INFLUENCES OF RANDOMNESS, AMPLIFICATION ERROR, AND BARCODE SEQUENCES ON MICROBIOTA STRUCTURE ANALYSIS THROUGH HIGH-THROUGHPUT SEQUENCING OF 16S rDNA AMPLICONS

WU, S.[1] – ZHANG, X.[2] – YANG, Y.[1] – NI, J.[3,4] – DING, W.[1*]

[1]*School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*

[2]*Zhejiang Mariculture Research Institute; Zhejiang Key Laboratory of Exploitation and Preservation of Coastal Bio-Resource; Wenzhou Key Laboratory of Marine Biological Genetics and Breeding, Wenzhou 325005, China*

[3]*Meilikang Research and Development Center, Guangdong Meilikang Bio-Science Ltd., Foshan 528315, China*

[4]*Dongguan Key Laboratory of Medical Bioactive Molecular Developmental and Translational Research, Guangdong Medical University, Dongguan 523808, China*

**Corresponding author*
e-mail: Dingweijun@cdutcm.edu.cn; phone: +86-286-180-0104

(Received 25th Jul 2022; accepted 14th Oct 2022)

Abstract. High-throughput sequencing of 16S rDNA amplicons is widely used to analyse prokaryotic community structure. However, influence of the error introduced by PCR amplification process on the results is not well evaluated. To evaluate the influences of randomness, amplification error, and barcode sequences on microbiota structure analysis through high-throughput sequencing of 16S rDNA amplicons, we used primers with three different barcode sequences to sequence and compare the high-throughput sequencing results of 16S rDNA amplicons of *Tegillarca granosa* gut microbiota. Our results showed that removing chimeric sequences, different barcoding sequences, and different sequencing depths did not fundamentally change the differences in the microbiota structure and microbiota with different treatments were still aggregated according to the samples. Removing chimeric sequences and adopting different barcode sequences did not obviously affect the composition of the *T. granosa* gut microbiota at the phylum and genus levels. Chimeric sequences mainly led to overestimation of the α-diversity of the microbiota and the number of rare OTUs, whereas their impact on the number of rare OTUs was insufficient, and random factors had a higher impact than chimeric sequences. When analysing the β-diversity of the microbiota and dominant OTUs, chimeric sequences had little impact on the results.
Keywords: *ribosomal small subunit DNA, microbial community, chimeric sequences, rare species, operational taxonomic units*

Introduction

Analysis of microbial community structure is the basis of microbial ecology research. High-throughput sequencing of 16S rDNA amplicons is widely used to analyse prokaryotic community structures existing in various habitats, such as soil (Rousk et al., 2010; Chu et al., 2020; Mao et al., 2020, 2022), sediment (Paul et al., 2016; Liu et al., 2022b), ponds (Ni et al., 2018; Gong et al., 2021), rivers (Chen et al., 2021; Liu et al., 2022a), lakes (Paul et al., 2016; Wu et al., 2019; Zhang et al., 2020), and vertebrate guts (Ni et al., 2021; Li et al., 2021, 2022a, b; Liu et al., 2022b; Wu et al., 2022; Zhu et al., 2022). However, due to the PCR amplification process, it is

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

generally believed that the result of high-throughput sequencing of 16S rDNA amplicons is inaccurate due to the different amplification efficiencies of various species, chimeric sequences, random errors, and amplification efficiency differences caused by barcode sequences (Acinas et al., 2005; Takahashi et al., 2014).

Although metagenomic sequencing can overcome the errors introduced by PCR amplification, due to the cost advantage of the high-throughput sequencing of 16S rDNA amplicons in the analysis of microbial community structure, metagenomic sequencing is still difficult to replace the high-throughput sequencing of 16S rDNA amplicons in the analysis of microbial community structure. Therefore, full evaluation of the influences of randomness, amplification error, and barcode sequences on microbiota structure analysis through high-throughput sequencing of 16S rDNA amplicons has important academic value for obtaining accurate and reliable information on microbial community structure (Sipos et al., 2007; Wu et al., 2010; Haas et al., 2011; Sze and Schloss, 2019).

Most large-scale, high-throughput sequencings use DNA barcodes for identification of individuals in pooled biomolecule populations (Hawkins et al., 2018). Considering DNA synthesis and sequencing errors confound the correct interpretation of observed barcodes, error-correcting codes, such as Hamming, Levenshtein codes were used to design the barcode sequences (Hamady et al., 2008; Hawkins et al., 2018). Chen et al. (2020) proposed a barcode construction method, which combines a block error-correction code with a predetermined pseudorandom sequence to generate a base sequence for labeling different samples based on the hidden Markov model and the forward-backword algorithm. However, the error-correcting codes do not properly account for insertions and deletions in DNA barcodes. Therefore, Hawkins et al. (2018) present and experimentally validate filled/truncated right end edit barcodes.

To fully evaluate the influences of randomness, amplification error, and barcode sequences on microbiota structure analysis through high-throughput sequencing of 16S rDNA amplicons, we introduced a novel evaluation method and used primers with three different barcode sequences to sequence and compare the high-throughput sequencing results of 16S rDNA amplicons of three samples of *Tegillarca granosa* gut microbiota. The barcode sequences in the forward primer were used 12-nt barcode sequences that are commonly used in microbiota analysis (Xiang et al., 2018; Ni et al., 2019) for distinguishing samples, and the barcode sequences in the reverse primer were used 8-nt barcode sequences designed by filled/truncated right end edit barcodes (Hawkins et al., 2018) for evaluating the effect of primers on the analysis results.

Materials and methods

Experimental design

To evaluate the impact of chimeric sequences on sequencing results, three barcode sequences were used for polymerase chain reaction amplification and HiSeq sequencing analysis of three samples (*Fig. 1*). We first compared the difference between operational taxonomic unit (OTU) results with chimeric sequences and those obtained by removing chimeric sequences with Uchime algorithm (Edgar et al., 2011). Then, we analysed the OTUs affected by barcodes, the error OTUs introduced by PCR amplification, and the proportion of OTUs affected by random factors.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

We defined the OTUs that were affected by barcodes as follows: in the three samples analysed, the OTU abundances corresponding to the barcode in each sample were lower or higher than 10% of their average abundance.

The error OTUs introduced by PCR amplification were defined as follows: in the single sample, the OTU abundances corresponding to one barcode sequence were obviously higher than those corresponding to the other two barcode sequences, the OTU abundances corresponding to the other two barcode sequences were not significantly different, and these results did not appear in other samples.

OTUs affected by random factors were defined as the abundances of the OTUs with significant differences between three barcoding sequence results from the same sample, or the singleton in the sample.

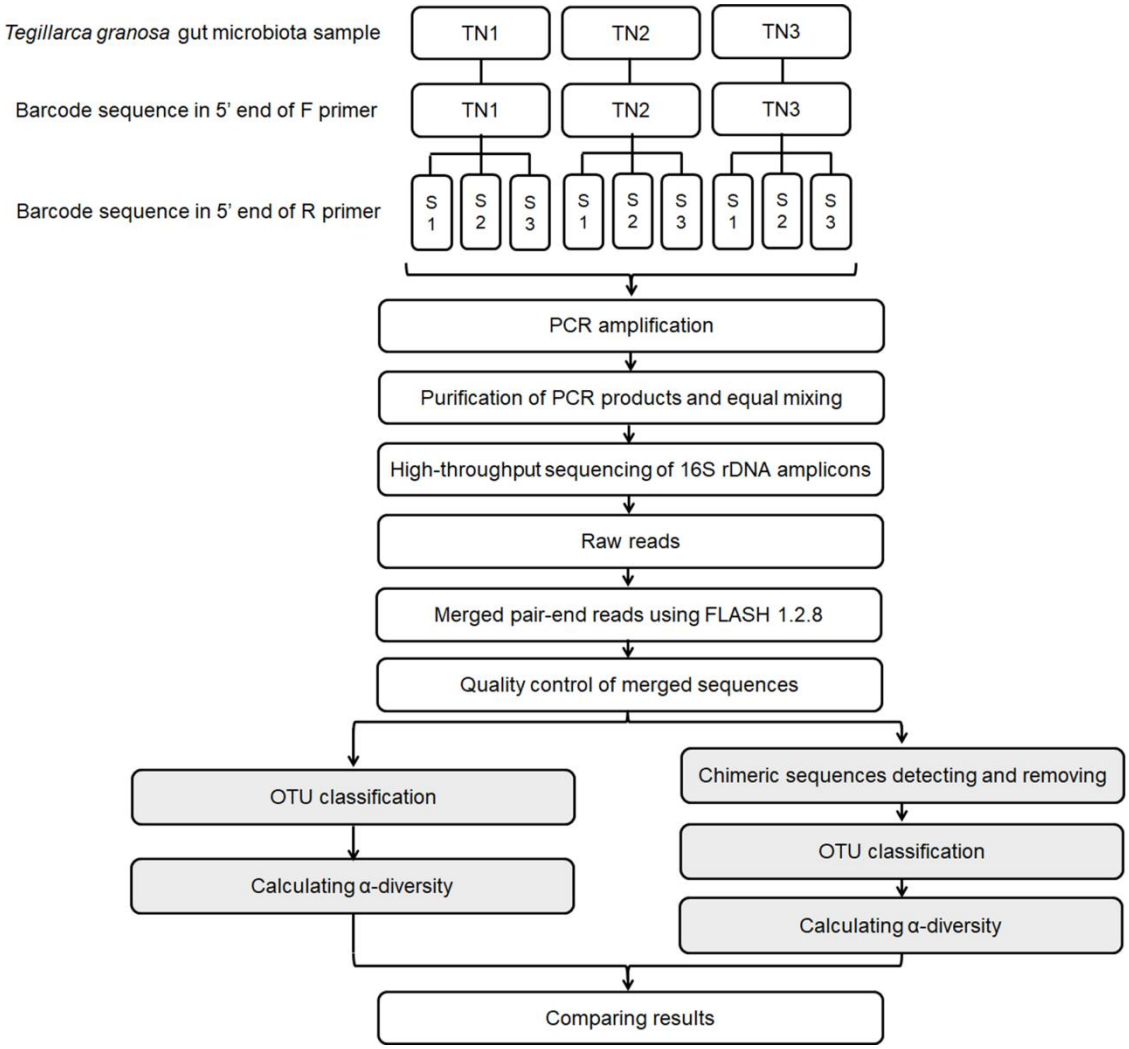


Figure 1. *Experimental frame*

Sample collection and high-throughput sequencing of 16S rDNA amplicons

T. granosa samples were collected from a farm in Wenzhou, China (28.09 N 121.05 E). The total DNA of gut microbiota was extracted using the PowerFecal DNA kit (QIAGEN, Germany) according to the manufacturer's introductions. The V4-V5

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

hypervariable region of prokaryotic 16S rDNA was amplified using the primer pair 515F and 909R as previously described (Xiang et al., 2018; Ni et al., 2019). The amplicons were quantified with a NanoDrop 2000 spectrophotometer (ThermoFisher, USA), and equimolar amounts of each sample were pooled and purified using the AxyPrep DNA gel extraction kit (Axygen, China). The purified DNA was sequenced using an Illumina HiSeq 3000 platform at Guangdong Meilikang Bio-Science Ltd., China.

Raw reads were merged using FLASH 1.2.8 (Magoc and Salzberg, 2011) and processed using QIIME 1.9.0 (Caporaso et al., 2010) as previously described (Ni et al., 2017a, 2021). The sequences were clustered into OTUs at 97% identity using UPARSE software (Edgar, 2013). Each OTU was taxonomically assigned using the RDP classifier (Wang et al., 2007) with the gg_13_8 dataset.

Data analysis

Data were showed as means ± standard errors. The Kruskal-Wallis H-test with Dunn's *post-hoc* test was conducted using R 4.2.0 (R Core Team, 2013). Nonparametric multivariate analysis of variance (PERMANOVA) (Anderson, 2001) was conducted using the vegan package (Dixon, 2003) in R 4.2.0. Principal coordinates analysis (PCoA) was conducted using the QIIME 1.9.0 pipeline (Caporaso et al., 2010). Differences with $p < 0.05$ were considered statistically significant.

Results

PCoA results showed that in general, whether chimeric sequences were removed, different barcoding sequences, and different sequencing depths did not fundamentally change the differences in the microbiota structure, and microbiota with different treatments were still aggregated according to the samples (*Fig. 2*). Moreover, even though only 300 16S rDNA sequences were used, the *T. granosa* gut microbiota structure could be effectively distinguished (*Fig. 2*). Correlation analysis showed that the weighted and unweighted UniFrac distances of the microbiota did not exhibit a consistent change pattern with the sequencing depth (*Fig. 2G, H*).

Removing chimeric sequences and adopting different barcode sequences did not obviously affect the composition of the *T. granosa* gut microbiota at the phylum level (*Fig. 3*). Tenericutes, Spirochaetes, Proteobacteria, Planctomycetes, OP3, Nitrospirae, Firmicutes, Elusimicrobia, Cyanobacteria, Chloroflexi, Chlorobi, Bacteroidetes, Actinobacteria, and Acidobacteria dominated the *T. granosa* gut microbiota (*Fig. 3A*). Heatmap combined with cluster analysis showed that removing the chimera sequences had less impact on the composition of the *T. granosa* gut microbiota at the phylum level than the results that were sequenced with different barcode sequences (*Fig. 3B*). These results indicated that the errors introduced by PCR amplification, the differences caused by barcode sequences, or random errors had more influence on the composition of the *T. granosa* gut microbiota at the phylum level than the chimeric sequences introduced by PCR amplification.

A total of 583 genera were detected in the *T. granosa* gut microbiota and 25 genera dominated the microbiota (*Fig. 4A*). *Mycoplasma* was the most abundant genus in the *T. granosa* gut microbiota. *Synechococcus*, *Vibrio*, *Treponema*, *Pseudoalteromonas*, *Crenothrix*, *Methylotenera*, and many unidentified genera were also detected as dominant genera (*Fig. 4A*).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

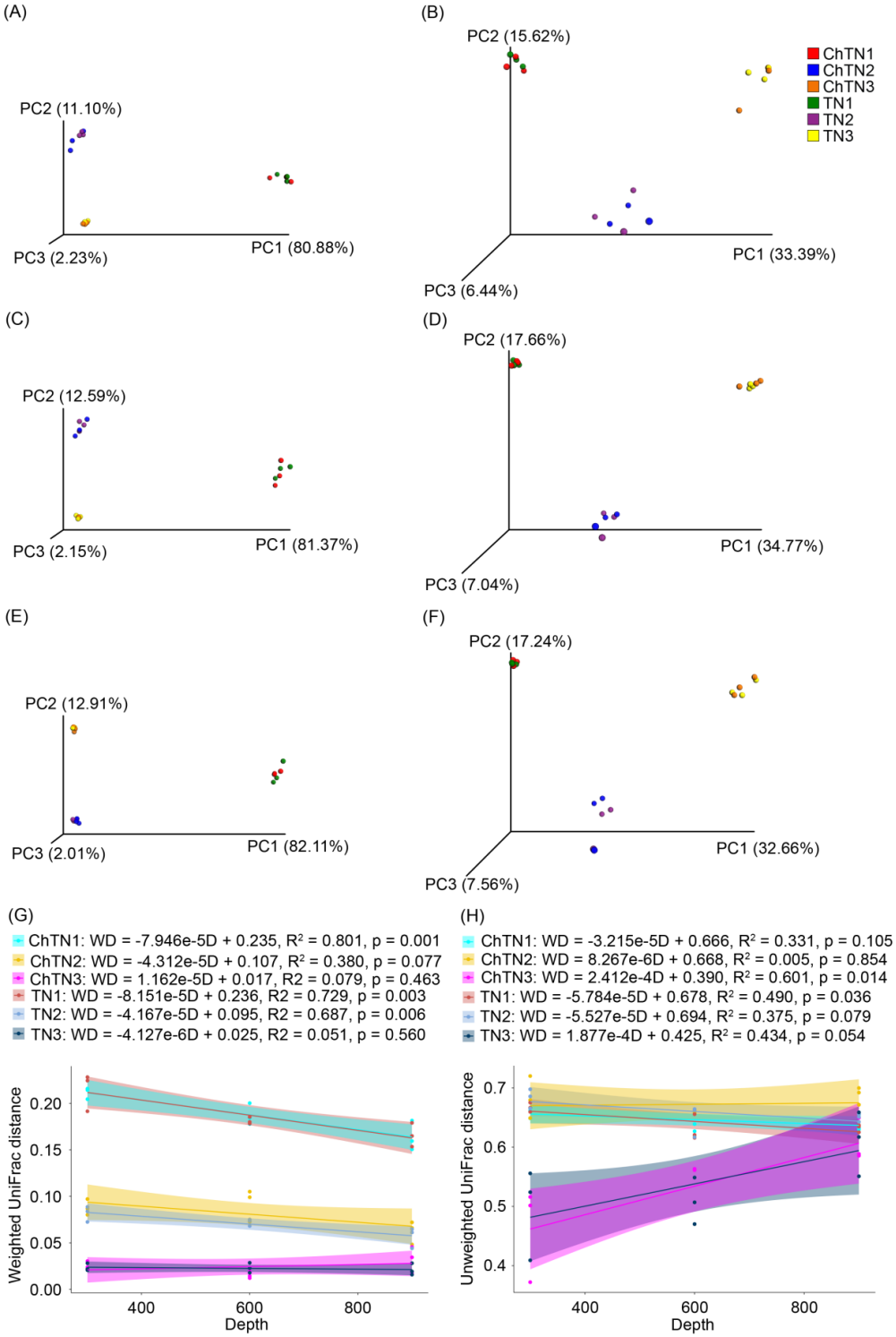


Figure 2. *Principal coordinate analysis (PCoA) profiles and correlations between sequencing depth and UniFrac distances of Tegillarca granosa gut microbiota based on different sequence treatments. (A) and (B) Weighted and unweighted PCoA profiles constructed based on 300 sequences per sample, respectively; (C) and (D) weighted and unweighted PCoA profiles constructed based on 600 sequences per sample, respectively; (E) and (F) weighted and unweighted PCoA profiles constructed based on 900 sequences per sample, respectively; (G) and (H) correlations between weighted and unweighted UniFrac distances of the microbiota and sequencing depths, respectively. The shaded part is the 95% confidence interval*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

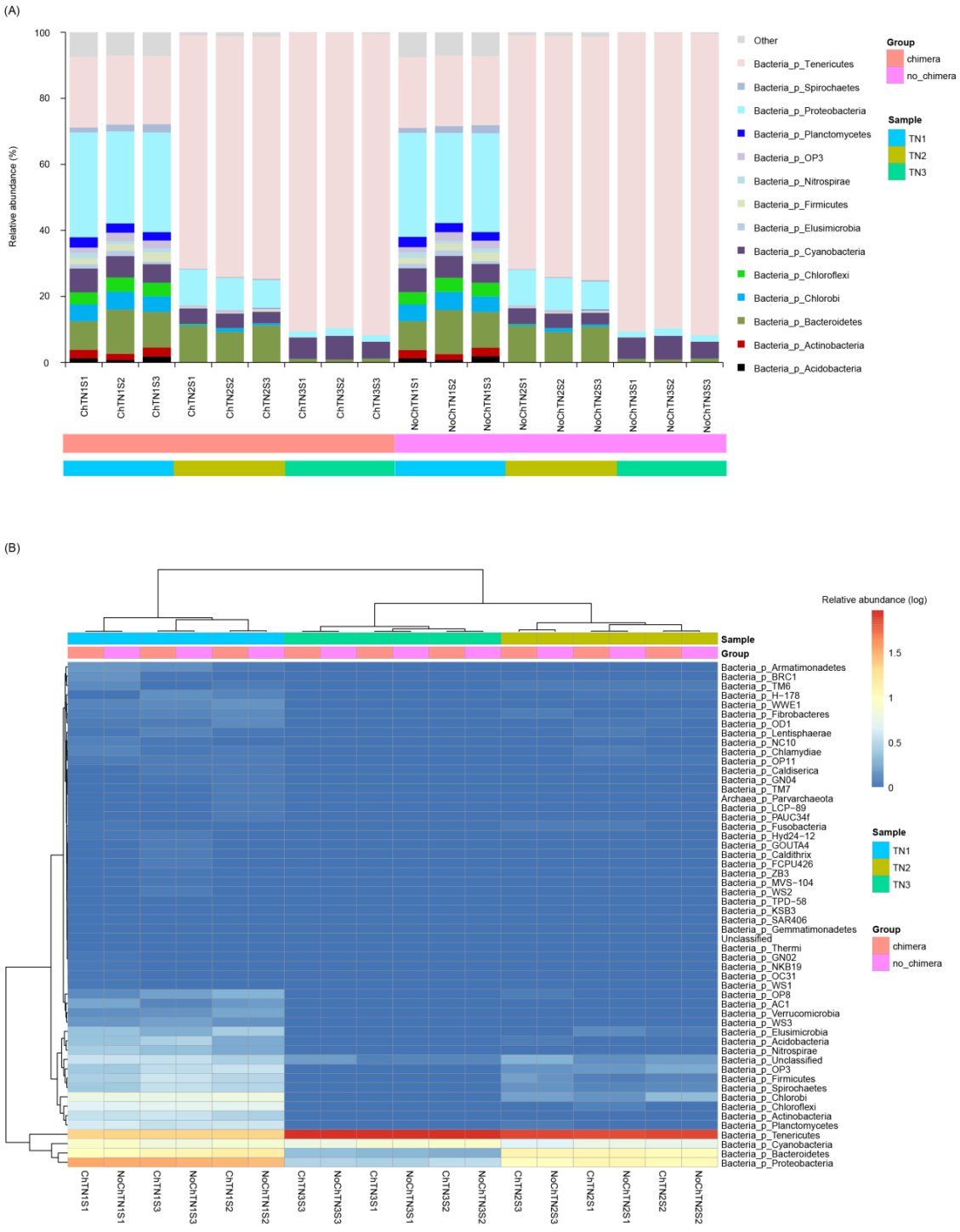


Figure 3. *Phylum compositions of Tegillarca granosa gut microbiota based on different sequence treatments. (A) Dominant phyla of T. granosa gut microbiota; (B) heatmap profile shows the relative abundance of phyla of T. granosa gut microbiota*

Heatmap combined with cluster analysis showed that removing the chimeric sequences did not influence the composition of dominant genera and had less impact on the composition of the *T. granosa* gut microbiota at the genus level than the results that

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

were sequenced with different barcode sequences (*Fig. 4A*). To analyse the influence of chimeric sequences on the genus composition of the *T. granosa* gut microbiota, we screened out the genera with differences between the results of removing and not removing chimeric sequences. Our results showed that the genera caused by chimeras were most abundant in TN1S2 and TN1S3, among which *Aquincola*, *Azovibrio*, *Arthrospira*, *Anaerospora*, and each unidentified genus of the family Acetobacteraceae and order Vibrionales were the genera caused by chimeras in TN1S2, and *Azohydromonas*, *Paenochrobactrum*, *Devosia*, *Azovibrio*, and an unidentified genus of order Vibrionales were the genera caused by chimeras in TN1S3. *Aquabacterium* and K82 of the family Rhodocyclaceae were the genera caused by chimeras in TN1S1. Moreover, *Lactobacillus*, *Acidovorax*, and *Kerstersia* were the genera caused by chimeras in TN2S1, TN3S2, and TN3S1, respectively. No genus caused by chimeras was detected in TN3S3, TN2S2, or TN2S3 (*Fig. 4B*). These results indicated that the occurrence of chimeras was completely random.

OTU is the most commonly used classification level for microbiota structure analysis, and is usually used to calculate the α-diversity indices (such as Shannon index and Chao1 index) of the microbiota, and perform PCoA or NMDS analysis (Ni et al., 2018; Wu et al., 2019, 2022; Chen et al., 2021). Therefore, we compared the effects of removing chimeric sequences on the α-diversity and OTU composition of the *T. granosa* gut microbiota. As we expected, chimeras significantly increased the OTU number, Shannon index and Simpson index of the microbiota (Wilcoxon test, $p < 0.05$; *Fig. 5A-C*), whereas they did not significantly increase the Chao1 index (Wilcoxon, $p = 0.055$; *Fig. 5D*). Among the 3087 OTUs obtained, there were 19 dominant OTUs. Removing chimeric sequences did not change the composition of the dominant OTUs of the *T. granosa* gut microbiota. Chimeric sequences only slightly reduced the relative abundances of the dominant OTUs in the communities (*Fig. 6A*). Furthermore, among the 19 dominant OTUs, since most OTUs were not detected in all samples, we only evaluated the barcode impact of seven dominant OTUs that were detected in all samples. Our results showed that among the seven dominant OTUs, the relative abundances of only two dominant OTUs were significantly underestimated in barcode B3 amplifications (<90% of their average abundances), although the relative abundances of these dominant OTUs in the samples corresponding to different barcodes varied greatly (most of the relative abundances differed by > 10%; *Appendix 1*). We believe that the difference introduced by nonbarcodes was mainly caused by random factors. These results indicated that although barcode sequences caused the relative abundances of specific dominant OTUs to be incorrectly estimated, random factors had a greater impact. Moreover, three dominant OTUs as the error OTUs introduced by PCR amplification were detected in sample TN1 (*Fig. 6A*).

For 3068 rare OTUs, only 73 OTUs were caused by chimeric sequences, whereas 1637 OTUs were singletons. Moreover, 116 rare OTUs avoided becoming singletons because we used three barcode sequences to analyse a sample simultaneously (*Table 1*). These results indicated that the interference of chimeric sequences with microbiota structure was very weak. Furthermore, the number of OTUs caused by chimeric sequences was significantly positively correlated with the Shannon index of the microbiota (Pearson's product-moment correlation, $R2 = 0.575$, $p = 0.018$; *Fig. 6B*), and the number of singletons was significantly positively correlated with the product of the Shannon index and sequencing depth of the communities (Pearson's product-moment correlation, $R2 = 0.671$, $p = 0.007$; *Fig. 6C*).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

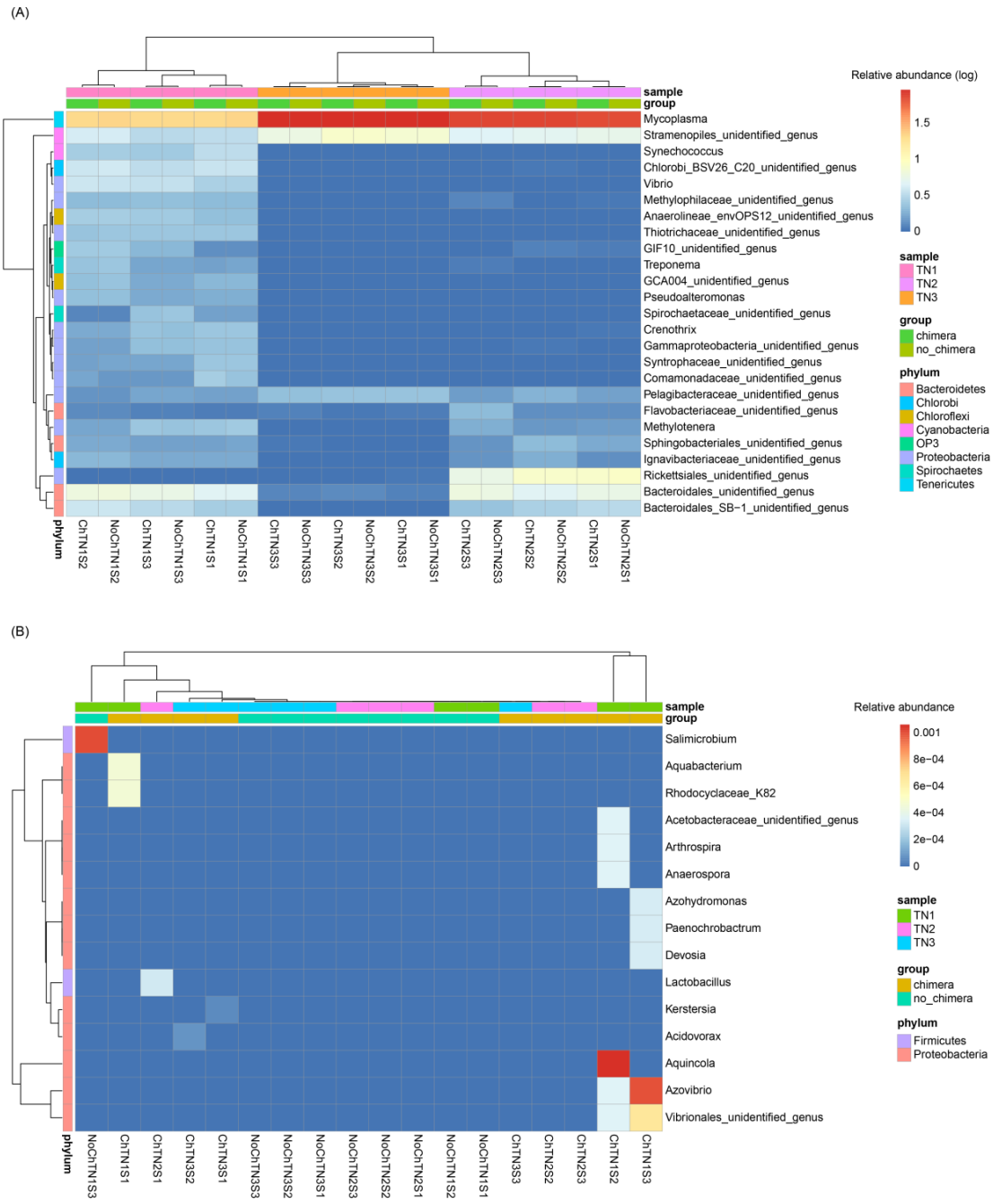


Figure 4. *Dominant genera of Tegillarca granosa gut microbiota (A) and the different genera caused by the chimera sequences in the samples (B)*

Table 1. *OTU number caused by chimeric sequences (OTU_CS), singletons and rare OTU number that avoid becoming singletons as repeat sequencing using different barcodes (OTU_R)*

OUT number	TN1	TN2	TN3
OUT_CS	17.67 ± 6.57	3.67 ± 0.88	4.00 ± 0.00
Singletons	259.33 ± 17.57	87.67 ± 26.40	198.67 ± 15.88
OUT_R	49.67 ± 4.67	9.00 ± 1.73	21.00 ± 1.53

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

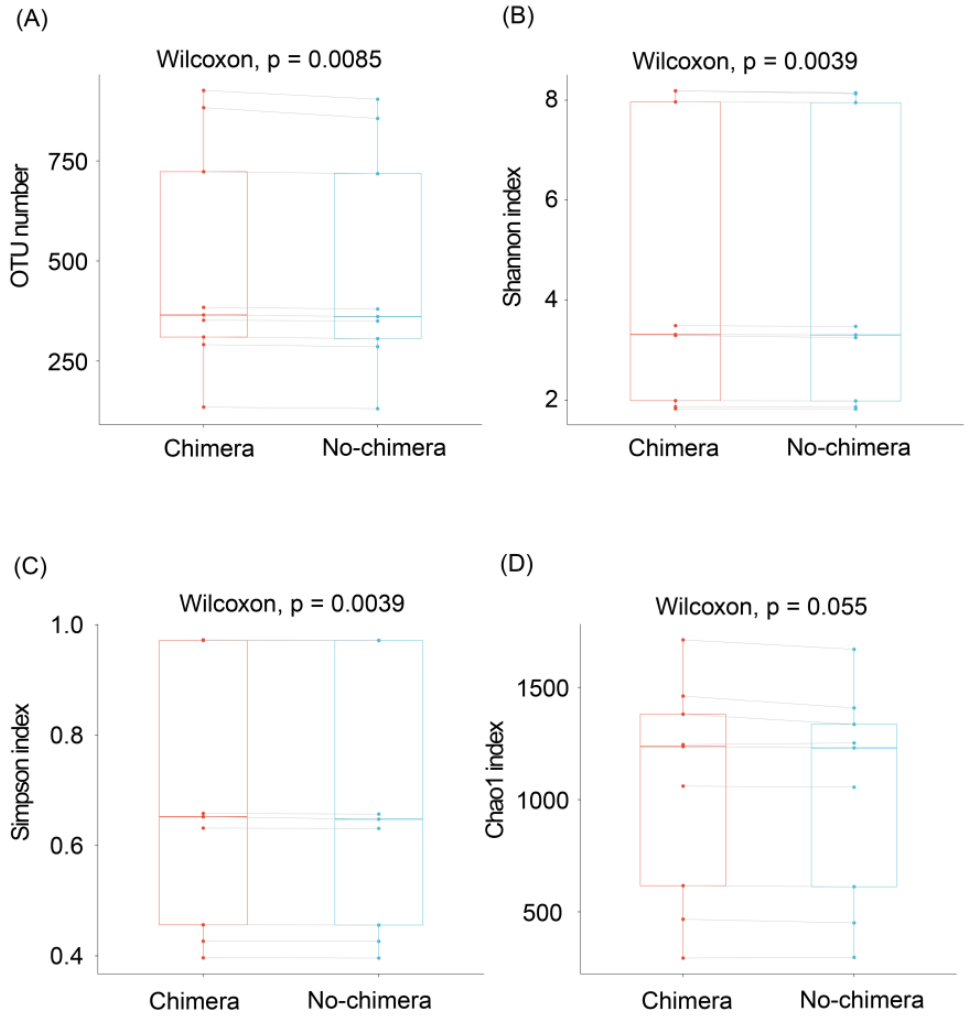


Figure 5. *Impact of chimeras on the α-diversity of the Tegillarca granosa gut microbiota. (A) OTU number; (B) Shannon index; (C) Simpson index; (D) Chao1 index*

Discussion

How to obtain "real" information on microbiota structure has always been the core issue of microbial ecology. Although culture-free molecular technologies have been widely used in the ecological investigation of microbiota structure (Lazarevic et al., 2009; Siddiqui et al., 2011; Ni et al., 2014a, b, 2017b; Doyle et al., 2016; Liu et al., 2022a, b), errors are introduced from sample collection, cell lysis and DNA extraction, PCR amplification to the analysis of amplification products and data, resulting in our interpretation of microbiota structure being misleading (Wintzingerode et al., 1997). Chimeric sequences are the most concerning problem in the high-throughput sequencing analysis of 16S rDNA amplicons of microbiota structures (Haas et al., 2011; Edgar et al., 2011). However, our results showed that chimeric sequences mainly led to the overestimation of the α-diversity of microbiota and affected the number or rare OTUs, whereas their impact on the number of rare OTUs was insufficient, and random factors had a greater impact on the rare OTUs than the chimeric sequences. When analysing the β-diversity of the *T. granosa* gut microbiota and using dominant OTUs, chimeric sequences had little impact on the results.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

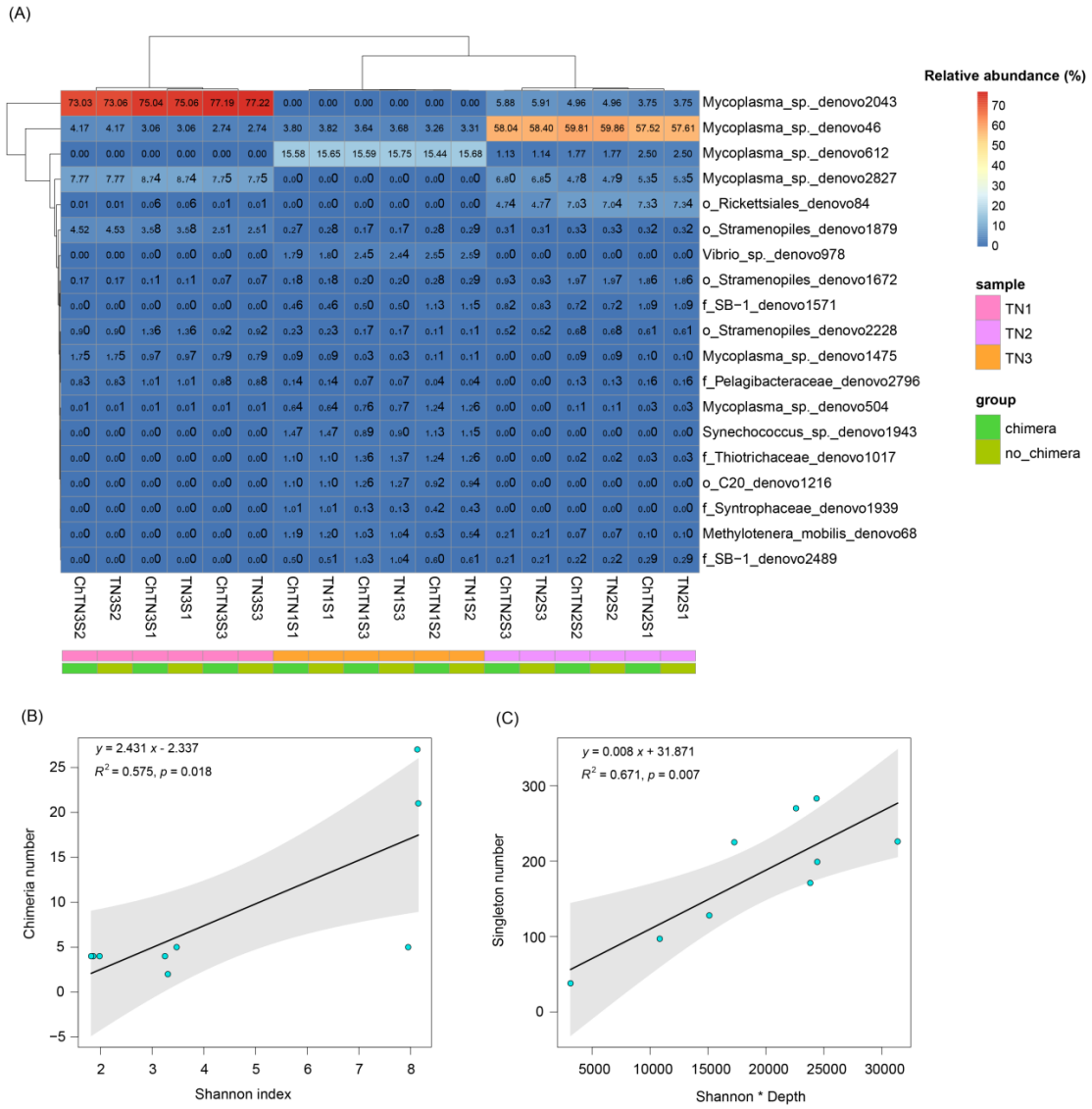


Figure 6. *Relative abundances of dominant OTUs of Tegillarca granosa gut microbiota (A), correlation between OTU number caused by chimeric sequences and Shannon index (B), and correlation between singleton number and Shannon index × sequencing depth (C). The shaded part is the 95% confidence interval*

Barcode sequences are usually introduced into primers for PCR amplification to distinguish amplified fragments from different sample sources (Hamady et al., 2008; Hawkins et al., 2018; Chen et al., 2020). However, to date, there has been little evaluation of the impact of different barcode sequences on multiple amplification results of microbiota structure analysis. Our results showed that although different barcode sequences did not affect the relative abundances of all OTUs, and their impact on the relative abundances of OTUs was not as obvious as that caused by random factors, different barcode sequences indeed led to the incorrect estimation of the relative abundances of some specific OTUs (*Appendix 1*). Therefore, to obtain more accurate relative abundance data, we suggest using different barcode sequences to amplify and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

sequence the same sample, and then using the average of the relative abundances of each OTU from the same samples as the relative abundance of the OTU in the sample.

Although rare species play an important role in the stability of microbiota structure (Jiao et al., 2017; Wu et al., 2017; Gotoh et al., 2019; Wang et al., 2022), our results showed that the lock of sequencing depth has a great impact on the relative abundances of rare OTUs when using high-throughput sequencing of 16S rDNA to analyse microbiota structure. Furthermore, random factors were also important factors that mainly affected the relative abundances of rare OTUs in the microbiota. Therefore, sufficient sequencing depth combined with repeated analysis can obtain more accurate relative abundance data of rare OTUs in the microbiota.

Conclusions

In general, removing chimeric sequences, different barcoding sequences, and different sequencing depths did not fundamentally change the differences in the microbiota structure, and microbiota with different treatments were still aggregated according to the samples. Removing chimeric sequences and adopting different barcode sequences did not obviously affect the composition of the *T. granosa* gut microbiota at the phylum and genus levels. Chimeric sequences mainly led to overestimation of the α-diversity of the *T. granosa* gut microbiota and the number of rare OTUs, whereas their impact on the number of rare OTUs was insufficient and random factors had a higher impact than chimeric sequences. When analysing the β-diversity of the microbiota and dominant OTUs, chimeric sequences had little impact on the results. Although different barcode sequences did not affect the relative abundances of all OTUs, and their impact on the relative abundances of OTUs was not as obvious as that caused by random factors, different barcode sequences indeed led to the incorrect estimation of the relative abundances of some specific OTUs. In this study, we only used three barcode sequences to analyse three *T. granosa* gut microbiota samples, and larger environmental microbiota samples need to be further analysed.

Acknowledgements. This research was funded by the Chengdu University of TCM Science Foundation (grant number BSH2019010).

REFERENCES

[1] Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., Polz, M. F. (2005): PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. – Applied and Environmental Microbiology 71(12): 8966-8969.

[2] Anderson, M. J. (2001): A new method for non-parametric multivariate analysis of variance. – Austral Ecology 26: 32-46.

[3] 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. (2010): QIIME allows analysis of high-throughput community sequencing data. – Nature Methods 7: 335-336.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

[4] Chen, W., Wang, L., Han, M., Han, C., Li, B. (2020): Sequencing barcode construction and identification methods based on block error-correction codes. – Science China Life Sciences 63: 1580-1592.
[5] Chen, X., He, D., Zhou, L., Cao, Y., Li, Z. (2021): Influence of hydropower stations on the water microbiota in the downstream of Jinsha River, China. – PeerJ 8: e9500.
[6] Chu, H., Gao, G.-F., Ma, Y., Fan, K., Delgado-Baquerizo, M. (2020): Soil microbial biogeography in a changing world: recent advances and future perspectives. – mSystems 5(2): e00803-19.
[7] Dixon, P. (2003): VEGAN, a package of R functions for community ecology. – Journal of Vegetation Science 14: 927-930.
[8] Doyle, C. J., Gleeson, D., O'Toole, P. W., Cotter, P. D. (2016): Impacts of seasonal housing and test preparation on raw milk microbiota: a high-throughput sequencing study. – Applied and Environmental Microbiology 83(2): e02694-16.
[9] Edgar, R. C. (2013): UPARSE: highly accurate OTU sequences from microbial amplicon reads. – Nature Methods 10: 996-998.
[10] Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., Knight, R. (2011): UCHIME improves sensitivity and speed of chimera detection. – Bioinformatics 27: 2194-2200.
[11] Gong, W., Gao, S., Zhu, Y., Wang, G., Zhang, K., Li, Z., Yu, E., Tian, J., Xia, Y., Xie, J., Ni, J. (2021): Effect of the aerobic denitrifying bacterium *Pseudomonas furukawaii* ZS1 on microbiota compositions in grass carp culture water. – Water 13: 1329.
[12] Gotoh, A., Ojima, M. N., Katayama, T. (2019): Minority species influences microbiota formation: the role of *Bifidobacterium* with extracellular glycosidases in bifidus flora formation in breastfed infant guts. – Microbial Biotechnology 12(2): 259-264.
[13] Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., The Human Microbiome Consortium, Petrosino, J. F., Knight, R., Birren, B. W. (2011): Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. – Genome Research 21: 494-504.
[14] Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., Knight, R. (2008): Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex. – Nature Methods 5(3): 235-237.
[15] Hawkins, J. A., Jones, S. K. Jr, Finkelstein, I. J., Press, W. H. (2018): Indel-correcting DNA barcodes for high-throughput sequencing. – Proceedings of the National Academy of Sciences of United States of American 115(27): E6217-E6226.
[16] Jiao, S., Luo, Y., Lu, M., Xiao, X., Lin, Y., Chen, W., Wei, G. (2017): Distinct succession patterns of abundant and rare bacteria in temporal microcosms with pollutants. – Environmental Pollution 225: 497-505.
[17] Lazarevic, V., Whiteson, K., Huse, S., Hernandez, D., Farinelli, L., Østerås, M., Schrenzel, J., François, P. (2009): Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. – Journal of Microbiological Methods 79(3): 266-271.
[18] Li, J. J., Ni, J. J., Wang, C., Yu, Y. H., Zhang, T. L. (2021): Different response patterns of fish foregut and hindgut microbiota to host habitats and genotypes. – Applied Ecology and Environmental Research 18(5): 7369-7378.
[19] Li, J., Li, J., Ni, J., Zhang, C., Jia, J., Wu, G., Sun, H., Wang, S. (2022a): Berberine relieves metabolic syndrome in mice by inhibiting liver inflammation caused by a high-fat diet and potential association with gut microbiota. – Frontiers in Microbiology 12: 752512.
[20] Li, J., Sun, H., Wang, C., Li, S., Cai, Y. (2022b): Subchronic toxicity of microcystin-LR on young frogs (*Xenopus laevis*) and their gut microbiota. – Frontiers in Microbiology 13: 895383.
[21] Liu, Q., Lai, Z., Wang, C., Ni, J., Gao, Y. (2022a): Seasonal variation significantly affected bacterioplankton and eukaryoplankton community composition in Xijiang River, China. – Environmental Monitoring and Assessment 194: 55.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

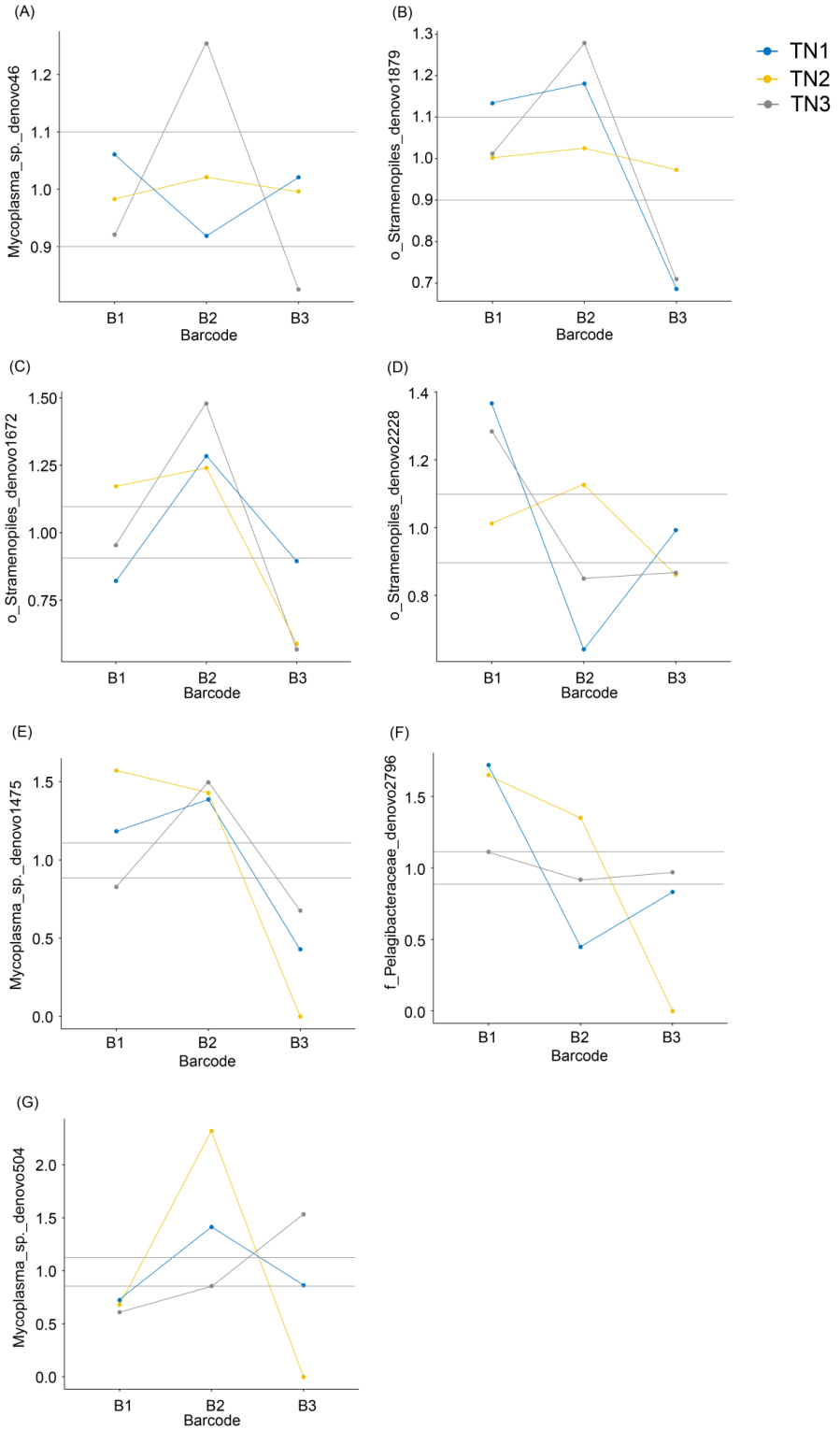
[22] Liu, Q., Gao, Y., Wang, C., Zeng, Y., Ni, J., Lai, Z. (2022b): Effects of ambient microbiota on the gut microbiota of *Macrobrachium rosenbergii*. – Water 14: 658.
[23] Magoc, T., Salzberg, S. L. (2011): FLASH: fast length adjustment of short reads to improve genome assemblies. – Bioinformatics 27: 2957-2963.
[24] Mao, L. T., Lai, L. E., Lin, G. G., Xu, L. X., Chen, Z. G., Zhu, W. J. (2020): Differences in rhizosphere microbiota compositions between healthy and diseased potato (*Solanum tuberosum*) in China. – Applied Ecology and Environmental Research 18(2): 3683-3691.
[25] Mao, L. T., Huang, Z. C., Lu, Z. Z., Lin, G. G., Chen, Z. F., Yang, Z. (2022): Multi-species *Bacillus* inoculant and its effect on potato growth and control of potato (*Solanum tuberosum* L.) blackleg disease. – Applied Ecology and Environmental Research 20(2): 1571-1583.
[26] Ni, J., Yan, Q., Yu, Y., Zhang, T. (2014a): Fish gut microecosystem: a model for detecting spatial pattern of microorganisms. – Chinese Journal of Oceanology and Limnology 32(1): 54-57.
[27] Ni, J., Yan, Q., Yu, Y., Zhang, T. (2014b): Factors influencing the grass carp gut microbiome and its effect on metabolism. – FEMS Microbiology Ecology 87(3): 704-714.
[28] Ni, J., Li, X., He, Z., Xu, M. (2017a): A novel method to determine the minimum number of sequences required for reliable microbial community analysis. – Journal of Microbiological Methods 139: 196-201.
[29] Ni, J. J., Yan, Q. Y., Yu, Y. H., Wu, H. H., Chen, F. (2017b): Dispersal patterns of endogenous bacteria among grass carp (*Ctenopharyngodon idellus*) guts. – Iranian Journal of Fisheries Sciences 16(2): 605-618.
[30] Ni, J. J., Li, X. J., Chen, F., Wu, H. H., Xu, M. Y. (2018): Community structure and potential nitrogen metabolisms of subtropical aquaculture pond microbiota. – Applied Ecology and Environmental Research 16(6): 7687-7697.
[31] Ni, J., Huang, R., Zhou, H., Xu, X., Li, Y., Cao, P., Zhong, K., Ge, M., Chen, X., Hou, B., Yu, M., Peng, B., Li, Q., Zhang, P., Gao, Y. (2019): Analysis of the relationship between the degree of dysbiosis in gut microbiota and prognosis at different stages of primary hepatocellular carcinoma. – Frontiers in Microbiology 10: 1458.
[32] Ni, J., Fu, C., Huang, R., Li, Z., Li, S., Cao, P., Zhong, K., Ge, M., Gao, Y. (2021): Metabolic syndrome cannot mask the changes of faecal microbiota compositions caused by primary hepatocellular carcinoma. – Letters in Applied Microbiology 73: 73-80.
[33] Paul, D., Kumbhare, S. V., Mhatre, S. S., Chowdhury, S. P., Shetty, S. A., Marathe, N. P., Bhute, S., Shouche, Y. S. (2016): Exploration of microbial diversity and community structure of Lonar Lake: The only hypersaline meteorite crater lake within basalt rock. – Frontiers in Microbiology 6: 1553.
[34] R Core Team. (2013): R: A Language and Environment for Statistical Computing. – R Foundation for Statistical Computing, Vienna.
[35] Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R., Fierer, N. (2010): Soil bacterial and fungal communities across a pH gradient in an arable soil. – The ISME Journal 4: 1340-1351.
[36] Siddiqui, H., Nederbragt, A. J., Lagesen, K., Jeansson, S. L., Jakobsen, K. S. (2011): Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. – BMC Microblogy 11: 244.
[37] Sze, M. A., Schloss, P. D. (2019): The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. – mSphere 4(3): e00163-19.
[38] Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., Nishijima, M. (2014): Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. – PloS ONE 9(8): e105592.
[39] Wang, Q., Garrity, G. M., Tiedje, J. M., Cole, J. R. (2007): Naïve Bayesian classifier for rapid assignment of rRNA sequences into new bacterial taxonomy. – Applied and Environmental Microbiology 73: 5261-5267.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

[40] Wang, H., Qi, Z., Zheng, P., Jiang, C., Diao, X. (2022): Abundant and rare microbiota assembly and driving factors between mangrove and intertidal mudflats. – Applied Soil Ecology 174: 104438.
[41] Wintzingerode, F. V., Göbel, U. B., Stackebrandt, E. (1997): Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. – FEMS Microbiology Reviews 21: 213-229.
[42] Wu, J.-Y., Jiang, X.-T., Jiang, Y.-X., Lu, S.-Y., Zou, F., Zhou, H.-W. (2010): Effects of polymerase, template dilution and cycle number on PCR based 16S rRNA diversity analysis using the deep sequencing method. – BMC Microbiology 10: 255.
[43] Wu, L., Sun, Q., Ni, J. (2017): Not all of the rare operational taxonomic units (OTUs) play the same role in maintaining community stability. – Applied Ecology and Environmental Research 15(1): 105-112.
[44] Wu, L., Shu, F., Ou, Z., Chen, Q., Wang, L., Wang, H., Xu, Z. (2019): Compositions of prokaryote communities and their relationship to physiochemical factors in December in Chaohu Lake and three urban rivers in China. – Applied Ecology and Environmental Research 17(4): 7265-7281.
[45] Wu, Y., Wang, B., Tang, L., Zhou, Y., Wang, Q., Gong, L., Ni, J., Li, W. (2022): Probiotic *Bacillus* alleviates oxidative stress-induced liver injury by modulating gut-liver axis in a rat model. – Antioxidants 11: 291.
[46] Xiang, J., He, T., Wang, P., Xie, M., Xiang, J., Ni, J. (2018): Opportunistic pathogens are abundant in gut of cultured giant spiny frog (*Paa spinosa*). – Aquaculture Research 49: 2022-2041.
[47] Zhang, X., Wang, H., Li, Z., Xie, J., Ni, J. (2020): Hydrological and soil physiochemical variables determine the rhizospheric microbiota in subtropical lakeshore areas. – PeerJ 8: e10078.
[48] Zhu, T., Li, X., Wu, X., Yang, D. (2022): Temperature acclimation alters the thermal tolerance and intestinal heat stress response in a Tabetan fish *Oxygymnocypris stewarti*. – Frontiers in Microbiology 13: 898145.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary

APPENDIX



Appendix 1. *Effects of different barcode sequences on the relative abundance of dominant OTUs. The ordinate is the ratio of the relative abundance corresponding to each barcode sequence to the average value obtained from three different barcode sequences in the same sample*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2006_53275341
© 2022, ALÖKI Kft., Budapest, Hungary