

University of Windsor

## Scholarship at UWindsor

---

Biological Sciences Publications

Department of Biological Sciences

---

2015

### Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities

Changsen Sun  
*University of Windsor*

Yan Zhao  
*University of Windsor*

Li, Haitao Li, Haitao  
*University of Windsor*

Yanhong Dong  
*University of Windsor*

Hugh J. Maclsaac  
*University of Windsor*

*See next page for additional authors*

Follow this and additional works at: <https://scholar.uwindsor.ca/biologypub>



Part of the [Biology Commons](#)

---

#### Recommended Citation

Sun, Changsen; Zhao, Yan; Li, Haitao, Li, Haitao; Dong, Yanhong; Maclsaac, Hugh J.; and Zhan, Aibin, "Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities" (2015). *Aquatic Biology*, 24, 1, 9-15.  
<https://scholar.uwindsor.ca/biologypub/1160>

This Article is brought to you for free and open access by the Department of Biological Sciences at Scholarship at UWindsor. It has been accepted for inclusion in Biological Sciences Publications by an authorized administrator of Scholarship at UWindsor. For more information, please contact [scholarship@uwindsor.ca](mailto:scholarship@uwindsor.ca).

---

**Authors**

Changsen Sun; Yan Zhao; Li, Haitao Li, Haitao; Yanhong Dong; Hugh J. MacIsaac; and Aibin Zhan



# Unreliable quantitation of species abundance based on high-throughput sequencing data of zooplankton communities

Changsen Sun<sup>1</sup>, Yan Zhao<sup>2</sup>, Haitao Li<sup>3</sup>, Yanhong Dong<sup>3</sup>, Hugh J. MacIsaac<sup>4</sup>,  
Aibin Zhan<sup>2,\*</sup>

<sup>1</sup>School of Life Science, Taizhou University, 1139 Shifu Avenue, Taizhou, Zhejiang 318000, PR China

<sup>2</sup>Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 18 Shuangqing Road, Haidian District, Beijing 100085, PR China

<sup>3</sup>South China Sea Environmental Monitoring Center, State Oceanic Administration, 155 Xingang Road West, Guangzhou, Guangdong 510300, PR China

<sup>4</sup>Great Lakes Institute for Environmental Research, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada

**ABSTRACT:** High-throughput sequencing (HTS) is rapidly becoming a popular and robust tool to characterize biodiversity of complex communities, especially for those dominated by microscopic species such as zooplankton. The popular use of HTS-based methods has prompted a possible method of inferring relative species abundance from sequencing data. However, these methods remain largely untested in many communities as to whether sequence data can reliably quantify relative species abundance. Here we tested the relationship between species abundance and sequence abundance in zooplankton using 2 methods: (1) spiking known amounts of indicator species into existing zooplankton communities, and (2) comparing results obtained from parallel replicates for the same natural zooplankton communities. Although we detected a general trend that low-abundance species usually corresponded to low-abundance sequence reads, further statistical analyses revealed that sequencing data could not reliably quantify relative species abundance, even for the same indicator species spiked into different zooplankton communities. The distribution of sequence reads statistically varied even between parallel replicates of the same natural zooplankton communities. Our study reveals that sequence abundance may generally qualitatively reflect species abundance as the general trend between these 2 variables exists; however, extra caution is required when using HTS-based approaches to make quantitative inferences regarding zooplankton communities.

**KEY WORDS:** Biodiversity · Biomass · High-throughput sequencing · HTS · Small subunit ribosomal DNA · SSU 18S rDNA · Species abundance · Zooplankton

## INTRODUCTION

Quantitation of species abundance in zooplankton communities represents one of the most crucial issues in many studies of aquatic ecology, such as ecosystem functioning, changes in seasonal community dynamics, and environmental quality assess-

ment (Nigam et al. 2006, Creer 2010, Weber & Pawlowski 2013). To date, determination of species abundance in zooplankton communities has largely relied on direct counting under microscopes or flow cytometry (Shi et al. 2011, Weber & Pawlowski 2013), although molecular methods such as quantitative PCR (e.g. Mackie & Geller 2010) have increasingly

\*Corresponding author: zhanaibin@hotmail.com

© The authors 2015. Open Access under Creative Commons by Attribution Licence. Use, distribution and reproduction are unrestricted. Authors and original publication must be credited.

been utilized. However, these methods suffer from drawbacks, including poor species resolution and detection errors for rare species for the former methods, and low throughput for quantitative PCR-based methods for the latter (i.e. difficulties in determining large numbers of species in zooplankton using a single effort; Weber & Pawlowski 2013).

Recently, high-throughput sequencing (HTS)-based approaches have become popular and robust tools for characterizing the biodiversity of complex communities, especially for those dominated by microscopic species or species that are difficult to identify, such as zooplankton (Lindeque et al. 2013, Zhan et al. 2013, Hirai et al. 2015). The use of HTS has allowed the exploration of complex communities at an unprecedented depth, identifying orders of magnitude of more biodiversity than was previously recognized (Creer 2010, Fonseca et al. 2010). Moreover, the use of HTS-based methods has highly improved detection sensitivity (e.g. as low as  $2.3 \times 10^{-5}\%$  of sample biomass; Zhan et al. 2013), allowing for detection of biodiversity at orders of magnitude of lower abundance than traditional morphological taxonomy-based methods (e.g. Lindeque et al. 2013, Zhan et al. 2014a).

The use of HTS-based methods has also prompted a possible way of inferring relative abundance of species from sequencing data, as HTS data derived from complex communities recovers not only species composition but also relative abundance based upon sequence reads for each taxon. However, it is unclear whether relative abundances of sequence reads in HTS datasets can reliably quantify species abundance in communities, that is, the possible relationship between biological abundance in communities and sequence abundance in sequencing data (Zhou et al. 2011, Weber & Pawlowski 2013, Dannemiller et al. 2014). Such a possible relationship remains largely untested in natural complex communities, especially in aquatic communities such as zooplankton.

Here we tested the relationship between biological abundance in zooplankton communities and sequence abundance in sequencing data using a series of biomass gradients of 4 known indicator species spiked into existing zooplankton communities. In addition, we examined parallel replicates of the same natural zooplankton communities to test whether quantitative inferences are reliable in comparative studies when using HTS-based methods.

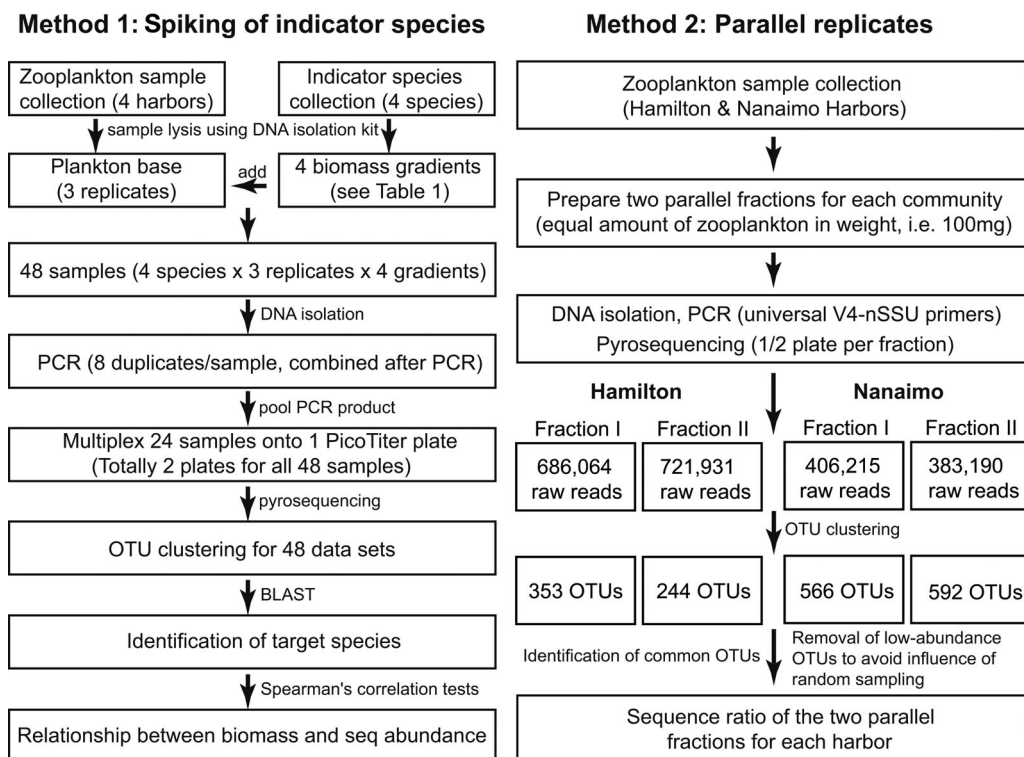


Fig. 1. Flow charts for testing the relationship between species abundance (i.e. biomass) and sequence abundance in zooplankton using 2 methods: (1) spiking known amounts of indicator species into existing zooplankton communities (left); and (2) comparing results obtained from parallel replicates for the same zooplankton communities (right). OTU: operational taxonomic unit

## MATERIALS AND METHODS

As zooplankton communities are extremely complex in species composition and biomass highly varies among species, the spiking of a large amount of foreign species into existing zooplankton communities may result in technical problems for PCR amplification/sequencing or loss of power to detect rare species. Thus, we used a relatively low level, but a wide range, of relative biomass of indicator species (see Table 1).

The 4 known indicator species, including bay scallop *Argopecten irradians*, Japanese sea cucumber *Apostichopus japonicus*, golden mussel *Limnoperna fortunei*, and water louse *Asellus aquaticus*, were spiked into existing zooplankton communities using a series of biomass gradients, following Zhan et al. (2013). These indicator species were chosen to cover common aquatic groups including Mollusca, Crustacea, and Echinodermata based on the availability of species in our laboratories. To avoid possible errors and confusion derived from spiked species, we spiked marine species into freshwater zooplankton samples and freshwater species into marine zooplankton samples (Fig. 1). Specifically, we spiked larvae of 2 marine species, the bay scallop and Japanese sea cucumber, into ethanol-preserved zooplankton samples collected from 2 freshwater harbors: Thunder Bay and Nanticoke in Ontario, Canada, while larvae/juveniles of 2 freshwater species, the golden mussel and water louse, were spiked into ethanol-preserved zooplankton samples collected from 2 marine harbors: Bayside and Hawksbury in Nova Scotia, Canada. The indicator species used here are typically either marine or freshwater organisms, and there is no report on habitat transitions from freshwater to marine water bodies and vice versa. In addition, these species have never been detected in zooplankton communities where they were spiked.

For zooplankton sample collection, we used 80  $\mu\text{m}$  oblique plankton nets to tow from the bottom to water surface in each port. All collected zooplankton samples were immediately preserved in 100% ethanol. Larvae of the 2 marine species were artificially cultured in the laboratory (Zhan et al. 2008), while the larvae/juveniles of golden mussel and water louse were collected from the wild in South America and Europe, respectively.

Depending on the available amount of zooplankton from each port, we used 50–150 mg of zooplankton samples for spiking indicator species. These preserved samples were well homogenized

and weighed for quantitative measurement of relative abundance of spiked indicator species. For each indicator species, we established 4 gradients, and for each gradient, we established 3 replicates (see Table 1). We assembled a total of 48 artificial communities (i.e. 4 species  $\times$  4 gradients  $\times$  3 replicates for each gradient; Fig. 1). All spiking procedures were performed before DNA extraction. For the gradients using  $>1$  larva, we spiked larvae directly into zooplankton samples, while for those  $<1$ , we lysed 1 larva/juvenile using 200  $\mu\text{l}$  DNA lysis buffer and then added different amounts of lysed larva/juvenile solution into corresponding lysed zooplankton samples based on dilution gradients (Zhan et al. 2013).

Total genomic DNA was extracted from the 48 artificially assembled communities using DNeasy Blood and Tissue Kit (Qiagen). This kit was approved to be robust for DNA isolation from zooplankton, as DNA was successfully isolated and large numbers of divergent taxonomic groups were detected from both marine and freshwater zooplankton communities using HTS (e.g. Zhan et al. 2014a). In addition, this kit worked well for a low level of biomass of the indicator species as the detection sensitivity was as low as  $2.3 \times 10^{-5}\%$  of sample biomass when these indicator species were spiked into existing zooplankton as rare species (Zhan et al. 2013). The quantity and quality of isolated DNA was assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific). PCRs were performed using the primer pair Uni18S-Uni18SR spanning the hypervariable V4 region of small subunit ribosomal DNA (SSU rDNA) (Zhan et al. 2013, 2014a). We prepared PCR mixtures (25  $\mu\text{l}$ ) in 8 replicates for each sample to avoid biased amplification (i.e. varied amplification efficiencies of particular molecules in different PCR mixtures). Each replicate consisted of 100 ng of genomic DNA,  $1\times$  PCR buffer, 2 mM of  $\text{Mg}^{2+}$ , 0.2 mM of each dNTP, 0.4  $\mu\text{M}$  of each primer, and 2 U of *Taq* DNA polymerase (Genscript). PCR cycling parameters consisted of an initial denaturation step at 95°C for 5 min, followed by 25 amplification cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 90 s, and a final elongation step at 72°C for 10 min. All replicated PCR products derived from the same samples were pooled and purified using the Solid Phase Reversible Immobilization (SPRI) paramagnetic bead-based method (Agencourt Bioscience).

For 454 pyrosequencing, we pooled PCR products derived from 24 artificially assembled communities to form 1 PicoTiter plate (2 plates total for all 48 assembled communities). Pyrosequencing was performed using 454 FLX Adaptor A on a GS-FLX Tita-

nium platform (454 Life Sciences) by Engencore at the University of South Carolina. After pyrosequencing, spiked indicator species were identified by local BLAST from each dataset using available reference sequences. Reference sequences for each indicator species were established by sequencing several individuals using traditional Sanger sequencing. The relationship between relative species abundance and sequence read abundance was assessed using Spearman's correlation tests implemented in Statistica v.6 (StatSoft). For non-normal distribution data, Spearman's correlation test is powerful enough to measure statistical dependence between 2 variables.

To test quantitative inferences for comparative studies when using HTS-based methods, we established 2 parallel fractions for each of the 2 zooplankton communities collected from Hamilton Harbor, Ontario and Nanaimo Harbor, British Columbia, Canada (Fig. 1). In general, we prepared an equal amount (100 mg) of 2 parallel zooplankton samples derived from each of the 2 harbors. For each parallel fraction, we performed DNA extraction, PCR, and pyrosequencing (1/2 PicoTiter plate for each parallel sample) using the same protocols as those for spiking of known indicator species mentioned above. After pyrosequencing, raw reads were denoised by Mothur v.1.31.2 (Schloss et al. 2009) using default settings implemented in the pipeline Seed v.1.1.35 (Větrovský & Baldrian 2013). Subsequently, we used the RDP pyrosequencing pipeline (<http://rdp.cme.msu.edu/>) to remove low-quality sequences that: (1) contained any mismatch for the forward primer; (2) contained any undetermined nucleotide; (3) were too short (i.e. <250 bp); (4) contained homopolymers greater than 8; or (5) had Phred scores ( $Q$ ) lower than 20. The processed sequences from the 2 fractions of each community were clustered into similarity-based operational taxonomic units (OTUs) at a commonly used similarity cut-off value of 97% (e.g. Kunin et al. 2010) using the CD-HIT method (Li & Godzik 2006) implemented in the pipeline CLOTU (Kumar et al. 2011). PCR-mediated recombinants in PCR amplification products (i.e. chimeras) were automatically removed during data processing using CLOTU.

Because the number of sequences for each shared OTU in the 2 fractions varied due to different sequencing depth, we used the sequence percentage of each OTU for statistical analyses. As the data had an L-shaped distribution (i.e. non-normal distribution), we used the Wilcoxon signed-rank test, a non-parametric statistical method implemented in Statistica v.6, to test whether the sequence percentage for each

shared OTU differed statistically in the 2 parallel fractions. In order to avoid influence of random sampling for low-abundance OTUs (see detail in Zhan et al. 2014b), we removed singletons, doubletons, and tripletons (i.e. OTUs represented by 1, 2, and 3 sequences, respectively) from statistical analyses.

## RESULTS AND DISCUSSION

After 454 pyrosequencing, ~25 000 sequence reads were obtained after error/artefact removal for each of the 48 artificially assembled communities. Indicator species were recovered in 31 assembled communities, with 6, 11, 6, and 8 for the bay scallop, Japanese sea cucumber, golden mussel, and water louse, respectively (Table 1). All failed cases involved samples spiked with low quantities of indicator species, suggesting that the biomass of spiked indicator species was below the detection threshold (Zhan et al. 2013). When relative abundance of indicator species (i.e. biomass) was plotted against their corresponding abundance of sequence reads, we detected a consistent trend that low-abundance species assembled into communities usually corresponded to low-abundance sequence reads after pyrosequencing (Table 1, Fig. 2). However, several exceptions were observed in 3 species: sea cucumber, bay scallop, and water louse (Table 1, Fig. 2). For example, in the water louse, the percentage of sequence reads reached 2.3% in an assembled community with biomass percentage of 0.16%, an order of magnitude higher than those (0.13% and 0.16%) at similar biomass percentage (0.18%) in other communities (Table 1, Fig. 2). We detected a significant correlation between the 2 variables in only 1 species, the golden mussel (Spearman's correlation  $r = 0.88$ ,  $p = 0.02$ ), while non-significant correlations were recorded for the remaining 3 species ( $r = 0.46$ – $0.58$ ,  $p > 0.15$ ; Fig. 2). These results suggest that sequencing data could not reliably quantify relative species abundance in zooplankton communities.

The findings here are consistent with those results in studies of microbial communities (Engelbrektson et al. 2010, Zhou et al. 2011). For example, Zhou et al. (2011) used control DNA spiked into PCRs to test quantitation of amplicon sequencing-based detection of soil microorganisms. Substantial variation was detected in their control experiments, even after efforts were taken to minimize detection biases, such as using PCR products from multiple amplifications and fewer PCR cycles. Using mock communities, Amend et al. (2010) found sequence read abundance

Table 1. Relative abundance of indicator species (i.e. biomass) spiked into existing zooplankton communities and their corresponding abundance of sequence reads after 454 pyrosequencing of assembled communities. The indicator species were successfully recovered in 31 out of 48 assembled communities. '-' indicates that spiked indicator species were not recovered mainly owing to the low biomass

Species	Relative biomass (%)	Relative abundance of sequence reads (%)
Bay scallop	$1.80 \times 10^{-6}$	-
<i>Argopecten</i>	$2.43 \times 10^{-6}$	-
<i>irradians</i>	$2.43 \times 10^{-6}$	-
	$2.73 \times 10^{-5}$	-
	$2.31 \times 10^{-5}$	$4.4 \times 10^{-3}$
	$2.00 \times 10^{-5}$	-
	$2.62 \times 10^{-4}$	$1.2 \times 10^{-3}$
	$2.33 \times 10^{-4}$	$3.6 \times 10^{-3}$
	$1.89 \times 10^{-4}$	-
	$1.69 \times 10^{-3}$	$1.2 \times 10^{-2}$
	$1.52 \times 10^{-3}$	$1.4 \times 10^{-2}$
	$1.51 \times 10^{-3}$	$1.9 \times 10^{-2}$
Sea cucumber	$6.12 \times 10^{-4}$	$9.99 \times 10^{-3}$
<i>Apostichopus</i>	$5.33 \times 10^{-4}$	$6.04 \times 10^{-3}$
<i>japonicus</i>	$5.13 \times 10^{-4}$	-
	$4.59 \times 10^{-3}$	0.23
	$5.11 \times 10^{-3}$	$4.70 \times 10^{-2}$
	$4.79 \times 10^{-3}$	0.22
	$4.82 \times 10^{-2}$	$2.50 \times 10^{-2}$
	$5.41 \times 10^{-2}$	$1.32 \times 10^{-2}$
	$1.31 \times 10^{-2}$	4.44
	0.23	0.23
	0.25	0.11
	0.24	0.49
Golden mussel	$4.00 \times 10^{-6}$	-
<i>Limnoperna</i>	$3.81 \times 10^{-6}$	-
<i>fortunei</i>	$3.74 \times 10^{-6}$	-
	$4.12 \times 10^{-5}$	-
	$4.30 \times 10^{-5}$	-
	$4.17 \times 10^{-5}$	-
	$4.71 \times 10^{-4}$	$1.02 \times 10^{-2}$
	$5.88 \times 10^{-4}$	$9.14 \times 10^{-3}$
	$5.88 \times 10^{-4}$	$3.26 \times 10^{-2}$
	$2.43 \times 10^{-3}$	0.18
	$2.59 \times 10^{-3}$	0.18
	$2.41 \times 10^{-3}$	0.14
Water louse	$1.42 \times 10^{-3}$	$6.42 \times 10^{-3}$
<i>Asellus</i>	$1.67 \times 10^{-3}$	-
<i>aquaticus</i>	$1.70 \times 10^{-3}$	-
	$1.79 \times 10^{-2}$	-
	$1.46 \times 10^{-2}$	-
	$2.11 \times 10^{-2}$	$3.75 \times 10^{-3}$
	$6.72 \times 10^{-2}$	1.89
	$9.09 \times 10^{-2}$	0.15
	$9.01 \times 10^{-2}$	0.11
	0.18	0.13
	0.18	0.16
	0.16	2.30

was approximately quantitative within species, though sequence abundances had 1 order of magnitude differences among species. However, we detected 1 order of magnitude differences in sequence read abundance in some cases when the same species were spiked into different existing zooplankton communities (Table 1, Fig. 2). Further statistical analysis showed no significant correlations between relative species abundance and sequence abundance for the same species (Table 1, Fig. 2). The observed patterns may be due to experimental biases such as DNA extraction, PCR, and sequencing, as well as different species composition (complexities) of zooplankton communities (see Amend et al. 2010, Ó Cuív et al. 2011, Zhan & MacIsaac 2014 and references therein). Collectively, these results suggest that these biases can lead to substantial variation when estimating species abundance in comparative studies.

To further illustrate how these biases can affect comparative studies, we analyzed parallel replicates of the 2 natural zooplankton communities. A total of 686 064 and 721 931 sequences were obtained for the 2 fractions for Hamilton, while 406 215 and 383 190 sequences were obtained for Nanaimo. After filtered and denoised sequences were clustered into similarity-based OTUs at a commonly used similarity cut-off value of 97%, 353 and 244 OTUs, and 566 and 592 OTUs were recovered for the 2 parallel fractions for Hamilton and Nanaimo, respectively. A total of 84 and 126 shared OTUs remained for Hamilton and Nanaimo after removing singletons, doubletons, and tripletons. The Wilcoxon signed-rank test revealed a significant difference ( $p = 0.0012$  and  $p < 0.0001$  for Hamilton and Nanaimo, respectively) in the sequence percentage for each shared OTU in both datasets. Such differences can be visually reflected by using the natural log (ln)-transformed ratio between sequence percentage of each OTU in Fraction I and Fraction II (Fig. 3), as this ln-transformed ratio is expected to be 0 if the sequence percentage for each OTU matches perfectly in the 2 fractions (i.e. 1:1 between 2 fractions). The percentage of sequence reads for some OTUs in one fraction was more than 8-fold higher than that in the other (Fig. 3).

The analyses presented here illustrate that the distribution of sequence reads varies greatly, even between parallel replicates of the same zooplankton samples. Although studies suggest that the variation of rDNA gene copy number may contribute to poor quantitation of species abundance when using HTS (e.g. Weber & Pawlowski 2013), this explanation did

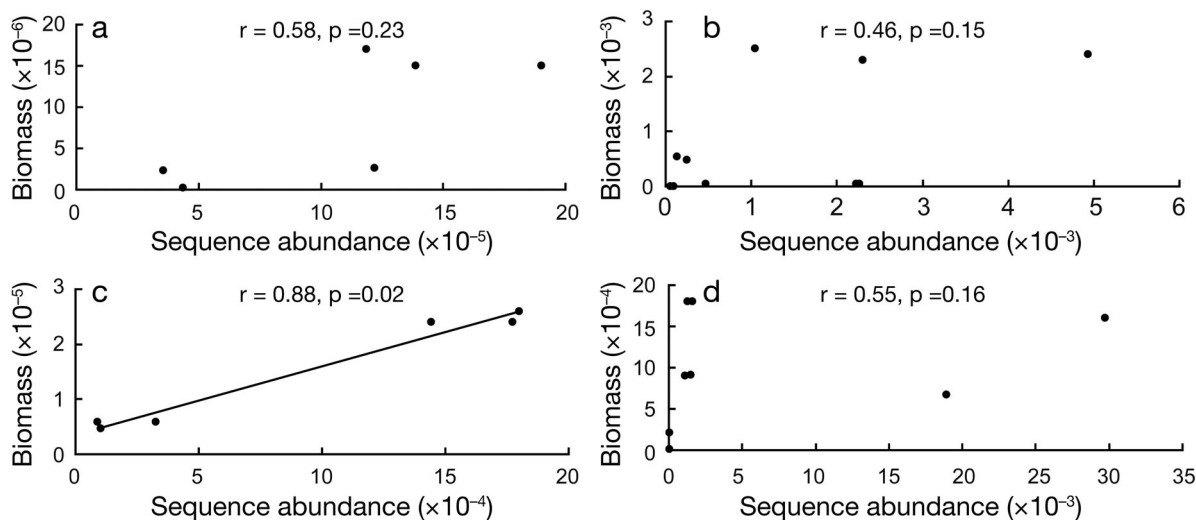


Fig. 2. Relative abundance (i.e. biomass) of the 4 indicator species (a: bay scallop *Argopecten irradians*, b: Japanese sea cucumber *Apostichopus japonicus*, c: golden mussel *Limnoperna fortunei*, d: water louse *Asellus aquaticus*) plotted against their corresponding abundance of sequence reads in 454 pyrosequencing data

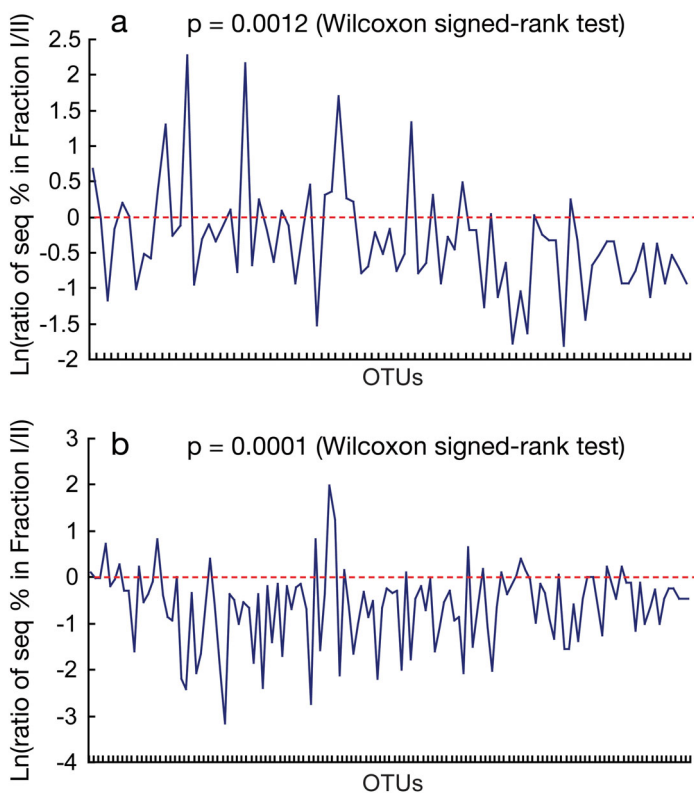


Fig. 3. Distribution of sequence reads for each shared operational taxonomic unit (OTU) in the 2 parallel zooplankton fractions for the 2 harbors (a: Hamilton, b: Nanaimo) as shown by the natural log ( $\ln$ )-transformed ratios between sequence percentage of each OTU in Fraction I and Fraction II. The  $\ln$ -transformed ratio is expected to be 0 (dotted line) if the distribution of sequence reads for each OTU matches perfectly in the 2 fractions (i.e. 1:1 in the 2 parallel fractions)

not apply to patterns observed within species or even within individuals here, as we spiked the lysed solution derived from the same individuals into different communities. The analyses of parallel replicates of the same communities suggest that experimental biases such as DNA extraction, PCR, and sequencing mainly lead to poor quantitation of species abundance (Amend et al. 2010, Ó Cuív et al. 2011). Such experimental biases result in varied distributions of sequence reads, as shown by both methods in this study (Table 1, Figs. 2 & 3). Interestingly, the ratios of sequence reads for relatively abundant species were also highly varied, as shown by establishing parallels for the same communities (Fig. 3). These biases, together with species composition (complexity) of communities, can potentially lead to biased, and sometimes even incorrect, conclusions in comparative studies when using HTS-based methods for characterization of zooplankton communities. Hence, we call for caution when using HTS-based approaches to make quantitative inferences in comparative studies.

**Acknowledgements.** This work was supported by the One-Three-Five Program (YSW2013B02) of the Research Center for Eco-Environmental Sciences and 100-Talent Program of the Chinese Academy of Sciences to A.Z.; by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), and Canada Research Chair to H.J.M.; by the grant of Comprehensive Evaluation Methods for Quality of Marine Environments in South China Sea (DOMEP(MEA)-01-03); and by the NSERC Canadian Aquatic Invasive Species Network (CAISN).



## LITERATURE CITED

- Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: Does read abundance count? *Mol Ecol* 19:5555–5565
- Creer S (2010) Second-generation sequencing derived insights into the temporal biodiversity dynamics of freshwater protists. *Mol Ecol* 19:2829–2831
- Dannemiller KC, Lang-Yona N, Yamamoto N, Rudich Y, Peccia J (2014) Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations. *Atmos Environ* 84:113–121
- Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* 4:642–647
- Fonseca VG, Carvalho GR, Sung W, Johnson HF and others (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat Commun* 1:98
- Hirai J, Kuriyama M, Ichikawa T, Hidaka K, Tsuda A (2015) A metagenetic approach for revealing community structure of marine planktonic copepods. *Mol Ecol Resour* 15:68–80
- Kumar S, Carlsen T, Mevik B, Enger P, Balaalid R, Shalchian-Tabrizi K, Kauserud H (2011) CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics* 12:182
- Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* 12:118–123
- Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659
- Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next generation sequencing reveals the hidden diversity of zooplankton assemblages. *PLoS ONE* 8:e81327
- Mackie JA, Geller J (2010) Experimental parameters affecting quantitative PCR of *Artemia franciscana*: a model for a marine zooplanktonic target in natural plankton samples. *Limnol Oceanogr Methods* 8:337–347
- Nigam R, Saraswat R, Panchang R (2006) Application of foraminifers in ecotoxicology: retrospect, prospect and prospect. *Environ Int* 32:273–283
- Ó Cuív P, de Cárcer DA, Jones M, Klaassens ES and others (2011) The effects from DNA extraction methods on the evaluation of microbial diversity associated with human colonic tissue. *Microb Ecol* 61:353–362
- Schloss PD, Westcott SL, Ryabin T, Hall JR and others (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541
- Shi XL, Lepère C, Scanlan DJ, Vaultot D (2011) Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLoS ONE* 6:e18979
- Větrovský T, Baldrian P (2013) Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. *Biol Fertil Soils* 49:1027–1037
- Weber AAT, Pawlowski J (2013) Can abundance of protists be inferred from sequence data: a case study of foraminifera. *PLoS ONE* 8:e56739
- Zhan A, MacIsaac HJ (2014) Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. *Conserv Genet*, doi:10.1007/s10592-014-0678-9
- Zhan A, Bao Z, Hu X, Lu W and others (2008) Accurate methods of DNA extraction and PCR-based genotyping for single scallop embryos/larvae long-preserved in ethanol. *Mol Ecol Resour* 8:790–795
- Zhan A, Hulák M, Sylvester F, Huang X and others (2013) High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods Ecol Evol* 4:558–565
- Zhan A, Bailey SA, Heath DD, MacIsaac HJ (2014a) Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Mol Ecol Resour* 14:1049–1059
- Zhan A, He S, Brown EA, Chain FJJ and others (2014b) Reproducibility of pyrosequencing data for biodiversity assessment in complex communities. *Methods Ecol Evol* 5:881–890
- Zhou J, Wu L, Deng Y, Zhi X and others (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *ISME J* 5:1303–1313

Editorial responsibility: Nikolaos Schizas, Mayagüez, Puerto Rico, USA

Submitted: October 27, 2014; Accepted: February 24, 2015  
Proofs received from author(s): April 21, 2015