Sewage-specific enterococcal bacteriophages and multiple water quality parameters for coastal water quality assessment
Akechai Kongprajug, Namfon Booncharoen, Kanyaluck Jantakee, Natcha Chyerochana, Skorn Mongkolsuk and Kwanrawee Sirikanchana

ABSTRACT
Coastal water quality is deteriorating worldwide. Water quality monitoring is therefore essential for public health risk evaluation and the management of water bodies. This study investigated the feasibility of using bacteriophages of Enterococcus faecalis as sewage-specific faecal indicators, together with physicochemical (dissolved oxygen, pH, temperature and total suspended solids) and biological parameters, to assess coastal water quality using multivariate analysis incorporating non-detects. The principal component and cluster analyses demonstrated that coastal water quality was mostly influenced by biological parameters, including Escherichia coli and total coliforms, which were found in all 31 sampling sites, and enterococci, which was found in all but two sampling sites. The enterococcal bacteriophages AIM06 and SR14 were detected in 17 and 18 samples at concentrations up to 1,815 and 2,790 PFU/100 mL, respectively. Both bacteriophages co-presented in approximately 80% of phage-positive samples, and the concentrations at each site were not significantly different. Overall, either bacteriophage could be used to differentiate high- and low-level coastal water pollution, as grouped by cluster analysis. This study is the first to investigate the suitability of sewage-specific bacteriophages of E. faecalis for monitoring coastal water quality and emphasises the importance of a multivariate analysis with non-detects to facilitate coastal water quality monitoring and management.

Key words | bacteriophage of enterococci, coastal water quality, microbial source tracking, multivariate analysis, non-detects, sewage pollution

INTRODUCTION
Coastal water pollution has become a significant problem in both developed and developing countries. Water quality deterioration can affect coastal water use, such as coastal aquaculture, fisheries, recreation and tourism. Diffuse sources from different land use adjacent to coastal areas can aggravate water quality deterioration already impacted by point sources (Molina et al. 2014; Yau et al. 2014). Monitoring coastal water quality can not only assess the risk to public health, but also identify the inefficient treatment of wastewater from communities or agriculture (Oun et al. 2017). Because a range of pollutants can contaminate receiving water bodies, water quality standards in many countries include multiple water quality parameters, including physicochemical and biological parameters (National Environment Board 1994; US EPA 2012). However, the assessment of physicochemical parameters is often insufficient for tracking the exact sources of pollution. Multivariate analysis has been proven appropriate for retrieving information about and associations among water quality parameters (Wangkahad et al. 2011). However, environmental monitoring frequently involves negative detections, meaning that the data are below the detection limits of the methods or instruments used; so-called non-detects, or censored data. Importantly, substitution of negative detections with zero or with the detection limits was reported to cause significant biases in subsequent data analysis; thus, this practice should be discouraged (Helsel 2012). An alternative analytical method of datasets containing non-detects is nonparametric survival
analysis, which processes the data based on the relative positions of the data (Helsel 2012; Wangkahad et al. 2017).

In addition to groups of physicochemical and biological parameters that reflect water quality, another set of biological parameters could differentiate pollution sources. Microbial source tracking (MST) is a frontier technology that utilizes intestinal microorganisms specific to each animal type (e.g., human, pig, cattle, fowl) to track sources of faecal contamination in polluted water. MST indicators provide information about pollution sources that could facilitate the management and restoration of polluted water bodies (Teaf et al. 2018). Bacteriophages of enterococci are bacterial viruses that specifically infect their enterococcus bacteria. They have been successfully used to track human sewage pollution in many countries, such as the UK (Purnell et al. 2018, 2011), Puerto Rico (Santiago-Rodriguez et al. 2010), and Thailand (Wangkahad et al. 2017; Booncharoen et al. 2018). The detection technique is simple and cost-effective, requiring only the culturing of the bacteriophage using a double-layer agar overlay (Wangkahad et al. 2018). Sewage-specific enterococcal bacteriophages AIM06 and SR14 isolated in Thailand are highly specific and sensitive, and have performed well as MST indicators in freshwater (Wangkahad et al. 2017); however, information about their performance in seawater is limited and, therefore, urgently needed. The objective of this study was to evaluate the performance of sewage-specific bacteriophages of enterococci for assessing coastal water quality using multivariate analysis incorporating non-detects, which used nonparametric survival analysis procedure.

MATERIALS AND METHODS

Water sample collection and quality control

A total of 31 sampling stations were located along the upper part of the Gulf of Thailand (Figure 1 and Table S1, available with the online version of this paper), from Trat province to Prachuab Kirikan province, matching the sampling stations of Thailand’s Pollution Control Department. From June to July 2015, at low tide, 3 L water samples were collected at

Figure 1 | Locations of coastal water sampling sites along the upper part of the Gulf of Thailand.
30 cm below the sea surface for biological parameters and 1 L was collected at the mid-depth of the water column for total suspended solids (TSS). The samples were stored on ice in sterile bottles before transportation back to the laboratory within 1 day. Dissolved oxygen (DO), pH, and water temperature were measured on site using YSI Pro2030 and YSI60 devices (YSI, Yellow Springs, OH, USA), according to the manufacturer’s instructions. Reproducibility of sample preparation and analysis was assessed using field duplicates at two sites, representing low- and high-level microbial pollution (i.e., sites AA and AE, respectively). Three field blank and three laboratory blank samples were also processed. Field blanks were prepared by transporting autoclaved reverse osmosis (RO) water to the field site and then following a protocol similar to that used for the sample collection. Laboratory blanks were prepared by following a protocol similar to that used for processing the sample at the laboratory with autoclaved RO water.

**Physicochemical and biological parameters**

TSS was analysed using the drying method at 103–105 °C, according to standard protocols (American Public Health Association et al. 2017a). Total coliforms and *Escherichia coli* were simultaneously measured using a membrane filtration method with MI medium, containing 4-methylumbelliferyl-β-d-galactopyranoside (MUGal) and indoxyl-β-d-glucuronide (IBDG) (US EPA 2002a). Enterococci were detected with a membrane filtration method using membrane-*Enterococcus* Indoxyl-Beta-d-Glucoside Agar (mEI) (US EPA 2002b). Plates were incubated at 37 °C for up to 24 hours, and the number of colony-forming units (CFU) was recorded separately.

**Sewage-specific bacteriophages**

Two litres of water sample were centrifuged at 4,000 g for 20 min to separate the supernatant and suspended solids. The supernatant was subsequently filtered with a 0.45 μm pore size polyvinylidene fluoride (PVDF) membrane filter to remove bacteria and suspended solids without removing the bacteriophage (Booncharoen et al. 2018). One litre each of the resulting filtrate was filtered with a 0.45 μm cellulose nitrate filter (CNF) to retain the bacteriophage on the membrane. Each membrane filter was placed face down on the bacterial lawn of *Enterococcus faecalis* strain AIM06 or strain SR14 (DSM100702 or DSM100701) in tryptic soy agar (Wangkahad et al. 2017, 2015). The plates were incubated at 37 °C for 24 hours before the membrane was removed and plaques were counted. Additionally, the sediment pellet resulting from the previous centrifuging step was eluted with 50 ml of the eluent solution (American Public Health Association et al. 2017b). The suspension was centrifuged at 4,000 g for 20 min, and the pellet was discarded. Each 8–12 mL sample of the supernatant was filtered with a 0.45 μm CNF membrane to retain the bacteriophage on the membrane. Following this protocol, the membrane filter was plated and plaques were counted. The number of plaque-forming units (PFU) for each water sample was the sum of the number from the supernatant portion and the number from the eluate of the sediment portion. To measure the recovery percentage of this bacteriophage detection method, two 1 L representative autoclaved water samples (Sites AF and AG as representatives of high and low bacterial contamination, respectively; Tables S1 and S2, available online) were spiked with 20,000 PFUs of bacteriophages A1 and S4 (Booncharoen et al. 2018), representative of bacteriophages infecting *E. faecalis* strains AIM06 and SR14, respectively. The resulting bacteriophages were measured after the water samples were processed as described above. The recovery percentage was calculated as the fraction of bacteriophages recovered.

**Statistical analysis**

Data analysis of nine parameters, including DO, pH, temperature, TSS, total coliforms, *E. coli*, enterococci, the AIM06 phage and the SR14 phage, was conducted for the 31 samples. The datasets for TSS, enterococci, the AIM06 phage and the SR14 phage contained values lower than the detection limits, or non-detects, requiring statistics for censored data (Helsel 2012). Statistical analysis was performed in the R program (R Core Team 2017). The descriptive statistics for the datasets containing non-detects were calculated using Kaplan-Meier estimates. Normality was tested using the maximum likelihood estimation method, using an adjusted Anderson-Darling value for datasets with non-detects and the Shapiro-Wilk test for those without non-detects. The Mann-Whitney test and Student's t-test were used to compare the two non-normal and normal datasets, respectively. Paired comparisons were conducted for datasets containing non-detects using the generalized Wilcoxon test. All correlation analyses were performed using the non-parametric Kendall's tau analysis using the psych package in R. Multivariate analysis was achieved by first assigning U-score ranks to both censored and uncensored data. Then, a principal component analysis (PCA) was conducted to explore the patterns of multiple variables, and a cluster analysis was performed using Euclidean distance.
measures with a k-means algorithm to cluster the group of sampling sites, using the FactoMineR and factoextra packages in R. The optimal number of clusters estimation was computed using the NbClust package in R.

RESULTS AND DISCUSSION

Physicochemical and biological water quality parameters

A total of nine water quality parameters were monitored at 31 sampling sites along the coastal area of Thailand (Table 1). Low DO concentrations (below 4.0 mg/L) were found mostly at sites near estuaries and open markets for fresh produce, meat and seafood. E. coli and total coliforms were found at more than 50,000 CFU/100 mL at sites near an estuary, recreational beaches, an open market, and piers, while the highest enterococci (5,750 CFU/100 mL), AIM06 phage and SR14 phage concentrations (1,815 and 2,790 PFU/100 mL, respectively) were detected at a pier. TSS and enterococci datasets contained one and two data points with negative test results, respectively. AIM06 and SR14 phages were detected in 55% (17/31) and 58% (18/31) of the total samples. Both bacteriophages were simultaneously detected in 14 samples, constituting

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Descriptive statistics of physicochemical and biological parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameters without non-detects</strong></td>
<td><strong>Parameters with non-detects</strong></td>
</tr>
<tr>
<td>DO (mg/L)</td>
<td>pH</td>
</tr>
<tr>
<td>Number of values (n)</td>
<td>31</td>
</tr>
<tr>
<td>Normality testb</td>
<td>0.033</td>
</tr>
<tr>
<td>Passed normality test</td>
<td>No</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.29</td>
</tr>
<tr>
<td>25th percentile</td>
<td>3.41</td>
</tr>
<tr>
<td>Median</td>
<td>6.07</td>
</tr>
<tr>
<td>75th percentile</td>
<td>6.70</td>
</tr>
<tr>
<td>Maximum</td>
<td>8.81</td>
</tr>
<tr>
<td>Mean</td>
<td>5.19</td>
</tr>
<tr>
<td>SD</td>
<td>2.16</td>
</tr>
</tbody>
</table>

*Concentration after multiplying by recovery percentages.

<table>
<thead>
<tr>
<th>Site ID/statistics</th>
<th>TSS (mg/L)</th>
<th>E. coli (CFU/100 mL)</th>
<th>Total coliforms (CFU/100 mL)</th>
<th>Enterococci (CFU/100 mL)</th>
<th>AIM06 phage* (PFU/100 mL)</th>
<th>SR14 phage* (PFU/100 mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA1</td>
<td>23.0</td>
<td>2,560</td>
<td>4,600</td>
<td>30</td>
<td>2.6</td>
<td>3.4</td>
</tr>
<tr>
<td>AA2</td>
<td>29.0</td>
<td>2,480</td>
<td>3,200</td>
<td>32</td>
<td>5.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Mean</td>
<td>26.0</td>
<td>2,520</td>
<td>3,900</td>
<td>31</td>
<td>3.8</td>
<td>3.4</td>
</tr>
<tr>
<td>SD</td>
<td>4.2</td>
<td>56.6</td>
<td>990.0</td>
<td>1.4</td>
<td>1.8</td>
<td>0.0</td>
</tr>
<tr>
<td>%CV</td>
<td>16.3</td>
<td>2.2</td>
<td>25.4</td>
<td>4.6</td>
<td>47.1</td>
<td>0.0</td>
</tr>
<tr>
<td>AE1</td>
<td>&lt;2.5</td>
<td>&gt;50,000</td>
<td>&gt;50,000</td>
<td>18,800</td>
<td>938.8</td>
<td>1,279.5</td>
</tr>
<tr>
<td>AE2</td>
<td>&lt;2.5</td>
<td>&gt;50,000</td>
<td>&gt;50,000</td>
<td>12,700</td>
<td>1,377.6</td>
<td>1,319.9</td>
</tr>
<tr>
<td>Mean</td>
<td>NAa</td>
<td>NA</td>
<td>NA</td>
<td>15,750</td>
<td>1,158.2</td>
<td>1,299.7</td>
</tr>
<tr>
<td>SD</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>4,313.4</td>
<td>310.3</td>
<td>28.6</td>
</tr>
<tr>
<td>%CV</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>27.4</td>
<td>26.8</td>
<td>2.2</td>
</tr>
</tbody>
</table>

*NA: not applicable.
78–82% of the phage-positive samples. The abundance of AIM06 and SR14 phages in seawater in this study is comparable to their previously reported concentrations in human sewage and freshwater at 2–4 and 1–3 orders of magnitude, respectively (Wangkahad et al. 2017). Both bacteriophages in treated wastewater effluents were not detectable, i.e., below a detection limit of 33 PFU/100 mL (Wangkahad et al. 2017), implying that the high abundance of bacteriophages in seawater as shown in this study could have derived from untreated sewage.

**Quality control and bacteriophage recovery**

The reproducibility of the laboratory measurement was investigated using field duplicates of two sampling sites for TSS and microbiological parameters (Table 2). The coefficient of variation (CV) showed values mostly up to 27.4%, with the exception of the AIM06 phage at site AA, for which the CV was 47.1% due to statistical variations in low PFU numbers. Moreover, the three field blanks and three laboratory blanks showed no false detection of any microbial or physicochemical parameters. Furthermore, the recovery percentages for the bacteriophage detection method were 3.92 ± 0.08% and 2.97 ± 0.31% for the AIM06 and SR14 phages, respectively. Although the recovery ratios were lower than those of other methods, this approach required less time and no elution step (Contreras-Coll et al. 2002; Méndez et al. 2004). The calculated recoveries demonstrated low variation between water samples with high and low bacterial contamination (Sites AF and AG; Table S2). Therefore, the average recovery percentages were used to calculate the actual concentrations of both bacteriophages in all water samples.

**Correlations among physicochemical and biological parameters**

Among the physicochemical parameters, DO and pH were moderately correlated \( (r = 0.48; p < 0.01) \) (Figure 2(a)). Among the biological parameters, *E. coli* showed significant correlations \( (p < 0.01) \) with total coliforms and enterococci \( (r = 0.58 \) and 0.47, respectively) (Figure 2(b)). Such correlations among bacterial indicators have also been reported in freshwater (Sirikanchana et al. 2014; Wangkahad et al. 2017, 2015). Moreover, the present study showed a strong correlation between the AIM06 and SR14 phages in seawater \( (r = 0.60; p < 0.001) \). This result agrees with previous findings for polluted freshwater, and both bacteriophages have been suggested for use as water quality parameters in such water (Wangkahad et al. 2017). In this study, the nonparametric paired Prentice-Wilcoxon test indicated no significant difference between the abundance of AIM06 and SR14 phages in all seawater samples \( (p > 0.1) \). Consequently, it is suggested that one of the two bacteriophages could be selected to monitor coastal water quality.
Multivariate and cluster analyses

PCA was performed to convert a set of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCs). The first four PCs showed the eigenvalues of 3.0551, 1.787, 1.2557 and 1.0542, respectively. An eigenvalue of more than 1 indicates that a PC accounts for more variance of all the variables than accounted for by one of the original variables. The cumulative percentage of variance of the first two PCs was 53.8%. The first PC displayed the highest variance of 33.9%, while the second PC showed 19.9% of the variance (Figure S1, available with the online version of this paper). The variables that heavily contributed to the first PC (>11.1% contribution) were enterococci, *E. coli*, AIM06 phage and total coliforms (Figure 3(a)). The variables that heavily contributed to the second PC were temperature, DO, pH and the SR14 phage (Figure 3(b)). In addition, the

![Graphs showing contributions of variables to PCs and cluster analysis results.](https://iwaponline.com/wst/article-pdf/79/5/799/562014/wst079050799.pdf)

**Figure 3** | Multivariate and cluster analyses. The contribution of the variables was extracted to the first PC (a) and the second PC (b). A PCA (c) and cluster analysis of water sampling sites (d) were constructed using Euclidean distance measures with a k-mean algorithm.
variables were classified into separate groups using the k-means clustering algorithm. The optimal number of clusters estimation revealed that two groups were optimal for this dataset (Figure S2, available online). Two groups were clearly classified into physicochemical and biological clusters (Figure 3(c)). The results indicated that both the AIM06 and SR14 phages were associated with the other biological indicators. Furthermore, the cluster analysis could classify the water samples into two groups, comprising 13 sites for the highly contaminated group and 18 sites for the low-contamination group (Figure 3(d)). Cluster centroids indicated that the categorisation of different levels of contamination was mainly affected by the biological parameters, including bacteria and bacteriophage indicators (Table S3, available online).

Comparison of water quality parameters between water samples with high and low levels of contamination

The physicochemical and biological parameters of the two groups with high- and low-level contamination were further analysed (Table S4, available online). DO and pH in the low-contamination group were significantly higher than in the highly contaminated group ($p = 0.0039$ for pH and $p = 0.0156$ for DO) (Figure 4(a) and 4(b)). Temperature and TSS were not significantly different between the two groups, implying that these parameters might not be major parameters indicating coastal water quality. On the other hand, all biological parameters were major parameters influencing coastal water quality. The highly contaminated group had a median $E. coli$ concentration of 5,900 CFU/100 mL, which was significantly higher than that of the low-contamination group (1,035 CFU/100 mL; $p = 0.0003$) (Figure 4(c)). A similar trend was observed for total coliforms, for which the highly contaminated group had a median concentration of 24,000 CFU/100 mL, compared to 3,400 CFU/100 mL for the low-contamination group ($p = 0.0048$) (Figure 4(d)).

There were also statistical differences between the groups with high and low levels of contamination for parameters containing non-detected data, comprising enterococci, the AIM06 phage and the SR14 phage. The results showed that the median concentration of enterococci in the highly contaminated group was 1,080 CFU/100 mL, whereas that in the low-contamination group was 81 CFU/100 mL ($p < 0.001$) (Figure 5(a)). The median concentrations of the AIM06 phage were 20.4 PFU/100 mL and below the detection limit (<2.6 PFU/100 mL) in the high- and low-contamination groups, respectively ($p < 0.001$) (Figure 5(b)). The median concentrations of SR14 were 13.5 PFU/100 mL and at the detection limit level (3.4 PFU/100 mL) in the groups with high and low levels of contamination,

![Figure 4](https://iwaponline.com/wst/article-pdf/79/5/799/562014/wst079050799.pdf)
respectively ($p = 0.001$) (Figure 5(c)). Although bacteriophages of enterococci have not been monitored in seawater in other geographical regions, sewage-specific enterococcal bacteriophages ENT-49 and ENT-55 were detected at freshwater beaches and harbours at 1–3 orders of magnitude (Vijayavel et al. 2014), similar to those found for AIM06 and SR14 phages in this study. The results from this study indicated that both bacteriophages, as well as other bacterial indicators, could be used to assess levels of pollution in coastal water. Because many faecal sources can contribute to deteriorating coastal water quality, the use of microbial source tracking markers to specify human-sewage-specific pollution is, therefore, very beneficial for the management of polluted water bodies (US EPA 2011).

**CONCLUSIONS**

The sewage-specific AIM06 and SR14 phages were evaluated for their suitability as coastal water quality indicators. Both bacteriophages were present at similar concentrations and were co-present in approximately 80% of the phage-positive samples. The results of this study indicate that microbial contamination is the predominant impact affecting coastal waters along the Gulf of Thailand. This study demonstrated that the sewage-specific AIM06 and SR14 phages were associated with total coliforms, *E. coli* and enterococci. Both bacteriophages could be used to differentiate coastal water with high levels of contamination from coastal water with low levels of contamination.

**ACKNOWLEDGEMENTS**

This research was financially supported by the Chulabhorn Research Institute. Miss Sirintip Ratthakit is gratefully acknowledged for her assistance in sampling and laboratory analysis.

**REFERENCES**


Booncharoen, N., Mongkolsuk, S. & Sirikanchana, K. 2018


First received 30 July 2018; accepted in revised form 22 October 2018. Available online 31 October 2018.