

# Time after time: detecting annual patterns in stream bacterial biofilm communities

Anju Gautam , Gavin Lear  and Gillian D. Lewis\*

School of Biological Sciences, The University of Auckland, Auckland, 1010, New Zealand.

## Summary

**To quantify the major environmental drivers of stream bacterial population dynamics, we modelled temporal differences in stream bacterial communities to quantify community shifts, including those relating to cyclical seasonal variation and more sporadic bloom events. We applied Illumina MiSeq 16S rRNA bacterial gene sequencing of 892 stream biofilm samples, collected monthly for 36-months from six streams. The streams were located a maximum of 118 km apart and drained three different catchment types (forest, urban and rural land uses). We identified repeatable seasonal patterns among bacterial taxa, allowing their separation into three ecological groupings, those following linear, bloom/trough and repeated, seasonal trends. Various physicochemical parameters (light, water and air temperature, pH, dissolved oxygen, nutrients) were linked to temporal community changes. Our models indicate that bloom events and seasonal episodes modify biofilm bacterial populations, suggesting that distinct microbial taxa thrive during these events including non-cyanobacterial community members. These models could aid in determining how temporal environmental changes affect community assembly and guide the selection of appropriate statistical models to capture future community responses to environmental change.**

## Introduction

Freshwater streams change over time, including in response to external anthropogenic pressures such as increased urbanization, land-use disturbances and related environmental changes (Jeffries *et al.*, 2015;

Labbate *et al.*, 2016; Qu *et al.*, 2017; Kim *et al.*, 2020). Due to stream water and associated microbial communities largely originating from the catchments they drain, the distribution and abundance of stream bacterial communities might provide insights into stream and catchment level ecosystem functional responses to environmental change (Chapin *et al.*, 2000). Indeed, freshwater microbial communities are highly diverse (Battin *et al.*, 2016), play a significant role in ecosystem functioning and are strongly influenced by changes in surrounding landscapes (Lear and Lewis, 2009; Kraemer *et al.*, 2020). Despite this, most studies describing freshwater ecosystem health to date have focused on communities of invertebrates and plants (Sanderson *et al.*, 2007; Wright and Ryan, 2016; Kim *et al.*, 2020). To understand and predict the impact of environmental changes (e.g. abiotic, biotic, anthropogenic, or other processes; Lear *et al.*, 2008) on bacterial community diversity (Loreau *et al.*, 2001) and ecosystem functioning, we first need a baseline understanding of the extent and nature of temporal patterns influencing their distribution and abundance.

Advances in molecular methods, including high-throughput sequencing techniques for determining bacterial community composition and diversity, based on 16S rRNA gene sequences (Good *et al.*, 2018), provide collective evidence that bacteria exhibit substantial spatial and temporal patterns in their distribution and abundance (Lear *et al.*, 2008; Esposito *et al.*, 2016; Gautam *et al.*, 2020). Highly dominant and diverse bacterial phyla reported among freshwater environments include Proteobacteria, Cyanobacteria, Bacteroidetes, Verrucomicrobia, Actinobacteria and Planctomycetes (Besemer *et al.*, 2012; Battin *et al.*, 2016), but their dominance varies greatly over time and geographic location. For example, a greater prevalence of phyla such as the Cyanobacteria is likely to result from warmer temperatures, reduced flows and high light availability (Salmaso *et al.*, 2018). Investigating the extent of repeating annual patterns for multiple bacterial taxa could reveal their potential to serve widely as predictors of environmental change, both over time and among catchments. Furthermore, it may aid in better prediction of harmful cyanobacterial blooms.

Received 16 September, 2021; revised 3 April, 2022; accepted 13 April, 2022. \*For correspondence. Tel. +64 9 9237396; E-mail [gd.lewis@auckland.ac.nz](mailto:g.d.lewis@auckland.ac.nz).

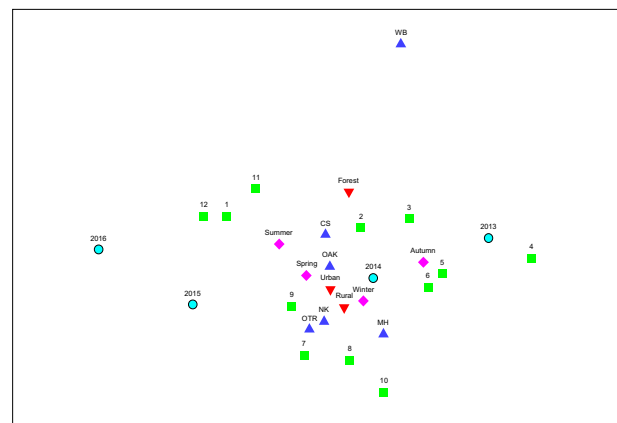
We investigated spatial and temporal patterns in freshwater stream bacterial community composition over multiple years. Building on the work of Gautam *et al.* (2020), we sampled stream biofilm communities monthly from six streams over 3 years (2013–2016), collecting a suite of environmental data from each stream, including (i) instream measurements such as water turbidity, temperature, pH, total suspended solids and dissolved oxygen concentration, (ii) local climate measurements such as air temperature, light and rainfall at each study site and (iii) average soil moisture deficit (i.e. the amount of rainfall, in mm, required to return the catchment soil to field capacity) and concentrations of nutrients such as nitrogen, phosphorus and ammonia in the upstream catchment. First, we used 16S rRNA gene sequencing analysis to reveal bacterial community assemblage patterns. Second, we analysed whether the observed patterns in composition and diversity are repeatable over annual seasonal cycles or if blooms in the abundance of certain taxa occur. Third, we assessed the extent to which environmental factors impact the observed temporal shifts in stream bacterial community dynamics. We set up a model framework to identify bacterial taxa responding to either bloom or seasonal patterns over time. We postulate that freshwater stream bacterial community dynamics can be predicted by seasonal change and provide *a priori* hypotheses for future experiments, facilitating an efficient approach to understand community assembly and succession in bacterial communities across freshwater ecosystems.

## Results

Our analysis generated a total of 123 121 quality-checked 16S rRNA gene sequences. These sequence data were then rarefied to 2000 reads per sample, after which we obtained 774 samples from six different streams. These high-quality sequences were clustered into 21 216 operational taxonomic units (OTUs) at 97% sequence similarity. The bacterial OTUs were classified into 61 phyla, 261 orders, 313 families and 473 genera. The PRIMER RELATE function found a correlation ( $Rho = 0.8$ ) between amplicon sequence variant (ASV) and OTU resemblance matrices at a significance level of 0.01%. This suggests that both methodologies provide broadly similar results (whether obtained from OTU or ASV files) and hereafter we report only the outputs of OTU table analyses. Diversity index analysis (Supplementary Material 1) showed that streams from highly impacted urban catchment of Oakley Creek had significantly higher relative bacterial richness (OTUs) and diversity compared to the streams draining predominantly native catchments (Tukey HSD multiple pairwise comparisons,  $p < 0.05$ ).

## Bacterial community variation (assessed using PERMANOVA models)

Multivariate analysis was used to elucidate and display spatial and temporal variations within the stream biofilm bacterial communities. Permutational multivariate analysis of variance (PERMANOVA) model 1 revealed that year is a significant individual factor correlated with the linear change observed in bacterial community composition ( $p < 0.01$ ) along with stream and months (Supplementary Material 2). The interactions between all the variance factors [ST(CA)xMO(YExSE)] correlate even more strongly with the observed changes in community composition, as shown by their highest square root contribution in this model (Supplementary Material 2). Likewise, PERMANOVA Model 2 (Supplementary Material 3) confirmed significant cyclical patterns in our extended dataset using monthly covariates with variance components (catchment, stream, year). Similar to our previous findings from this model, our extended dataset captured change occurring across repeated annual cycles, by sampling every month, over multiple years. The associations of catchment type, year and stream were also found to correlate significantly with the monthly covariates (Supplementary Material 3; PERMANOVA;  $p < 0.01$ ). An nMDS (Fig. 1) ordination based on relative abundance data was plotted to illustrate the patterns in these data, where large variability in bacterial community composition is observed over the years compared to among streams and catchment types. There was a clear separation of data collected at different seasons and months, including cyclical patterns in the data (i.e. in average monthly data, Fig. 1).



**Fig. 1.** nMDS ordination showing relative relationships among the bacterial community data averaged by (green square) month (where the numbers 1–12 represent the months January to December) respectively, (blue circle) year, (pink rhombus) season, (inverted red triangle) catchment and (blue triangle) stream using Bray–Curtis similarities. 2D Stress value: 0.15.

*Bacterial community diversity and composition*

The highest proportion of OTUs belonged to the phylum Proteobacteria across all samples (32.5%), followed by Planctomycetes (11.2%), Bacteroidetes (9%), Verrucomicrobia (5.5%), Chloroflexi (4.6%), Acidobacteria (4.3%), Actinobacteria (4%), Firmicutes (3.6%), Cyanobacteria (3.6%) and Chlamydiae (1%).

The cyanobacterial order Oscillatoriales was present in all the streams (except for Wairoa and Oakley creek) during summer 2014 and spring 2015, with a high abundance in the rural Mahurangi stream (>90% in October) (Supplementary Material 4). Planctomycetes made up 9% of the community overall (orders Gemmatales, Pirellulales, Planctomycetales) largely from February 2013 to September 2014 and then from May 2015 until the end of the sampling period, with greater representation during autumn and winter. The phylum Firmicutes (orders Bacillales, Lactobacillales, Clostridiales) made up 6.3% of the community. Firmicutes were less abundant in the early sampling period, but their relative abundance increased significantly in the later summer sampling period (December 2015–February 2016). This sudden increase in the order Bacillales appeared to be influenced by a significant increase of pH and dissolved oxygen during that time (Supplementary Material 5). The phylum Bacteroidetes (orders Saprospirales, Bacteroidales, Cytophagales, Flavobacteriales, Sphingobacteriales) contributed 7% of the community and was present in all the streams. Other well-represented phyla present in the dataset were Actinobacteria (4.3%) and Verrucomicrobia (4.2%) (Fig. 2, Supplementary Material 4).

*Temporal patterns in bacterial community composition (time series models)*

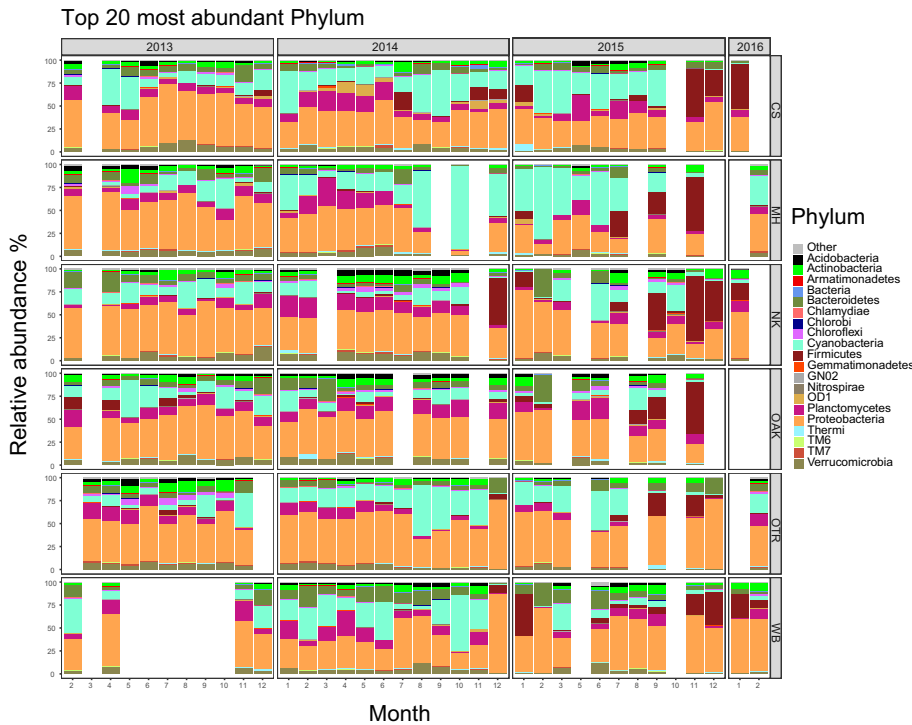
Examination of temporal trends exhibited by the top 50 most abundant bacterial orders using three conceptual time series generalized linear models showed significant differences among the bacterial taxa. In terms of time series model 1 (linear), five bacterial orders fit this model well (Table 1). Twenty-six bacterial orders significantly fit patterns indicative of time series model 2a (bloom) and 2b (trough) occurring sporadically since the beginning of sampling; 19 orders followed significant repetitive seasonal patterns, based on time series model 3 (seasonal). A complete list of orders identified by time series generalized linear model analysis is provided in Table 1.

We further visualized shifts and differences in the abundance of bacterial orders (Fig. 3) relating to each of our time series models by assembling these data into three groups, represented within a shade plot of taxon abundances (Fig. 4). Orders identified from Model 1 (Table 1)

were maximally abundant from 2013 to 2014, decreasing in abundance in late 2015–2016 (Supplementary Material 4, Fig. 3A). We plotted subsets of the bacterial communities recognized by model 2 against months to distinguish between communities developing from blooms and troughs over the years (Fig. 3B). Results from time-series models were similar to the PERMANOVA models examined in the previous section, in terms of differences in the relative abundance of bacterial communities over time (years), where some orders were more or less dominant with environmental and seasonal changes. For example, the orders Bacillales, Actinomycetales, Lactobacillales, Pseudomonadales, Deinococcales, SM2F11 and Nostocales were more abundant in 2014–2016, whereas Saprospirales, Myxococcales, Methylococcales, Stramenopiles were more prominent in 2013 during November, December, January and February. There were distinct differences between autumn and winter bloom communities of 2013, 2014 and 2015 (Fig. 3B). The outcomes of these inferences suggest that large variations exist in the conditions required for different taxa to increase in abundance. Further analysis of seasonal communities from time series model 3 revealed a greater abundance of Flavobacteriales and Enterobacteriales during summer; Acidimicrobiales, Pirellulales, Gemmatales, Clostridiales and Planctomycetales during winter and autumn, and Cyanobacteria, Caulobacteriales and Methylophilales during spring (Fig. 3C).

*Influence of physicochemical parameters on bacterial community composition*

Distance-based linear models (DistLM) analysis was performed to examine the relative influence of physicochemical parameters on subgroups of the bacterial communities identified from analysis of the three time series GLM models. Light, pH, nitrates, total phosphorus and water temperature (Table 2) combined were able to explain 8% (adjusted  $R^2$ ) of the total variation in the composition of the subset of taxa identified by model 1 (linear). The subset of the bacterial community identified by model 2 (bloom/trough) was found to be significantly affected by light, pH, deficit, total nitrogen, water temperature, turbidity, total phosphorus, concentration of dissolved oxygen, total suspended solids, nitrates, ammonia and maximum air temperature (Table 2), which explained 12% (adjusted  $R^2$ ) of the total observed variation. Similarly, light, deficit, pH, total nitrogen, ammonia, soluble phosphorus, nitrates, TKN (total Kjeldahl nitrogen), turbidity, total suspended solids, water temperature and dissolved oxygen were significantly correlated with the subset of the bacterial community identified by model 3 (seasonal), explaining 10% (adjusted  $R^2$ ) of the total variation (Table 2).



**Fig. 2.** Relative abundances of the top 20 most abundant bacterial phyla across our stream biofilm dataset. The barplot is partitioned into six small plots (rows) representing six different streams from the beginning of sampling in February 2013 until February 2016. The x-axis represents time and each bar represents the relative abundance of the top 20 most abundant phyla in each stream. Bars are ordered from left to right by date. Each colour corresponds to a specific phylum.

**Discussion**

Long-term investigations and temporal predictions of stream microbiology require new and improved statistical tools to comprehend temporal variability in freshwater bacterial community composition and potential environmental drivers that influence these patterns. Here, we provide a robust statistical approach for analysing bacterial community compositional shifts over time in freshwater stream biofilms, using data collected monthly. We developed and tested three time series generalized linear models by exploiting stream bacterial communities to investigate temporal patterns (month, year) and seasonal shifts from our monthly sampling period (February 2013–February 2016) for 3 years. Our results indicated that bacterial taxa respond to monthly and seasonal environmental changes, and some may be categorized as bloom and seasonally responsive taxa. Our models may be used to reduce the complexity of identifying differences in bacterial community composition, based on their responsiveness to the changing environment and assist with the prediction of community patterns over the years.

*Freshwater stream biofilm bacterial community composition*

PERMANOVA models used previously (Gautam *et al.*, 2020), but incorporating substantial additional data, confirmed that the bacterial community composition varied

between streams and over time, supporting observations from previous studies (Lear *et al.*, 2008; Hassell *et al.*, 2018). Proteobacteria formed the largest relative fraction of the bacterial community in our dataset; they are recognized as the most common dominant phylum in benthic and hyporheic biofilms overall (Battin *et al.*, 2016). Abundant cyanobacterial communities (mainly Oscillatoriales) in our dataset were correlated with changes in light intensity and nutrients, as similarly observed by others (McCall *et al.*, 2017). We also observed a relatively high abundance of the phylum Firmicutes (order Bacillales) during spring and summer 2015 and 2016 when dissolved oxygen, water temperature and pH were higher, creating ideal conditions for Firmicutes growth (Zhao *et al.*, 2017), since dissolved oxygen is typically elevated during summer (Schmidt *et al.*, 2017). We observed a considerable relative abundance of Planctomycetes (11.2%) in our community dataset despite them being recognized as a less-abundant bacterial phylum in freshwater ecosystems. It has been speculated that composition of Planctomycetes in freshwater ecosystems is associated with seasonal variations (Brümmer *et al.*, 2004) and influenced by parameters including water temperature, pH and nutrient concentrations (Pollet *et al.*, 2011).

*Temporal patterns in bacterial community diversity*

PERMANOVA models and time series generalized linear models designed and tested in this study provide

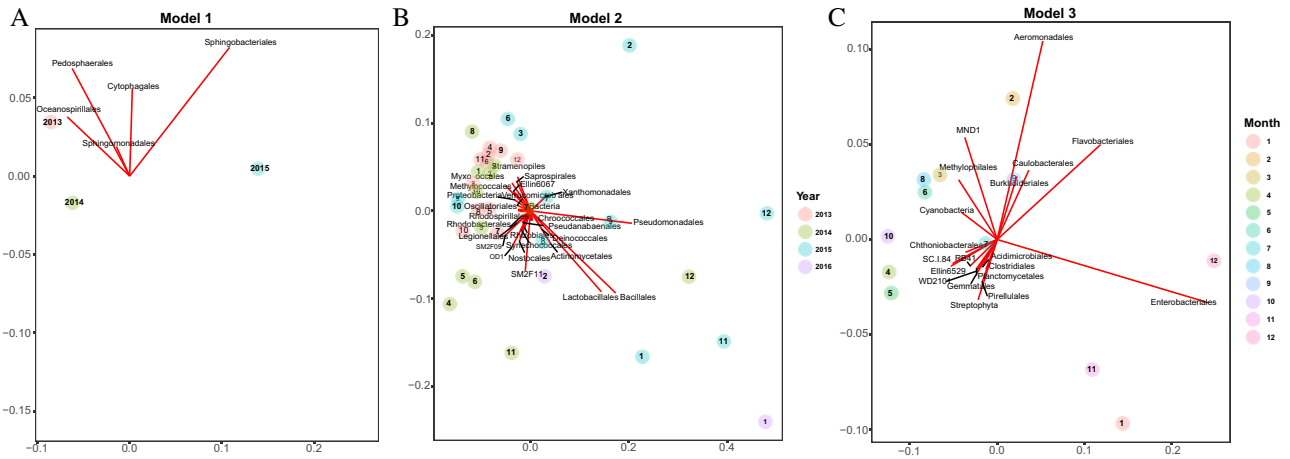
**Table 1.** Summary of time series generalized linear model analysis obtained from three conceptual models assessing the top 50 most abundant stream bacterial orders, based on their relative abundance.

Order	Model 1 AIC	Model 1 (linear) AICcWt	Model 2 AIC	Model 2 (bloom) AICcWt	Model 3 AIC	Model 3 (seasonal) AICcWt
Cytophagales	<b>724.04</b>	<b>0.58</b>	724.85	0.34	727.51	0.08
Sphingomonadales	<b>952.1</b>	<b>0.61</b>	953.96	0.21	954.06	0.18
Sphingobacteriales	<b>465.4</b>	<b>0.48</b>	467.16	0.18	465.62	0.34
Pedospaerales	<b>287.13</b>	<b>0.44</b>	287.42	0.34	288.03	0.22
Oceanospirillales	<b>209.21</b>	<b>0.46</b>	209.83	0.3	210.11	0.23
Rhizobiales	1131.7	0.01	<b>1123.3</b>	<b>0.88</b>	1127.4	0.1
Bacillales	1357.1	0	<b>1330.4</b>	<b>1</b>	1347.5	0
Verrucomicrobiales	760.11	0.02	<b>751.75</b>	<b>0.97</b>	760.27	0.01
Saprospirales	774.78	0.06	<b>769.36</b>	<b>0.86</b>	773.88	0.08
Actinomycetales	680.14	0.03	<b>673.15</b>	<b>0.95</b>	681.1	0.02
Xanthomonadales	632.07	0.05	<b>626.13</b>	<b>0.94</b>	635.44	0.01
Pseudanabaenales	692.91	0	<b>681.38</b>	<b>1</b>	696.37	0
Synechococcales	797.42	0.07	<b>792.04</b>	<b>0.92</b>	800.32	0.01
Chroococcales	697.75	0.02	<b>689.66</b>	<b>0.97</b>	698.83	0.01
SM2F11	580.45	0.14	<b>576.61</b>	<b>0.84</b>	583.88	0.02
Myxococcales	185.37	0.27	<b>183.54</b>	<b>0.6</b>	186.33	0.13
Lactobacillales	220.22	0	<b>202.42</b>	<b>1</b>	220.1	0
Rhodobacteriales	974.18	0.09	<b>970.33</b>	<b>0.53</b>	970.77	0.38
Pseudomonadales	1257.3	0.01	<b>1248.1</b>	<b>0.69</b>	1249.5	0.3
Legionellales	384.56	0.02	<b>377.77</b>	<b>0.55</b>	378.05	0.43
Oscillatoriales	1342	0.42	<b>1341.2</b>	<b>0.54</b>	1346	0.04
Rhodospirillales	357.49	0.34	<b>356.23</b>	<b>0.57</b>	359.83	0.08
Deinococcales	471.66	0.32	<b>470.12</b>	<b>0.62</b>	474.45	0.06
Methylococcales	655.06	0.33	<b>653.61</b>	<b>0.61</b>	658.09	0.06
Nostocales	639.18	0.29	<b>637.66</b>	<b>0.56</b>	640.03	0.15
Bacteria	156.51	0.34	<b>155.28</b>	<b>0.57</b>	158.68	0.09
Proteobacteria	350.48	0.32	<b>348.93</b>	<b>0.62</b>	353.39	0.06
Ellin6067	116.8	0.23	<b>114.4</b>	<b>0.68</b>	118.29	0.09
SM2F09	382.05	0.29	<b>380.33</b>	<b>0.61</b>	383.69	0.1
OD1	219.86	0.27	<b>217.87</b>	<b>0.64</b>	221.48	0.09
Stramenopiles	1290.6	0.33	<b>1289.5</b>	<b>0.52</b>	1291.7	0.15
Flavobacteriales	913.24	0.16	914.33	0.08	<b>909.71</b>	<b>0.75</b>
Acidimicrobiales	600.99	0	602.74	0	<b>583.64</b>	<b>1</b>
Burkholderiales	978.72	0.12	979.28	0.08	<b>974.43</b>	<b>0.8</b>
Pirellulales	923.75	0.02	924.22	0.02	<b>915.76</b>	<b>0.96</b>
Gemmatales	788.95	0.01	786.16	0.03	<b>778.78</b>	<b>0.97</b>
RB41	389.19	0	386.69	0	<b>374.15</b>	<b>1</b>
Chthoniobacteriales	309.13	0.01	302.66	0.18	<b>299.33</b>	<b>0.82</b>
SC.I.84	472.2	0.01	471.87	0.01	<b>461.54</b>	<b>0.99</b>
WD2101	538.32	0	540.21	0	<b>525.88</b>	<b>1</b>
Ellin6529	422.7	0.09	423.1	0.07	<b>417.72</b>	<b>0.85</b>
Cyanobacteria	244.26	0.13	245.46	0.06	<b>240.1</b>	<b>0.81</b>
Clostridiales	258.44	0.11	259.65	0.06	<b>253.98</b>	<b>0.83</b>
Aeromonadales	182.38	0.06	183.37	0.03	<b>176.42</b>	<b>0.91</b>
MND1	15.202	0.23	16.245	0.12	<b>12.698</b>	<b>0.64</b>
Planctomycetales	379.52	0.02	381.33	0.01	<b>370.78</b>	<b>0.98</b>
Enterobacteriales	1065.3	0.04	1060.9	0.31	<b>1059.2</b>	<b>0.65</b>
Caulobacteriales	261.72	0.31	263.62	0.11	<b>260.03</b>	<b>0.58</b>
Methylophilales	198.53	0.17	197.45	0.26	<b>195.57</b>	<b>0.58</b>
Streptophyta	248.25	0.23	247.52	0.3	<b>246.37</b>	<b>0.47</b>

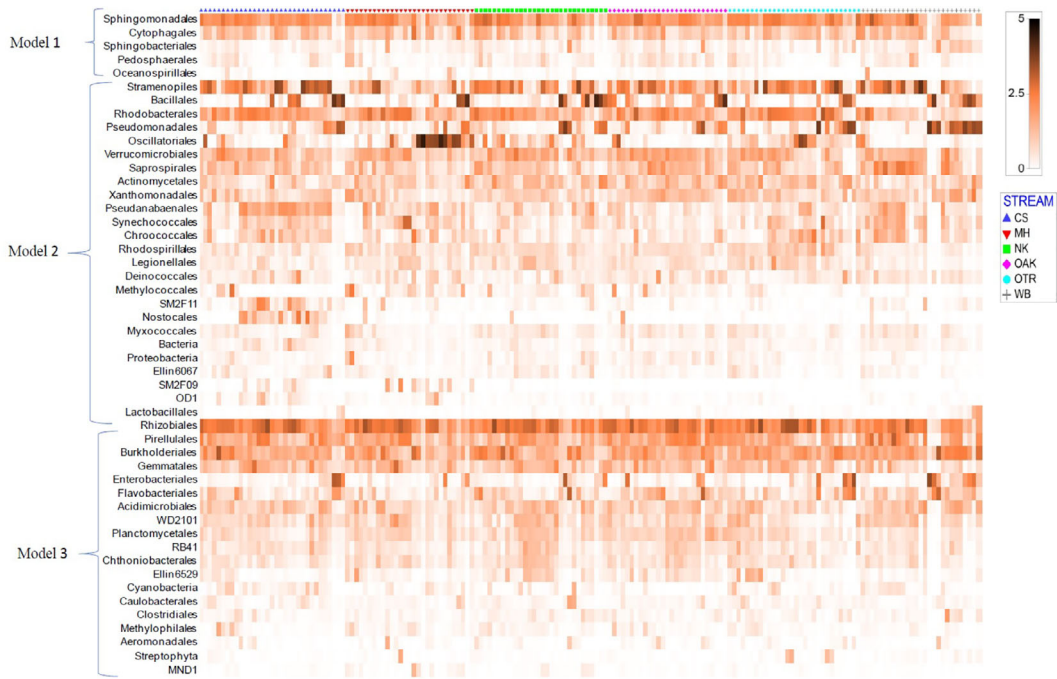
Bold numbers represent the best fit of each order analysed by time series models; (AIC) Akaike's information criterion; (AICcWt) the probability that a given model is the best model in the candidate set.

evidence that stream bacterial community composition is highly responsive to seasonal and environmental changes. The time series model 1, which identified bacterial taxa responding to linear trends over time, comprised the least taxa but included some common Proteobacteria and Bacteroidetes orders (Battin *et al.*, 2016). With increasing temperature and pH, the order Sphingomonadales (phylum proteobacteria) and

Cytophagales (phylum Bacteroidetes) were observed at increased abundance. Members of the phylum Bacteroidetes are reported to break down biopolymers that contribute to the high molecular weight fraction of dissolved organic debris in streams (Battin *et al.*, 2016) and so increased water temperature may have significant impacts on Bacteroidete-associated transformations of organic matter within freshwater systems.



**Fig. 3.** Bacterial community differences constructed using subsets of the taxa for which their abundance was significantly explained by the respective time series GLM models. Data are averaged by (A) year (Model 1-linear; Year 2016 was removed due to having only two data points), (B) month and year (Model 2-bloom), (C) month (Model 3-seasonal). Trajectories show the relationship between the data points and abundances of the respective bacterial orders, selected for each model using AIC criteria. nMDS Plot in Supplementary Material 8 represents annual and seasonal trends using significant physiochemical parameters highlighted in Table 2 using same dataset.



**Fig. 4.** Shade plot showing differences in the relative abundance of the 50 most abundant orders of taxa across our stream biofilm dataset over time (year\_month), clustered and categorized based on time series generalized linear model prediction [model 1 (linear), model 2 (bloom/trough), model 3 (seasonal)]. Shading intensity within the matrix indicates the log ( $n + 1$ ) transformed relative abundance of each Order (as represented by the legend in the upper right of the plot). On the x-axis, data from different streams are indicated by symbols, CS (blue triangle), MH (red inverted triangle), NK (green square), OAK (pink rhombus), OTR (blue circle), WB (plus symbol) from the beginning of sampling in February 2013 until February 2016.

The relative abundance of cyanobacterial and proteobacterial orders adhered most closely to a blooming pattern of temporal change (time series model 2). Indeed, the progression of diatoms in spring and a more diverse community dominated by chlorophytes and

cyanobacteria in spring and/or autumn is a widely identifiable trend in freshwater streams (Salmaso, 2000). With increase in the relative abundances of cyanobacteria particularly in spring 2014 (Fig. 2), the selection of cyanobacteria by time series model 2 is not surprising.

**Table 2.** Summary of the relationship between bacterial community composition and physicochemical variables using DistLM analysis of subsets of the community data, as selected by comparison of three time series GLM models.

Model 1 (variable)	Adjusted $R^2$ (cumulative)	SS (trace)	Pseudo-F	$p$
Light	0.040462	8732.6	8.3795	0.0003
pH	0.07145	6864.5	6.8068	0.0006
Nitrate + nitrite	0.086744	3865.4	3.8971	0.0095
Total phosphorus	0.088667	1349	1.363	0.2401
Water temperature	0.089744	1188.5	1.2022	0.2849
<b>Model 2 (variable)</b>				
Light	0.041976	14,320	8.6676	0.0001
pH	0.068711	9628.5	5.9953	0.0003
Deficit	0.080587	5128.4	3.2345	0.0067
Total N	0.092735	5167.9	3.3031	0.0056
Water temperature	0.10319	4630	2.9938	0.012
Turbidity	0.11018	3584.4	2.3359	0.0339
Total phosphorus	0.11478	2866.4	1.8777	0.0817
DO % Sat	0.11663	2057.8	1.3508	0.2101
Total suspended solids	0.11813	1954.6	1.2853	0.2446
Nitrate + nitrite	0.11877	1701.7	1.1198	0.3265
Ammonia	0.12112	2184.1	1.441	0.1782
Mean_Tmax	0.12117	1530.3	1.0097	0.3954
<b>Model 3 (variable)</b>				
Light	0.047729	12 414	9.7712	0.0001
Deficit	0.067584	5853	4.7052	0.0005
pH	0.080412	4187.6	3.4133	0.005
Total N	0.084612	2185	1.7892	0.0985
Ammonia	0.087952	1978.7	1.6262	0.1338
Soluble phosphorus	0.090824	1864.3	1.537	0.1723
Nitrate + nitrite	0.093122	1727.9	1.4282	0.1946
TKN	0.097542	2194.7	1.8229	0.0957
Turbidity	0.098554	1428.2	1.1875	0.2856
Total suspended solids	0.10125	1795.3	1.4973	0.1754
Water temperature	0.10166	1290.3	1.0766	0.3466
DO % Sat	0.10221	1318.3	1.1007	0.3444

A forward selection approach was utilized to identify environmental variables correlated significantly with observed changes in bacterial community composition (based on Bray–Curtis dissimilarity). These variables were fitted sequentially in order, with each subsequent variable being included in the model based on those which precede it in the table.  $p$  values were obtained using 999 permutations of the data. Mean Tmax refers to average daily maximum air temperature.

Representative cyanobacterial orders such as Pseudanabaenales, Synechococcales, Chroococcales, Oscillatoriales and Nostocales dominate during spring and autumn blooms, indicating favourable growth conditions for cyanobacterial blooms in reduced flows and increased residence time with high temperature (Steffen *et al.*, 2012). Nostocales' ability to fix nitrogen is thought to be a crucial component in their dominance during high turbidity and bloom conditions, indicative of regional changes in nutrient loads, as well as due to changes in various biotic conditions (Wu *et al.*, 2012). Nostocales has been identified as an indicator of global warming and is regarded as an effective nutrient competitor (Sukenik *et al.*, 2012) while being toxic to humans and animals (Stewart *et al.*, 2011). Pseudanabaenales and Chroococcales, on the other hand, possess the ability to create a favourable environment for their growth and metabolic activities by exchanging specialized chemical signals (Rasmussen and Givskov, 2006; Muñoz-García and Ares, 2016). For example, orders Chroococcales (also regarded as *Microcystis* species), Oscillatoriales

and Synechococcales are positively associated with nutrient enrichment. Warming of surface waters, promoting more substantial and longer thermal stratification periods (Kormas and Lymperopoulou, 2013; Ma *et al.*, 2016; Salmaso, 2000; Zhu *et al.*, 2019), can also increase the abundance of Oscillatoriales in late summer, perhaps as they occupy new ecological niches along vertically segregated environmental gradients (Salmaso, 2000). Environmental disturbances, for instance, elevated temperature, high light and eutrophic conditions (Olapade and Leff, 2005; Simonin *et al.*, 2019; Zhu *et al.*, 2019) provide ample opportunities that can be easily and competitively utilized by opportunistic bacterial taxa, including bloom-forming bacterial communities (Piehler *et al.*, 2009). Research shows that cyanobacteria interact with their biotic environment in various ways, from impacting predator–prey interactions to forming mutualistic relationships with micro, macroalgae and non-photosynthetic protists (Bauer and Forchhammer, 2021; Ma *et al.*, 2017; Mutalipassi *et al.*, 2021; Usher *et al.*, 2007). Cyanobacteria and proteobacteria coexist in

>90% of bloom-forming bacterial communities, contributing equally to the overwhelming majority of functional genes detected during bloom formation (Steffen *et al.*, 2012). Indeed, mesocosm studies confirm associations of various proteobacterial orders, such as Legionellales, Rhizobiales and Rhodobacterales, within freshwater cyanobacterial blooms (especially *Microcystis*) (Li *et al.*, 2011). Such interactions make sense because variables like grazing and viral lysis can cause blooms to decline (Riemann *et al.*, 2000; Yager *et al.*, 2001; Löder *et al.*, 2011; Buchan *et al.*, 2014), thereby supplying a diverse range of nutrient resources and organic matter that promotes proteobacterial growth and production (Bird and Karl, 1999). Firmicutes are typically underrepresented in freshwater habitats and exhibit low abundance during cyanobacterial blooms. Nevertheless, they possess numerous biodegradation competencies that are shown to be associated with cyanobacterial degradation (Kormas and Lymperopoulou, 2013). With these findings, we suggest that future bloom predictions should be based on an improved knowledge of bacterial community interactions, as well as physicochemical parameters such as differences in temperature, light, pH and nutrients, all of which have been shown to influence bloom-forming bacterial communities in freshwater. To investigate potential bloom trends in freshwater streams, regular sampling throughout time is essential, as changes in the bacterial populations may precede bloom events.

Our 'seasonal' time series model (model 3) was defined mainly by Planctomycetes (Pirellulales, Gemmatales, WD2101, Planctomycetales) and Proteobacterial (Burkholderiales, Enterobacteriales, SC-I-84, Aeromonadales, Enterobacteriales, Caulobacterales, Methylophilales) orders. Orders Planctomycetes Pirellulales and Gemmatales, as well as orders WD2101, Planctomycetales, Methylophilales, Chthoniobacterales and Acidimicrobiales, were abundant in winter and autumn, presumably due to seasonal succession from the ground up bacterial community composition. In contrast, Pollet *et al.* (2011) and Jinjun *et al.* (2006) discovered a greater abundance of Planctomycetes during the spring and summer seasons. Here, we confirmed that Caulobacterales, Flavobacteriales, Aeromonadales and Enterobacteriales dominated summer and spring communities; high Enterobacteriales abundances were positively linked with bloom-forming Bacillales and Pseudomonadales orders. Bacillales and Enterobacteriales are facultative anaerobes that have been previously detected in heavy metal and faecal contaminated sites and can survive extreme weather conditions (Humayoun *et al.*, 2003). We propose that many changes in bacterial community composition are not continuous over time and that

continual alteration in community composition is induced by slow environmental change and disparities in resource availability.

Aquatic microbial communities have been demonstrated to be temporally dynamic, with repeating patterns of community structure (Fuhrman *et al.*, 2006; Shade *et al.*, 2007; Cram *et al.*, 2015; Fuhrman *et al.*, 2015; Yan *et al.*, 2017; Capo *et al.*, 2019). Various studies have used culture-independent sequencing approaches and high-throughput 16S amplicon sequencing to confirm repeated changes of microbial communities in bloom-affected freshwater bodies (Eiler *et al.*, 2012; Li *et al.*, 2015; Woodhouse *et al.*, 2016; Tromas *et al.*, 2017). However, it is difficult to extrapolate the predictions from bloom and seasonal effects robustly if the studies are carried out for only short periods of time. Therefore, information on whether freshwater bacterial population dynamics are repeatable and predictable at different spatial and temporal scales, including incorporating samples collected over several consecutive years is highly advantageous.

Light, temperature and nutrients are growth-limiting factors for freshwater bacterial communities, essentially bloom-forming and seasonal bacterial communities (Olapade and Leff, 2005; Paerl *et al.*, 2011; Zhu *et al.*, 2019). Here, supporting the results of previous studies, water temperature and average maximum daily air temperature, along with concentrations of ammonia, nitrates, phosphorus, dissolved oxygen, light, pH, moisture deficit, total suspended solids and turbidity were found to be significantly associated with bloom and seasonal patterns (time series model 2, 3) explaining 12% and 10% of the observed variation [adjusted  $R^2$  (cumulative)] respectively. Water temperature, light, pH, total phosphorus and nitrates were significantly correlated with abundances of bacterial orders selected by the time series model 1, explaining 8% of the total variation observed. These results suggest that seasonal changes in freshwater, as well as differences in nutrient levels, provide ideal conditions for the growth of specific bacterial communities to form blooms that can dominate for short periods while also influencing competition for available nutrients among other organisms (Mur *et al.*, 1999; Stewart *et al.*, 2011; Zhu *et al.*, 2019). We demonstrate that the relative abundances of some bacterial community members are cyclically repetitive (Table 2). However, as we can only explain 12% or less of the total observed variation, more investigation is required to understand the underlying cause of these variations, for example relating to spatiotemporal changes in environmental drivers. Unmeasured abiotic factors such as water conductivity, salinity, anthropogenic activities and other biotic factors such as competition, mutualism and succession could explain the rest of the variation. Additionally, stochastic



factors, dispersal limitation, mass immigration effects, as well as deterministic factors such as species sorting likely influence bacterial community differences over time. Future research is needed to confirm community responses to a greater diversity of environmental, seasonal and successional changes and thus to incorporate them into conceptual models. An increased sampling frequency including measurements of biotic and abiotic environmental drivers might enable more accurate predictions of bloom, pre-bloom and seasonal specific communities. This would further help to predict repetitive patterns within stream biofilms, and the extent to which these predictors are stream specific.

## Conclusions

Given that there are unexplained drivers of differences in freshwater stream bacterial community composition, we recognized that spatially and temporally heterogeneous bacterial communities cannot be adequately defined by any one simple statistical model. Using multiple time series models, we have made first attempts to conceptualize the role of local and regional environmental drivers of stream bacterial community composition to provide a more flexible methodology for freshwater ecologists to understand bacterial community assembly and succession. Overall, we demonstrate a conceptual framework to better understand bacterial community assembly, learn the reasons for changes in the relative abundance of specific taxa, and quantify their response to environmental changes.

## Experimental procedures

### *Sample collection, processing and DNA extraction*

The six principal sampling locations for this study were based in the Auckland region of New Zealand, draining three different catchment types. These consisted of streams draining two urban, two rural and two native forest catchments. Cascades and Wairoa streams (Supplementary Material 6) drain densely vegetated native forest reserves and have consistently good water quality (Foley *et al.*, 2018; Buckthought *et al.*, 2020). Oakley and Otara Stream catchments drain mostly urban, bare and lightly vegetated surfaces with significant human populations and residential and commercial activities. Mahurangi and Ngakaroa streams are rural streams and drain catchments dominated by improved exotic grassland and other herbaceous vegetation used for grazing livestock. A summary of each stream's catchment area, land cover distribution and location is provided in Supplementary Material 7.

We collected five stream biofilm samples within ~10 m reach of each stream, monthly for 36 months (February 2013–February 2016). Biofilm biomass was removed from five submerged rocks by scraping the upper surface of each using a separate, sterile Speci-Sponge™ (Nasco, USA) before placement into individual Whirl-Pak™ bags (Nasco, USA). Bags were sealed and held at 4°C before being transported back to the laboratory and frozen (–20°C) until further analysis (Lear and Lewis, 2009). Auckland Council collected physicochemical stream water attributes, adhering to clear data collection standards (APHA, 1998) and collated them in a National Institute of Water and Atmospheric Research (NIWA; <https://cliflo.niwa.co.nz/>) database. At each location and sample date, the data for this study were collated as described by Gautam *et al.* (2020).

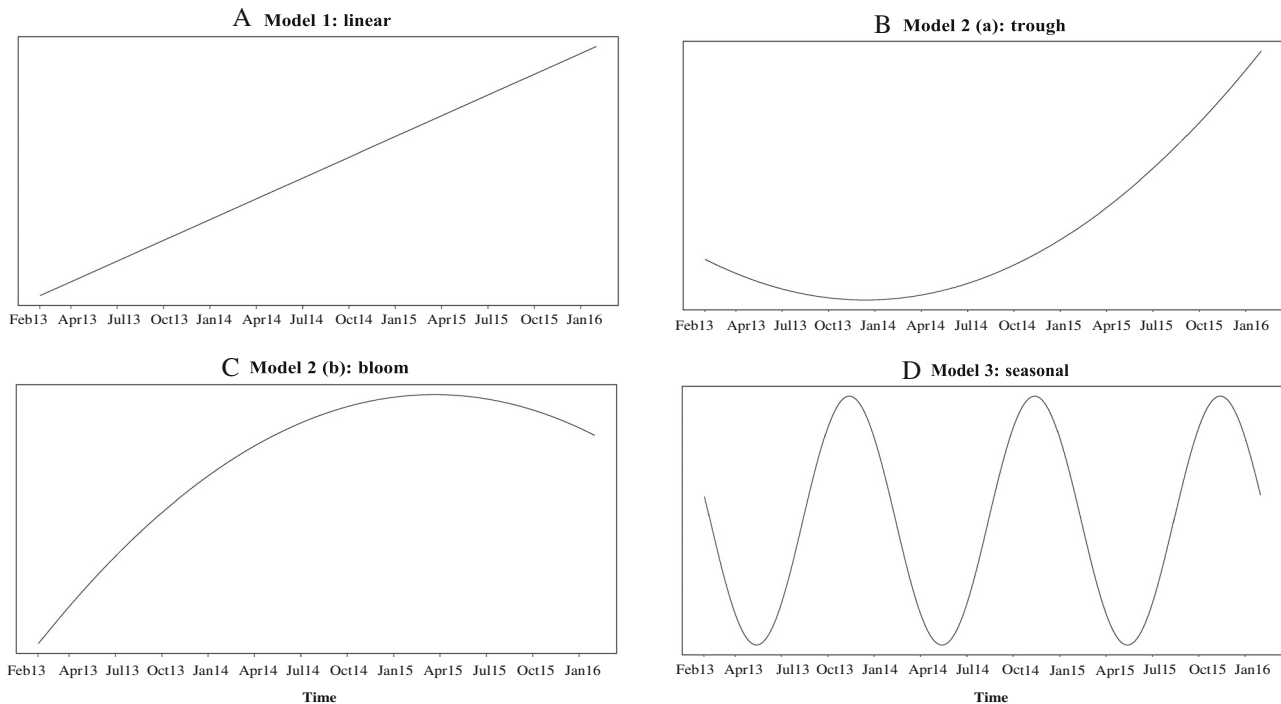
Each biofilm sample was thawed and thoroughly macerated to separate the biofilm samples from the sponges using a Lab Stomacher 400 device (Seward, Norfolk, UK). We extracted DNA from each pelleted biofilm sample using the method described by Miller *et al.* (1999). This method utilizes a bead-beating methodology with chloroform–isoamyl alcohol extraction (Miller *et al.*, 1999). The protocol has been described previously for biofilm sample processing and DNA extraction (Gautam *et al.*, 2020).

### *PCR amplification and sequencing*

To characterize the bacterial communities present in each sample, we amplified V3/V4 regions of bacterial 16S rRNA genes in each biofilm DNA extract using modifications of the Universal 16S rRNA amplicon primers 341F (5'-**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CCT-ACGGGNGGCWGCAG-3') and 785R (5'-**GTCTCGTGGGCTCGGAGATGT** **GTATAAGAGACAG**GACTACHVGGGTATCTAATCC-3') (Edgar, 2013). The primers include Illumina adapter sequences (bold and underlined) that are required for downstream sequencing. PCR amplification and library preparation were performed as described in Gautam *et al.* (2020). Sequencing was conducted by Auckland Genomics (University of Auckland, New Zealand) on an Illumina MiSeq instrument using 2 × 300 bp chemistry. Before sequencing, the sequencing provider attached a unique combination of Nextera XT dual indices (Illumina, USA) to the DNA from each sample to allow multiplex processing (Illumina, 2013). We deposited all raw amplicon sequence data in the NCBI Sequence Read Archive under accession number PRJNA643645.

### *Bioinformatics and statistical analyses*

DNA sequence data were processed as described previously (Hermans *et al.*, 2020) using the VSEARCH v 7.0 (Edgar, 2013) pipeline. Forward and reverse reads were



**Fig. 5.** Conceptual diagrams of three time series generalized linear models used to investigate temporal cyclicality in stream bacterial community dynamics across multiple seasons.

(A) Model 1 is based on a linear trend as shown by a straight line; (B, C) Model 2 is based on blooms/troughs as interpreted by a curve; (D) Model 3 is based on repetitive seasonal trends represented by periodic waves.

merged using the `fastq_mergepairs` command. Sequences with quality scores lower than 20 were trimmed and those with a length shorter than 150 bp were filtered out of all the merged sequences using the `fastq_filter` command. Finally, we dereplicated sequence data (`derep_fulllength`) and removed singleton reads (`-sort by size`). The remaining sequences were then clustered into species-level OTUs at 97% similarity cut-off in QIIME (Quantitative Insights into Microbial Ecology, version 1.8) and classified against the Greengenes reference database v13.8 (McDonald *et al.*, 2012). We also used DADA2 in R (Callahan *et al.*, 2016) to construct taxon tables of exact ASVs for comparison with our OTU data outputs.

Statistical analyses were performed in Primer-E with the additional add-on PERMANOVA+ (Version 7; Plymouth, UK; Clarke and Gorley, 2015) and using the R software environment for statistical computing and graphics (Version 1.3.959; R Core Team, 2020). The OTU table was rarefied to a sequencing depth of 2000 sequences per sample to maintain the highest number of samples possible after removing those poor numbers of DNA sequence reads, achieved using the `'rarefy'` function in the `'vegan'` package (Oksanen *et al.*, 2013) in R to achieve an equal and comparable sequence depth across all samples. After rarefaction, there were 34 440

OTUs across all the samples. The `'vegan'` package was also used to compute a Bray–Curtis dissimilarity matrix to compare bacterial communities' alpha diversity (OTU richness, Shannon, Simpson) among all sites. We used the RELATE routine in PRIMER7 to compare the resemblance matrix obtained from both ASV and OTU tables, based on Spearman rank correlation coefficients, using 9999 permutations.

A total of 892 samples were analysed before rarefaction, but we removed 117 samples from our analysis because they had fewer than 2000 DNA sequence reads. After this processing, any site/timepoint combinations represented by less than three replicates were excluded to achieve a fair comparison of data collected among different sites and times. Variations in bacterial community composition were evaluated using pairwise similarities among samples by calculating Bray–Curtis dissimilarities from the OTU table. Dissimilarity matrices were visualized using non-metric multidimensional scaling (nMDS), generating a graphical representation of bacterial community composition's spatial and temporal patterns. The Bray–Curtis dissimilarity data were then statistically scrutinized using PERMANOVA models in PRIMER as described previously (Gautam *et al.*, 2020). Factors were nested where appropriate and these factors are highlighted in parentheses [e.g. where ST(CA) denotes

the factor 'stream' is nested within the factor 'catchment']. The dynamics of the bacterial taxonomy structure and composition across streams and years were plotted according to their relative abundance (percentage) and visualized using the package 'ggplot' in R, at phylum and order levels.

A shade plot/heat map based on the Bray–Curtis similarity matrix was constructed to delineate the abundance and distribution of the top 50 most abundant bacterial orders from our dataset across streams and time. The relative abundances of bacterial community data (orders) were standardized by  $\log(n + 1)$  before clustering them via a Bray–Curtis similarity matrix, done using Primer-E, v.7.

We developed three time series models to identify significant correlations and temporal patterns of bacterial community distribution and abundance following environmental change over time using Generalized linear modelling (GLM) (Myers and Montgomery, 1997) in R. In these models, both time and space (i.e. streams) were applied as covariables and as fixed factors. Model 1 (linear) was constructed to assess linear trends (either increasing or decreasing) in the data based on 'days\_since' the start of the experiment. Model 2 (blooms/troughs) is based on 'days\_since + l(days\_since^2)' to identify blooms or non-linear relationships in the data since first sampling. Model 3 (seasonal trends) is based on 'days\_since + sin(2 × pi × days\_since/365) + cos(2 × pi × days\_since/365)' to recognize seasonal patterns since first sampling, using sine and cosine variables. To visually inspect the outcomes of these GLM functions, each bacterial order was compared to our conceptual models (Fig. 5) and visualized using the function 'ggplot'. These models enabled us to identify statistically significant patterns in bacterial community data over time, quantified by  $R^2$  adjusted values; ggplots provided a visual inspection of their relationship. Selection of the most plausible model for each order was carried out by calculating and comparing Akaike's information criterion (AIC) (Akaike, 1998) followed by model averaging (multimodel inference) using AICcmodavg (Mazerolle, 2020). Later, we used these outcomes to understand the differential patterns in individual orders (i.e. to which of the three distinct model categories data corresponding to these taxa most closely adhere to). Various physicochemical factors were recorded in-stream at the time of sampling (i.e. turbidity, water temperature, pH, total suspended solids, nitrate + nitrite, ammonia as N, TKN (Total Kjeldahl Nitrogen), TN (Total Nitrogen), DO% sat (Dissolved Oxygen, saturated), TP (Total Phosphorus) and SP (Soluble Phosphorus). We obtained average daily measurements of mean\_Tmax (mean air temperature maximum), mean\_Tmin (mean air temperature minimum), light, rain and soil moisture deficit from a NIWA Cliflo database, which were averaged per month to use as physicochemical factors. We also coupled our bacterial data with detailed

upstream catchment information. Catchment boundaries upstream of each river sampling location were delineated within ARCGIS (ESRI, 2010) (Supplementary Material 6) based on a river layer from the New Zealand River Environment Classification system (REC) (Snelder *et al.*, 2004).

A non-parametric DistLM analysis (Anderson *et al.*, 2008; Chu *et al.*, 2009; Botwe *et al.*, 2015) was performed to explore the relationships between the various physicochemical variables and bacterial community composition based upon Bray–Curtis dissimilarities among the latter using PRIMER. Each of these environmental variables was tested for significant correlation with the subgroups of bacterial community data identified from the analysis of each of the three time series GLM models. We used a forward selection procedure and adjusted  $R^2$  to identify variables that explain the greatest amount of variation in the bacterial community, using 999 permutations of the data.

### Acknowledgements

We gratefully acknowledge the role of the Auckland Council in providing funding for this work and biofilm sample collection. We are grateful to Associate Professor Hannah Buckley and Katie Miaow (PhD student), Auckland University of Technology, for their suggestions and advice in GLM models. We are grateful to the Centre for Biodiversity and Biosecurity, the University of Auckland for providing financial support. Open access publishing facilitated by The University of Auckland, as part of the Wiley - The University of Auckland agreement via the Council of Australian University Librarians. [Correction added on 25 May 2022, after first online publication: CAUL funding statement has been added.]

### References

- Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, Parzen, E., Tanabe, K., and Kitagawa, G. (eds). New York: Springer, pp. 199–213.
- Anderson, M., Clarke, K., and Gorley, R. (2008) *PERMANOVA+ for Primer. Guide to Software and Statistical Methods*. Plymouth, UK: University of Auckland and PRIMER-E.
- APHA. (1998) *Standard Methods for the Examination of Water and Wastewater Analysis*. Washington, DC: American Public Health Association.
- Battin, T.J., Besemer, K., Bengtsson, M.M., Romani, A.M., and Packmann, A.I. (2016) The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* **14**: 251–263.
- Bauer, A., and Forchhammer, K. (2021) Bacterial predation on cyanobacteria. *Microb Phys Ther* **31**: 99–108.
- Besemer, K., Peter, H., Logue, J.B., Langenheder, S., Lindström, E.S., Tranvik, L.J., and Battin, T.J. (2012) Unraveling assembly of stream biofilm communities. *ISME J* **6**: 1459–1468.
- Bird, D.F., and Karl, D.M. (1999) Uncoupling of bacteria and phytoplankton during the austral spring bloom in Gerlache Strait, Antarctic peninsula. *Aquat Microb Ecol* **19**: 13–27.

- Botwe, P.K., Barmuta, L.A., Magierowski, R., McEvoy, P., Goonan, P., and Carver, S. (2015) Temporal patterns and environmental correlates of macroinvertebrate communities in temporary streams. *PLoS One* **10**: e0142370.
- Brümmer, I.H.M., Felske, A.D.M., and Wagner-Döbler, I. (2004) Diversity and seasonal changes of uncultured Planctomycetales in river biofilms. *Appl Environ Microbiol* **70**: 5094–5101.
- Buchan, A., LeClerc, G.R., Gulvik, C.A., and González, J.M. (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12**: 686–698.
- Buckthought, L., Ingley, R., and Grant, C. (2020) Auckland River Water Quality: Annual Report and National Policy Statement for Freshwater Management, Current State Assessment, 2018. Technical Report 2020/014.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.
- Capo, E., Rydberg, J., Tolu, J., Domaizon, I., Debroas, D., Bindler, R., and Bigler, C. (2019) How does environmental inter-annual variability shape aquatic microbial communities? a 40-year annual record of sedimentary DNA from a boreal lake (Nylandssjön, Sweden). *Front Ecol Evol* **7**: 245.
- Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., *et al.* (2000) Consequences of changing biodiversity. *Nature* **405**: 234–242.
- Chu, C., Wang, Y., Li, Q., Zhao, L., Ren, Z., Xiao, S., *et al.* (2009) Effects of traits, species identity and local environmental conditions on the assessment of interactions: insights from an alpine meadow community. *J Plant Ecol* **2**: 135–141.
- Clarke, K., and Gorley, R. (2015) *Getting started with PRIMER v7.PRIMER-E: Plymouth: Ivybridge: PRIMER-E Ltd.*, p. 20.
- Cram, J.A., Chow, C.-E.T., Sachdeva, R., Needham, D.M., Parada, A.E., Steele, J.A., and Fuhrman, J.A. (2015) Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J* **9**: 563–580.
- Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**: 996–998.
- Eiler, A., Heinrich, F., and Bertilsson, S. (2012) Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* **6**: 330–342.
- Esposito, A., Engel, M., Ciccazzo, S., Daprà, L., Penna, D., Comiti, F., *et al.* (2016) Spatial and temporal variability of bacterial communities in high alpine water spring sediments. *Res Microbiol* **167**: 325–333.
- ESRI. (2010) *ArcGIS 9.3*. Redlands, CA: Environmental Systems Research Institute.
- Foley, M. M., Walker, J., Vaughan, M. R., Stanley, C. L., and Hussain, E. (2018) Marine Water Quality State and Trends in the Auckland Region From 2007 to 2016. Technical Report 2018/015.
- Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13**: 133–146.
- Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* **103**: 13104–13109.
- Gautam, A., Lear, G., and Lewis, G.D. (2020) Analysis of spatial and temporal variations in bacterial community dynamics within stream biofilms. *N Z J Mar Freshw Res* **55**: 505–523.
- Good, S.P., URycki, D.R., and Crump, B.C. (2018) Predicting hydrologic function with aquatic gene fragments. *Water Resour Res* **54**: 2424–2435.
- Hassell, N., Tinker, K.A., Moore, T., and Ottesen, E.A. (2018) Temporal and spatial dynamics in microbial community composition within a temperate stream network. *Environ Microbiol* **20**: 3560–3572.
- Hermans, S.M., Buckley, H.L., Case, B.S., Curran-Cournane, F., Taylor, M., and Lear, G. (2020) Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome* **8**: 79.
- Humayoun, S.B., Bano, N., and Hollibaugh, J.T. (2003) Depth distribution of microbial diversity in mono Lake, a meromictic soda Lake in California. *Appl Environ Microbiol* **69**: 1030–1042.
- Illumina. (2013) 16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System.
- Jeffries, T.C., Schmitz Fontes, M.L., Harrison, D.P., Van-Dongen-Vogels, V., Eyre, B.D., Ralph, P.J., and Seymour, J.R. (2015) Bacterioplankton dynamics within a large anthropogenically impacted urban estuary. *Front Microbiol* **6**: 1438.
- Jinjun, K., Kui, W., and Feng, C. (2006) Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). *Aquat Microb Ecol* **42**: 7–18.
- Kim, H.G., Song, S.J., Bae, H., Noh, J., Lee, C., Kwon, B.O., *et al.* (2020) Natural and anthropogenic impacts on long-term meiobenthic communities in two contrasting near-shore habitats. *Environ Int* **134**: 105200.
- Kormas, K.A., and Lymperopoulou, D.S. (2013) Cyanobacterial toxin degrading bacteria: who are they? *Biomed Res Int* **2013**: 463894.
- Kraemer, S.A., Barbosa da Costa, N., Shapiro, B.J., Fradette, M., Huot, Y., and Walsh, D.A. (2020) A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. *ISME J* **14**: 3011–3023.
- Labbate, M., Seymour, J.R., Lauro, F., and Brown, M.V. (2016) Editorial: anthropogenic impacts on the microbial ecology and function of aquatic environments. *Front Microbiol* **7**: 6–8.
- Lear, G., Anderson, M.J., Smith, J.P., Boxen, K., and Lewis, G.D. (2008) Spatial and temporal heterogeneity of the bacterial communities in stream epilithic biofilms. *FEMS Microbiol Ecol* **65**: 463–473.
- Lear, G., and Lewis, G.D. (2009) Impact of catchment land use on bacterial communities within stream biofilms. *Ecol Indic* **9**: 848–855.
- Li, H., Xing, P., Chen, M., Bian, Y., and Wu, Q.L. (2011) Short-term bacterial community composition dynamics in response to accumulation and breakdown of *Microcystis* blooms. *Water Res* **45**: 1702–1710.
- Li, J., Zhang, J., Liu, L., Fan, Y., Li, L., Yang, Y., *et al.* (2015) Annual periodicity in planktonic bacterial and

- archaeal community composition of eutrophic Lake Taihu. *Sci Rep* **5**: 15488.
- Löder, M.G.J., Meunier, C., Wiltshire, K.H., Boersma, M., and Aberle, N. (2011) The role of ciliates, heterotrophic dinoflagellates and copepods in structuring spring plankton communities at Helgoland roads, North Sea. *Mar Biol* **158**: 1551–1580.
- Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, A., *et al.* (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**: 804–808.
- Ma, A.T., Beld, J., Brahmasha, B., and Drake, H.L. (2017) An amoebal grazer of cyanobacteria requires cobalamin produced by heterotrophic bacteria. *Appl Environ Microbiol* **83**: e00035-00017.
- Ma, J., Qin, B., Paerl, H.W., Brookes, J.D., Hall, N.S., Shi, K., *et al.* (2016) The persistence of cyanobacterial (*Microcystis* spp.) blooms throughout winter in Lake Taihu, China. *Limnol Oceanogr* **61**: 711–722.
- Mazerolle, M. (2020) *AICcmodavg: Model Selection and Multimodel Inference Based on (Q) AIC (c)*. URL <https://cran.rproject.org/web/packages/AICcmodavg/index.html>.
- McCall, S.J., Hale, M.S., Smith, J.T., Read, D.S., and Bowes, M.J. (2017) Impacts of phosphorus concentration and light intensity on river periphyton biomass and community structure. *Hydrobiologia* **792**: 315–330.
- McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**: 610–618.
- Miller, D.N., Bryant, J.E., Madsen, E.L., and Ghiorse, W.C. (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* **65**: 4715–4724.
- Muñoz-García, J., and Ares, S. (2016) Formation and maintenance of nitrogen-fixing cell patterns in filamentous cyanobacteria. *Proc Natl Acad Sci U S A* **113**: 6218–6223.
- Mur, L., Skulberg, O., and Utkilen, H. (1999) Cyanobacteria in the environment. In *Toxic Cyanobacteria in Water: A Guide to their Public Health Consequences, Monitoring and Management*, Chorus, I., and Bartram, J. (eds). London: E & FN Spon, pp. 15–40.
- Mutalipassi, M., Riccio, G., Mazzella, V., Galasso, C., Somma, E., Chiarore, A., *et al.* (2021) Symbioses of cyanobacteria in marine environments: ecological insights and biotechnological perspectives. *Mar Drugs* **19**: 227.
- Myers, R.H., and Montgomery, D.C. (1997) A tutorial on generalized linear models. *J Qual Technol* **29**: 274–291.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., *et al.* (2013) Package 'vegan'. Community Ecology Package, Version, 2(9), 1-295.
- Olapade, O.A., and Leff, L.G. (2005) Seasonal response of stream biofilm communities to dissolved organic matter and nutrient enrichments. *Appl Environ Microbiol* **71**: 2278–2287.
- Paerl, H.W., Xu, H., McCarthy, M.J., Zhu, G., Qin, B., Li, Y., and Gardner, W.S. (2011) Controlling harmful cyanobacterial blooms in a hyper-eutrophic Lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. *Water Res* **45**: 1973–1983.
- Piehler, M.F., Dyble, J., Moisander, P.H., Chapman, A.D., Hendrickson, J., and Paerl, H.W. (2009) Interactions between nitrogen dynamics and the phytoplankton community in Lake George, Florida, USA. *Lake Reservoir Manage* **25**: 1–14.
- Pollet, T., Tadonlécé, R.D., and Humbert, J.F. (2011) Spatio-temporal changes in the structure and composition of a less-abundant bacterial phylum (Planctomycetes) in two Perialpine Lakes. *Appl Environ Microbiol* **77**: 4811–4821.
- Qu, X., Ren, Z., Zhang, H., Zhang, M., Zhang, Y., Liu, X., and Peng, W. (2017) Influences of anthropogenic land use on microbial community structure and functional potentials of stream benthic biofilms. *Sci Rep* **7**: 1–12.
- R Core Team. (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org/>.
- Rasmussen, T.B., and Givskov, M. (2006) Quorum sensing inhibitors: a bargain of effects. *Microbiology* **152**: 895–904.
- Riemann, L., Steward, G.F., and Azam, F. (2000) Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol* **66**: 2282.
- Salmaso, N. (2000) Factors affecting the seasonality and distribution of cyanobacteria and chlorophytes: a case study from the large lakes south of the Alps, with special reference to Lake Garda. *Hydrobiologia* **438**: 43–63.
- Salmaso, N., Albanese, D., Capelli, C., Boscaini, A., Pindo, M., and Donati, C. (2018) Diversity and cyclical seasonal transitions in the bacterial community in a large and deep perialpine Lake. *Microb Ecol* **76**: 125–143.
- Sanderson, M.A., Goslee, S.C., Soder, K.J., Skinner, R.H., Tracy, B.F., and Deak, A. (2007) Plant species diversity, ecosystem function, and pasture management—a perspective. *Can J Plant Sci* **87**: 479–487.
- Schmidt, S., Bernard, C., Escalier, J.-M., Etcheber, H., and Lamouroux, M. (2017) Assessing and managing the risks of hypoxia in transitional waters: a case study in the tidal Garonne River (South-West France). *Environ Sci Pollut Res* **24**: 3251–3259.
- Shade, A., Kent, A.D., Jones, S.E., Newton, R.J., Triplett, E. W., and McMahan, K.D. (2007) Interannual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol Oceanogr* **52**: 487–494.
- Simonin, M., Voss, K.A., Hassett, B.A., Rocca, J.D., Wang, S. Y., Bier, R.L., *et al.* (2019) In search of microbial indicator taxa: shifts in stream bacterial communities along an urbanization gradient. *Environ Microbiol* **21**: 3653–3668.
- Snelder, T., Biggs, B., and Weatherhead, M. (2004) *New Zealand River Environment Classification User Guide*. Wellington, New Zealand: Ministry for the Environment.
- Steffen, M.M., Li, Z., Effler, T.C., Hauser, L.J., Boyer, G.L., and Wilhelm, S.W. (2012) Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS One* **7**: e44002.
- Stewart, I., Carmichael, W.W., and Backer, L.C. (2011) Toxic cyanobacteria. In *Water and Sanitation-Related Diseases and the Environment*, Selendy, J.M.H. (ed): Hoboken, NJ: John Wiley and Sons, pp. 95–109.
- Sukenik, A., Hadas, O., Kaplan, A., and Quesada, A. (2012) Invasion of Nostocales (cyanobacteria) to subtropical and temperate freshwater lakes – physiological, regional, and global driving forces. *Front Microbiol* **3**: 86.

- Tromas, N., Fortin, N., Bedrani, L., Terrat, Y., Cardoso, P., Bird, D., *et al.* (2017) Characterising and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course. *ISME J* **11**: 1746–1763.
- Usher, K.M., Bergman, B., and Raven, J.A. (2007) Exploring cyanobacterial mutualisms. *Annu Rev Ecol Evol Syst* **38**: 255–273.
- Woodhouse, J.N., Kinsela, A.S., Collins, R.N., Bowling, L.C., Honeyman, G.L., Holliday, J.K., and Neilan, B.A. (2016) Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. *ISME J* **10**: 1337–1351.
- Wright, I.A., and Ryan, M.M. (2016) Impact of mining and industrial pollution on stream macroinvertebrates: importance of taxonomic resolution, water geochemistry and EPT indices for impact detection. *Hydrobiologia* **772**: 103–115.
- Wu, Z., Zeng, B., Li, R., and Song, L. (2012) Physiological regulation of *Cylindrospermopsis raciborskii* (Nostocales, Cyanobacteria) in response to inorganic phosphorus limitation. *Harmful Algae* **15**: 53–58.
- Yager, P.L., Connelly, T.L., Mortazavi, B., Wommack, K.E., Bano, N., Bauer, J.E., *et al.* (2001) Dynamic bacterial and viral response to an algal bloom at subzero temperatures. *Limnol Oceanogr* **46**: 790–801.
- Yan, Q., Stegen, J.C., Yu, Y., Deng, Y., Li, X., Wu, S., *et al.* (2017) Nearly a decade-long repeatable seasonal diversity patterns of bacterioplankton communities in the eutrophic Lake Donghu (Wuhan, China). *Mol Ecol* **26**: 3839–3850.
- Zhao, D., Cao, X., Huang, R., Zeng, J., and Wu, Q.L. (2017) Variation of bacterial communities in water and sediments during the decomposition of *Microcystis* biomass. *PLoS One* **12**: e0176397.
- Zhu, C., Zhang, J., Nawaz, M.Z., Mahboob, S., Al-Ghanim, K.A., Khan, I.A., *et al.* (2019) Seasonal succession and spatial distribution of bacterial community structure in a eutrophic freshwater Lake, Lake Taihu. *Sci Total Environ* **669**: 29–40.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Supplementary Material 1.** Box plots for OTU richness, Shannon and Simpson diversity indices across streams. Each colour represents samples from an individual stream. Native streams: CS (Cascades) and WB (Wairoa); Moderately impacted: MH (Mahurangi) and NK (Ngakaraoa); Highly impacted: OTR (Otara) and OAK (Oakley). A Tukey post hoc test revealed significantly different data groups, represented by different letters assigned to each stream; the same letter represents no significant difference between streams (two factor ANOVA, Tukey HSD;  $p < 0.05$ ). The centre line in the boxplot represents median data value, top and bottom lines in the box represent the 25th and 75th percentiles; the top and bottom whiskers mark the 10th and 90th percentiles respectively. The points at the top and bottom of the box indicate outlying data points

**Supplementary Material 2.** PERMANOVA table of results of model '1' (see Fig. 1), based on Bray–Curtis dissimilarity

distance showing the partitioning of variance factors and tests for the factors of (CA) Catchment, (ST) Stream, (YE) Year, (MO) Month and (SE) Season and, where possible, their interactions.

**Supplementary Material 3.** PERMANOVA table of model '2' results generated by periodic regression analysis of Bray–Curtis dissimilarity distance comparisons of bacterial community data and accompanying tests for seasonality on the bacterial community. Covariates (CoV 1 and CoV2) generated from the regression model of the 12-month cycle were used as ANOVA estimators in this model. Factors used were (CA) Dominant catchment land-use, (YE) Year, (ST) Stream, CoV1 and Cov2, and their interaction terms where possible.

**Supplementary Material 4.** Relative abundances of the top 40 most abundant bacterial orders across our stream biofilm dataset. Barplot is partitioned into six small plots (rows) representing six different streams from the beginning of sampling in February 2013 until February 2016. The x-axis represents time and each bar represents the relative abundance of the top 40 most abundant orders in each stream. Bars ordered from left to right by date. Each colour corresponds to a specific order

**Supplementary Material 5.** Shade plot of measured environmental factors from Feb. 2013–Feb. 2016 from our stream biofilm sites. Data were normalized for each site and constrained by month\_year before making the shade plot. Each data point was normalized by subtracting the mean (across all samples) and dividing by the standard deviations of that variable using Euclidean distance, so that all variables have values within a similar range. The normalization scale is shown on the upper right of the plot. Shading intensity within the matrix indicates the transformed value of each parameter as represented by the legend on the right side of the plot. On the y-axis, streams are indicated by symbols, CS (▲), MH (▼), NK (■), OAK (◆), OTR (●), WB (⊕) from the beginning of sampling February 2013 until February 2016. Lines are added to the figure to separate data from each stream.

**Supplementary Material 6.** Map showing the study region and locations of sampling sites within the Auckland Region of New Zealand. The geographic location of each stream is as follows: Cascades Stream, Latitude: –36.89, Longitude: 174.51; Wairoa Stream, Latitude: –37.08, Longitude: 174.95; Mahurangi Stream, Latitude: –36.47, Longitude: 174.73; Ngakaraoa Stream, Latitude: –37.11, Longitude: 174.94; Oakley Creek, Latitude: –36.88, Longitude: 174.70; Otara Stream, Latitude: –36.96, Longitude: 174.87.

**Supplementary Material 7.** Site description, location, catchment area, land use distribution of the upstream catchment area (%), and closest climate stations to each sampling site in the Auckland Region (according to the climate database [www.climflo.niwa.co.nz](http://www.climflo.niwa.co.nz))

**Supplementary Material 8.** nMDS plot showing annual and seasonal trends followed by bacterial communities using significant physiochemical parameters highlighted in Table 2. Bacterial community data averaged by (a) year (Model 1-linear; Year 2016 was removed due to having only two data points), (b) month and year (Model 2-bloom), (c) month (Model 3-seasonal).