

DOI: 10.5846/stxb201505130981

李杰, 杨婧, 黄原. 应用 DNA 复合条形码技术研究秦岭水生动物多样性. 生态学报, 2016, 36(19): 6103-6112.

Li J, Yang J, Huang Y. Characterizing the aquatic biodiversity of the Qinling Mountains using DNA metabarcoding approach. Acta Ecologica Sinica, 2016, 36(19): 6103-6112.

应用 DNA 复合条形码技术研究秦岭水生动物多样性

李 杰, 杨 婧, 黄 原*

陕西师范大学生命科学学院, 西安 710062

摘要: 基于新一代测序技术的 DNA 复合条形码技术, 以其简便快捷和省时省力的特点, 已经成为监测生物多样性的主要方法。采用这一技术, 选取 CO I 和 18S rRNA 两个条形码标记, 对秦岭 5 个淡水水域内, 不同生境下 10 个样品的水生动物多样性进行了初步调查。区系组成结果显示: 18S rRNA 基因鉴定了 9 门 42 纲 52 目, CO I 基因鉴定了 5 门 11 纲 36 目, 而两个条形码标记共鉴定了 10 门 48 纲 89 目。群落组成分析结果显示: CO I 分析得到样本中含量相对较高的类群是双翅目、毛翅目、基眼目、鞘翅目、蜉蝣目等; 而 18S rRNA 分析得到样本中主要有节肢动物门、软体动物门和扁形动物门三个大门。此外, 两者的分析结果也表明所选样地下游的类群数要高于上游。 α 多样性分析结果显示: 金龙峡、泮峪口和石砭峪 3 个受人类影响较大样地的群落丰富度和群落多样性相对较高, 而五台山和子午峪两个自然样地的物种丰富度和群落多样性相对较低, 表明一定程度的外来干扰会对一个地方的水生生物多样性有明显的提高。 β 多样性分析结果显示, 不同环境因素下样品的群落结构差异性较大, 而相似环境因素下样品的群落结构差异性则相对较小。此外, 在聚类分析时, 环境相似性较高的样品首先聚集在一起, 而且群落相似性也相对较高。

关键词: DNA 复合条形码; 秦岭; 水生动物; 生物多样性

Characterizing the aquatic biodiversity of the Qinling Mountains using DNA metabarcoding approach

LI Jie, YANG Jing, HUANG Yuan*

College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China

Abstract: DNA metabarcoding, a technique based on next-generation genetic sequencing, enables the rapid characterization of species composition in bulk biodiversity samples or when analyzing environmental DNA, and thus facilitates comprehensive, large-scale biodiversity assessments and monitoring. The Qinling Mountain range, extending in an east-west direction across central China, includes a series of valleys traversed by mountain streams of various sizes. Although these streams potentially contain an array of aquatic organisms, the complexity of the environment and the presence of small, cryptic, rare, or poorly characterized species makes studying the aquatic biodiversity of these streams a challenge. The goal of this study was to use DNA metabarcoding to examine the composition of both the aquatic fauna and of the aquatic communities as a whole in the Qinling Mountain streams, employing alpha diversity, beta diversity and cluster analyses to evaluate differences in biodiversity from samples collected from different locations. For this purpose, 10 samples containing both zooplankton and zoobenthos were collected from downstream and upstream locations of five streams (Jin Longxia, Shi Bianyu, Feng Yukou, Wutai Mountain and Meridian Valley) in the Qinling Mountains. The cytochrome c oxidase subunit I (CO I) and 18S ribosomal RNA (18S rRNA) genes were selected as barcoding sequences. Following DNA extraction

基金项目: 国家自然科学基金 (31372192)

收稿日期: 2015-05-13; 修订日期: 2016-03-21

* 通讯作者 Corresponding author. E-mail: yuanh@snnu.edu.cn

and PCR amplification using degenerate primers, the amplicons were sequenced on an Illumina MiSeq platform, and Qiime and Mothur software were used to analyze the raw data and to obtain an operational taxonomic unit (OTU) list. Ecological analysis was subsequently performed using Excel, R, and Qiime software. Analysis of the fauna composition revealed that a total of 89 orders, from 48 classes, and 10 phyla were identifiable in the total group of samples using the two gene markers. Individually, 52 orders, from 42 classes, and 9 phyla were identified using the 18S rRNA sequences, and 36 orders, from 11 classes, and 5 phyla were identified using the CO I sequences, demonstrating that the two gene markers together resulted in a higher rate of identification than either marker alone. With regard to community composition, analysis of CO I gene sequences revealed that the Arthropod orders Diptera, Trichoptera, Ephemeroptera and Coleoptera were the most common taxa, with lower occurrences of Protozoa and Rotifera. Analysis of community composition using the 18S rRNA gene sequences, on the other hand, indicated three main groups in the samples, namely the Arthropoda, Mollusca and Platyhelminthes. Both the fauna composition and community composition analyses showed that the number of groups in the downstream samples was higher than that in the upstream samples. Moreover, alpha diversity analysis revealed that the three sampling plots (Jin Longxia, Shi Bianyu and Feng Yukou) most intensively affected by human activities had relatively high values of community richness and diversity compared with the two more natural sampling plots (Wutai Mountain and Meridian Valley), suggesting that aquatic biodiversity may be improved if a location has a certain degree of external interference. Finally, beta diversity analysis demonstrated that, although community variations may be very obvious when samples collected from different environments are compared, such variations may not be so obvious in samples collected from similar environments, with cluster analysis showing that community similarity values in such samples are relatively high.

Key Words: DNA metabarcoding; Qinling Mountains; aquatic animal; biodiversity

DNA 复合条形码技术(DNA metabarcoding)是在 DNA 条形码基础上结合高通量测序技术和分子分类学技术兴起的一种用于生物多样性检测的方法。该技术既可以在短时间内对大量样本的物种组成进行识别,也可以对大批量的多物种混合环境样本进行快速的 DNA 条形码鉴定,避免了单物种 DNA 条形码鉴定的低效率^[1-3]。随着高通量测序技术的进步、条形码参考数据库的完善和分析软件的开发,DNA 复合条形码技术已经成为生物多样性检测最常用的方法。

复合条形码技术从诞生以来广泛应用于多方面的研究,如 Yang 等采用 18S rDNA 作为标记基因分析了土壤小型动物的多样性^[4];Hajibabaei 等对格兰德河水生生物多样性进行了研究^[5];Nolte 等比较了奥地利一个湖中不同季节原生动物多样性^[6];Porazinska 等比较了哥斯达黎加不同热带雨林微生物多样性^[7]。此外,DNA 复合条形码在微生物和真菌多样性^[8-9]、动物排泄物以及食性结构^[10]和古细菌^[11]的研究上都有很广泛的应用。

自然界内江河、湖泊、水库众多,水生生物种类繁多,其中水生动物是最主要的类群,包括浮游动物和底栖动物。浮游动物包括原生动物、轮虫、枝角类和桡足类等,底栖动物包括扁形动物门、环节动物门、节肢动物门等^[12]。由于广泛的水域面积和复杂的水生环境,关于水生动物多样性的研究,虽然有很多发现,但仍然有一些科学问题亟待解决。秦岭是横亘于我国中部东西走向的巨大山脉,是我国的南北分界线^[13]。其内部山谷众多,大小溪流贯穿其中,水生生态环境复杂,但目前还缺乏对水生动物区系和群落的研究。因此,本文选取 CO I 和 18S rRNA 基因作为条形码标记,用 DNA 复合条形码技术对秦岭水生动物区系组成和群落结构组成进行了初步研究,同时用 α 和 β 多样性以及聚类分析来比较不同生境和取样环境下,水生动物多样性组成的差别,为秦岭生物多样性的保护和研究提供数据支持。

1 材料与方法

1.1 样品采集与准备

从秦岭中选择金龙峡、石砭峪、五台山、子午峪和沔峪口 5 个淡水水域,分上下游取样,在上下游分别设置

3 个采样点,每个采样点用 50mL 冻存管取 30mL 水样,同时用毛笔和镊子在采样点的石头上或底层采集肉眼可见底栖动物,放入 50mL 冻存管中。采样地点和环境如表 1 所示。

将采得的水样带回实验室用真空抽滤机浓缩,一个样地 3 个取样点的水样浓缩为一个浮游动物水样。然后将浓缩水样和底栖动物用 100%乙醇保存,放入-20℃的冰箱,用于提取总 DNA。

1.2 总 DNA 的提取

将浓缩水样和底栖动物,混合放入一个经高压灭菌的 1.5mL 离心管中,用 DNeasy Tissue and blood kit (Qiagen Inc)试剂盒提取总 DNA,得到 10 个 DNA 样品,编号如表 1 所示。

表 1 样品采集的地点及环境
Table 1 The location and environment of sampling

样本 Sample	取样地 Location	环境 Environment	样本 Sample	取样地 Location	环境 Environment
J1	金龙峡下游	静水,阴暗,深	W2	五台山上游	静水,阴暗,浅
J2	金龙峡上游	流水,阴暗,浅	Z1	子午峪下游	静水,阳光,深
D1	石砭峪下游	静水,阴暗,浅	Z2	子午峪上游	流水,阴暗,浅
D2	石砭峪上游	流水,阴暗,深	F1	沔峪口下游	流水,阳光,深
W1	五台山下	静水,阳光,浅	F2	沔峪口上游	流水,阳光,浅

1.3 PCR 扩增以及测序

根据 Hajibabaei 等^[5]和 Nolte 等^[6]类似研究采用的引物,实验扩增 CO I 序列的正向引物为 LepF1:5'-ATTCACCAATCATAAAGATATTGG-3',反向引物为 EPT-long-univR: 5'-AARAAAATYATAAYAAANGCGTGN ANNGT-3',扩增长度约为 130bp;扩增 18S rRNA 序列的正向引物为 fw:5'-ATTAGGGTTCGATTCCGGAGAGG-3',反向引物为 rv:5'-CTGGAATTACCGCGGCTGCTG-3',扩增长度约为 180-200bp。为了能够同时对多个样品进行测序,以及在后续分析中区分每个样品序列,在每个样品的 5'端都加了一个 7bp 的标签序列。

实验中 PCR 体系为 25μL,包括 12.5μL 的 2×Taq PCR StarMix with Loading Dye,1.0μL 的 DNA 模版,上游引物(10μmol/L)和下游引物(10μmol/L)各 1μL,9.5μL 的双蒸水。CO I 基因的扩增循环程序是:95℃ 5min 预热,35 个循环:95℃ 40s,43.5℃ 1min,72℃ 30s,72℃ 5min 延伸,4℃ 保温;18S rRNA 基因的扩增循环程序是:98℃ 2min 预热,35 个循环:98℃ 30s,65℃ 45s,72℃ 90s,72℃ 7min 延伸,4℃ 保温。

PCR 产物送往上海派森诺生物科技有限公司进行高通量测序,采用 Illumina MiSeq 平台,进行双末端测序。最后将得到的测序数据返回,进行数据处理和生物多样性分析。

2 数据处理和分析

2.1 原始序列处理:质量过滤和嵌合体的去除

首先,将返回的双末端 Fataq 序列用 fastqc 软件对序列文件进行质量检测;然后用 trimmomatic 0.33 软件进行质量过滤,要求步长为 1 的 5bp 窗口从第 1 个碱基开始移动,窗口中碱基的平均质量≥Q20(即碱基准确率为 99%),从第 1 个低于 Q20 处截断序列,最终要求序列长度≥150bp,且不容许有模糊碱基;然后利用 Flash 软件对质量过滤的序列进行拼接,要求两个序列数据的最大重叠片段≥10bp,且不容许有碱基错配。

由于 PCR 扩增可能会产生嵌合体,测序过程中会产生点突变等测序错误,因此还要对拼接后的序列进一步的过滤和去除嵌合体。实验用 Qiime^[14]软件进行序列过滤,用 mothur^[15]软件去除嵌合体序列。

2.2 OTU 分类和注释

实验用 Qiime 软件中的 uclust^[16]对所得到的优质序列按照 0.97 的相似度进行 OTU 聚类,选取每个 OTU 中最长序列为代表序列;用 Qiime 软件中的 blast,将得到的代表序列与数据库比对,获得 OTU 分类学信息。其中 18S rRNA 基因与 Silva^[17]数据库比对,CO I 基因与 BOLD 数据库比对。

然后,根据注释得到的精简 OTU 列表,用 excel 和 mothur 软件做目水平上的群落组成图和样品间 venn 图。

2.3 群落多样性分析

根据物种丰度,用 mothur 软件中的 summary.single 命令,求出每个样品的 Chao,ACE,Shannon 和 Simpson 指数。其中 Chao 和 ACE 指数是群落丰富度(Community richness)指数,二者的值越大,说明群落丰富度越高;Shannon 和 Simpson 指数是群落多样性(Community diversity)指数,Simpson 指数越大,说明群落多样性越低,Shannon 指数越大,说明群落多样性越高。

用 Qiime 软件,根据各样品的物种进化和丰度信息,进行 Unifrac 分析^[18-19],得到样品间差异距离矩阵,然后进行 PCoA 分析。

2.4 聚类分析

根据精简后的 OTU 列表,用 R 软件中的 pheatmap 程序包进行属水平上的聚类分析,并绘制出 heatmap^[20-21]图。

3 结果

3.1 测序结果

原始序列经过滤和嵌合体的去除,两个标记基因共得到 1120240 条有效序列,971319 条优质序列,其中 18S rRNA 的有效序列有 767441 条,优质序列有 640700 条,而 CO I 的有效序列有 352799 条,优质序列有 330619 条,它们的优质序列所占比例分别为:83.49%和 93.71%。

3.2 区系组成和群落结构

将优质序列与数据库比对,共得到水生动物 10 门 48 纲 89 目,其中 18S rRNA 鉴定了 9 门 42 纲 52 目,CO I 鉴定了 5 门 11 纲 36 目。其中 18S rRNA 鉴定的各样品组成情况是:J1 包含 31 目 55 科,J2 包含 23 目 41 科,D1 包含 28 目 49 科,D2 包含 24 目 44 科,W1 包含 27 目 45 科,W2 包含 25 目 41 科,Z1 包含 23 目 47 科,Z2 包含 20 目 46 科,F1 包含 20 目 35 科,F2 包含 19 目 30 科。

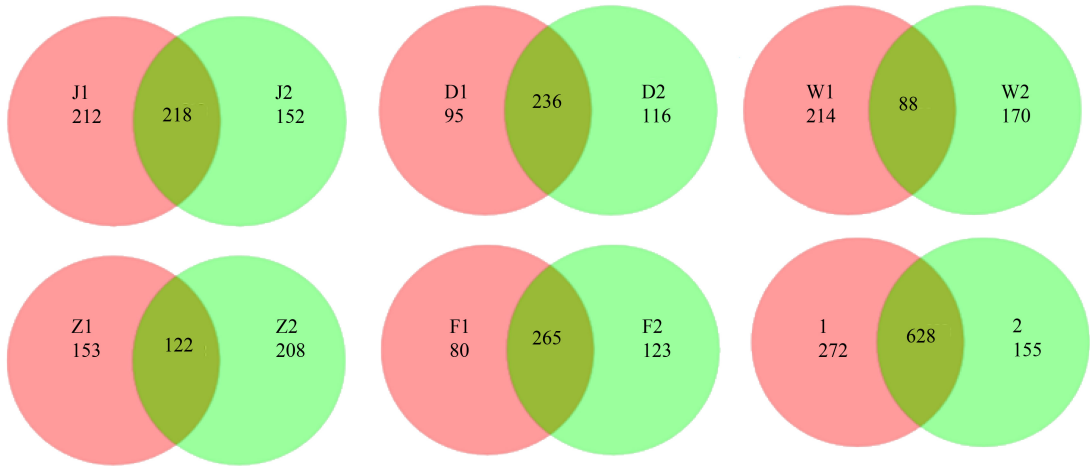


图 1 CO I 基因在种水平上的 venn 图(图中 1 和 2 分别代表所有取样点的下游和上游)

Fig.1 The venn diagram for CO I gene in the species level (1 and 2 represents the downstream and upstream of all the sample plots, respectively)

根据 OTU 列表做 venn 图,如图 1 和图 2 所示,CO I 鉴定得到:J1 与 J2 的共有物种数是 218,特有物种数分别为 212 和 152;D1 与 D2 的共有物种数是 236,特有物种数分别为 95 和 116;W1 与 W2 的共有物种数是 88,特有物种数分别为 214 和 170;Z1 与 Z2 的共有物种数是 122,特有物种数分别为 153 和 208;F1 与 F2 的共有物种数是 265,特有物种数分别为 80 和 123;所有取样点下游与上游的共有物种数是 628,特有物种数分别

是 272 和 155;D、F、J、W 和 Z 的总物种数分别是:683、733、800、560 和 605;18S rRNA 鉴定得到:J1 与 J2 的共有物种数是 154,特有物种数分别为 65 和 47;D1 与 D2 的共有物种数是 158,特有物种数分别为 74 和 28;W1 与 W2 的共有物种数是 131,特有物种数分别为 56 和 67;Z1 与 Z2 的共有物种数是 160,特有物种数分别为 49 和 68;F1 与 F2 的共有物种数是 139,特有物种数分别为 43 和 53;所有取样点下游与上游的共有物种数是 305,特有物种数分别是 11 和 10;J、D、W、Z 和 F 的共有物种数是 140,特有物种数分别为 3、2、9、2 和 0,D、F、J、W 和 Z 的总物种数分别是:418、374、420、375 和 437。

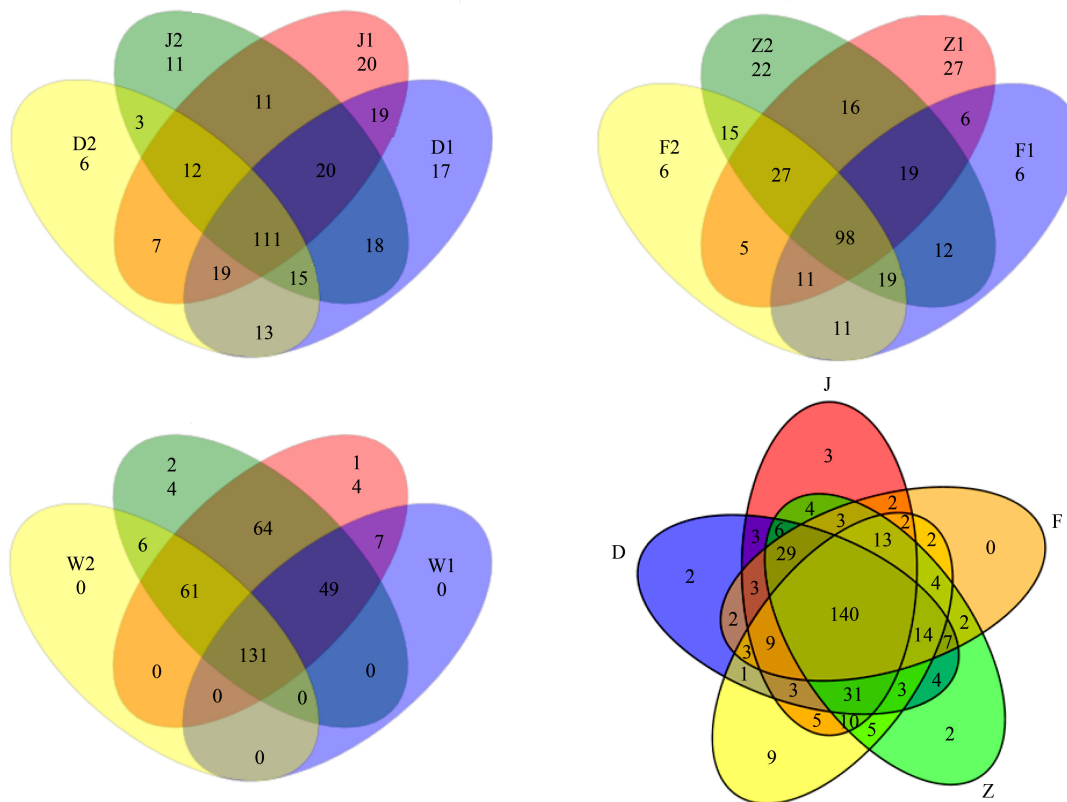


图 2 18S rRNA 基因在种水平上的 venn 图

Fig.2 The venn diagram for 18S rRNA gene in the species level

图中 1 和 2 分别代表所有取样点的下游和上游,J、D、W、Z 和 F 分别代表每个取样地上游和下游的总和

根据 OTU 列表,可以用样品在各分类水平(门、纲、目、科、属)上的物种组成比例情况,来反应样品在不同分类水平上的群落结构。CO I 从目水平上分析结果如图 3 所示:D1、D2、F1、F2、J1、J2 和 W1 中双翅目的含量相对较高,所占比例分别为 45.2%、46.3%、46.9%、27.1%、83.1%、53.9%和 48.9%;W2 中毛翅目的含量较高,所占比例为 56.6%;Z1 中基眼目的含量较高,所占比例为 29.3%;Z2 中鞘翅目的含量相对较高,所占比例为 45.3%。18S rRNA 从门水平上分析结果如图 4 所示:D1、D2、J1、J2、W1、Z1 和 Z2 中节肢动物门的含量相对较高,所占比例分别为 59.2%、84.5%、92.7%、65.8%、91.6%、53.5%和 66.9%;F1 和 F2 中软体动物门的含量相对较高,所占比例分别为 55.2 和 58.6%;W2 中扁形动物门的含量相对较高,所占比例为 54.5%。可见,CO I 调查表明:样本中含量相对较高的类群是双翅目、毛翅目、基眼目、鞘翅目、蜉蝣目等;而 18S rRNA 调查表明:样本中主要有节肢动物门、软体动物门和扁形动物门三个大门。

3.3 α 多样性分析

CO I 基因的 Chao 和 ACE 指数表明 J1、J2 和 F2 的群落丰富度相对较高,Z1、W2 和 Z2 的则相对较低;Simpson 和 Shannon 指数表明 J1、Z2 和 F2 的群落多样性相对较高,而 J2、D1 和 W2 的则相对较低。18S rRNA 基因的 Chao 和 ACE 指数表明 J1、D1 和 J2 的群落丰富度相对较高,而 F1、F2 和 W1 的则相对较低,Simpson 和

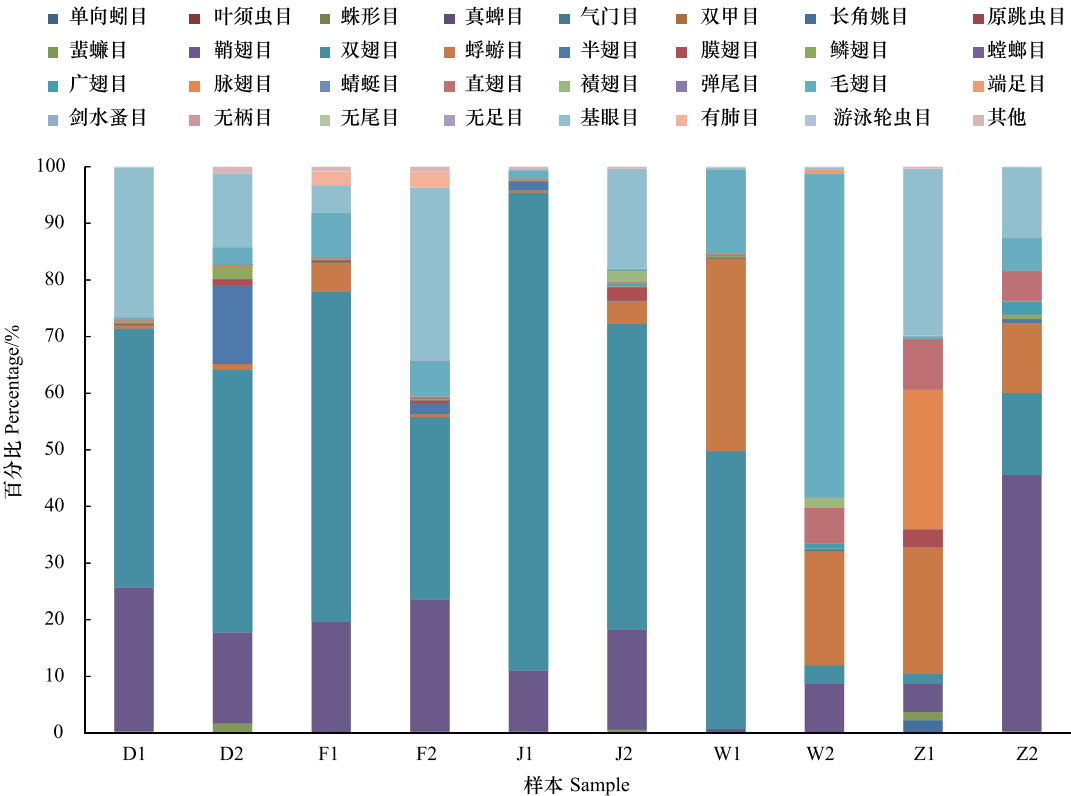


图3 CO I 基因目水平上群落组成图

Fig.3 The community composition diagram for CO I gene on the order level

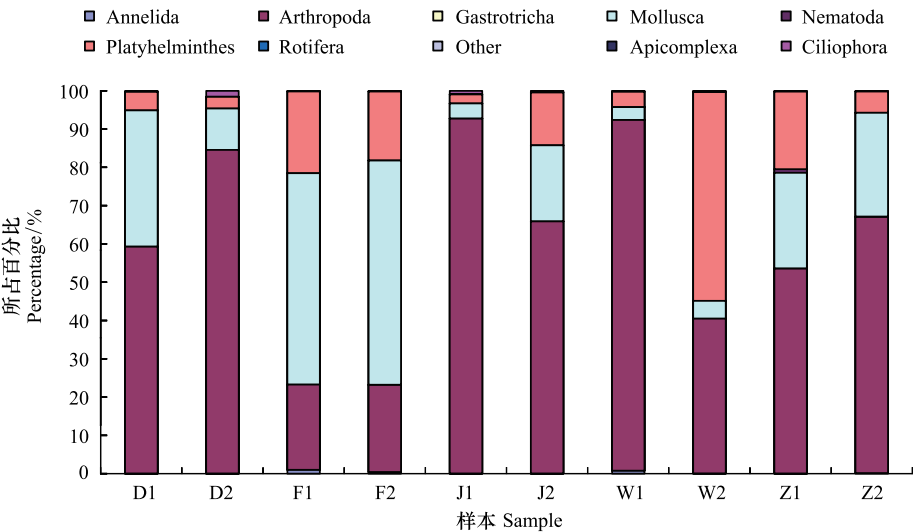


图4 18S rRNA 基因门水平上群落组成图

Fig.4 The community composition diagram for 18S rRNA gene on the phyla level

Shannon 指数表明 J1、Z1 和 W2 的群落多样性相对较高,而 F2、F1、和 D2 的则相对较低。整体来看,J、F、D 的群落丰富度和群落多样性相对较高,Z、W 的则相对较低。

3.4 β 多样性分析

用样品间物种进化和丰度信息,进行 Unifrac 分析,得到样品间差异距离矩阵,然后用差异距离矩阵进行

PCoA 分析,得到 PCoA 图,如图 5 所示,图中两点之间的距离越近说明两者之间的生物群落差异性越小。18S rRNA 基因的 PCoA 图显示;Z1、D2 和 J1、J2 和 D1 之间的生物群落结构差异性较小;CO I 基因的 PCoA 图显示;D1、Z1、D2 和 J2、Z2 和 W2 之间的生物群落差异性较小。

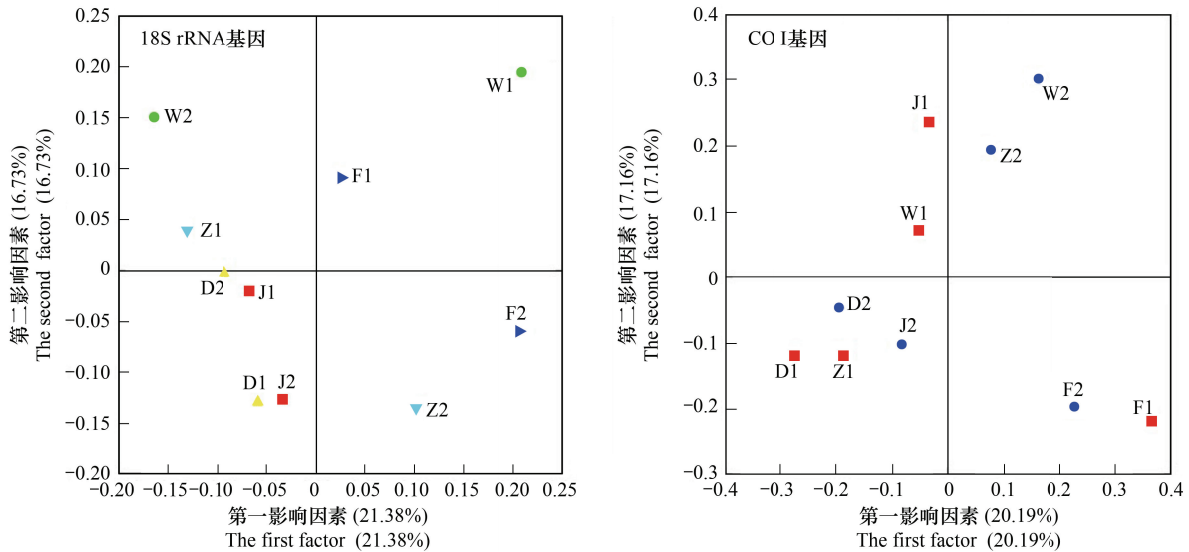


图 5 各样品在对群落结构影响最大两因素下的 PCoA 图

Fig.5 The PCoA diagram for each sample in the two most important factors that affect the community structure

3.5 聚类分析

18S rRNA 基因的聚类情况显示 Z2、J2 和 D2、F2 和 F1、W1 和 W2、Z1 和 D1 在属水平上群落相似性相对较高(图 6);CO I 基因的聚类情况显示 J2、D2、D1 和 Z2、J1 和 W1、F1 和 F2、Z1 和 W2 在属水平上的群落相似性相对较高(图 7)。

4 分析与讨论

4.1 环境因素与结果讨论

由于水环境的复杂性和多变性,使得水生生物多样性受多方面影响,其可以分为自然因素和人为因素两个方面^[12]。自然因素主要包含上下游、水体深浅以及流动性等,而人为因素主要是指人类活动对生物多样性的影响。本文主要从以上这些方面对秦岭水域的水生生物多样性做了初步调查。

对于水流来说,水的流动性会让上游的营养物质向下游流动,使得下游的营养高于上游,从而导致下游的物种丰富度高于上游^[22]。从 venn 图可以看到,各样地下游的物种丰度普遍高于上游;群落结构图显示,各样地下游的群落结构比上游的复杂,并且下游群落结构中优势物种的含量也高于上游;从 α 多样性指数可以看到,虽然个别样品上游的群落丰富度和群落多样性较下游的高,但绝大多数样品下游的群落丰富度和群落多样性还是较上游的高。可见,本次调查也表明下游的物种丰富度要高于上游。

实验选择了秦岭中 5 个淡水水域:沔峪口、五台山、子午峪、金龙峡和石砭峪,其中沔峪口,金龙峡和石砭峪在景区内,人为干扰较大,而子午峪和五台山相对来说人为干扰较小,基本上处于自然状态。从实验结果可以看出各样地之间的不同:如各样地物种丰度的大小顺序是:W<Z<D<F<J,表明受人为影响较大的 J、F 和 D 3 个取样地的物种丰度大于受人为影响较小的 W、Z 两个取样地;群落结构图显示,D1、D2、F1、F2、J1 和 J2 的主要类群是双翅目,W1 和 W2 的主要类群是双翅目和毛翅目,Z1 和 Z2 的主要类群是基眼目和鞘翅目; α 多样性指数表明,J、F 和 D 的群落丰富度和群落多样性都相对较高,而 Z 和 W 的则相对较低。这些都表明一个样地的生物多样性会受到人为活动的影响。“放牧效应”^[23-24]在一般生态系统中都是适合的,一定程度的外来干扰会对一个地方的生物多样性有明显的提高,如本实验中有人为干扰的 3 个取样地的物种丰度、群落丰富

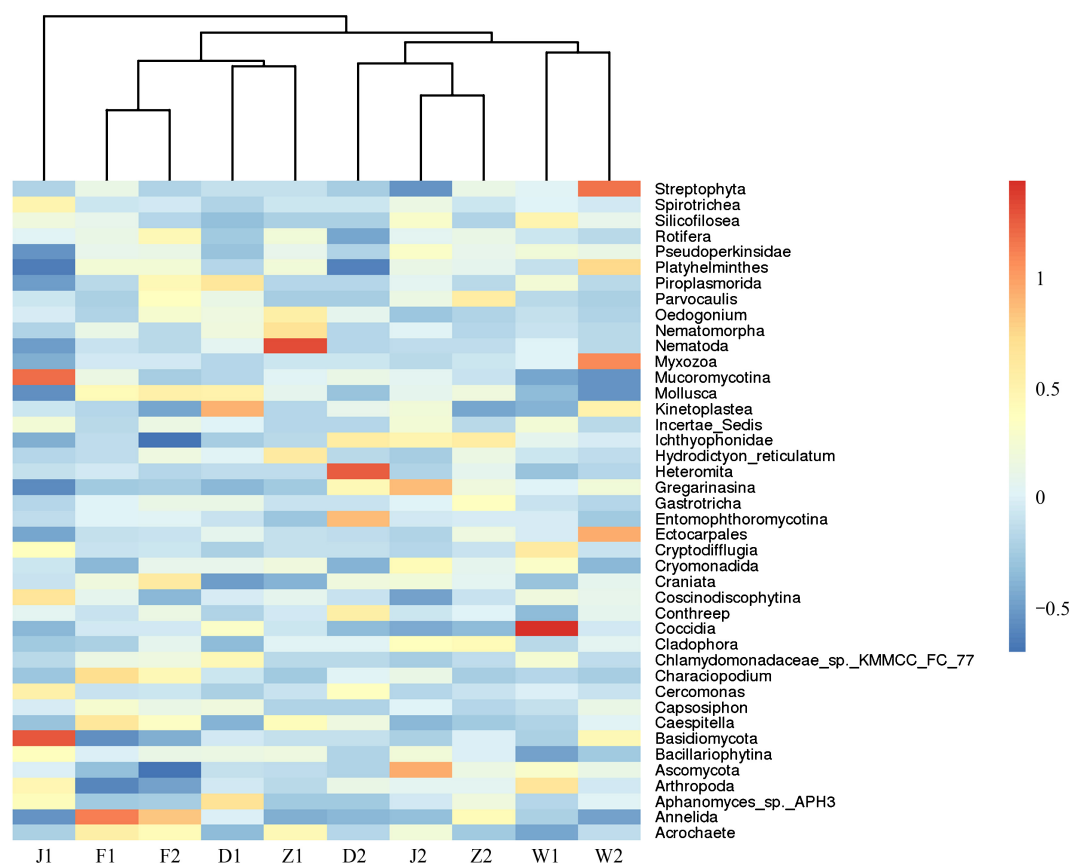


图6 18S rRNA 基因在属水平上的 Heatmap 图

Fig.6 The heatmap diagram for 18S rRNA gene on the genus level

度和群落多样性都比自然状态下的两个取样地要高。但是当外来干扰超过一定程度之后,就会使该地的生物多样性急剧下降,对原来的生态系统造成严重的破坏。

对一个生态系统来说,影响生物多样性的因素是多方面的,除了前面所述的上下游和人为活动的影响外,本实验还关注了水的深浅、流动性以及阳光的有无对生物多样性的影响。如取自深水的 D2 和 F1 样品与取自浅水的 D1 和 F2 样品,跟其他都取自浅水的样品相比,有相对较多的共有物种;PCoA 图分析表明不同环境因素下样品的群落结构差异性较大,而相似环境因素下样品的群落结构差异性则相对较小,如 J2 和 D2 都是取自较阴暗的上游流动水,取样环境相似,因此它们之间的群落结构差异性相对较小。此外,Heatmap 图也显示,环境相似性较高的样品首先聚集在一起,并且群落相似性相对较高,如 Z1 和 D1 都是取自较浅的下游静水,取样环境相似,因此它们首先聚类在一起,表明它们之间的群落相似性相对较高。

4.2 实验方法与结果讨论

实验结果显示,两个分子标记所得的实验结果并不完全一致。这主要是由于每个基因的通用引物在不同类群中的扩增能力和在数据库中的覆盖范围不同导致的。对基于高通量测序的复合条形码技术来说,由于样品的复杂性和多重性,在研究时如果选择多个分子标记基因可能会得到更全面的实验结果,如本论文中 18S rRNA 基因鉴定了 9 门 42 纲 52 目,CO I 基因鉴定了 5 门 11 纲 36 目,而两个分子标记基因共同鉴定了 10 门 48 纲 89 目。Carew^[25]等人的研究结果也表明,使用两个或两个以上的分子标记可以更全面的鉴定到样本中所包含的物种。这使得一些依赖于类群数计算的各种生物多样性指数,将变得更加准确。但是在使用多个分子标记进行分析时,由于每个分子标记在进化速率和单独鉴定的类群上不同,可能使得构建的系统发育树之间有区别,从而影响了依赖于系统发育树计算的 unifrac 分析结果。

目前,应用 DNA 复合条形码技术对生物多样性的研究越来越多,但是关于复合条形码技术的一些本质问

题仍然没有得到很好的解决。其中一个最主要的问题就是:复合条形码技术只能做定性分析,而不能很准确的进行定量分析。在研究中,一直用序列数的多少来表示物种个体数量的多少,虽然有些研究认为序列数与物种的个体数之间存在相关性,但也仅仅是它们的变化趋势相对一致,而并不是很完美的契合。因此,在分析的过程中,只能用序列数大致的估计物种的个体数,而不能完全用序列数来代替物种的个体数。Piñol^[26]等人的研究也表明复合条形码技术的定量研究能力有限,只有定性分析的结果才是可靠的。对于本研究来说,有些分析会不可避免的应用到物种个体数的多少,如群落结构图中各样本中类群的含量,以及 α 多样性中的 Shannon 指数等,所以对于这些结果,只能定性来看,如果想要得到更加精确的结果,还须进一步的验证。当然,在本实验中也避免了一些用物种个体数的分析,如没有使用 Bray-Curis 指数进行 β 多样性的分析,而是使用了依赖于物种数和进化信息的 jaccard 和 unweighted_unifrac 指数进行 β 多样性和后续的分析。但是,生态学研究中的群落组成和群落结构以及群落之间的关系,都是依赖于物种的个体数量,所以 DNA 复合条形码技术还不能很准确的应用到生态学研究中的每一个方面。

虽然复合条形码技术还不能更精确的进行生物多样性研究,但相比于传统耗时耗力以及单一的物种鉴定和生物多样性研究,复合条形码技术有明显的优势:它能在短时间对大量复合样本进行物种鉴定和生物多样性研究,既省时又省力。随着测序技术的进步和提高,条形码数据库的全面和完善,生物信息学软件的开发和使用,复合条形码技术将会成为生物多样性研究的主要方法,并且准确性也会大幅度提高,更重要的是它将会能够进行更大尺度和更大规模的生物多样性研究。

参考文献 (References):

- [1] Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 2012, 21(8): 2045-2050.
- [2] Ji Y Q, Ashton L, Pedley S M, Edwards D P, Tang Y, Nakamura A, Kitching R, Dolman P M, Woodcock P, Edwards F A, Larsen T H, Hsu W W, Benedick S, Hamer K C, Wilcove D S, Bruce C, Wang X Y, Levi T, Lott M, Emerson B C, Yu D W. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 2013, 16(10): 1245-1257.
- [3] Coissac E, Riaz T, Puillandre N. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, 2012, 21(8): 1834-1847.
- [4] Yang C X, Ji Y Q, Wang X Y, Yang C Y, Yu D W. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. *Science China Life Sciences*, 2013, 56(1): 73-81.
- [5] Hajibabaei M, Shokralla S, Zhou X, Singer G A C, Baird D J. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*, 2011, 6(4): e17497.

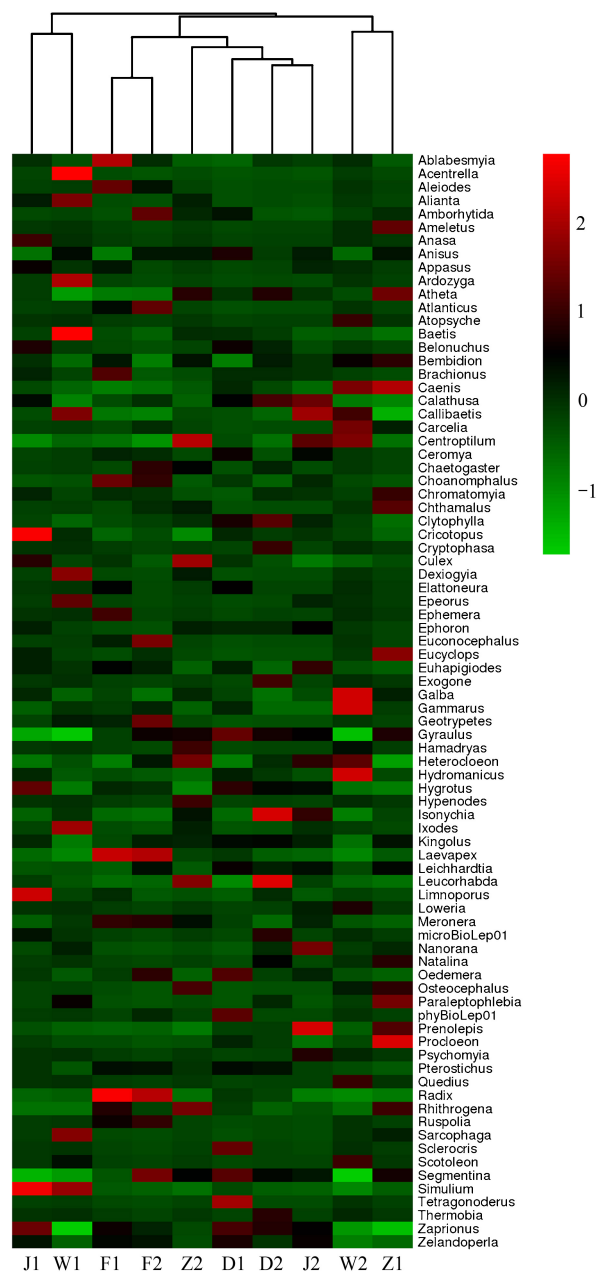


图 7 各样品 CO I 基因在属水平上的 Heatmap 图

Fig.7 The heatmap diagram for CO I gene on the genus level

- [6] Nolte V, Pandey R V, Jost S, Medinger R, Ottenwälder B, Boenigk J, Schlötterer C. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Molecular Ecology*, 2010, 19(14): 2908-2915.
- [7] Poretsky R S, Hewson I, Sun S L, Allen A E, Zehr J P, Moran M A. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology*, 2009, 11(6): 1358-1375.
- [8] Schmidt P A, Bólint M, Greshake B, Bandow C, Römbke J, Schmitt I. Illumina metabarcoding of a soil fungal community. *Soil Biology & Biochemistry*, 2013, 65: 128-132.
- [9] Bólint M, Schmidt P A, Sharma R, Thines M, Schmitt I. An Illumina metabarcoding pipeline for fungi. *Ecology and Evolution*, 2014, 4(13): 2642-2653.
- [10] Soininen E M, Zinger L, Gielly L, Bellemain E, Brathen K A, Brächnmann C, Epp L S, Gussarova G, Hassel K, Henden J A, Killengreen S T, Rämä T, Stenøien H K, Yoccoz N G, Ims R A. Shedding new light on the diet of Norwegian lemmings: DNA metabarcoding of stomach content. *Polar Biology*, 2013, 36(7): 1069-1076.
- [11] Jørgensen T, Kjær K H, Haile J, Rasmussen M, Boessenkool S, Andersen K, Coissac E, Taberlet P, Brochmann C, Orlando L, Gilbert M T P, Willerslev E. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Metabarcoding. *Molecular Ecology*, 2012, 21(8): 1980-1988.
- [12] 章继华, 何永进. 我国水生生物多样性及其研究进展. *南方水产*, 2005, 1(3): 69-72.
- [13] 刘晓清, 张霞, 王亚萍, 成西娟. 秦岭地区生物多样性及其保护对策. *安徽农业科学*, 2012, 40(12): 7365-7367, 7496.
- [14] Caporaso J G, Kuczynski J, Stombaugh J, Bittinger K, Bushman F D, Costello E K, Fierer N, Peña A G, Goodrich J K, Gordon J I, Huttley G A, Kelley S T, Knights D, Koenig J E, Ley R E, Lozupone C A, McDonald D, Muegge B D, Pirrung M, Reeder J, Sevinsky J R, Turnbaugh P J, Walters W A, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 2010, 7(5): 335-336.
- [15] Schloss P D, Westcott S L, Ryabin T, Hall J R, Hartmann M, Hollister E B, Lesniewski R A, Oakley B B, Parks D H, Robinson C J, Robinson C J, Sahl J W, Stres B, Thallinger G G, Van Horn D J, Weber C E. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 2009, 75(23): 7537-7541.
- [16] Edgar R C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 2010, 26(19): 2460-2461.
- [17] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner F O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 2013, 41(D1): D590-596.
- [18] Lozupone C, Knight R. Unifrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 2005, 71(12): 8228-8235.
- [19] Lozupone C, Hamady M, Knight R. Unifrac-An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*, 2006, 7:371.
- [20] Škuta C, Bartůňek P, Svozil D. InChlib interactive cluster heatmap for web applications. *Journal of Cheminformatics*, 2014, 6:44.
- [21] Deng W K, Wang Y B, Liu Z X, Cheng H, Xue Y. HemI: A Toolkit for Illustrating Heatmaps. *PLoS One*, 2014, 9(11): e111988.
- [22] 周桔, 雷霆. 土壤微生物多样性影响因素及研究方法的现状与展望. *生物多样性*, 2007, 15(3): 306-311.
- [23] 罗天相, 刘莎. 中度放牧干扰对草地生物多样性影响的思考. *安徽农业科学*, 2007, 35(21): 6567-6568, 6612.
- [24] 何飞, 何亚平, 黄登才, 刘兴良, 费世民, 慕长龙, 蒋俊明, 陈秀明, 隆廷伦, 张旭东. 放牧的环境效应及其对环境保护的负面影响. *四川林业科技*, 2009, 30(3): 43-54.
- [25] Crarew M E, Pettigrove V J, Metzeling L, Hoffmann A A. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, 2013, 10(1): 45.
- [26] Piñol J, Mir G, Gomez Polo P, Agusti N. 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.