Spatial Modelling of Bacterial Diversity over the Selected Regions in Bangladesh by Next-Generation Sequencing: Role of Water Temperature

Nabila Akter 1,2, Md Wahiduzzaman 1,3,*, Alea Yeasmin 4, Kazi Saiful Islam 2,5 and Jing-Jia Luo 1,*

1 Institute for Climate and Application Research, Nanjing University of Information Science and Technology, Nanjing 210000, China; nabilabiochemistry@gmail.com
2 Department of Biochemistry and Molecular Biology, Jahangirnagar University, Dhaka 1342, Bangladesh; Kazi.islam@sydney.edu.au
3 Institute for Marine and Antarctic Studies, University of Tasmania, Hobart 7000, Australia
4 School of Science, Information and Technology, Federation University, Ballarat 3350, Australia; a.yeasmin@federation.edu.au
5 School of Life and Environmental Sciences, The University of Sydney, New South Wales 2006, Australia
* Correspondence: wahid.zaman@nuist.edu.cn or md.wahiduzzaman@utas.edu.au (M.W.); jjluo@nuist.edu.cn (J.-J.L.)

Received: 12 February 2020; Accepted: 3 April 2020; Published: 7 April 2020

Abstract: In this study, a spatial model has been developed to investigate the role of water temperature to the distribution of bacteria over the selected regions in the Bay of Bengal, located in the southern region of Bangladesh using next-generation sequencing. Bacterial concentration, quantitative polymerase chain reactions, and sequencing were performed on water samples and identified Acidobacteria, Actinobacteria, Bacteroidetes, Chlorobi, Chloroflexi, Cyanobacteria, Firmicutes, Nitrospirae, Planctomycetes, Proteobacteria, and Verrucomicrobia. The spatial model tessellated the parts of the Bay of Bengal with hexagons and analyzed the relationship between the distribution of bacteria and water temperature. A geographically weighted regression was used to observe whether water temperature contributed strongly or weakly to the distribution of bacteria. The residuals were examined to assess the model’s fitness. The spatial model has the potential to predict the bacterial diversity in the selected regions of Bangladesh.

Keywords: spatial modelling; bacterial distribution; next generation sequencing; water temperature; Bangladesh

1. Introduction

Bacterial communities play a pivotal role in the welfare of water environments [1]. However, they vary spatially, depending on different aspects like environmental factors, community adaptation and human activities [2–4]. The dynamics of bacterial communities in developing countries are poorly understood [5,6] including the present study area: selected regions of the Bay of Bengal, Bangladesh.

Bangladesh, located in the north of the Bay of Bengal, is known as a deltaic country and 70% of the area is covered by water. Bangladesh is an agricultural country that is located between the 20°34′ and 26°38′ N latitudes and 88°01′ and 92°41′ E longitudes in South Asia. Like Bangladesh, the Bay of Bengal rim region is currently under pressure due to invasion of human pollutants [7]. Wastes in this zone affect the physical, biological and chemical qualities of the water environment, and as such have negative impacts on human health for (e.g., pathogenic bacteria have caused typhoid fever and pediatric bloodstream infections in this area, which have been documented in hospitals [8]).
Microbiota are the base of biogeochemical cycles [9] and act as a pollutant biodegradation, so it is essential to assess a microbial community structure’s changing nature in aquatic systems. The structures of bacterial communities are affected by the hydrodynamics of storm events [10], physiochemical parameters, longitudinal distance, upstream inputs affects and fecal indicator bacteria concentrations. Further, bacterivory, substrate availability, variation in hydrological and nutrient conditions [11], land use changes, distance from the pristine headwaters, rainfall and pH [12] influence the structure of microbial communities in riverine environments.

Bacterial community diversity and dynamics, including pathogens [13] and their impacts on water [4], have been characterized through different molecular methods [14,15]. Among different methods, the intensive application of next-generation sequencing (NGS) helps to explore the diversity, dynamics and sensitivity of microorganisms [12,14,16–18]. [7] studied the different bacterial diversities in physical ecosystems using 16S rRNA-based and independent culture studies approach where bacterial lineages were found without cultivated representatives.

NGS provides a powerful approach by which to explore the structure of bacterial communities [19]. Without prior cultivation, it is possible to get information about bacteria (phyla and family) from environmental samples, whereas NGS is a diagnostic method for direct detection [20]. In addition, NGS can be used for metagenomics studies [21] and assessing the presence of microbiota in large rivers [12]. NGS methods could be a valuable tool in various water quality monitoring applications. They are useful for the development and assessment of molecular diagnostic tools to find the abundant bacterial indicators in water resources [22].

Season-wise (e.g., winter, pre-monsoon, monsoon and post-monsoon) the water cycle in the Bangladesh region of the Bay of Bengal varies and shows different characteristics [5]. There are many people living around this area and, as such, this zone is controlled by both natural and human pressures. Most importantly, the diversity and dynamics of bacterial communities in this area are still not known [8].

We directed this research to examine the distribution of bacterial communities and the impact of water temperature. To determine the bacterial distribution through the region, Illumina next-generation sequencing analyses [23–25] have been used to identify bacteria from water samples in the study region.

2. Methodology

2.1. Collection of Water Samples

Sampling sites were selected across the northern part of the Bay of Bengal, south of Bangladesh (Figure 1), based on geographical patterns and climate. The samples were collected from surface water and sub-layer water depth. The water samples were collected using a water sampler. Water (0.25–200 m depth) was selected for this study during the monsoon period (temp: 25–40 °C) from five geographically independent area (the western, middle and eastern sides of southern Bangladesh). These areas are different in terms of climate. Each water sample was bottled on the same day 0.5-L bottles. Water bottles were transported to the Laboratory of Microbiology, Jahangirnagar University, Dhaka, Bangladesh within 24 h and stored under dark conditions at 20 ± 2 °C. We collected water samples from different areas and transported them via Eskey, in which environmental temperature is maintained. After arrival of samples at the laboratory, we processed them as quickly as possible under an aseptic condition in the bacterial culture media, then incubated them overnight. On the following day, we picked a single colony from the culture plate and kept it in a bacterial culture broth for large growth; this was left overnight in a water bath shaker at 37 °C, then further processed according to bacterial concentration, quantitative polymerase chain reactions (qPCR) and NGS.
Figure 1. Bacterial diversity (as number) over the Bay of Bengal region, Bangladesh. The highest and lowest numbers of bacteria are shown as red and blue, respectively.

2.2. DNA Extraction and Next-Generation Sequencing

Firstly, water samples (500 mL each from five sites) were collected from the selected regions in the Bay of Bengal. These samples were then sent to the laboratory and concentrated. The DNA extraction process (the DNA kit was formulated to isolate genomic DNA from water samples) continued about 3–4 h, then bacterial cells were collected. The concentration and purity of the template DNA were measured by a spectrophotometer). On average, 270 ng/µL of DNA were extracted from each sample and placed directly on the spectrophotometer.

The following steps were used for DNA extraction, following the method by [26]:

- a tube (2 mL) containing glass beads (ø0.1 and ø0.5 mm) was selected to store the water samples (membranes in 70% ethanol);
- suspension was centrifuged (at 8000 g) for 10 minutes;
- after centrifugation, the supernatant was redundant and the pellet was re-proscribed in pure water (200 µL);
- the glass bead tube containing the membrane received resuspension;
- phenol-chloroform protocol was followed to perform the DNA extraction process as previously described by [13];
- isopropanol containing 1 µL glycogen precipitated the template DNA at −20 °C for 30 minutes;
- the pellet was washed with 70% ethanol after centrifugation at 15,000 g for 15 minutes;
- after drying, the template DNA was finally dissolved in 30 µL TE buffer, and the concentration and purity of the template DNA were measured by a spectrophotometer (NanoDrop 2000, Thermo Fisher Scientific, Waltham, MA, USA) then stored at −20 °C for further analysis.

The bacterial concentration was calculated by considering all the dilution steps during the sample preparation and expressed as copies number per mL. A t-test was applied to determine the significance difference of the total 16S rRNA gene copy number in the water. The V3 hypervariable region of the 16S rRNA gene was amplified by nested PCR.

Only primers with an overall coverage above 75% for Bacteria were considered for primer pair analysis. All primers are available in probeBase, a comprehensive online database for rRNA-targeted oligonucleotides (www.microbial-ecology.net/probebase). PCR primer pairs had previously been validated in the literature [27,28], and we used 341F- 5’CCTACGGGAGGCAGCAG3’ and 534R-5’ATTACCG CGGCTGCTGG3’.
Next-generation sequencing protocol were been followed for preparing the gene libraries, and the samples were quantified by the real-time quantitative PCR at a 95% confidence level. The preparation for this experiment followed the study of [12], and the concentration of bacteria was calculated following the dilution steps during the preparation of samples.

The V3-V4 hypervariable region of the bacterial 16S rRNA gene (460-bp size on average) was amplified using primers. Each water bottle contained a distinct community structure at the phylum, class, genus and operational taxonomic unit (OTU) levels one day after bottling, and were class-dominated by γ-Proteobacteria. At the genus level, the dominant genera were Acinetobacter. A total of 141,497 qualified reads (ranging from 19,241 to 34,323 per DNA sample) were recovered. The predominant phylum was the Proteobacteria (43.6% of total 16s rRNA gene sequences).

On the Illumina MiSeq platform, samples were paired-end sequenced (low-quality sequences were removed) at a read length of 300 nt. Amplification and sequencing were achieved under negative control.

2.3. Spatial Modelling and Geographical Weighted Regression for Bacterial Distribution

A spatial model has been developed through hexagons which shows a number of grids in an area. In this research, we used hexagon tessellates, because many hexagons will cover a surface without overlapping or leaving any gaps. The most basic reason that regular hexagons tessellate is that the degree measure of the angle at each vertex is a divisor of 360.

Geographically weighted regression (GWR) is a technique that analyzes the variables in spatially and models their local relationships [29,30]. GWR identifies spatial heterogeneities in regression models [31–34] of geo-referenced data. The development of GWR includes a number of techniques like local regression and smoothing, spatial autocorrelation, kernel bandwidths, generalized linear modelling and Bayesian GWR [35]. It is a combination of a generalized linear model and a standard linear model (SLM). SLM assumes that the response $y$ is normally distributed for its expected value $\mu$ with constant variance $\sigma^2$

$$y \sim N(\mu, \sigma^2)$$

while in turn, $\mu$ can be expressed as a linear combination of $p$ predictor variables $x_1, x_2, \cdots, x_p$

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$

where $\beta_0, \beta_1, \cdots, \beta_p$ are regression coefficients.

The GWR extends the linear model

$$\mu = \beta_0 W_0 + \beta_1 x_1 w_1 + \beta_2 x_2 w_2 + \cdots + \beta_p x_p w_p.$$

where $W$ is a matrix of location-specific weights.

In this paper, the relationship between bacterial diversity and water temperature across the set of hexagons is shown, with providing a smooth fit.

2.4. Model Validation

We examined the residuals to assess the model’s fitness. The difference between observed and predicted diversities determined the residuals in each hexagon, and residuals above or below zero indicated that the model under- or over-predicted the diversity, respectively.

3. Results

3.1. Identification of Bacteria

A number of five geographically independent areas (the western, middle and eastern sides of southern Bangladesh) were selected to identify the bacteria. The highest abundance of bacterial
groups found were based on the methodology described in Section 2, namely, Acidobacteria [N1], Bacteroidetes [N2], Actinobacteria [N3], Cyanobacteria [N4], Chlorobi [N5], Firmicutes [N6], Nitrospirae [N7], Chloroflexi [N8], Proteobacteria [N9], Planctomycetes [N10] and Verrucomicrobia [N11] (Table 1). This bacteria are included in a spatial model to observe the influence of water temperature for their distribution.

Table 1. Distribution of bacteria in survey site. N1 … N11 denote the name of the bacteria.

<table>
<thead>
<tr>
<th>Survey Zone</th>
<th>Name of Bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone-1 (Western side-Khulna)</td>
<td>N1, N3, N2, N4, N6, N7, N10, N9, and N11.</td>
</tr>
<tr>
<td>Zone-2 (Western side-Satkhira)</td>
<td>N1, N4, N7, N9, N10, and N11.</td>
</tr>
<tr>
<td>Zone-3 (Middle side-Bhola)</td>
<td>N1, N3, N2, N5, N8, N4, N6, N7, and N10.</td>
</tr>
<tr>
<td>Zone-4 (Eastern side-Chittagong)</td>
<td>N1, N2, N3, N5, N4, N7, N10, N9, and N11.</td>
</tr>
<tr>
<td>Zone-5 (Eastern side-Cox’s Bazar)</td>
<td>N1, N3, N2, N4, N5, and N11.</td>
</tr>
</tbody>
</table>

We created a spatial data framework in a statistical ‘R’ environment that tessellated the study area with hexagons and populated them with bacterial diversity. Hexagon tessellation was constructed using grid points with hexagon boundaries around each one. The area of each hexagon depended on the number of points (for example, more points means less hexagons). We obtained the number of observations per hexagon. Some hexagons showed many bacteria, while others contained few. The numbers of bacteria are shown in Figure 1. The highest numbers were observed in four hexagons, while the lowest was seen in six hexagons (Figure 1).

3.2. Spatial Modelling of Bacteria Distribution and Water Temperature

Several studies have attempted to discover the relationship between bacterial distribution [3,36] and water temperature, but so far we are unaware of any studies that have attempted to perform an analysis for the selected Bay of Bengal region, Bangladesh, using a spatial model. For this reason, we attempted to investigate the relationship between bacterial distribution and water temperature by developing hexagons through a geographically weighted regression model. Firstly, we added water temperature data in each hexagon. Temperatures exceeding 29 °C were seen mostly in the study area. Figure 2 shows the relationship between bacterial distribution and water temperature in each hexagon, revealing the average water temperature within each hexagon. Water temperature and bacterial number data were analyzed and modelled to create a spatial polygon data frame. Importantly, the spatial information allowed relationships to be analyzed on a map. Here, the map displayed hexagons that were colored according to groups defined by a two-way table of bacterial diversity and water temperature using above- and below-median values. The median water temperature and number of bacteria calculated from the data were 28.2 °C and two, respectively. Red and blue hexagons indicated regions of high or low diversity and high or low water temperature, respectively. Magenta hexagons showed low diversity coupled with a relatively high water temperature, indicating regions with ‘under-performing’ bacteria numbers. Cyan hexagons showed high diversity coupled with relatively low temperatures, indicating regions with ‘over-performing’ bacterial numbers.

Relationships were detected between bacterial diversity and water temperature through the spatial GWR model, in particular with the local regression and the relationship between bacterial diversity and water temperature changes for different water temperature values (Figure 3).

To assess the model fit, we examined the residuals. A residual was the difference between the observed and predicted diversity in each hexagon. Residuals above zero indicated that the model under-predicted the diversity, whereas residuals below zero indicated that the model over-predicted diversity. A histogram (Figure 4) of the residuals showed the values were centered on zero. The left tails indicate that the model over-predicted diversity in some hexagons. The over-prediction can be associated with other environmental factors.
Figure 2. Water temperature and bacterial diversity over the Bay of Bengal region, Bangladesh. Low and high water temperatures with low and high bacterial diversity regions are shown in blue and red, respectively, whereas low and high water temperatures with high bacterial diversity are shown in cyan.

Figure 3. Linear (black line) and local linear regressions (red line) between bacterial diversity and water temperature. The relationship between bacterial diversity (x-axis) and water temperature (y-axis) changes for different values of water temperature.

Figure 4. Histogram of geographically weighted regression (GWR) residuals. A residual (x-axis) is the difference between the observed and predicted diversity number (y-axis) in each hexagon. Residuals above or below two indicate that the model under-predicted or over-predicts the diversity, respectively.
4. Discussion and Conclusions

The bacterial community in the selected regions of the Bangladesh region of the Bay of Bengal experienced eutrophication and stressors from the biological and chemical pollutants due to anthropogenic pressure. On the other hand, the physical singularity of the water cycle in this region had several benefits [2]. High abundances of Acidobacteria, Bacteroidetes, Actinobacteria, Cyanobacteria, Chlorobi, Firmicutes, Nitrospirae, Chloroflexi, Proteobacteria, Planctomycetes and Verrucomicrobia were found in the water samples in the study area. These bacteria play an important role in maintaining the region; for example, Actinobacteria maintains the quality of the water by molding organic carbon and remediating pollutants, while Cyanobacteria keeps the water environment safe [5].

The diversity of the bacterial community in the Bay of Bengal, Bangladesh, was quantified in this study according to bacterial concentration, qPCR and NGS. After extraction, the raw sequences were checked for initial quality and reads in fast QC pipelines. For the quality trimming of reads, Sickle software was used. Assembling of 16S rRNA sequences into operational taxonomic units (OTUs) and removal of singleton OTUs were performed using MICCA OTU [37]. We used a spatial model through a geographically weighted regression approach to fit the bacterial diversity into hexagons. The distribution of bacterial diversity and water temperature was approximated using hexagon tessellation. The relationship between bacterial diversity and water temperature was determined through GWR. Furthermore, GWR was used to predict the diversity of the study area. Calculation of residuals was applied to assess the model’s fitness.

We allocated a number of 13 hexagons over the study area to determine the relationship between water temperature and bacterial diversity. We found that the highest 31% of hexagons consisted of a high water temperature and high diversity. On the other hand, low water temperature and low intensity were seen in 23% of hexagons. The model fitness was assessed by calculating residuals.

There is currently no spatial model work available utilizing hexagons to determine diversity in the Bay of Bengal, Bangladesh, so a statistical spatial forecast model using a geographical weighted regression approach was applied here and found that this model has potential in forecasting bacterial diversity.


Funding: This research received no external funding. M. Wahiduzzaman has been supported by a Postdoctoral Fellowship from Nanjing University of Information Science and Technology (NUIST), China. A Jahangirnagar University scholarship supported Nabila Akter and data are available from her upon request. Jing-Jia Luo was supported by “The Startup Foundation for Introducing Talent of NUIST”.

Acknowledgments: Authors are grateful to two anonymous reviewers for their constructive review that significantly improved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References


32. Wahiduzzaman, M.; Yeasmin, A. Statistical forecasting of tropical cyclone landfall activities over the North Indian Ocean rim countries. *Atmos. Res.* **2019**, *227*, 89–100. [CrossRef]


