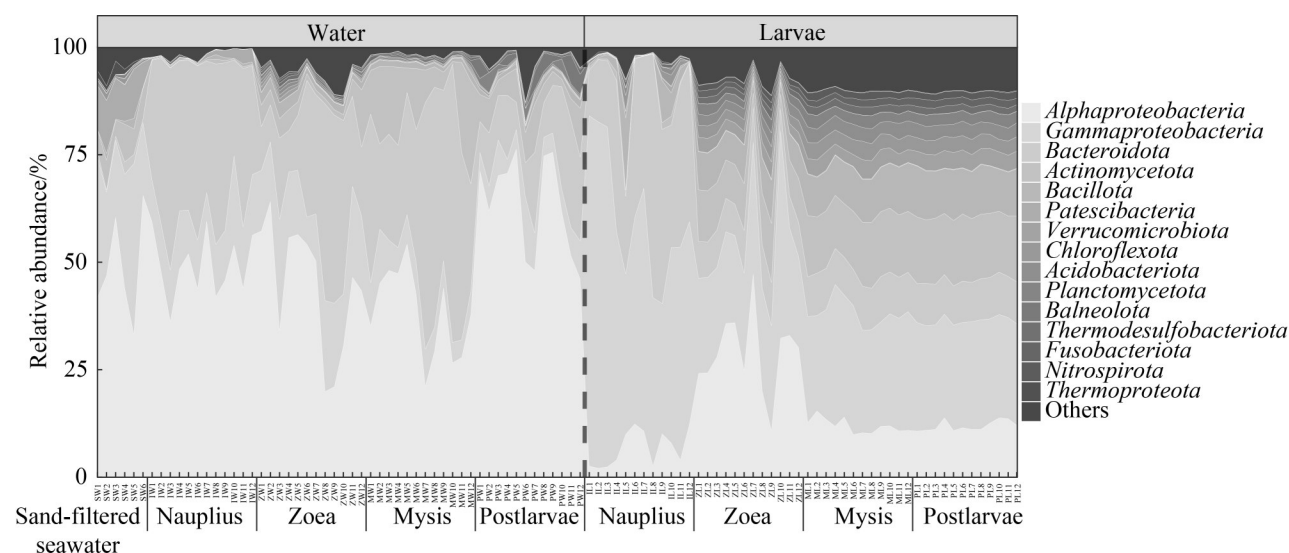


图4 对虾幼体和水体样品细菌群落的主要物种组成

Figure 4 Taxonomic composition of bacterial communities in shrimp larvae and water samples.

6 结论

本数据集聚焦凡纳滨对虾苗种培育全程，系统收录 102 个样品的 16S rRNA 基因扩增子双端测序数据，构建了包括环境本底、对虾幼体与环境动态同步的精细监测链条：以砂滤海水为环境微生物本底，在受精卵孵化后的 4 个关键发育阶段(无节幼体、蚤状幼体、糠虾幼体、仔虾期)同步采集对虾幼体与养殖水体样本，完整覆盖了微生物从水源输入至幼体迁徙和定殖的全过程，提升细菌群落演替数据的可靠性。数据分析流程严格遵循微生物组学技术规范，配套发布统一格式的元数据表(含样本类型、发育阶段等信息)、标准化分析参数卡及可复现分析流程，增强了数据的可信度与可追溯性。该数据集不仅补充了对虾苗种培育完整阶段的微生态研究数据，而且为解析菌群“环境-宿主”传播路径与演替规律提供了高质量的数据支撑，同时为幼虾培育阶段的微生态调控技术研发提供了可复用范例。上述成果对推动对虾等水产动物微生物组学数据规范与研究范式创新具有重要学术价值与应用前景。

7 数据(集)可用性声明

本研究产生的凡纳滨对虾苗种培育系统微生物群落数据集(包含幼体及水体样本的测序原始数据、标准化参数卡 metadata.txt 及完整样本元数据)已完整上传至 ScienceDB 科学数据银行。数据集编号为 DOI: 10.57760/sciencedb.j00231.00047, 可通过 <https://doi.org/10.57760/sciencedb.j00231.00047> 免费获取全部资源。

为提升研究过程的透明化及数据的高效复用，课题组构建了规范的数据分析流程及配套文档资源，将 Dix-seq 扩增子数据分析流程的安装及使用指南、参数卡文件的详细示例上传至以下 3 个主流代码与文档共享平台。

快速入门版(CSDN): https://blog.csdn.net/dpsh_6/article/details/146340577。

详细完整版(Gitee & GitHub): [https://gitee.com/dpsh/dix-seq/wikis\(Gitee\)](https://gitee.com/dpsh/dix-seq/wikis(Gitee)); [https://github.com/jameslz/dix-seq/wiki\(GitHub\)](https://github.com/jameslz/dix-seq/wiki(GitHub))。

metadata.txt 完整示例文件: <https://gitee.com/dpsh/dix-seq/tree/master/example>。

作者贡献声明

王艳婷: 样品收集, DNA 提取, 数据论文撰写及数据质量控制, 论文校稿; 王可: 样品收集, DNA 提取, 数据质量控制; 高婧: DNA 提取, 数据质量控制; 刘小爽: 样品收集, 数据质量控制; 於俊琦: 养殖实验, 样品收集, 数据质量控制; 李明: 术语表述规范, 数据质量控制; 董鹏生: 论文构架, 数据论文撰写及数据质量控制, 论文校稿。

作者利益冲突公开声明

作者声明无任何与本研究相关的利益冲突。

参考文献

- [1] 刘梅, 王宝杰, 蒋克勇, 王雷. 对虾病害防控的理论与实践[J]. 海洋科学, 2020, 44(7): 16-25.
Liu M, Wang BJ, Jiang KY, Wang L. Disease control in shrimps: theory and practice[J]. Marine Sciences, 2020, 44(7): 16-25 (in Chinese).
- [2] Thitamadee S, Prachumwat A, Srisala J, Jaroenlak P, Salachan PV, Sritunyaluksana K, Flegel TW, Itsathitphaisarn O. Review of current disease threats for cultivated penaeid shrimp in Asia[J]. Aquaculture, 2016, 452: 69-87.
- [3] Xiong JB, Dai WF, Li CH. Advances, challenges, and directions in shrimp disease control: the guidelines from an ecological perspective[J]. Applied Microbiology and Biotechnology, 2016, 100(16): 6947-6954.
- [4] De Schryver P, Vadstein O. Ecological theory as a foundation to control pathogenic invasion in aquaculture[J]. The ISME Journal, 2014, 8(12): 2360-2368.
- [5] Moss KRK, Moss SM. Effects of artificial substrate and stocking density on the nursery production of Pacific white shrimp *Litopenaeus vannamei*[J]. Journal of the World Aquaculture Society, 2004, 35(4): 536-542.
- [6] Shan XY, Li KY, Stadler P, Borbor M, Reyes G, Solórzano R, Chamorro E, Bayot B, Cordero OX. Microbiome determinants of productivity in aquaculture of whiteleg shrimp[J]. Applied and Environmental Microbiology, 2025, 91(5): e02420-24.
- [7] Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease[J]. Nature Reviews Immunology, 2009, 9(5): 313-323.
- [8] Akshintala VS, Talukdar R, Singh VK, Goggins M. The gut microbiome in pancreatic disease[J]. Clinical Gastroenterology and Hepatology, 2019, 17(2): 290-295.
- [9] Xiong JB, Wang K, Wu JF, Qiuqian LL, Yang KJ, Qian YX, Zhang DM. Changes in intestinal bacterial communities are closely associated with shrimp disease severity[J]. Applied Microbiology and Biotechnology, 2015, 99(16): 6911-6919.
- [10] Xiong JB, Zhu JY, Dai WF, Dong CM, Qiu QF, Li CH. Integrating gut microbiota immaturity and disease-discriminatory taxa to diagnose the initiation and severity of shrimp disease[J]. Environmental Microbiology, 2017, 19(4): 1490-1501.
- [11] Rungrassamee W, Klanchui A, Chaiyapechara S, Maibunkaew S, Tangphatsornruang S, Jiravanichpaisal P, Karoonuthaisiri N. Bacterial population in intestines of the black tiger shrimp (*Penaeus monodon*) under different growth stages[J]. PLoS One, 2013, 8(4): e60802.
- [12] Zeng SZ, Huang ZJ, Hou DW, Liu J, Weng SP, He JG. Composition, diversity and function of intestinal microbiota in Pacific white shrimp (*Litopenaeus vannamei*) at different culture stages[J]. PeerJ, 2017, 5: e3986.
- [13] Callac N, Giraud C, Boulo V, Wabete N, Pham D. Microbial biomarker detection in shrimp larvae rearing water as putative bio-surveillance proxies in shrimp aquaculture[J]. PeerJ, 2023, 11: e15201.
- [14] Dong PS, Guo HP, Wang YT, Wang RY, Chen HP, Zhao YJ, Wang K, Zhang DM. Gastrointestinal microbiota imbalance is triggered by the enrichment of *Vibrio* in subadult *Litopenaeus vannamei* with acute hepatopancreatic necrosis disease[J]. Aquaculture, 2021, 533: 736199.
- [15] 张家松, 段亚飞, 张真真, 董宏标, 李卓佳. 对虾肠道微生物菌群的研究进展[J]. 南方水产科学, 2015, 11(6): 114-119.
Zhang JS, Duan YF, Zhang ZZ, Dong HB, Li ZJ. Research progress of intestinal microbial flora in shrimp[J]. South China Fisheries Science, 2015, 11(6): 114-119 (in Chinese).
- [16] Zhang FD, Chen H, Yu JQ, Chen RD, Zhang DM, Chen C, Wang K. Source tracking of larval bacterial community of Pacific white shrimp across the developmental cycle: Ecological insights for microbial management and pathogen prevention in larviculture[J]. Aquaculture, 2025, 609: 742800.
- [17] Callac N, Boulo V, Giraud C, Beauvais M, Ansquer D, Ballan V, Maillez JR, Wabete N, Pham D. Microbiota of the rearing water of *Penaeus stylirostris* larvae influenced by lagoon seawater and specific key microbial lineages of larval stage and survival[J]. Microbiology Spectrum, 2022, 10(6): e04241-22.
- [18] Dong PS, Chen YB, Wei YJ, Zhao XY, Wang T, Jiang S, Xu JJ, Ren TQ, Li M, Zhang L. Dix-seq: an integrated pipeline for fast amplicon data analysis[J]. The Innovation Life, 2025, 3(1): 100120.
- [19] Xue M, Wu LY, He YY, Liang HF, Wen CQ. Biases during DNA extraction affect characterization of the microbiota associated with larvae of the Pacific white shrimp, *Litopenaeus vannamei*[J]. PeerJ, 2018, 6: e5257.
- [20] Zheng YF, Yu M, Liu JW, Qiao YL, Wang L, Li ZT, Zhang XH, Yu MC. Bacterial community associated with healthy and diseased Pacific white shrimp (*Litopenaeus vannamei*) larvae and rearing water across different growth stages[J]. Frontiers in Microbiology, 2017, 8: 1362.

- [21] Zheng YF, Yu M, Liu Y, Su Y, Xu T, Yu MC, Zhang XH. Comparison of cultivable bacterial communities associated with Pacific white shrimp (*Litopenaeus vannamei*) larvae at different health statuses and growth stages[J]. *Aquaculture*, 2016, 451: 163-169.
- [22] Wang YT, Wang K, Huang L, Dong PS, Wang SP, Chen HP, Lu Z, Hou DD, Zhang DM. Fine-scale succession patterns and assembly mechanisms of bacterial community of *Litopenaeus vannamei* larvae across the developmental cycle[J]. *Microbiome*, 2020, 8: 106.
- [23] Chen H, Zhang FD, Yu JQ, Chen RD, Zhang DM, Chen C, Wang K. Divergence patterns of bacterial communities between larviculture systems of two *Penaeus vannamei* strains with distinct culture traits[J]. *Aquaculture*, 2025, 606: 742572.
- [24] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [25] Edgar RC. Search and clustering orders of magnitude faster than BLAST[J]. *Bioinformatics*, 2010, 26(19): 2460-2461.
- [26] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools[J]. *Nucleic Acids Research*, 2012, 41(D1): D590-D596.
- [27] Edgar RC. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequence[J/OL]. *bioRxiv*, 2016. DOI: <https://doi.org/10.1101/074161>.
- [28] 董鹏生, 黄雷, 王艳婷, 郭海朋, 张德民. 对虾养殖系统微生物组样品的采集与制备[J/OL]. *Bio-Protocol*, 2021. DOI: 10.21769/BioProtoc.2003389.
- Dong PS, Huang L, Wang YT, Guo HP, Zhang DM. Collection and preparation of microbiota samples in shrimp culture system[J/OL]. *Bio-Protocol*, 2021. DOI: 10.21769/BioProtoc.2003389 (in Chinese).