



# Integrating Stochastic and Deterministic Process in the Biogeography of N<sub>2</sub>-Fixing Cyanobacterium *Candidatus Atelocyanobacterium Thalassa*

Liuyang Li<sup>1,2</sup>, Chao Wu<sup>3</sup>, Danyue Huang<sup>1,4</sup>, Changling Ding<sup>3,5</sup>, Yuqiu Wei<sup>6</sup> and Jun Sun<sup>1\*</sup>

<sup>1</sup> College of Marine Science and Technology, China University of Geosciences, Wuhan, China, <sup>2</sup> State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, <sup>3</sup> Research Centre for Indian Ocean Ecosystem, Tianjin University of Science and Technology, Tianjin, China, <sup>4</sup> School of Oceanography, Shanghai Jiao Tong University, Shanghai, China, <sup>5</sup> College of Biotechnology, Tianjin University of Science and Technology, Tianjin, China, <sup>6</sup> Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China

## OPEN ACCESS

### Edited by:

Julie LaRoche,  
Dalhousie University, Canada

### Reviewed by:

Peng Xing,  
Nanjing Institute of Geography  
and Limnology Chinese Academy  
of Sciences (CAS), China  
Phoebe Dreux Chappell,  
Old Dominion University,  
United States

### \*Correspondence:

Jun Sun  
phytoplankton@163.com

### Specialty section:

This article was submitted to  
Aquatic Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 17 January 2021

**Accepted:** 22 September 2021

**Published:** 21 October 2021

### Citation:

Li L, Wu C, Huang D, Ding C,  
Wei Y and Sun J (2021) Integrating  
Stochastic and Deterministic Process  
in the Biogeography of N<sub>2</sub>-Fixing  
Cyanobacterium *Candidatus*  
*Atelocyanobacterium Thalassa*.  
*Front. Microbiol.* 12:654646.  
doi: 10.3389/fmicb.2021.654646

UCYN-A is one of the most widespread and important marine diazotrophs. Its unusual distribution in both cold/warm and coastal/oceanic waters challenges current understanding about what drives the biogeography of diazotrophs. This study assessed the community assembly processes of the nitrogen-fixing cyanobacterium UCYN-A, developing a framework of assembly processes underpinning the microbial biogeography and diversity. High-throughput sequencing and a qPCR approach targeting the *nifH* gene were used to investigate three tropical seas: the Bay of Bengal, the Western Pacific Ocean, and the South China Sea. Based on the neutral community model and two types of null models calculating the  $\beta$ -nearest taxon index and the normalized stochasticity ratio, we found that stochastic assembly processes could explain 66–92% of the community assembly; thus, they exert overwhelming influence on UCYN-A biogeography and diversity. Among the deterministic processes, temperature and coastal/oceanic position appeared to be the principal environmental factors driving UCYN-A diversity. In addition, a close linkage between assembly processes and UCYN-A abundance/diversity/drivers can provide clues for the unusual global distribution of UCYN-A.

**Keywords:** cyanobacteria, *nifH*, UCYN-A, community assembly, stochasticity, determinism, temperature

## INTRODUCTION

Diazotrophs play a fundamental role in biogeochemical processes by providing bioavailable nitrogen (Sohm et al., 2011; Karl et al., 2012), which determines the primary production in the ocean (Postgate, 1982; Gruber and Sarmiento, 1997; Karl et al., 1997). UCYN-A (*Candidatus Atelocyanobacterium thalassa*) is considered to be the most abundant unicellular diazotroph in the

global ocean (Moisander et al., 2010; Luo et al., 2012; Martinez-Perez et al., 2016; Turk-Kubo et al., 2018; Zehr and Capone, 2020). It has been shown that UCYN-A exchanges fixed nitrogen for fixed carbon from its haptophyte host (Thompson et al., 2012; Zehr, 2015; Martinez-Perez et al., 2016). The combination of an obligate symbiosis with haptophyte algae (Thompson et al., 2012; Krupke et al., 2014, 2015; Zehr, 2015; Cabello et al., 2016; Cornejo-Castillo et al., 2016; Martinez-Perez et al., 2016; Zehr et al., 2016; Munoz-Marin et al., 2019; Zehr and Capone, 2020; Farnelid et al., 2021) and cosmopolitan distribution (Moisander et al., 2010; Farnelid et al., 2016; Harding et al., 2018; Shiozaki et al., 2018; Cabello et al., 2020; Tang et al., 2020) makes UCYN-A unusual in the world's ocean. A recent study has suggested that  $N_2$  fixation, instead of nitrate ( $NO_3^-$ ) or ammonium ( $NH_4^+$ ) utilization, is important for the UCYN-A/haptophyte symbiosis to meet their nitrogen demands under conditions of replete dissolved inorganic nitrogen (DIN) (Mills et al., 2020). These findings expanded our knowledge of the spatial distribution and scale of habitat preference of UCYN-A, which were different from the well-known cyanobacterium *Trichodesmium* and the diatom symbiont *Richelia* that prefer warm (Breitbarth et al., 2007) and oligotrophic seawater (Tyrrell, 1999). Although UCYN-A was originally thought to be of low genetic diversity (Tripp et al., 2010), it is currently known to have formed at least four sublineages (e.g., UCYN-A1, UCYN-A2, UCYN-A3, and UCYN-A4) (Turk-Kubo et al., 2017; Henke et al., 2018), according to the nitrogenase (*nifH*) phylogeny of nucleotide sequences (Thompson et al., 2014; Farnelid et al., 2016). Additionally, different sublineages were inferred as being potentially distinct ecotypes (Turk-Kubo et al., 2017). The question of whether symbionts with such self-sufficiency and ability to survive over a wide range of temperatures were less susceptible to niche-based selection, or instead they formed versatile ecotypes to adapt to various environments owing to niche differentiation, leads us to reconsider how the biogeography and diversity of UCYN-A are assembled.

Community assembly processes are divided into two categories, namely, deterministic (niche-based) and stochastic (neutral-based) processes (Stegen et al., 2012; Wang et al., 2013; Dini-Andreote et al., 2015; Tripathi et al., 2018). Niche theory assumes that microbial communities are controlled by deterministic processes consisting of abiotic and biotic screening, owing to habitat preferences and species fitness (Fargione et al., 2003; Bahram et al., 2016). On the contrary, neutral theory supposes that ecologically equivalent members in microbial communities are shaped by stochastic processes including random birth, death, and dispersal events (Chave, 2004; Bahram et al., 2016; Zhou and Ning, 2017; Tripathi et al., 2018). These ecological processes can drive community assembly simultaneously (Fargione et al., 2003; Langenheder and Szekely, 2011; Zhou and Ning, 2017; Jiao and Lu, 2020). Furthermore, some factors are important in mediating the balance between stochastic and deterministic assembly of microbial communities, such as the pH and temperature of the soil environment (Tripathi et al., 2018; Jiao and Lu, 2020). However, it is still challenging to uncover the community assembly process and the factors

mediating community determinism/stochasticity in various ecosystems (e.g., marine habitats).

Despite the increasing data available from next-generation sequencing studies of nitrogenase genes (Turk-Kubo et al., 2017; Henke et al., 2018; Cornejo-Castillo et al., 2019; Farnelid et al., 2019; Moreira-Coello et al., 2019; Cabello et al., 2020; Gradoville et al., 2020), the biogeography and diversity of UCYN-A have been analyzed mainly from a deterministic perspective, and the stochasticity in UCYN-A community assembly is still unclear. Accordingly, our limited understanding of the underlying mechanism in biogeographic variations of UCYN-A constrains our ability to incorporate them into ecological models. The Bay of Bengal (BOB) has received far less attention in research on UCYN-A than has the South China Sea (SCS) and the Western Pacific Ocean (WPO). As a crossroads of the BOB and WPO, the SCS is undoubtedly influenced by neighboring physical processes, such as ocean currents (e.g., the Kuroshio Current) (Shiozaki et al., 2013), and other physical processes including upwellings (Voss et al., 2006; Zhang et al., 2015), mesoscale eddies (Li et al., 2021a), or river plumes (Zhang et al., 2014). In particular, high temperature, stratification, and oligotrophic seawater make these waters an ideal habitat for supporting the growth of diazotrophs (Kong et al., 2011). On the other hand, UCYN-A has also been observed active in N-replete conditions (Mills et al., 2020; Tang et al., 2020). To date, however, the biogeography of UCYN-A remains patchy (Turk-Kubo et al., 2017), and there is a lack of comprehensive knowledge on the assembly mechanism of microbial communities throughout the heterogenous BOB, WPO, and SCS.

Here, we performed ecological surveys in the BOB, WPO, and SCS using *nifH* gene-based high-throughput sequencing and qPCR. Our aims were to (I) develop a comprehensive picture of large-scale variability of UCYN-A in the tropical seas; (II) uncover the relative contribution of stochastic and deterministic processes of the UCYN-A community in tropical seas; and (III) reveal the drivers mediating the assembly processes. We hope that this study will provide a novel context for the spatial heterogeneity of UCYN-A in the BOB, WPO, and SCS, as well as providing some inspirations for researchers to explore the mechanisms underlying the unusual distribution of UCYN-A.

## MATERIALS AND METHODS

### Sample Collection and Analysis of Environmental Parameters

Surface seawater samples (5 m) were collected from the BOB cruise (onboard R/V “*Dongfanghong 2*” from November to December 2016; latitude, 5–17°N), WPO cruise (onboard R/V “*Kexue*” from October to November 2017, 2–18°N), western South China Sea (wSCS), and central South China Sea (cSCS) cruises (onboard R/V “*Shiyan 3*” from September to October 2016 and “*Shiyan 1*” from March to May 2016, respectively; 10–22°N). Three samples from the cSCS cruise were strongly affected by the Pearl River plume, and they were grouped into the northern SCS region (nSCS) in subsequent analyses. Seawater samples were collected using Polyethylene (PE) buckets

that were previously rinsed with 10% HCl and Milli-Q water. To obtain the material for *nifH* gene-based high-throughput sequencing and qPCR analysis, 0.5–4 L of seawater were filtered through a filter of 0.22- $\mu\text{m}$  pore size (47 mm in diameter, Millipore, Eschborn, Germany) with low pressure ( $> -0.03$  Mp). The filters were flash-frozen in liquid nitrogen on board, transported to the laboratory, and stored at  $-80^\circ\text{C}$  until DNA extraction.

Sea surface temperature and salinity were determined and recorded using a conductivity–temperature–depth (CTD) system on board (Seabird SBE 911Plus, Sea-Bird Electronics, WA, United States). Sample analysis for other physicochemical parameters, such as inorganic nutrients and Chlorophyll *a* (Chl *a*), were consistent with our previous work (Wu et al., 2019; Li et al., 2021b). Briefly, we used 100-ml PE bottles to collect seawater for nutrient analysis. The nutrients, including nitrate ( $\text{NO}_3^-$ ) or nitrite ( $\text{NO}_2^-$ ), ammonium ( $\text{NH}_4^+$ ), silicate ( $\text{SiO}_3^{2-}$ ), and phosphate ( $\text{PO}_4^{3-}$ ), were measured using the copper–cadmium column reduction methods, indophenol blue method, silico-molybdate complex methods, and phosphor-molybdate complex methods, respectively (Pai et al., 2001; Guo et al., 2014). Furthermore, we used a 0.7- $\mu\text{m}$  pore size GF/F filter (25 mm diameter, Waterman, Florham Park, NJ, United States) to collect subsamples for Chl *a* analysis under a pressure higher than  $-10$  mmHg. The filters were subsequently stored in aluminum foil paper bags at  $-20^\circ\text{C}$ . Chl *a* was extracted with 90% acetone and measured using a Turner-Designs Trilogy<sup>®</sup> fluorometer (San Jose, CA, United States) in the laboratory (Welschmeyer, 1994; Liu et al., 2015). We used the distance from the sea surface to the bottom (DSB) and salinity as indirect indicators associated with coastal/oceanic habitats.

## DNA Extraction and Sequencing

DNA were extracted and purified using DNeasy PowerWater<sup>®</sup> Kit (Qiagen, Hilden, Germany) as the manufacturer's protocol described. We used an ND-2000 Nanodrop spectrometer (Thermal Scientific, Wilmington, DE, United States) to check the quantity and quality of genomic DNA. In the nested polymerase chain reaction (nested PCR), universal *nifH* primers *nifH3/nifH4* (Zehr and Turner, 2001) and UCYN-A-specific *nifH* primers with 5' common sequence linkers (Turk-Kubo et al., 2017) were used, respectively, in the first amplification and second amplification (**Supplementary Table 1**). PCRs were conducted in quintuplicate with a Veriti 9902 thermocycler (Applied Biosystems, Foster City, CA, United States). The reaction system, reaction conditions, and thermocycling parameters for nested PCR analysis are described in **Supplementary Tables 2, 3** in the Supporting Information. The libraries were constructed and sequenced using paired-end chemistry (PE300) on the Illumina MiseqPE300 platform (Illumina, San Diego, CA, United States).

## Quality Control and Sequencing Data Processing

The raw sequencing reads were demultiplexed and quality filtered by their barcode sequences under the limit of no more

than one mismatch. The open source bioinformatics pipeline QIIME was applied to preliminarily analyze and process the *nifH* gene sequences (Caporaso et al., 2010). The paired-ended sequences were merged *via* FLASH (v1.2.7) software (Magoc and Salzberg, 2011), and the generated raw tags were quality filtered using Trimmomatic (v0.33) software (Bolger et al., 2014). After undertaking the above processes, we obtained high-quality clean tags without the barcodes, linker sequences, and primers. Chimeras were also removed to get effective tags by using UCHIME (v4.2) software (Edgar et al., 2011). Furthermore, singletons were eliminated to minimize sequencing errors. We set a 99% similarity level to cluster the remaining tags to different operational taxonomic units (OTUs) (Cornejo-Castillo et al., 2019) using software USEARCH (v10.0) (Edgar, 2010). After normalization (1,395,208 tags of all samples), 10,640 tags for each sample were finally generated. The ultimate total dataset retained 166 OTUs at 99% similarity level. For taxonomically classified OTUs of UCYN-A sublineages, representative sequences and alignment sequences were aligned with ClustalW in MEGA v7.0. A maximum likelihood (ML) tree based on the Tamura-Nei model was then constructed. The node support of the ML tree was determined with 1,000 bootstrap replicates, and bootstrap values greater than 50% were shown.

## Determining UCYN-A Abundances Using Quantitative Polymerase Chain Reaction

The *nifH* genes for UCYN-A1 (Church et al., 2005) and UCYN-A2 (Thompson et al., 2014) were quantified using qPCR assays with TaqMan<sup>®</sup> chemistry. It is notable that the UCYN-A2 assay is now known to also amplify the UCYN-A2/A3/A4 sublineages due to cross-reactivity between assays (Farnelid et al., 2016). Primer and probe sequences for the UCYN-A1 *nifH* gene assay are as follows: Forward primer (5'–3') AGCTATAACAACGTTTTATGCGTTGA; Reverse primer (5'–3') ACCACGACCAGCACATCCA; Probe, 5'-FAM-TCTGGTGGTCCCTGAGCCTGGA-TAMRA-3'. For the UCYN-A2/A3/A4 *nifH* assay, the same reverse primer is used, and the forward primer and probe sequences are as follows: Forward primer (5'–3') GGTTACAACAACGTTTTATGTGTTGA; Probe, 5'-FAM-TCTGGTGGTCCCTGAGCCCGGA-TAMRA-3'. Preparation of standards, reaction conditions, and thermocycling parameters are described in Goebel et al. (2010), but it is important that primer probe specificity is ensured by setting different annealing temperatures of  $60^\circ\text{C}$  (UCYN-A1) and  $64^\circ\text{C}$  (UCYN-A2/A3/A4). The  $r^2$  values of the standard curves ranged from 0.98 to 1.00, and PCR efficiency ranged from 96 to 110%. All samples, standards, and inhibition tests were 10- $\mu\text{l}$  qPCR reactions in duplicate or triplicate, and non-target templates were also tested under the same conditions. Based on the volumes of seawater filtered, as well as the DNA extract and DNA template volumes in the qPCR reactions, the limit of detection (LOD) and limit of quantitation (LOQ) in this study ranged between 25 and 154 *nifH* copies  $\text{L}^{-1}$ , and between 200 and 1231 *nifH* copies  $\text{L}^{-1}$ , respectively (Goebel et al., 2010). Samples with abundances greater than



the LOD but less than the LOQ were labeled as “detected not quantified” (DNQ).

## Statistical Analysis

All statistical analyses were performed in R (v3.6.1<sup>1</sup>), using the “vegan” (Oksanen et al., 2015), “stats” (Field et al., 2012), “picante” (Kembel et al., 2010), “NST” (Ning et al., 2019), “hmisc” (Harrell and Dupont, 2013), “igraph” (Csardi and Nepusz, 2006), “gbmpls” (De’ath, 2007), and “ggplot2” (Wickham, 2016) packages.

## Spatial Variation Based on UCYN-A Community Diversity Indices

Alpha-diversity indices including richness, Chao1 richness, Ace richness, Pielou’s diversity, Shannon diversity, and the Simpson diversity index, were calculated at the OTU level using the “vegan” package (Oksanen et al., 2015). Kruskal–Wallis tests (Field et al., 2012) were used to compare the spatial differences of alpha-diversity indices. To interpret the spatial distribution patterns of UCYN-A communities and the differences among samples, non-metric multidimensional scaling (NMDS) was performed to represent UCYN-A community characters according to Bray–Curtis dissimilarities of the sequencing data. Differences in UCYN-A communities between region-based groups were investigated using permutational multivariate analysis of variance (PERMANOVA, permutations = 9999).

## The Assembly Processes of UCYN-A Communities

To evaluate the relative importance of deterministic or stochastic process on UCYN-A communities, we performed various statistical methods, including the neutral community model (NCM) (Sloan et al., 2006), and two types of null models from Stegen et al. (2013) and Ning et al. (2019) using the *nifH* amplicon dataset.

The NCM was applied to assess the contribution of stochastic processes to community assembly. This model predicts OTUs occurrence frequency in local communities (i.e., the UCYN-A community at each station) and their relative abundance across the wider metacommunities (i.e., the UCYN-A communities of total amplicon dataset) by the parameter  $m$ , which assesses the probability of a random loss being replaced by dispersal (Sloan et al., 2006; Burns et al., 2016). In this model, abundant taxa tend to be dispersed randomly and widespread across metacommunity, while rare taxa tend to disappear from local communities as a result of ecological drift. The parameter  $R^2$  reflects the overall fit of UCYN-A OTUs to NCM. The 95% confidence interval around the NCM prediction was used to identify which UCYN-A taxa were well predicted by this model (Burns et al., 2016).

In addition, a null modeling approach was applied using the framework as Stegen et al. (2013) described. Phylogenetic depths of environmental responses of the OTUs in the UCYN-A community were first estimated using Blomberg’s  $K$  via the

“multiPhyloSignal” function. Higher  $K$  values indicate a better correlation between species and phylogeny as expected by a Brownian motion-based metric of the strength of phylogenetic signal (Blomberg et al., 2003).  $P$  values for each  $K$  value were also calculated to test for significant phylogenetic signal, according to the variance of the phylogenetically independent comparisons deviated from a null model that randomly reshuffles trait values, based on 999 simulations. Subsequently, the pairwise phylogenetic turnover ( $\beta$ MNTD, quantifying the phylogenetic distance for each OTU in the community) was calculated using the “comdistnt” function in the “picante” package (Kembel et al., 2010).  $\beta$ -nearest taxon index ( $\beta$ NTI) is introduced to quantify the degree to which  $\beta$ MNTD deviates from the turnover expected, using a separate null model repeating the randomization 999 times (Stegen et al., 2012).  $\beta$ NTI measures the phylogenetic turnover given stochastic and deterministic ecological processes (Stegen et al., 2012; Wang et al., 2013; Dini-Andreote et al., 2015).  $|\beta$ NTI| > 2 indicates the dominance of niche-based processes, whereas  $|\beta$ NTI| < 2 indicates the influence of neutral processes. For the deterministic processes,  $\beta$ NTI > +2 indicates variable selection processes and  $\beta$ NTI < -2 indicates homogeneous selection processes. Bray–Curtis-based Raup–Crick ( $RC_{\text{Bray}}$ ) was further employed to quantify neutral processes when  $|\beta$ NTI| < 2 (Chase et al., 2011; Stegen et al., 2013).  $RC_{\text{Bray}} > +0.95$  indicates a dominant role for dispersal limitation, whereas  $RC_{\text{Bray}} < -0.95$  indicates a dominant process of homogenizing dispersal. The fractions of  $|\beta$ NTI| < 2 and  $|RC_{\text{Bray}}| < 0.95$  indicate “undominated” processes, including weak selection, weak dispersal, and diversification or drift processes (Zhou and Ning, 2017; Tripathi et al., 2018).

To further enhance our understanding of the assembly process in microbial communities, another null model was applied using the framework as Ning et al. (2019) reported. In this framework, deterministic factors can drive the communities to be more similar or dissimilar than the null expectation, resulting in two types of selection processes. An index, named the normalized stochasticity ratio (NST), was inferred to quantify the deterministic (<50%) and stochastic assembly (>50%) by setting 50% as the boundary point. Multiple null model algorithms and community similarity metrics were applied in the current study to obtain a more comprehensive assessment.

## The Drivers Mediating UCYN-A Assembly

To evaluate the importance of each environmental variable on the UCYN-A community, we performed the aggregated boosted tree (ABT) analysis (De’ath, 2007), the Mantel test, and the co-occurrence network (CON) analyses using the *nifH* amplicon dataset.

The ABT analysis was employed to quantify the effect of each environmental variable on the UCYN-A community composition using the “gbmpls” package with 500 trees for boosting (De’ath, 2007). The Mantel test was also used to clarify significant environmental traits influencing the UCYN-A community via the “vegan” package (Oksanen et al., 2015).

Subsequently, CON including the correlations of OTUs vs. OTUs and OTUs vs. environmental factors was constructed.

<sup>1</sup><http://www.r-project.org/>

A total of 99 OTUs (more than 10 sequences) and 12 detected variables were selected to calculate pairwise Spearman's rank correlations ( $r$ ) within the "hmisc" R package (Harrell and Dupont, 2013). The final network consisted of all robust ( $r > 0.5$  or  $r < -0.5$ ) and statistically significant ( $p < 0.05$ ) correlations. To obtain the pure species coexistence relationship in the UCYN-A communities, subnetwork analyses of OTUs belonging to total UCYN-A community or each sublineage were performed separately using the `induced_subgraph` function in the "igraph" package (Csardi and Nepusz, 2006). The node-level topological features, such as betweenness centrality, closeness centrality, degree, and eigenvector centrality, were calculated for all nodes. Betweenness centrality indicates the potential influence of a particular node on the connections of other nodes; the closeness centrality suggests the average distance of a node to any other node; degree represents the number of connections for a particular node with all neighboring nodes; and the eigenvector centrality assigns relative scores to all nodes in the network according to the concept that compared with low-scoring nodes, the connection of nodes with high scores contributes more to the score of a node in question (Greenblum et al., 2012; Berry and Widder, 2014). A core position of a node in the network is measured by high values of these topological features, while a peripheral position is reflected by low values of topological features (Ma et al., 2016; Jiao et al., 2017). Environmental variables that significantly correlated with OTUs could be seen as important niche-based factors. Networks were visualized based on Cytoscape 3.8.2.

## RESULTS

### Biogeographic Diversity of UCYN-A From the Amplicon Dataset

An amplicon dataset with 1,395,208 high-quality sequences was obtained. The flat rarefaction curve indicates that the sequencing depths were sufficient to represent the diversity of local UCYN-A communities (Supplementary Figure 1). OTUs with more than 10 sequences were selected to represent UCYN-A communities. Thirty-six OTUs representing 23.81% of all sequences were clustered as UCYN-A1, and 51 OTUs contributing 72.30% of the dataset were affiliated to UCYN-A2. In fact, several OTUs together accounted for 96.6% of all sequences recovered, clustering closely within UCYN-A1/A2/A3/A4 (Figure 1A).

The distribution patterns of distinct sublineages were obviously different in the tropical seas. UCYN-A1/A2 dominated the dataset, while high relative abundances of UCYN-A3/A4 were only observed at several stations, such as station cSCS3 (Figure 1B). OTU 1 and OTU 16 dominated UCYN-A2, accounting for 42.0 and 28.2% of the UCYN-A *nifH* amplicon dataset, respectively. The OTUs of UCYN-A2 had a high proportion of relative abundance (average 72.3%), possessing a central position in the ternary plot (Figure 1C). Furthermore, UCYN-A2 was particularly abundant (relative abundance > 95%) at some stations in the BOB and wSCS.

The phylogenetic tree contained seven OTUs (OTU 2, OTU 4, etc.) closely related with UCYN-A1 reference sequences

(bootstrap value 67%). OTU 2 and OTU 9 dominated UCYN-A1, accounting for 13.2 and 8.1% of all sequences, respectively. UCYN-A3 was widely distributed, and its relative abundance was 22% in a coastal station near Hainan Island (cSCS10) (Supplementary Figure 2). Meanwhile, at this station, the relative abundance of UCYN-A4 (6.7%) was about 20 times higher than that in other regions.

Distinct biogeographic patterns emerged when the distribution pattern of NMDS and diversity indices were divided by regions (Figure 2). PERMANOVA confirmed the statistical significance of the region-based group (permutations = 9,999,  $p < 0.05$ ). The difference of beta-diversity implied the spatial differentiation of UCYN-A. Alpha-diversity indices in the WPO and cSCS were significantly higher than those in the wSCS and BOB (Wilcoxon rank-sum test,  $p < 0.01$ ). Moreover, high-latitude areas presented higher alpha-diversity indices than did low-latitude areas (Supplementary Figure 3).

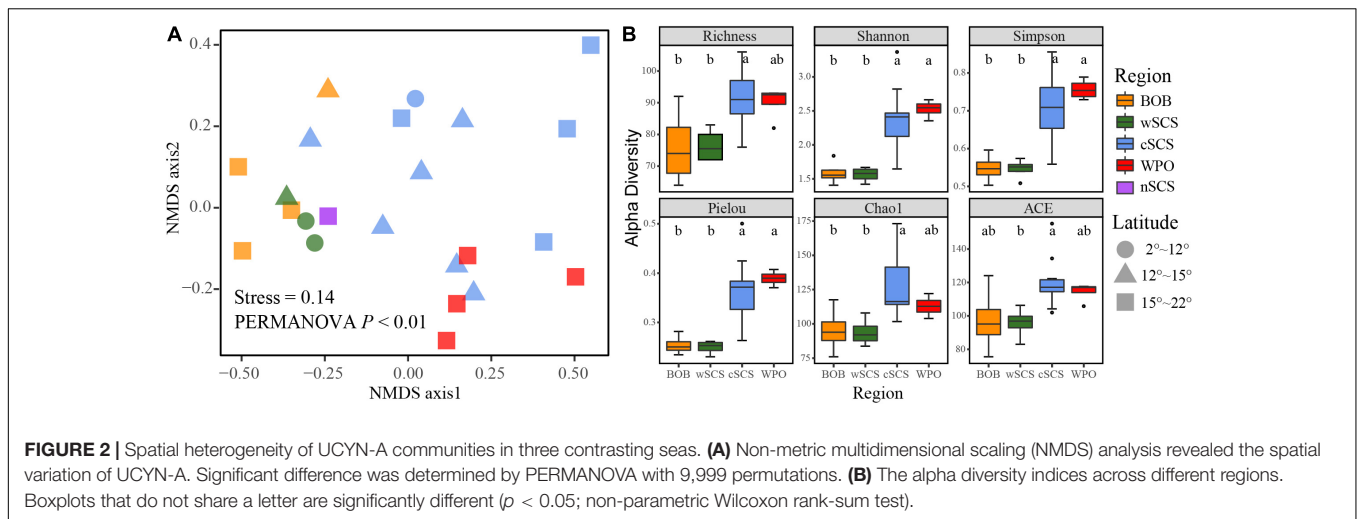
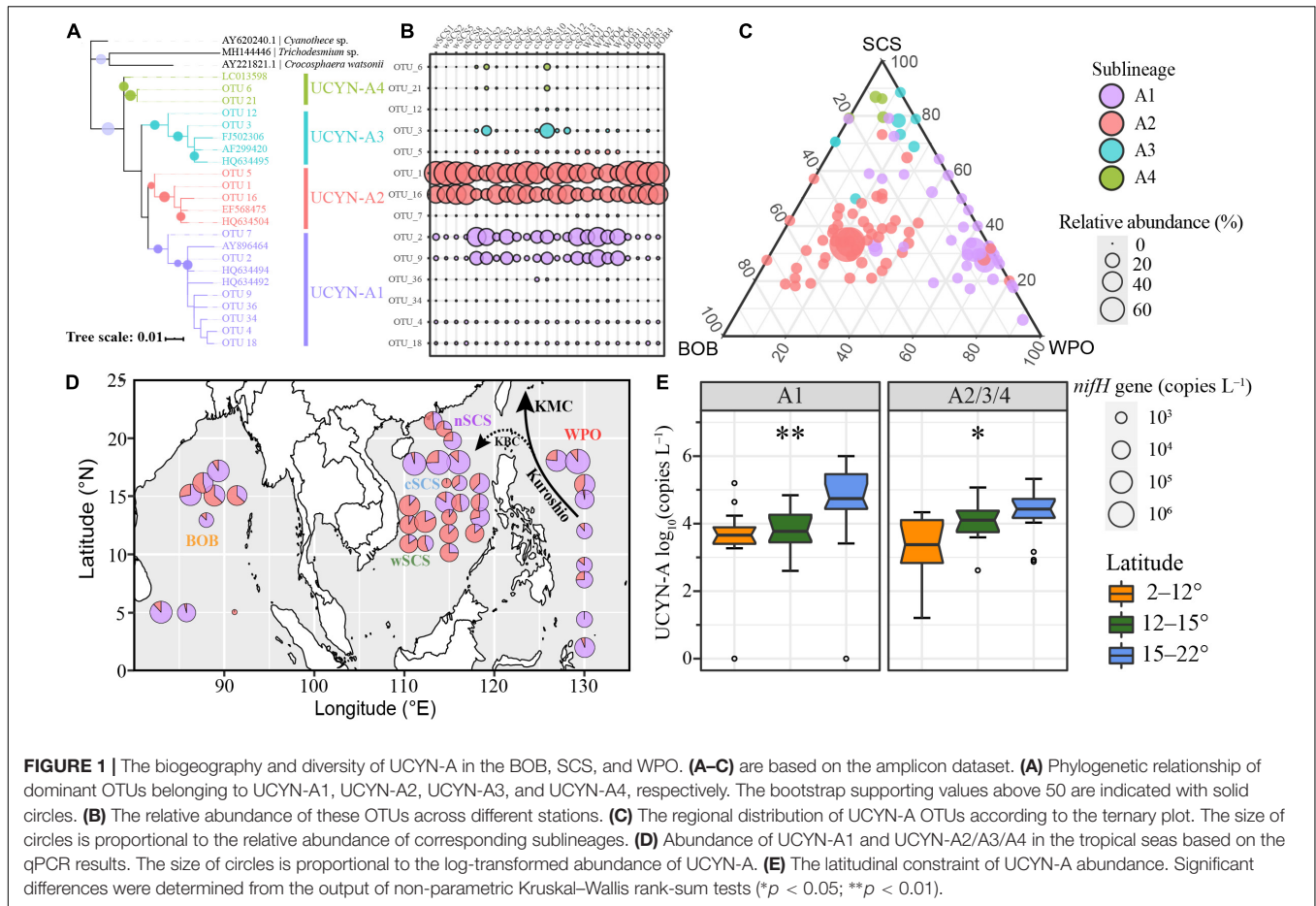
### Biogeographic Diversity of UCYN-A From qPCR

Significant variation in UCYN-A1/A2 abundances was observed among distinct latitudes rather than different regions (Figures 1D,E and Supplementary Figure 4). UCYN-A1 had higher mean qPCR-based abundances than did UCYN-A2/A3/A4 ( $8.4 \times 10^4$  *nifH* copies  $L^{-1}$ , including non-detects), with two samples below the LOD and one sample below the LOQ. The highest UCYN-A1 abundance, at  $9.98 \times 10^6$  *nifH* copies  $L^{-1}$ , occurred at station WPO-6. UCYN-A2/A3/A4 mean abundances were lower ( $2.5 \times 10^4$  *nifH* copies  $L^{-1}$ , including non-detects) than those of UCYN-A1, with only one sample below the LOD. The highest UCYN-A2/A3/A4 abundances, at  $2.1 \times 10^6$  *nifH* copies  $L^{-1}$ , occurred at station cSCS-2. Generally, UCYN-A abundances peaked in several stations close to 18°N, such as station WPO-6, and were lower toward the lower latitudes.

The ratio of UCYN-A1/A2 based on abundances and relative abundance shared a similar trend (Pearson's  $r = 0.82$ ) (Figure 1 and Supplementary Figure 2). Both methods indicated that the proportion of UCYN-A1 tended to be higher near the Kuroshio Current and lower after dispersal to remote areas or mixing with local waters. This biogeographic pattern suggested a potential stochastic effect on UCYN-A distribution and diversity.

### Quantifying Influences of Ecological Processes in Shaping Community Structure

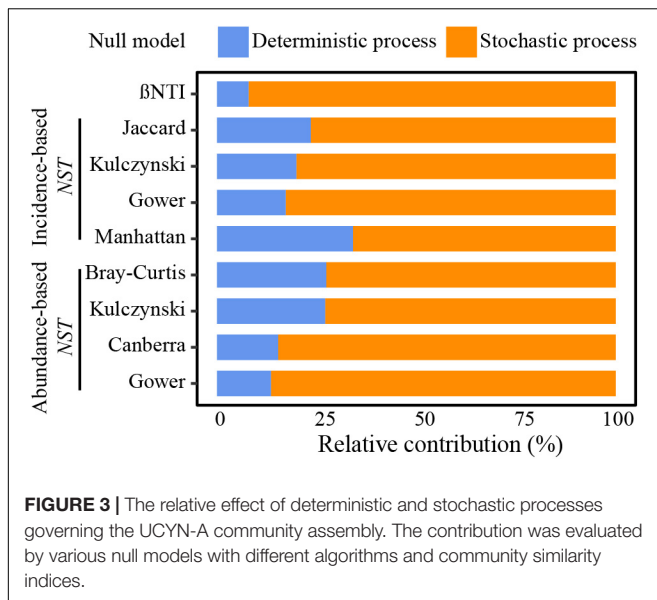
A large proportion of  $|\beta NTI|$  values were lower than 2 and all *NST* values from different algorithms and community similarity metrics were >50%, indicating that stochastic processes mainly governed community dynamics (Figure 3). The relative contribution of stochasticity was 79–91% for UCYN-A1 and 80–95% for UCYN-A2 (Supplementary Figure 5 and Supplementary Table 4). Additionally, there was a high proportion of homogenizing dispersal (49.8%) and weak-ecological processes (40.7%) in the overall community. Variable selection processes (7.9%) seemed to be more important than



dispersal limitation (1.6%) in UCYN-A communities, and a relatively high proportion of variable selection processes was estimated in UCYN-A2 (6.7%) compared with UCYN-A1 (5.5%). In addition, a relatively small proportion of homogeneous selection processes was calculated in UCYN-A2 (non-detected) than in UCYN-A1 (3.2%).

### The Fit of Neutral Processes

With 76.0% of community variance explained, the NCM successfully estimated a large proportion of the relationship between OTUs occurrence frequency and their variation in relative abundance (Figure 4A), indicating the importance of stochastic processes in shaping the community assembly.

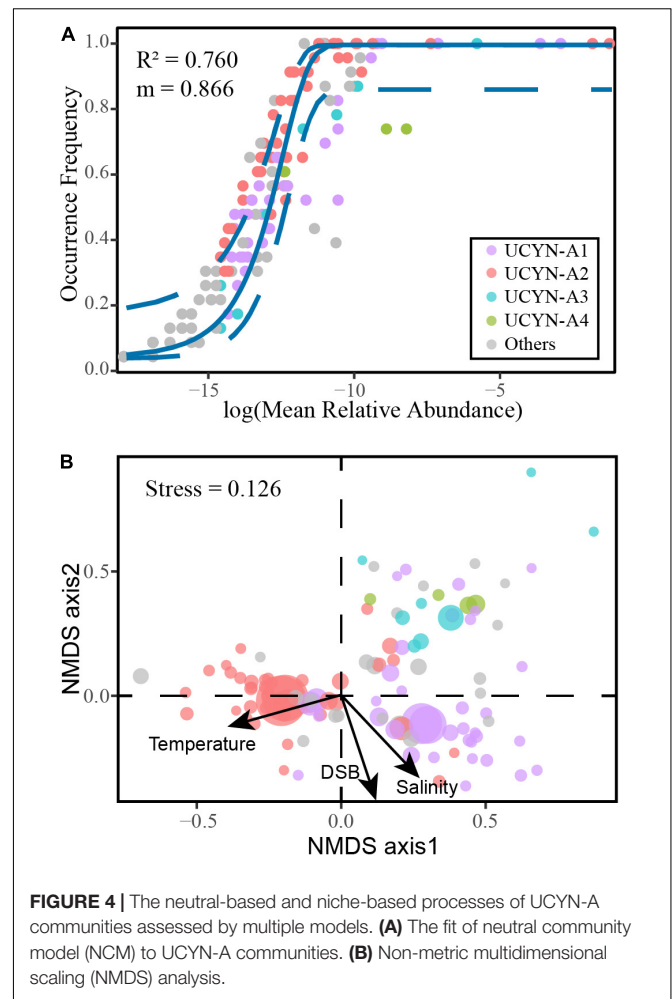


Moreover, neutral and non-neutral partitions are genetically and ecologically distinct (**Supplementary Table 5**). Interestingly, two OTUs below the NCM confidence interval accounted for 93.4% of UCYN-A4 relative abundance. The strong deviation of UCYN-A4 indicated that the assembly mechanism of UCYN-A4 was significantly different from other sublineages. Furthermore, the  $m$  value was estimated to be 0.866, suggesting strong species dispersal processes.

## The Niche-Based Drivers Shaping the Biogeography of UCYN-A1 and UCYN-A2

Some environmental traits (e.g., nutrients) among different regions were significant in the current study (**Figure 5**). Although niche-based processes played a minor role in community assembly in this study, several environmental factors significantly altered the community structure. Based on the species scores in NMDS, closely related UCYN-A taxa had more similar niche preferences (**Figure 4B**). We further evaluated trait conservatism in response to environmental changes (**Supplementary Table 6**). The UCYN-A community exhibited a strong and significant phylogenetic signal for temperature ( $K = 0.567$ ;  $p < 0.001$ ). This indicates that within a UCYN-A community, closely related taxa had more similar responses to temperature.

The ABT model was further applied to interpret the relative importance of environmental parameters on the composition of UCYN-A communities. The results were consistent with the NMDS ordination and the Mantel test, where temperature (Mantel test,  $R = 0.486$ ,  $p < 0.001$ ), coastal/oceanic habitats associated with salinity (Mantel test,  $R = 0.211$ ,  $p < 0.05$ ), and DSB (Mantel test,  $R = 0.330$ ,  $p < 0.01$ ) were identified as primary factors directly or indirectly correlated with the changes in community composition (**Figures 4, 6** and **Supplementary Table 7**). ABT verified that sea surface temperature was the most important factor influencing the diversity of the UCYN-A community. ABT also clearly showed that salinity and



DSB exerted significant influence on the UCYN-A community. Meanwhile, UCYN-A abundances decreased as temperature (26.9–30.0°C) increased ( $r = -0.451$ ,  $p < 0.001$ ) in the tropical seas (**Figure 7**). Additionally, the ratio of UCYN-A1/A2 abundance was significantly positively correlated with salinity ( $r = 0.387$ ,  $p < 0.05$ ) and DSB ( $r = 0.335$ ,  $p < 0.05$ ). Interestingly, the most important drivers varied with different sublineages, implying that UCYN-A1 and UCYN-A2 are distinct ecotypes.

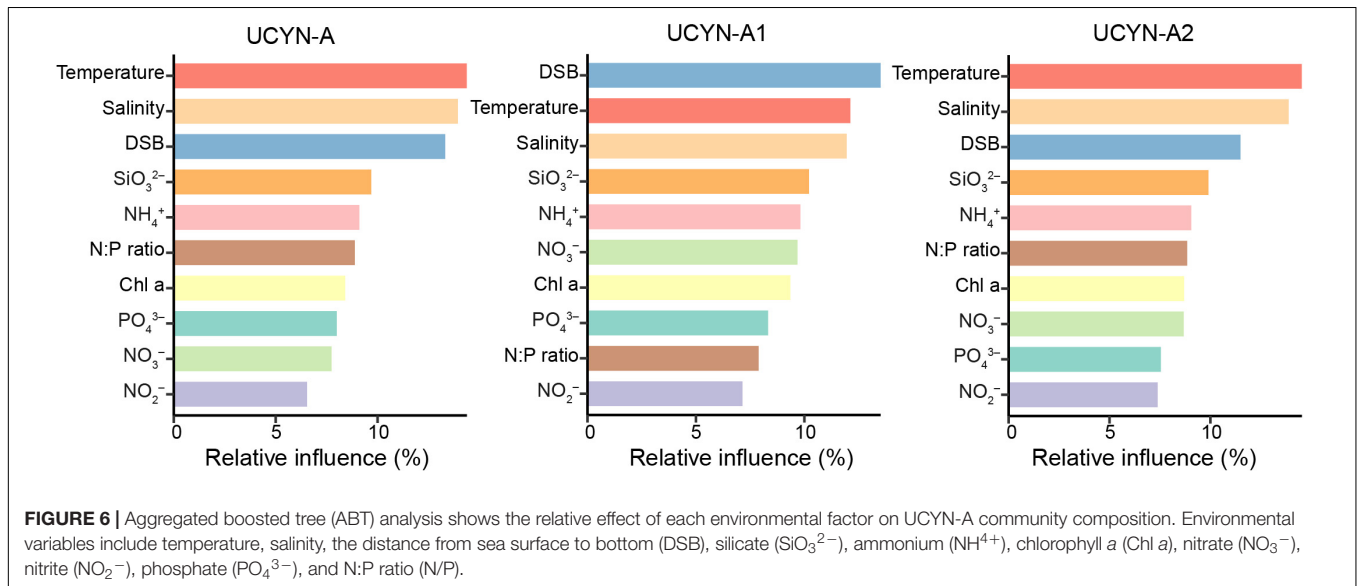
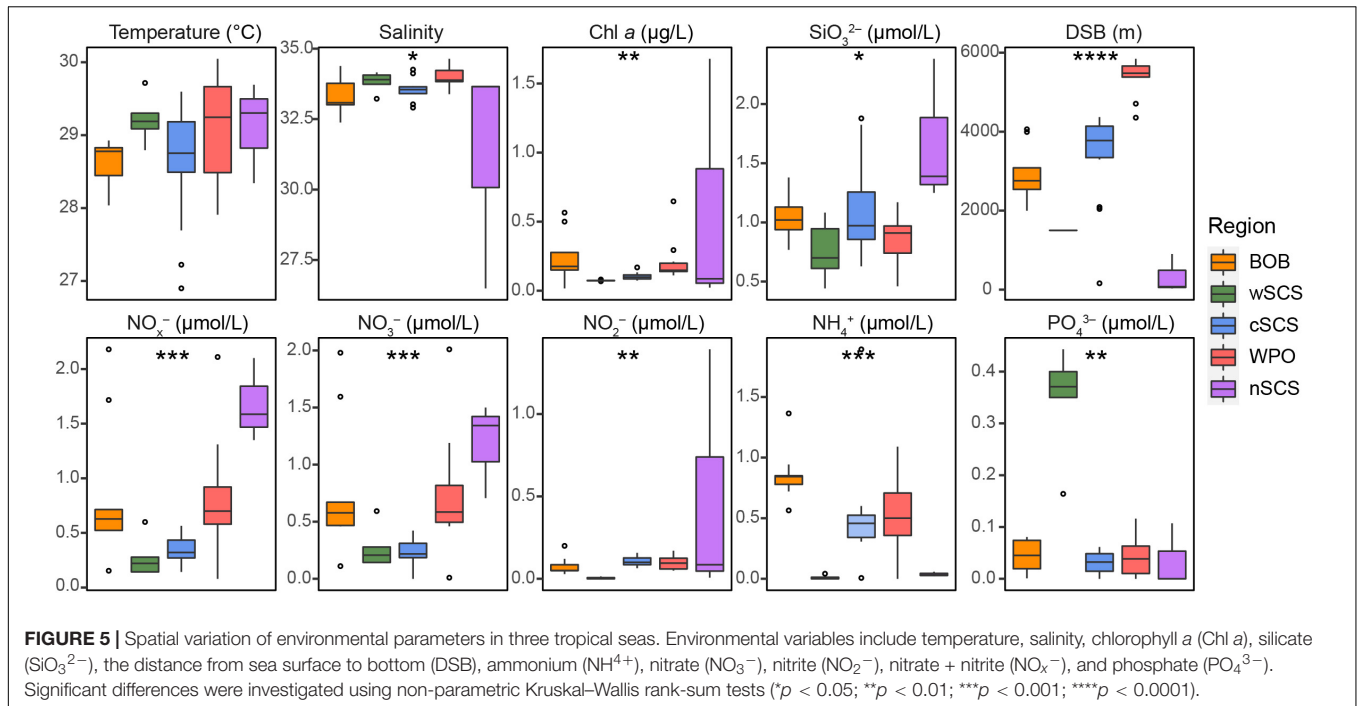
## DISCUSSION

### Distinct Assembly Mechanisms Shaping the Spatial Heterogeneity of UCYN-A

Little is known about the assembly mechanisms underpinning the diversity and biogeography of UCYN-A communities. The NMDS and PERMANOVA analyses showed that samples from the same region ( $p < 0.05$ ) clustered more closely than those from the same latitude ( $P = 0.14$ ), suggesting the significant genetic diversity of UCYN-A from different tropical seas.

This distinct distribution pattern can be the result of stochastic community assembly processes. The null models were employed

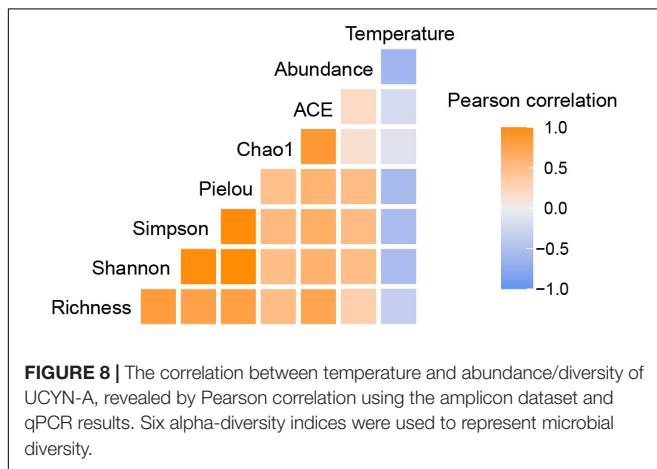
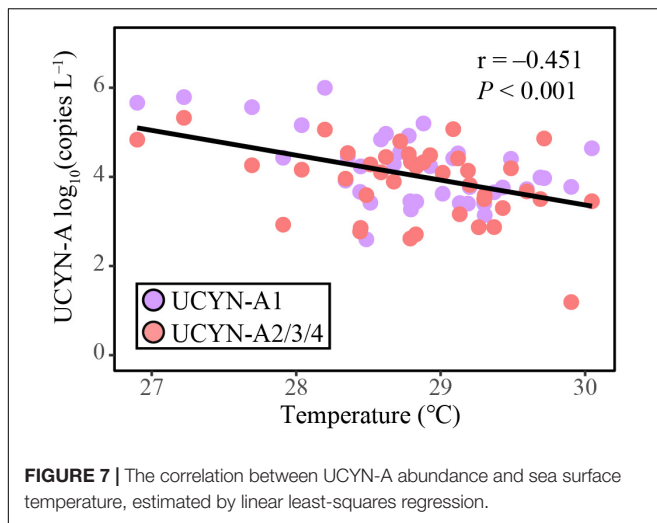




to quantify the ecological processes, and the findings that stochastic processes dominated UCYN-A community assembly were consistent with the results of a soil study conducted in bacterial communities across the North China Plain (Shi et al., 2018), as well as a study focused on methanotrophs communities in the Qinghai-Tibetan Plateau (Zhang et al., 2019). Due to geographical isolation and/or large-scale geographical distance, dispersal limitation processes may influence the generation and maintenance of microbial community diversity (Zhou and Ning, 2017) across different seas. Despite the low contribution of dispersal limitation in the current study, the significantly

increased  $\beta$ NTI values with increasing geographical distance (**Supplementary Figure 6**) suggest that determinism may overwhelm stochasticity across a broad spatial scale of more than 5,000 km, which is consistent with a previous assumption that spatial scale influences the importance of niche and neutral processes (Chase, 2014). This large scale is significantly greater than the turning point of determinism/stochasticity reported previously in plateau (130 km) and plain (900 km) environments (Shi et al., 2018). This difference may come from the more similar environmental pressures in the waters compared with those in the highly variable soil habitats across the same spatial scale, as

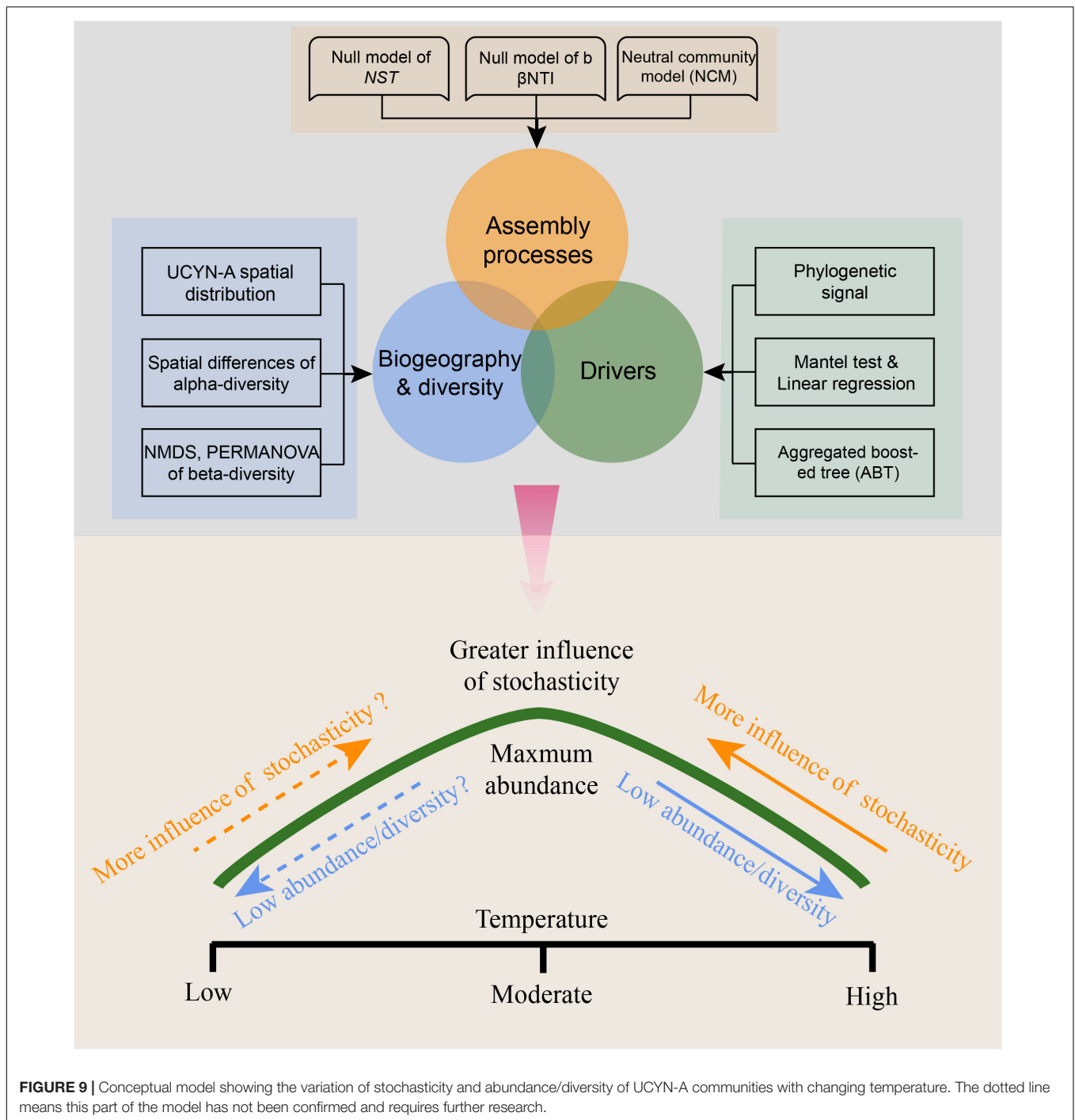




well as the distinct dispersal rates under the combined effect of a different dissemination medium (waters or soils) and microbial carrier (bacteria or symbionts). Meanwhile, the transportation of ocean currents and water masses can deliver microbes to distant regions and cause high levels of microbial dispersal, alleviating the dispersal limitation along the current path. Both of the qPCR-derived UCYN-A *nifH* gene abundances and *nifH* amplicon assays indicate a declining trend of UCYN-A1 proportion along the Kuroshio Current intrusion, suggesting a significant role of ocean currents on the spatial distribution pattern of microbes in these areas (Figure 1; Shiozaki et al., 2013, 2018). Indeed, it has been suggested that diazotrophs can remain active after being transported over great distances with ocean currents (Shiozaki et al., 2013). In addition, wind direction and speed have been identified as important environmental factors influencing the composition and abundance of UCYN-A sublineages in the Western Tropical South Pacific (Henke et al., 2018). Although it is challenging to evaluate the influences of stochasticity, it is of vital importance to incorporate physical parameters (e.g., the direction, velocity, and volume of ocean currents) into marine ecological frameworks in the future.

The NCM is an extensively applied method for inferring stochastic processes acting on community assembly (Burns et al., 2016; Zhou and Ning, 2017; Chen W. et al., 2019). This model is used to depict the importance of neutral processes, which are difficult to observe directly but have significant influences on microbial communities (Tong et al., 2019). A major part of the community variation is estimated by the NCM, suggesting that stochastic processes (e.g., random births, deaths, and immigrations) are critical in shaping the community assembly. Several important observations based on the NCM also revealed similar results to our finding. For instance, Chen W. et al. (2019) investigated the microeukaryotic communities during wet and dry seasons in the Tingjiang River, and found that microeukaryotic communities were strongly driven by stochastic processes ( $R^2 = 0.90$  explained by the NCM). Moreover, Roguet et al. (2015) revealed that bacterial community structure was mainly driven by stochastic processes ( $R^2 = 0.76$  explained by NCM) in 49 lakes in Paris, France (Roguet et al., 2015). Accordingly, the stochastic processes are powerful enough to generate significant community diversity among tropical seas.

Our study highlights the importance of stochastic processes on microbial community assembly in marine ecosystems. However, both null model and neutral community model methods have an intrinsic problem, that an obtained result fitting a null/neutral model cannot reject the importance of deterministic processes, since a niche model could present a result resembling that of null/neutral models (Zhou and Ning, 2017). Therefore, habitat filtering, to some extent, structures the microbial community. The difference in some environmental traits (e.g., Chl *a* and nutrients, bottom-up control) among different regions was significant in the present study (Figure 5). For instance, the wSCS region was affected by upwelling (Voss et al., 2006). The high concentrations of phosphate and low N/P ratios here may both directly and indirectly influence the diazotrophic community composition (Zhang et al., 2015). We also detected strong correlations between the UCYN-A community structure and the position of coastal/oceanic habitats across different stations. In addition, biotic interactions can shape a distinct community structure. For example, the C-N exchange of UCYN-A/haptophytes may enhance their ability to survive in a variety of habitats. Recently, a co-occurrence relationship between UCYN-A1 and UCYN-A3 has been reported (Turk-Kubo et al., 2017). Our NMDS and CCN results further confirmed this co-existence, as UCYN-A1 and UCYN-A3 clustered closely with significantly positive correlations among all corresponding OTUs (Figure 4B; Supplementary Figure 7; and Supplementary Table 8). All correlations between UCYN-A2 and UCYN-A4 were also positive, which is consistent with the potential co-occurring relationship between UCYN-A2 and UCYN-A4 as previously reported (Turk-Kubo et al., 2017). In fact, UCYN-A3 and UCYN-A4 are rare taxa compared with UCYN-A1 and UCYN-A2, thus likely possessing narrower niches in the global ocean. However, the ecological roles of these rare species should not be neglected because they can provide cover for the selection from a “seed bank” (Caron and Countway, 2009; Konopka et al., 2015). Moreover, grazing (Scavotto et al., 2015) and viral infection (Wilhelm and Suttle, 1999) (top-down control) may influence



the UCYN-A symbionts. Therefore, the combination of different abiotic and biotic drivers or stochasticity can shape the distinct community diversity and biogeography of UCYN-A in the tropical seas.

### The Drivers Mediating UCYN-A Assembly

High UCYN-A abundances were detected in the tropical seas, yet the abundances were higher near the subtropical area. This result agrees with a previous study in the

WPO and SCS that higher UCYN-A abundances mainly occurred between 18 and 20°N (Chen M. et al., 2019). This slight but statistically significant latitudinal constraint on UCYN-A abundances suggest that although niche-based processes are relatively weak in the tropical seas, their effects are also important in driving the biogeography of UCYN-A communities.

Temperature may be responsible for the latitudinal constraint on the abundance of UCYN-A as discussed above. As a

**TABLE 1** | Stochasticity in UCYN-A community assembly estimated by different indices based on various similarity metrics.

	Similarity metrics	High	Moderate
Incidence-based	Jaccard	<b>0.608 ± 0.052</b>	<b>0.798 ± 0.038</b>
	Kulczynski	<b>0.622 ± 0.062</b>	<b>0.819 ± 0.04</b>
	Gower	<b>0.677 ± 0.054</b>	<b>0.846 ± 0.037</b>
	Manhattan	<b>0.482 ± 0.05</b>	<b>0.694 ± 0.051</b>
Abundance-based	Bray-Curtis	0.565 ± 0.154	0.675 ± 0.084
	Kulczynski	0.563 ± 0.159	0.675 ± 0.084
	Canberra	0.769 ± 0.062	0.843 ± 0.039
	Gower	0.694 ± 0.077	0.793 ± 0.082

The distribution of NST at high temperatures and near-optimal temperatures was obtained from 999 permutations. The significance of NST differences between two groups was tested for using permutational ANOVA test with 999 permutations. Bold values indicate significant differences ( $p < 0.01$ ).

fundamental environmental variable, temperature can determine the composition and function of microbial communities. It has been suggested that the global biogeographical distribution of marine diazotrophic cyanobacteria reflects the physiologically preferential temperature range at which they can perform nitrogen fixation (Brauer et al., 2013). However, it has not been clear how temperature mediates diazotrophic community assembly processes across the global ocean. Here, we reported an interesting taxa, UCYN-A, the temperature optima of which was previously inferred to be near 24°C (Moisander et al., 2010), to partly address this mystery. A phylogenetic signal in the ecological niches of OTUs was detected, suggesting that more closely related UCYN-A taxa have more similar niche preferences related to temperature (Losos, 2008). Moreover, the ABT model, NMDS ordination and Mantel test demonstrated that temperature was the most important environmental factor in driving the UCYN-A community variation observed in the current study. Correlation analysis also showed that the UCYN-A abundances and diversity indices were negatively correlated with temperature (26.9–30.0°C) (Figure 8), which is consistent with previous studies (Moisander et al., 2010; Bonnet et al., 2015; Turk-Kubo et al., 2015).

Null models using various community similarity algorithms showed similar results of inferred ecological processes (Table 1), indicating that temperature mediates the degree to which deterministic vs. stochastic processes shape UCYN-A community assembly. Our results are conceptually consistent with the reported findings of Tripathi et al. (2018) on bacteria and some studies on macro-organisms (Chase, 2007; Helmus et al., 2010; Myers and Harms, 2011). These results, coupled with previous paradigms (Tripathi et al., 2018), lead us to propose a conceptual model (Figure 9), which depicts how stochastic processes and abundance/diversity of UCYN-A communities change with sea surface temperature in the ocean. From the spatial perspective, a near-optimal temperature will lead to more stochastic assemblies and higher abundance/diversity in moderate latitude areas, and a shift in temperature toward relatively extreme conditions at low latitude will result in more deterministic assemblies and less abundance/diversity.

While from the temporal perspective, extreme temperature in warmer seasons will lead to more deterministic assemblies and less abundance/diversity, and progressive shifts in temperature toward optima conditions will generate weaker selection, more stochasticity and higher abundance/diversity. However, although we suggest that the importance of neutral vs. niche-based assembly of UCYN-A communities is significantly influenced by temperature, the generality of this framework is largely limited by the narrow range of temperatures in the tropical seas and the uncertainty of temperature optima; there have been many detections of abundant UCYN-A in cold waters (Shiozaki et al., 2017, 2018; Harding et al., 2018). Therefore, it is necessary to test this conceptual model in different ecosystems spanning a large temperature range (e.g., including ecosystems in both high- and low-latitude regions).

Ecologists and biogeographers might also focus on other factors affecting UCYN-A communities. There is no doubt that distinct habitats exert an effect on the processes of microbial community assembly. Mills et al. (2020) recently reported an unusual mechanism that UCYN-A/prymnesiophyte symbiosis relies on N<sub>2</sub> fixation, meets little of its N demands via ammonium uptake, and does not assimilate nitrate even in N-rich environments. The implication is that N<sub>2</sub> fixation of UCYN-A is important in global coastal and oceanic habitats. This unique N-utilization strategy also partly explained the high N<sub>2</sub> fixation rates and UCYN-A abundance in the coastal region where nitrate concentrations were extremely high as Tang et al. (2020) reported. Moreover, the cross-feeding between UCYN-A and their hosts (Thompson et al., 2012; Zehr, 2015; Martinez-Perez et al., 2016) allows them to survive in a variety of habitats even under nutrient limitation and competitive conditions, so that the distribution of UCYN-A is more influenced by neutral processes.

Interestingly, in the current study, the ratio of UCYN-A1/A2 abundance (and relative abundance) positively correlated with salinity and DSB, suggesting the preference of UCYN-A2 (UCYN-A1) in coastal (oceanic) habitats. The UCYN-A2/B. *bigelowii* association (4–20 μm) is known to be larger than the UCYN-A1 association (1–3 μm) (Thompson et al., 2012, 2014; Cabello et al., 2016; Farnelid et al., 2016; Zehr et al., 2016; Cornejo-Castillo et al., 2019). These size differences are consistent with the theory that larger phytoplankton with higher abundances tend to be found in eutrophic conditions (Irwin et al., 2006). However, our findings of higher stochasticity for UCYN-A2 compared with UCYN-A1 (Supplementary Table 4) contradict a recent study that small microorganisms are relatively more influenced by stochastic processes (Luan et al., 2020). This contradiction may arise from the fact that the body size of UCYN-A symbionts is not a proper driver governing UCYN-A community assembly. Apart from temperature, other important deterministic parameters, such as light, iron availability, biotic grazing, and viral infection, have been shown to be of vital importance in the growth and metabolism of microbes (Wilhelm and Suttle, 1999; Ribalet et al., 2015; Scavotto et al., 2015; Bonnain et al., 2016).

Determining how limiting effects of these factors definitively drive community variability cannot be discussed elaborately in the present study, yet quantifying the contribution of these factors using mathematical models has profound implications for further investigations.

## CONCLUSION

Our study provides the first example of research revealing both stochastic processes and deterministic processes driving the biogeography and diversity of marine UCYN-A communities. Dispersal limitation can cause distinct microbial biogeography, influencing the generation and maintenance of microbial community diversity. Meanwhile, the transportation of ocean currents and water masses may bring the high possibility of microbial immigration and promote the cross-regional distribution of microbial communities in the ocean. These stochastic processes hint at the urgency of incorporating parameters of physical processes into marine ecological frameworks in the future, and clarify the balance between microbial dispersal and dispersal limitations mediating microbial biogeography. On the other hand, we cannot reject the vital roles of deterministic processes given that a niche model could show a pattern similar to that of null/neutral models (Zhou and Ning, 2017). The ABT model, NMDS ordination, and Mantel test concurrently demonstrated that temperature and coastal/oceanic habitats associated with salinity and DSB were important environmental factors in shaping UCYN-A community assembly. We further propose a conceptual framework that temperature mediates species abundance/diversity and the degree of stochasticity/determinism. Although we suggest that our framework provides novel insights, it is important to recognize its limitations including that the framework was concluded based on the narrow temperature range of the tropical seas. It is promising to disentangle the relative importance of stochastic processes and deterministic processes and further propose a broader framework by integrating the global data including tropical, temperate, and polar environments into ecological models.

## DATA AVAILABILITY STATEMENT

The raw data of *nifH* gene sequences have been submitted to the NCBI Sequence Read Archive (SRA) under Bioproject

## REFERENCES

- Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K., and Tedersoo, L. (2016). Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME J.* 10, 885–896. doi: 10.1038/ismej.2015.164
- Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5:219. doi: 10.3389/fmicb.2014.00219
- Blomberg, S. P., Theodore Garland, J., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more reliable. *Evolution* 57, 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x

PRJNA650247. The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a manuscript that does not adhere to our open data policies.

## AUTHOR CONTRIBUTIONS

CW, JS, and LL conceived and designed the experiments. LL, CW, and CD carried out the sample collection, extracted DNA, and performed the experiments. LL and DH analyzed the data. LL wrote the manuscript. LL, JS, CW, DH, and YW contributed to the manuscript development and revisions. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (41876134, 41676112, and 41276124) and the Changjiang Scholar Program of Chinese Ministry of Education (T2014253 to JS).

## ACKNOWLEDGMENTS

We would like to thank the Open Cruise Projects of National Nature Science Foundation of China (NORC2016-06, NORC2016-07, NORC2016-10, and NORC2017-09) for sharing their ship time. We also gratefully acknowledge the crew of the R/V “*Shiyan 1*,” “*Shiyan 3*,” “*Dongfang2*,” and “*Kexue*” for their assistance and all the participants for their inputs and contributions during the cruises.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.654646/full#supplementary-material>

- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bonnain, C., Breitbart, M., and Buck, K. N. (2016). The Ferrojan horse hypothesis: iron-virus interactions in the ocean. *Front. Mar. Sci.* 3:82. doi: 10.3389/fmars.2016.00082
- Bonnet, S., Rodier, M., Turk-Kubo, K. A., Germineaud, C., Menkes, C., Ganachaud, A., et al. (2015). Contrasted geographical distribution of N<sub>2</sub> fixation rates and *nifH* phylotypes in the Coral and Solomon Seas (southwestern Pacific) during austral winter conditions. *Glob. Biogeochem. Cycles* 29, 1874–1892. doi: 10.1002/2015GB005117



- Brauer, V. S., Stomp, M., Rosso, C., van Beusekom, S. A., Emmerich, B., Stal, L. J., et al. (2013). Low temperature delays timing and enhances the cost of nitrogen fixation in the unicellular cyanobacterium *Cyanothece*. *ISME J.* 7, 2105–2115. doi: 10.1038/ismej.2013.103
- Breitbarth, E., Oschlies, A., and LaRoche, J. (2007). Physiological constraints on the global distribution of *Trichodesmium*—effect of temperature on diazotrophy. *Biogeosciences* 4, 53–61. doi: 10.5194/bg-4-53-2007
- Burns, A. R., Stephens, W. Z., Stagaman, K., Wong, S., Rawls, J. F., Guillemin, K., et al. (2016). Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J.* 10, 655–664. doi: 10.1038/ismej.2015.142
- Cabello, A. M., Cornejo-Castillo, F. M., Raho, N., Blasco, D., Vidal, M., Audic, S., et al. (2016). Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J.* 10, 693–706. doi: 10.1038/ismej.2015.147
- Cabello, A. M., Turk-Kubo, K. A., Hayashi, K., Jacobs, L., Kudela, R. M., and Zehr, J. P. (2020). Unexpected presence of the nitrogen-fixing symbiotic cyanobacterium UCYN-A in Monterey Bay, California. *J. Phycol.* 56, 1521–1533. doi: 10.1111/jpy.13045
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Caron, D. A., and Countway, P. D. (2009). Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat. Microb. Ecol.* 57, 227–238. doi: 10.3354/ame01352
- Chase, J. M. (2007). Drought mediates the importance of stochastic community assembly. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17430–17434. doi: 10.1073/pnas.0704350104
- Chase, J. M. (2014). Spatial scale resolves the niche versus neutral theory debate. *J. Veg. Sci.* 25, 319–322. doi: 10.1111/jvs.12159
- Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., and Inouye, B. D. (2011). Using null models to disentangle variation in community dissimilarity from variation in  $\alpha$ -diversity. *Ecosphere* 2, 1–11. doi: 10.1890/ES10-00117.1
- Chave, J. (2004). Neutral theory and community ecology. *Ecol. Lett.* 7, 241–253. doi: 10.1111/j.1461-0248.2003.00566.x
- Chen, M., Lu, Y., Jiao, N., Tian, J., Kao, S. J., and Zhang, Y. (2019). Biogeographic drivers of diazotrophs in the western Pacific Ocean. *Limnol. Oceanogr.* 64, 1403–1421. doi: 10.1002/lno.11123
- Chen, W., Ren, K., Isabwe, A., Chen, H., Liu, M., and Yang, J. (2019). Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* 7:138. doi: 10.1186/s40168-019-0749-8
- Church, M. J., Short, C. M., Jenkins, B. D., Karl, D. M., and Zehr, J. P. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl. Environ. Microbiol.* 71, 5362–5370. doi: 10.1128/AEM.71.9.5362-5370.2005
- Cornejo-Castillo, F. M., Cabello, A. M., Salazar, G., Sánchez-Baracaldo, P., Lima-Mendez, G., Hingamp, P., et al. (2016). Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat. Commun.* 7:11071. doi: 10.1038/ncomms11071
- Cornejo-Castillo, F. M., Muñoz-Marin, M. D. C., Turk-Kubo, K. A., Royo-Llonch, M., Farnelid, H., Acinas, S. G., et al. (2019). UCYN-A3, a newly characterized open ocean sublineage of the symbiotic  $N_2$ -fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa*. *Environ. Microbiol.* 21, 111–124. doi: 10.1111/1462-2920.14429
- Csardi, G., and Nepusz, T. J. I. (2006). The igraph software package for complex network research. *Complex Syst.* 1695, 1–9.
- De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251. doi: 10.1890/0012-9658(2007)88[243:BTFFEMA]2.0.CO;2
- Dini-Andreote, F., Stegen, J. C., Van Elsas, J. D., and Salles, J. F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1326–E1332. doi: 10.1073/pnas.1414261112
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Fargione, J., Brown, C. S., and Tilman, D. (2003). Community assembly and invasion: an experimental test of neutral versus niche processes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8916–8920. doi: 10.1073/pnas.1033107100
- Farnelid, H., Turk-Kubo, K., and Zehr, J. P. (2021). Cell sorting reveals few novel prokaryote and photosynthetic picoeukaryote associations in the oligotrophic ocean. *Environ. Microbiol.* 23, 1469–1480. doi: 10.1111/1462-2920.15351
- Farnelid, H., Turk-Kubo, K., Muñoz-Marin, M. C., and Zehr, J. P. (2016). New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. *Aquat. Microb. Ecol.* 77, 125–138. doi: 10.3354/ame01794
- Farnelid, H., Turk-Kubo, K., Ploug, H., Ossolinski, J. E., Collins, J. R., Van Mooy, B. A. S., et al. (2019). Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *ISME J.* 13, 170–182. doi: 10.1038/s41396-018-0259-x
- Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. Thousand Oaks, CA: Sage publications.
- Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., et al. (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to  $N_2$  fixation in the tropical Atlantic Ocean. *Environ. Microbiol.* 12, 3272–3289. doi: 10.1111/j.1462-2920.2010.02303.x
- Gradoville, M. R., Farnelid, H., White, A. E., Turk-Kubo, K. A., Stewart, B., Ribalet, F., et al. (2020). Latitudinal constraints on the abundance and activity of the cyanobacterium UCYN-A and other marine diazotrophs in the North Pacific. *Limnol. Oceanogr.* 65, 1858–1875. doi: 10.1002/lno.11423
- Greenblum, S., Turnbaugh, P. J., and Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. U.S.A.* 109, 594–599. doi: 10.1073/pnas.1116053109
- Gruber, N., and Sarmiento, J. L. (1997). Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* 11, 235–266. doi: 10.1029/97GB00077
- Guo, S., Feng, Y., Wang, L., Dai, M., Liu, Z., Bai, Y., et al. (2014). Seasonal variation in the phytoplankton community of a continental-shelf sea: the East China Sea. *Mar. Ecol. Prog. Ser.* 516, 103–126. doi: 10.3354/meps10952
- Harding, K., Turk-Kubo, K. A., Sipler, R. E., Mills, M. M., Bronk, D. A., and Zehr, J. P. (2018). Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. *Proc. Natl. Acad. Sci. U.S.A.* 115, 13371–13375. doi: 10.1073/pnas.1813658115
- Harrell, F. E., and Dupont, C. (2013). *Hmisc: Harrell Miscellaneous. R Package Hmisc, Version 3.0-12*. 3.
- Helmus, M. R., Keller, W. B., Paterson, M. J., Yan, N. D., Cannon, C. H., and Rusak, J. A. (2010). Communities contain closely related species during ecosystem disturbance. *Ecol. Lett.* 13, 162–174. doi: 10.1111/j.1461-0248.2009.01411.x
- Henke, B. A., Turk-Kubo, K. A., Bonnet, S., and Zehr, J. P. (2018). Distributions and abundances of sublineages of the  $N_2$ -fixing cyanobacterium *Candidatus atelocyanobacterium thalassa* (UCYN-A) in the new caledonian coral lagoon. *Front. Microbiol.* 9:554. doi: 10.3389/fmicb.2018.00554
- Irwin, A. J., Finkel, Z. V., Schofield, O. M. E., and Falkowski, P. G. (2006). Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *J. Plankton Res.* 28, 459–471. doi: 10.1093/plankt/fbi148
- Jiao, S., and Lu, Y. (2020). Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environ. Microbiol.* 22, 1052–1065. doi: 10.1111/1462-2920.14815
- Jiao, S., Chen, W., and Wei, G. (2017). Biogeography and ecological diversity patterns of rare and abundant bacteria in oil-contaminated soils. *Mol. Ecol.* 26, 5305–5317. doi: 10.1111/mec.14218
- Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M., and Mahaffey, C. (2012). Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1842–1849. doi: 10.1073/pnas.1120312109
- Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J., and Hebel, D. (1997). The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* 388, 533–538. doi: 10.1038/41474

- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. doi: 10.1093/bioinformatics/btq166
- Kong, L., Jing, H., Kataoka, T., Sun, J., and Liu, H. (2011). Phylogenetic diversity and spatio-temporal distribution of nitrogenase genes (*nifH*) in the northern South China Sea. *Aquat. Microb. Ecol.* 65, 15–27. doi: 10.3354/ame01531
- Konopka, A., Lindemann, S., and Fredrickson, J. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. *ISME J.* 9, 1488–1495. doi: 10.1038/ismej.2014.251
- Krupke, A., Lavik, G., Halm, H., Fuchs, B. M., Amann, R. I., and Kuypers, M. M. (2014). Distribution of a consortium between unicellular algae and the N<sub>2</sub> fixing cyanobacterium UCYN-A in the North Atlantic Ocean. *Environ. Microbiol.* 16, 3153–3167. doi: 10.1111/1462-2920.12431
- Krupke, A., Mohr, W., LaRoche, J., Fuchs, B. M., Amann, R. I., and Kuypers, M. M. (2015). The effect of nutrients on carbon and nitrogen fixation by the UCYN-A-haptophyte symbiosis. *ISME J.* 9, 1635–1647. doi: 10.1038/ismej.2014.253
- Langenheder, S., and Szekely, A. J. (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J.* 5, 1086–1094. doi: 10.1038/ismej.2010.207
- Li, L., Wu, C., Sun, J., Song, S., Ding, C., Huang, D., et al. (2021a). Nitrogen fixation driven by mesoscale eddies and the Kuroshio current in the northern South China Sea and the East China Sea. *Acta Oceanol. Sin.* 39, 30–41. doi: 10.1007/s13131-020-1691-0
- Li, L., Pujari, L., Wu, C., Huang, D., Wei, Y., Guo, C., et al. (2021b). Assembly processes and co-occurrence patterns of abundant and rare bacterial community in the Eastern Indian Ocean. *Front. Microbiol.* 12:616956. doi: 10.3389/fmicb.2021.616956
- Liu, H., Xue, B., Feng, Y., Zhang, R., Chen, M., and Sun, J. (2015). Size-fractionated Chlorophyll a biomass in the northern South China Sea in summer 2014. *Chin. J. Oceanol. Limnol.* 34, 672–682. doi: 10.1007/s00343-016-5017-1
- Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* 11, 995–1003. doi: 10.1111/j.1461-0248.2008.01229.x
- Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q., et al. (2020). Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. *Nat. Commun.* 11:6406. doi: 10.1038/s41467-020-20271-4
- Luo, Y. W., Doney, S. C., Anderson, L. A., Benavides, M., Berman-Frank, I., Bode, A., et al. (2012). Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data* 4, 47–73. doi: 10.5194/essd-4-47-2012
- Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., et al. (2016). Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* 10, 1891–1901. doi: 10.1038/ismej.2015.261
- Magoc, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Martinez-Perez, C., Mohr, W., Loscher, C. R., Dekaezemaeker, J., Littmann, S., Yilmaz, P., et al. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol.* 1:16163. doi: 10.1038/nmicrobiol.2016.163
- Mills, M. M., Turk-Kubo, K. A., van Dijken, G. L., Henke, B. A., Harding, K., Wilson, S. T., et al. (2020). Unusual marine cyanobacteria/haptophyte symbiosis relies on N<sub>2</sub> fixation even in N-rich environments. *ISME J.* 14, 2395–2406. doi: 10.1038/s41396-020-0691-6
- Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., et al. (2010). Unicellular cyanobacterial distributions broaden the oceanic N<sub>2</sub> fixation domain. *Science* 327, 1512–1514. doi: 10.1126/science.1185468
- Moreira-Coello, V., Mourino-Carballido, B., Maranon, E., Fernandez-Carrera, A., Bode, A., Sintès, E., et al. (2019). Temporal variability of diazotroph community composition in the upwelling region off NW Iberia. *Sci. Rep.* 9:3737. doi: 10.1038/s41598-019-39586-4
- Munoz-Marin, M. D. C., Shilova, I. N., Shi, T., Farnelid, H., Cabello, A. M., and Zehr, J. P. (2019). The transcriptional cycle is suited to daytime N<sub>2</sub> fixation in the unicellular cyanobacterium “*Candidatus* atelocyanobacterium thalassa” (UCYN-A). *MBio* 10:e02495-18. doi: 10.1128/mBio.02495-18
- Myers, J. A., and Harms, K. E. (2011). Seed arrival and ecological filters interact to assemble high-diversity plant communities. *Ecology* 92, 676–686. doi: 10.1890/10-1001.1
- Ning, D., Deng, Y., Tiedje, J. M., and Zhou, J. (2019). A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl. Acad. Sci. U.S.A.* 116, 16892–16898. doi: 10.1073/pnas.1904623116
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P., and O’Hara, R. B. (2015). *Vegan: Community Ecology Package. R Package Vegan, Version 2.2-1*.
- Pai, S.-C., Tsau, Y.-J., and Yang, T.-I. (2001). pH and buffering capacity problems involved in the determination of ammonia in saline water using the indophenol blue spectrophotometric method. *Anal. Chim. Acta* 434, 209–216. doi: 10.1016/S0003-2670(01)00851-0
- Postgate, J. R. (1982). *The Fundamentals of Nitrogen Fixation*. London: Cambridge University Press.
- Ribalet, F., Swalwell, J., Clayton, S., Jimenez, V., Sudek, S., Lin, Y., et al. (2015). Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8008–8012. doi: 10.1073/pnas.1424279112
- Roguet, A., Laigle, G. S., Therial, C., Bressy, A., Soullignac, F., Catherine, A., et al. (2015). Neutral community model explains the bacterial community assembly in freshwater lakes. *FEMS Microbiol. Ecol.* 91, 1–20. doi: 10.1093/femsec/fiv125
- Scavotto, R. E., Dziallas, C., Bentzon-Tilia, M., Riemann, L., and Moisaner, P. H. (2015). Nitrogen-fixing bacteria associated with copepods in coastal waters of the North Atlantic Ocean. *Environ. Microbiol.* 17, 3754–3765. doi: 10.1111/1462-2920.12777
- Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., et al. (2018). Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome* 6:27. doi: 10.1186/s40168-018-0409-4
- Shiozaki, T., Bombar, D., Riemann, L., Hashihama, F., Takeda, S., Yamaguchi, T., et al. (2017). Basin scale variability of active diazotrophs and nitrogen fixation in the North Pacific, from the tropics to the subarctic Bering Sea. *Glob. Biogeochem. Cycles* 31, 996–1009. doi: 10.1002/2017GB005681
- Shiozaki, T., Fujiwara, A., Ijichi, M., Harada, N., Nishino, S., Nishi, S., et al. (2018). Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnol. Oceanogr.* 63, 2191–2205. doi: 10.1002/lno.10933
- Shiozaki, T., Kodama, T., Kitajima, S., Sato, M., and Furuya, K. (2013). Advective transport of diazotrophs and importance of their nitrogen fixation on new and primary production in the western Pacific warm pool. *Limnol. Oceanogr.* 58, 49–60. doi: 10.4319/lno.2013.58.1.0049
- Sloan, W. T., Lunn, M., Woodcock, S., Head, I. M., Nee, S., and Curtis, T. P. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* 8, 732–740. doi: 10.1111/j.1462-2920.2005.00956.x
- Sohm, J. A., Webb, E. A., and Capone, D. G. (2011). Emerging patterns of marine nitrogen fixation. *Nat. Rev. Microbiol.* 9, 499–508. doi: 10.1038/nrmicro.2594
- Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., et al. (2013). Quantifying community assembly processes and identifying features that impose them. *ISME J.* 7, 2069–2079. doi: 10.1038/ismej.2013.93
- Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 6, 1653–1664. doi: 10.1038/ismej.2012.22
- Tang, W., Cerdan-Garcia, E., Berthelot, H., Polyviou, D., Wang, S., Baylay, A., et al. (2020). New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *ISME J.* 14, 2514–2526. doi: 10.1038/s41396-020-0703-6
- Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaulot, D., et al. (2012). Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* 337, 1546–1550. doi: 10.1126/science.1222700
- Thompson, A., Carter, B. J., Turk-Kubo, K., Malfatti, F., Azam, F., and Zehr, J. P. (2014). Genetic diversity of the unicellular nitrogen-fixing cyanobacteria UCYN-A and its prymnesiophyte host. *Environ. Microbiol.* 16, 3238–3249. doi: 10.1111/1462-2920.12490
- Tong, X., Leung, M. H. Y., Wilkins, D., Cheung, H. H. L., and Lee, P. K. H. (2019). Neutral processes drive seasonal assembly of the skin microbiome. *mSystems* 4:e00004-19. doi: 10.1128/mSystems.00004-19

- Tripathi, B. M., Stegen, J. C., Kim, M., Dong, K., Adams, J. M., and Lee, Y. K. (2018). Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* 12, 1072–1083. doi: 10.1038/s41396-018-0082-4
- Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., et al. (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* 464, 90–94. doi: 10.1038/nature08786
- Turk-Kubo, K. A., Connell, P., Caron, D., Hogan, M. E., Farnelid, H. M., and Zehr, J. P. (2018). In situ diazotroph population dynamics under different resource ratios in the North Pacific subtropical Gyre. *Front. Microbiol.* 9:1616. doi: 10.3389/fmicb.2018.01616
- Turk-Kubo, K. A., Farnelid, H. M., Shilova, I. N., Henke, B., and Zehr, J. P. (2017). Distinct ecological niches of marine symbiotic N<sub>2</sub>-fixing cyanobacterium *Candidatus atelocyanobacterium thalassa* sublineages. *J. Phycol.* 53, 451–461. doi: 10.1111/jpy.12505
- Turk-Kubo, K. A., Frank, I. E., Hogan, M. E., Desnues, A., Bonnet, S., and Zehr, J. P. (2015). Diazotroph community succession during the VAHINE mesocosm experiment (New Caledonia lagoon). *Biogeosciences* 12, 7435–7452. doi: 10.5194/bg-12-7435-2015
- Tyrrell, T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* 400, 525–531. doi: 10.1038/22941
- Voss, M., Bombar, D., Loick, N., and Dippner, J. W. (2006). Riverine influence on nitrogen fixation in the upwelling region off Vietnam, South China Sea. *Geophys. Res. Lett.* 33:L07604. doi: 10.1029/2005GL025569
- Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J. C., et al. (2013). Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J.* 7, 1310–1321. doi: 10.1038/ismej.2013.30
- Welschmeyer, N. A. (1994). Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll *b* and pheopigments. *Limnol. Oceanogr.* 39, 1985–1992. doi: 10.4319/lo.1994.39.8.1985
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer. doi: 10.1007/978-3-319-24277-4
- Wilhelm, S. W., and Suttle, C. A. (1999). Viruses and nutrient cycles in the sea. *BioScience* 49, 781–788. doi: 10.2307/1313569
- Wu, C., Kan, J., Liu, H., Pujari, L., Guo, C., Wang, X., et al. (2019). Heterotrophic bacteria dominate the diazotrophic community in the Eastern Indian Ocean (EIO) during pre-southwest monsoon. *Microb. Ecol.* 78, 804–819. doi: 10.1007/s00248-019-01355-1
- Zehr, J. P. (2015). How single cells work together. *Science* 349, 1163–1164. doi: 10.1126/science.aac9752
- Zehr, J. P., and Capone, D. G. (2020). Changing perspectives in marine nitrogen fixation. *Science* 368:eaay9514. doi: 10.1126/science.aay9514
- Zehr, J. P., Shilova, I. N., Farnelid, H. M., Muñoz-Marin, M. D. C., and Turk-Kubo, K. A. (2016). Unusual marine unicellular symbiosis with the nitrogen-fixing cyanobacterium UCYN-A. *Nat. Microbiol.* 2, 1–11. doi: 10.1038/nmicrobiol.2016.214
- Zehr, J., and Turner, P. (2001). Nitrogen fixation: nitrogenase genes and gene expression. *Methods Microbiol.* 30, 271–286. doi: 10.1016/S0580-9517(01)30049-1
- Zhang, L., Adams, J. M., Dumont, M. G., Li, Y., Shi, Y., He, D., et al. (2019). Distinct methanotrophic communities exist in habitats with different soil water contents. *Soil Biol. Biochem.* 132, 143–152. doi: 10.1016/j.soilbio.2019.02.007
- Zhang, R., Chen, M., Yang, Q., Lin, Y., Mao, H., Qiu, Y., et al. (2015). Physical–biological coupling of N<sub>2</sub> fixation in the northwestern South China Sea coastal upwelling during summer. *Limnol. Oceanogr.* 60, 1411–1425. doi: 10.1002/lno.10111
- Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G. J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol. Ecol.* 23, 2260–2274. doi: 10.1111/mec.12739
- Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* 81:e00002-17. doi: 10.1128/MMBR.00002-17

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Wu, Huang, Ding, Wei and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.