



OPEN

DATA DESCRIPTOR

# A novel real-world ecotoxicological dataset of pelagic microbial community responses to wastewater

J. E. Ruprecht<sup>1</sup>✉, W. C. Glamore<sup>1</sup>, K. A. Dafforn<sup>2,3</sup>, F. Wemheuer<sup>2</sup>, S. L. Crane<sup>4</sup>, J. van Dorst<sup>4</sup>, E. L. Johnston<sup>2</sup>, S. M. Mitrovic<sup>5</sup>, I. L. Turner<sup>1</sup>, B. C. Ferrari<sup>4</sup> & S. C. Birrer<sup>2</sup>

Real-world observational datasets that record and quantify pressure-stressor-response linkages between effluent discharges and natural aquatic systems are rare. With global wastewater volumes increasing at unprecedented rates, it is urgent that the present dataset is available to provide the necessary information about microbial community structure and functioning. Field studies were performed at two time-points in the Austral summer. Single-species and microbial community whole effluent toxicity (WET) testing was performed at a complete range of effluent concentrations and two salinities, with accompanying environmental data to provide new insights into nutrient and organic matter cycling, and to identify ecotoxicological tipping points. The two salinity regimes were chosen to investigate future scenarios based on a predicted salinity increase at the study site, typical of coastal regions with rising sea levels globally. Flow cytometry, amplicon sequencing of 16S and 18S rRNA genes and micro-fluidic quantitative polymerase-chain reactions (MFQPCR) were used to determine chlorophyll-a and total bacterial cell numbers and size, as well as taxonomic and functional diversity of pelagic microbial communities. This strong pilot dataset could be replicated in other regions globally and would be of high value to scientists and engineers to support the next advances in microbial ecotoxicology, environmental biomonitoring and estuarine water quality modelling.

## Background & Summary

The world is facing a global water quality crisis<sup>1,2</sup>. The vast majority (more than 80%) of global wastewater is released directly into natural waterways, resulting in widespread pollution<sup>3</sup>. The most frequent contaminants are domestic waste (~2 million tonnes per day), industrial wastes and chemicals, agricultural pesticides and fertilizers<sup>2,4</sup>. The implications of wastewater discharges include, but are not limited to: degraded aquatic ecosystems<sup>5,6</sup>; decreased biodiversity<sup>7</sup>; increased greenhouse gas emissions<sup>8,9</sup>; and a wide range of detrimental impacts to human health<sup>10</sup>. Global wastewater volumes are increasing at unprecedented rates as a result of population growth, rapid urbanisation and economic development, and these drivers are concentrated in coastal regions<sup>11–14</sup>. This worldwide trend poses immediate management challenges if we are to prevent further damage to sensitive aquatic ecosystems, human health and water security<sup>15</sup>.

Globally, comprehensive datasets that characterise the impacts of wastewater discharges to water quality in natural aquatic environments are generally lacking<sup>2</sup>. Based on a comprehensive review of data published in 181 countries<sup>11</sup>, the authors found that only 55 countries had any data available on wastewater generation, treatment and use, and much of this information was dated (i.e., pre-2008). Significant data gaps exist on the linkages between the physical, chemical and biological characteristics of many urban surface water environments that receive wastewater discharges<sup>16–18</sup>. In highly developed countries where the largest percentage of treated

<sup>1</sup>Water Research Laboratory, School of Civil & Environmental Engineering, UNSW Sydney, Sydney, NSW, 2052, Australia. <sup>2</sup>Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney, NSW, 2052, Australia. <sup>3</sup>Department of Earth and Environmental Sciences, Macquarie University, Macquarie Park, NSW, 2109, Australia. <sup>4</sup>Ferrari Lab, School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney, NSW, 2052, Australia. <sup>5</sup>Freshwater and Estuarine Research Group, School of Life Sciences, University of Technology Sydney, Ultimo, 2007, NSW, Australia. ✉e-mail: [j.ruprecht@wrl.unsw.edu.au](mailto:j.ruprecht@wrl.unsw.edu.au)

domestic wastewater is currently discharged directly to natural waterways – for example: Australia (85%; 1,234 treatment plants), North America (75%; 14,748 treatment plants) and Europe (71%; >18,000 treatment plants) – an understanding of the impacts of effluent discharges on ecosystem health is still in its relative infancy<sup>11</sup>. These pressure-stressor-response relationships are particularly difficult to disentangle in estuaries, due to highly variable physio-chemical conditions<sup>19</sup>. To effectively address increasing concerns regarding wastewater discharge to natural aquatic systems worldwide, comprehensive data is urgently needed from real-world observations. These data can be used to investigate and understand the pressure-stressor-response linkages between treated effluent discharges and natural aquatic environments.

Pelagic microbial communities are extremely sensitive to rapid changes in their environment making them ideal indicators of water quality processes and functioning<sup>19–21</sup>. They are also ubiquitous in natural aquatic environments and play an important role in nutrient and organic matter cycling<sup>22</sup>. Traditional microbial ecotoxicological studies have relied on the combination of single algal species toxicity testing and chemical surveys to ascertain the aggregate toxic effect of whole effluent wastewater discharge on microalgae, rather than attempting to quantify both diversity and function of entire microbial communities at the same time<sup>21</sup>. The combination of community-level testing and environmental 'omics moves beyond the scenario of single species toxicity testing and provides the opportunity to determine real-world community interactions and shifts in response to wastewater.

In this study we have incorporated recent advances in water quality science and assessment techniques to characterise the ecotoxicological response of tertiary-level treated effluent following discharge in temperate estuarine environments. The integrated field and laboratory assessments were completed on the Hunter River estuary located on the New South Wales (NSW) coastline in southeast Australia. The dataset obtained is specifically targeted at the dynamics and health of pelagic microbial communities at a practical scale, with these novel observations having the potential to provide new insights and understanding of nutrient and organic matter cycling. Whole Effluent Toxicity (WET) testing was performed to highlight the effects of mixing treated effluent within freshwater and saltwater environments and to identify potential tipping points that both inhibit and stimulate the growth of microalgae and microbial communities. Two salinity regimes were chosen to include future scenarios based on a predicted increasing salinity of the Hunter River with rising sea levels. All water samples were subjected to sequencing for 16S and 18S rRNA genes to measure changes in microbial community structure. Flow cytometry was used to enumerate chlorophyll-*a* and total bacterial cells and to estimate their size. The abundance of genes associated with nutrient cycling, antibiotic resistance and the identification of pathogens that would be harmful to human health were determined using microfluidic quantitative polymerase-chain reactions (MFQPCR) at the microbial community-level.

## Methods

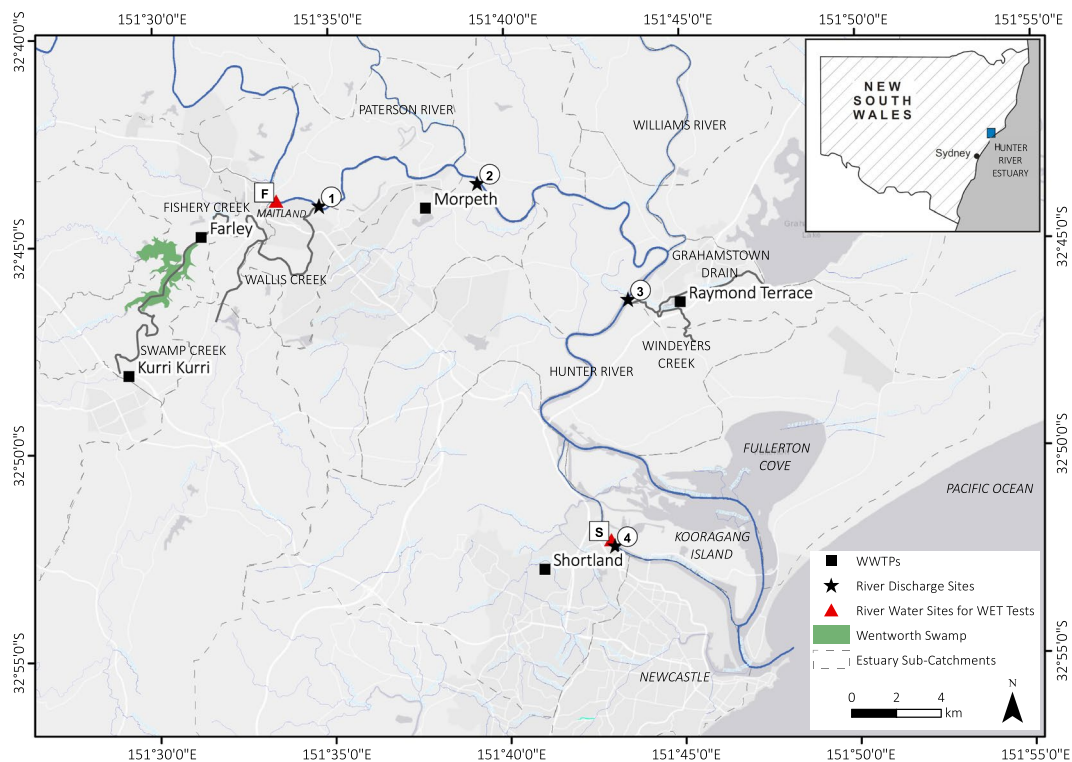
**Study area.** The Hunter River estuary (151.8°E, 32.9°S) is situated on the temperate southeast coastline of NSW, Australia (Fig. 1). The Hunter River estuary is a typical wave-dominated, mature barrier estuary<sup>23</sup> with a large tidal pool that extends 60 km inland to its tidal limits. The Hunter River estuary has semi-diurnal tides and a mean tidal range of approximately 1.2 m. Catchment inflows to the estuary via the Hunter River and its two main tributaries – the Paterson and Williams Rivers – are regulated by dams and weirs.

The Hunter River catchment covers an area over 22,000 km<sup>2</sup> and is typical of many other developed coastal regions globally in that it has been extensively modified by human activity and multiple land uses<sup>24</sup>. The upper catchment is predominately agricultural land, whereas the lower catchment around the Port of Newcastle includes extensive urban and industrial areas, entrance dredging and training, the world's largest coal export terminal and a growing multi-purpose cargo hub. Despite these pressures, the Hunter River estuary still supports significant areas of estuarine habitat such as mangroves, saltmarsh, inter-tidal and sub-tidal soft sediment shoals, as well as a Ramsar listed wetland of international significance<sup>25</sup>.

The Hunter River estuary receives diffuse water pollution and nutrients from the catchment, as well as high nutrient point loads from several major wastewater treatment plants (WWTP) capable of servicing a population of approximately 200,000 people across the floodplain. Generally, these WWTP provide tertiary-level wastewater treatment designed to remove excess nutrients of nitrogen and phosphorus. Unallocated treated effluent is discharged either directly, or indirectly via tributary channels, to the tidal zone of the Hunter River. The tributary channels in this study are freshwater creek systems as they are excluded from tidal flows via one-way floodgates and form part of the expansive Lower Hunter Flood Mitigation Scheme.

The study sampling points were located in the upper, mid and lower portions of the Hunter River estuary (Fig. 1). In the upper Hunter River estuary, the Swamp-Fishery-Wallis Creek system near Maitland, receives unallocated treated effluent from the townships of Kurri Kurri (3.4 ML/day) and Farley (5.6 ML/day). Kurri Kurri WWTP discharges into Swamp Creek which flows into Wentworth Swamp, a large, low-lying permanent waterbody, which in turn discharges into Fishery Creek. Farley WWTP discharges into Fishery Creek downstream of its confluence with Wentworth Swamp, and this flows into Wallis Creek which discharges into the Hunter River estuary (Fig. 1, river discharge site 1). Unallocated treated effluent from Morpeth WWTP (10 ML/day) is discharged directly into the Hunter River, approximately 3 km downstream of Wallis Creek (Fig. 1, river discharge site 2). Unallocated treated effluent from the Raymond Terrace WWTP (7.3 ML/day) is discharged to the Hunter River via Grahamstown Drain and Windeyers Creek (Fig. 1, river discharge site 3). The Shortland WWTP (9.6 ML/day) discharges unallocated treated effluent directly to the Hunter River South Arm (Fig. 1, river discharge site 4). A description of site characteristics for each WWTP river discharge site (Fig. 1) is provided in Table 1.

**Microbial field surveys.** Water samples were taken on two (2) separate occasions in the Austral summer (November 2016 and February 2017) from five (5) WWTP outfalls along the Hunter River estuary, including



**Fig. 1** Hunter River estuary study area and WWTP river discharge sites for Kurri Kurri, Farley, Morpeth, Raymond Terrace and Shortland shown as circled numbers. River water collection sites used for WET test dilutions are shown as boxed letters.

Discharge site	Distance from ocean entrance (km)	Estimated residence time (days) <sup>*</sup>	Salinity range (ppt) <sup>**</sup>	50 <sup>th</sup> percentile salinity (ppt)
1	55	5	0.3–0.4	0.3
2	45	8	0.2–1.8	0.3
3	29	6	0.1–17.2	1.5
4	11	6	0.4–33.4	26.0

**Table 1.** Description of WWTP river discharge sites. <sup>\*</sup>Based on 50<sup>th</sup> percentile modelled flows and dispersion coefficients<sup>24</sup>. <sup>\*\*</sup>5<sup>th</sup> and 95<sup>th</sup> percentiles calculated from a 110-year salinity timeseries simulated using a calibrated and validated hydrodynamic and salinity model of the Hunter River estuary.

Kurri Kurri, Farley, Morpeth, Raymond Terrace and Shortland. Water samples were collected upstream and downstream of each outfall at different distances totalling 20 sampling points (see Table 2). Note that water at Shortland was only sampled from the river immediately upstream of the Shortland WWTP outfall – downstream sampling was not logistically possible and not relevant as the WWTP was not discharging at the time of the investigation. All 20 sites were sampled within the same two-day time window on both sampling occasions with three (3) replicates collected per site for sequencing of microbial (prokaryotic and eukaryotic) communities and a single replicate for water quality measurements.

Specifically, surface water was collected in 2 L sterile Whirl-Paks<sup>®</sup> stored on ice in the dark until filtering within 24 hours. To capture all microbial cells and fragments in the water samples for DNA extraction and sequencing, samples were homogenised by repeated inversion and 500 mL of water was filtered through a 0.22  $\mu$ m Express PLUS Polyethersulfone membrane (Millipore) using a hydraulic pump. Filter units were sterilised before use and rinsed with ethanol between water samples. In some cases, when the water samples contained excess organic material, the filters were clogged before 500 mL could be filtered, and the volume that had been filtered was noted for later adjustment of the data. Filters were rolled, inserted into bead tubes from the DNeasy PowerWater Kit (Qiagen) and frozen at  $-80^{\circ}\text{C}$  until DNA extraction and sequencing.

For water quality measurements, water samples were simultaneously obtained using standard bottles provided by a NATA accredited facility for analysis of total suspended solids (TSS), total organic carbon (TOC), total nitrogen (TN), total phosphorous (TP), ammonia ( $\text{NH}_3$ ), nitrate and nitrite ( $\text{NO}_x$ ), biological oxygen demand (BOD), chlorophyll-a, and hydrogen sulphide ( $\text{H}_2\text{S}$ ). Water samples were stored on ice in the field and then sent for testing on the same day they were sampled. Additional environmental data, including pH, dissolved oxygen

WWTP	Site name	Sample ID	Type	Distance from WWTP
	Kurri Kurri	KK-UP (BG)	Creek	50 m upstream of outfall (background)
Kurri Kurri	Kurri Kurri	KK-DN	Creek	50 m downstream of outfall
	Farley	FAR-UP (BG)	Creek	50 m upstream of outfall (background)
Farley	Farley	FAR-DN	Creek	50 m downstream of outfall
Kurri Kurri + Farley	Wallis Creek	WLC (WLC-DN)	Creek	1500 m upstream of river discharge site 1
	Wallis Creek	WLC-UP (BG)	River	50 m upstream of river discharge site 1 (background)
Kurri Kurri + Farley	Wallis Creek	WLC-0m	River	River discharge site 1
Kurri Kurri + Farley	Wallis Creek	WLC-50m	River	50 m downstream of river discharge site 1
Kurri Kurri + Farley	Wallis Creek	WLC-500m	River	500 m downstream of river discharge site 1
	Morpeth	MOR-UP (BG)	River	50 m upstream of river discharge site 2 (background)
Morpeth	Morpeth	MOR-0m	River	River discharge site 2
Morpeth	Morpeth	MOR-50m	River	50 m downstream of river discharge site 2
Morpeth	Morpeth	MOR-500m	River	500 m downstream of river discharge site 2
	Windeyers Creek	WC-UP (BG)	Creek	50 m upstream of outfall (background)
Raymond Terrace	Windeyers Creek	WC-DN	Creek	50 m downstream of outfall
	Raymond Terrace	RT-UP (BG)	River	50 m upstream of river discharge site 3 (background)
Raymond Terrace	Raymond Terrace	RT-0m	River	River discharge site 3
Raymond Terrace	Raymond Terrace	RT-50m	River	50 m downstream of river discharge site 3
Raymond Terrace	Raymond Terrace	RT-500m	River	500 m downstream of river discharge site 3
Shortland	Shortland	SL-0m	River	River discharge site 4

**Table 2.** Microbial field survey site details, including sample IDs, site names, the main WWTP contributing to the water collected and whether the site was in a creek or the main channel of the Hunter River. Sites are listed from highest to lowest chainage from the ocean entrance, with increasing salinity in the river. Note that the WWTP has been abbreviated in the data records as follows: Farley = FAR, Kurri Kurri = KK, Morpeth = MOR, Raymond Terrace = RT, Shortland = SL. Additional abbreviations include: Wallis Creek = WLC, Windeyers Creek = WC, Background = BG, Downstream = DN.

(DO), and electrical conductivity (EC) were measured using a calibrated water quality unit (Horiba) at each site. The water quality unit was calibrated before and after each sampling trip.

**Whole Effluent Toxicity (WET) testing.** Whole Effluent Toxicity (WET) testing was conducted within the laboratory facilities of the Sydney Institute of Marine Science (SIMS) in Chowder Bay, Sydney, Australia. WET tests were completed for algal single-species and whole microbial communities in a fully-crossed experimental design created using UV-disinfected effluent from the five (5) WWTP sampled. Following standard protocols for WET testing as published by the US EPA and the ANZECC/ARMCANZ water quality guidelines (2000), five effluent concentrations were selected along with a control. The concentrations comprising 0% effluent (control), 0.1% effluent, 1% effluent, 10% effluent, 50% effluent, 90% effluent and 100% effluent.

Sample collection, preparation and testing was completed over three (3) weeks in May 2017. Disinfected effluent from each WWTP and river water samples were collected on the first day of each test week and stored in a dark constant temperature (25 °C) room O/N to allow for water temperature adjustment to the testing conditions. On each sampling occasion, a total of 30 L of disinfected effluent from each WWTP and 60 L from each river site was collected to create the dilutions for the WET tests.

**Single species.** The single-species WET tests were completed based on standard procedures – US EPA Test Method 1003.0<sup>26</sup> and the Environment Canada test method<sup>27</sup> – using a 4 to 7-day old culture of the freshwater unicellular green algae *Raphidocelis subcapitata* (formerly also known as *Selenastrum capricornutum* and *Pseudokirchneriella subcapitata*) obtained from the NSW Department of Planning, Industry and Environment (formerly NSW Office of Environment and Heritage). Prior to WET testing, cells of *R. subcapitata* were washed three (3) times with artificial soft-water to remove culture media and extracellular substances. For this step, the algal culture was centrifuged at 700 g for seven (7) minutes, the supernatant discarded, and the pellet resuspended in artificial soft-water. Algal cells were counted using a haemocytometer and each test sample was inoculated with approximately  $3 \times 10^4$  microalgal cells/mL.

For the single-species WET tests, 200 mL dilutions were prepared using UV-disinfected effluent (filtered at 0.45 µm) from all five (5) WWTP and artificial soft-water (filtered at 0.22 µm). Artificial soft-water used for the tests was created by the addition of sodium bicarbonate ( $\text{NaHCO}_3$ ), calcium sulfate dihydrate ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ), magnesium sulfate heptahydrate ( $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ ), and potassium chloride ( $\text{KCl}$ ) to milli-Q water. The artificial soft-water was adjusted to the hardness of each effluent water type on the starting day of the WET tests. Total hardness as  $\text{CaCO}_3$  for effluents from each WWTP included: Kurri Kurri (106 mg/L); Farley (257 mg/L); Morpeth (78 mg/L); Raymond Terrace (65 mg/L); and Shortland (168 mg/L). Three (3) replicates of 50 mL each were added to 100 mL conical flasks and one (1) replicate of 50 mL was added to a 100 mL glass beaker. The flasks and beakers were covered with a 0.4 mm mesh to avoid any airborne particles contaminating the samples. Further, 200 µL of



2.1 g/L sodium nitrate (1.5 mg  $\text{NO}_3^-/\text{L}$ ) and 200  $\mu\text{L}$  of 0.22 g/L potassium dihydrogen phosphate (0.15 mg  $\text{PO}_4^{3-}/\text{L}$ ) was added to each sample to provide sufficient nutrients for exponential algal growth.

The test samples were incubated for 72 hours at continuous daylight conditions. During the tests, full-spectrum daylight fluorescent lighting (36 W) provided a light intensity of 71.6  $\text{W}/\text{m}^2$ . The shelves used for the tests were lined with aluminium foil and samples were randomised once a day to maximise light exposure. At the time of inoculation (time point 0 h), the concentrations of  $\text{NH}_3$ ,  $\text{NO}_x$ , TKN, TN, and TP, as well as water hardness were measured in each effluent type using standard inorganics and nutrients water testing suites. At the end of every nominal 24-hour period, test samples were mixed by swirling (conical flasks) or with pipette tips (beakers). 100  $\mu\text{L}$  from each conical flask was taken to determine cell counts using flow cytometry, and water quality microsensors (Unisense) were used to measure pH and DO levels in the beakers.

**Whole microbial community.** For the community-level microbial WET tests, dilutions were prepared using unfiltered UV-disinfected effluent from all five (5) WWTP and unfiltered river water from two (2) locations in the Hunter River estuary, including a freshwater site (salinity < 1 PSU, river site 'F' in Fig. 1) and a saltwater site (salinity > 30 PSU, river site 'S' in Fig. 1). These two (2) salinity regimes were chosen to include future scenarios based on a predicted increasing salinity of the Hunter River with rising sea levels. The community-level WET tests were run in triplicate 2 L plastic beakers for 72 hours at a 12/12 hour day/night light cycle. During the tests, LED-600 aquarium lights (BeamsWork) provided conditions suitable for the growth of photoautotrophic microbes, having a light frequency of 10,000 K (white LED) and 460 nm (blue LED), equivalent to 1,340 lumens and a light intensity of 65.1  $\text{W}/\text{m}^2$  (assuming a luminous efficacy of 40.98  $\text{lm}/\text{W}$ ). As for the single-species tests, the shelves used for the tests were lined with aluminium foil and the beaker positions on the shelves were randomised once a day to maximise the light exposure in each beaker.

At the time of inoculation (time point 0 h), concentrations of  $\text{NH}_3$ ,  $\text{NO}_x$ , TKN, TN, and TP were measured in each effluent type and river site sample. At the end of every nominal 24-hour period, beakers were mixed with a glass stirring rod and samples from each beaker were taken for further analysis, including 1 mL of water for microbial cell counts using flow cytometry, and approximately 500 mL of water, collected in sterile Whirl-Paks and stored on ice in the dark until filtering, for DNA extraction and sequencing. Note that filtering for DNA extraction during the WET tests followed the same protocol previously described for the microbial field surveys. Further, water quality microsensors (Unisense) were used to measure pH, DO and  $\text{H}_2\text{S}$  at the sampling times (i.e. every 24 hours) for each effluent type and dilution, except Shortland WWTP due to technical difficulties. Note that DO at time point 0 h was only measured for Raymond Terrace and Morpeth effluent types due to a temporary fault in the DO sensor.

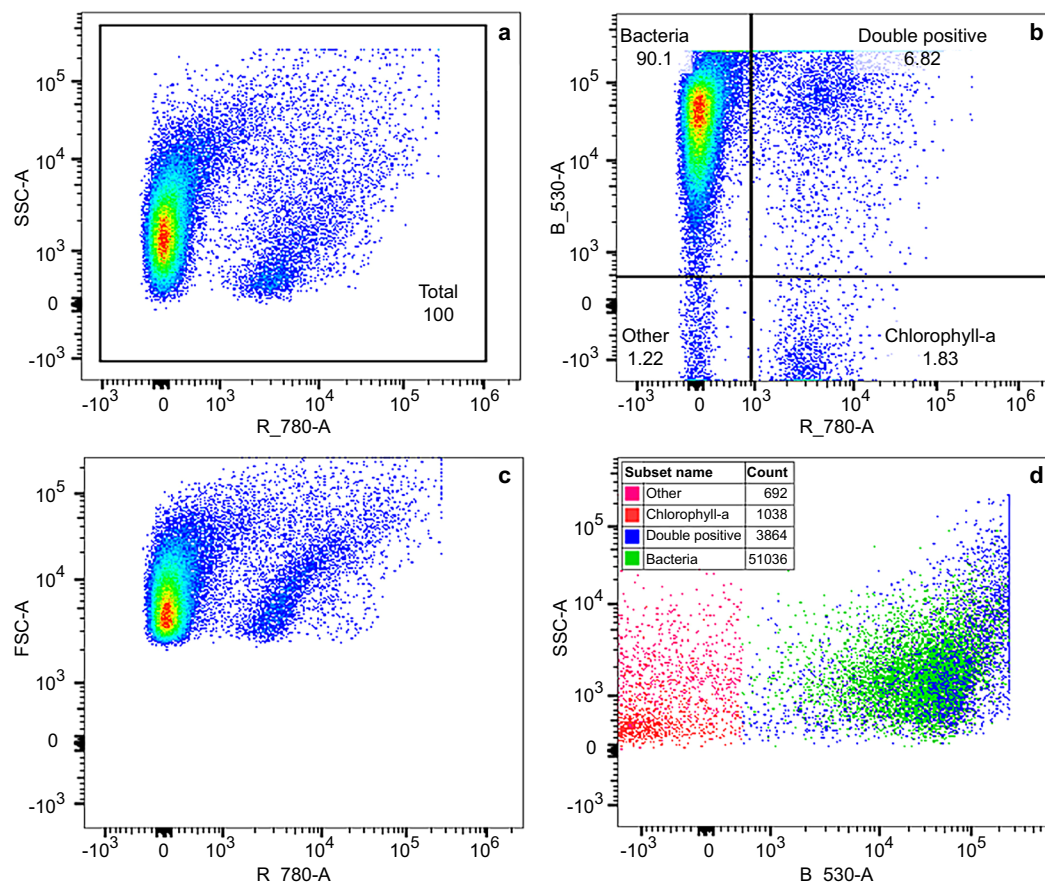
**Biological oxygen demand.** At the end of the community-level WET tests (time point 72 h), water was sampled from two (2) replicates of each effluent type and dilution for nutrient ( $\text{NH}_3$ ,  $\text{NO}_x$ , TKN, TN, TP) analysis, and 200 mL of the remaining water was used to estimate the BOD of the microbial communities. The water sampled for the BOD tests were capped in standard BOD bottles and incubated for 2.5 hours in the dark. DO measurements were taken at the start of the BOD tests (T0) and after 2.5 hours incubation in the dark (T1).

**Flow cytometry. Staining optimization.** The nucleic acid dye SYTO9 (ThermoFisher Scientific), which has been shown to preferentially stain live and dead total bacteria cells, was tested at varying dilutions (1:80, 1:500, 1:1000 v/v of the dye's stock solution and dimethyl sulfoxide (DMSO))<sup>28,29</sup>. All dilutions were tested using triplicate positive (stained) and negative controls (no stain). After stain optimisation using an analogue river water sample, all aliquots were subsequently stained using 0.2  $\mu\text{L}$  of SYTO9 (1:1000 v/v dilution of commercial stock solution with DMSO) and incubated for 15 minutes in the dark at room temperature prior to analysis<sup>28</sup>. Note the differentiation and quantification of both live and dead cells, using Propidium Iodide, was purposely omitted, since sequencing typically incorporates both live and dead cells, and the results were aimed at determining the total number of bacteria cells (live and dead) present in the sample.

**Flow cytometry analysis.** All cell count samples were analysed using a LSRFortessa SORP (BD Biosciences) flow cytometer with the FACSDiva software (v8.0.1) in the Mark Wainwright Analytical Centre (Flow Cytometry Facility) at UNSW Sydney. Cell count samples from the single-species and microbial community WET tests were stored on ice until flow cytometry analysis. The samples analysed in May 2017 ('Fresh') were left unstained and the autofluorescence from chlorophyll-a present in the live algal cells was used to determine cell densities. The chlorophyll-a fluorescence assays were designed to compare the growth response of single-celled algae between the two (2) WET setups. Additional water samples from the WET tests were stored at  $-20^\circ\text{C}$  and later used to quantify the total bacterial cell densities in February and March 2019 ('Frozen'). This was done to distinguish between the bacterial and algal (chlorophyll-a) cell populations in the water samples.

All cell count samples were separated into 200  $\mu\text{L}$  aliquots across eight (8) microplates (96-wells) by effluent type (Kurri Kurri and Farley, Morpeth and Raymond Terrace) and sampling time (0 h, 24 h, 48 h, 72 h). The Frozen samples were thawed quickly (at  $37^\circ\text{C}$ ) to avoid cell loss during the thawing process<sup>30</sup>, and after thawing, were mixed via manual shaking for 10 s. The defrosted 200  $\mu\text{L}$  aliquots in the microplate wells were stained and analysed automatically in standard-throughput mode.

A threshold of 200 cell counts was set on the side-scattered (SSC) light detector to exclude noise from non-algal and unstained bacterial cell particles. All readings were collected as logarithmic signals at a flow rate of 2.0  $\mu\text{L}/\text{s}$ . Maximum events were set to 10,000,000 to ensure the counting of all cells within a 60  $\mu\text{L}$  sample (100  $\mu\text{L}$  for the Fresh samples). Before cell counting, all samples were mixed twice-through by mechanical pipetting up and down of 100  $\mu\text{L}$  of sample at a mixing rate of 180  $\mu\text{L}/\text{s}$ . Note there was no cell count data available for the



**Fig. 2** Example cytograms of bacterial and chlorophyll-a populations resolved by fluorescence for 50% dilution of Kurri Kurri effluent type in freshwater (replicate 1) at 48 hours, February and March 2019 dataset. A bivariate density plot of SSC vs red fluorescence (chlorophyll) shows the total population (a). Regions on the green vs red fluorescence plot were defined for bacteria (Bacteria), cells containing chlorophyll and SYTO9 stain (Double positive), cells containing only chlorophyll (Chlorophyll-a) and all other cells (Other) (b). Double positives were counted as total bacteria due to the separation of cells shown in SSC vs green fluorescence (d).

community-level WET tests containing the Shortland effluent type due to complications with the flow cytometer at the time of the analysis.

**Enumeration and size.** The 488 nm laser was used for excitation of both SYTO9 stain and chlorophyll-a fluorescence. The detection was measured using emission filters for green (530 nm  $\pm$  15 nm) and red (780 nm  $\pm$  60 nm) fluorescence. Determination of the region within the SSC vs chlorophyll bivariate density plot, which included live algal cells containing chlorophyll-a, was done prior to the WET tests, using algal cell cultures and negative controls. Further, determination of the region defined for total bacteria within the SSC vs green fluorescence plot, was done during staining optimisation, using positive (stained) and negative (unstained) controls. Example cytograms of bacterial and chlorophyll-a populations resolved by fluorescence for 50% dilution of Kurri Kurri effluent type in freshwater (replicate 1) at 48 hours are provided in Fig. 2. Figure 2a shows the total population which is separated into quadrants within the green vs red fluorescence plot (Fig. 2b). The region defined for total bacteria is labelled “Bacteria”, while the region for cells containing chlorophyll and SYTO9 stain is labelled “Double positive”. The double positives were counted as bacteria due to the separation shown in SSC vs green fluorescence (SYTO9 stain) (Fig. 2d). Final cell counts were determined as the number of events within the previously selected regions. Presentation of the data as bivariate density plots (Fig. 2) enabled the best distinction between stained bacteria and chlorophyll-a. Following enumeration of the microbial cells, size was estimated using CountBright absolute counting beads (200 nm, 500 nm, 800 nm, 1  $\mu$ m, 3  $\mu$ m, 6  $\mu$ m) as volumetric standards. There may be some variation in the size estimates caused by the way cells are placed during the analysis.

**DNA extraction and sequencing.** DNA was extracted using the DNeasy PowerWater Kit (Qiagen) following manufacturer protocols and stored at  $-80^{\circ}\text{C}$  until sequencing. Bacterial communities were identified using the 27f (5'-3': AGAGTTTGATCMTGGCTCAG<sup>31</sup>) and 519r (5'-3': GWATTACCGCGGCKGCTG<sup>32</sup>) primers for the V1-V3 region of the 16S rRNA gene and Illumina<sup>TM</sup> MiSeq, 2  $\times$  300 base pair (bp), paired-end v2 sequencing runs across two (2) lanes. Eukaryotic composition was determined using the 1391 f (5'-3': GTACACACCGCCGTC) and EukBr (5'-3': TGATCCTTCTGCAGGTTACCTAC) primer set<sup>33</sup> for the V9 region of the 18S rRNA gene and Illumina<sup>TM</sup> MiSeq, 2  $\times$  150 bp paired-end v3 sequencing runs across two (2)

Primer pair	Category	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	Reference	Standard	Length (bp)
16S	bacteria	GTGSTGCAYGGYTGTCGTCA	ACGTCRTCCMCACCTTCCTC	48	AM910662.1	145
18S	eukaryotes	AAGGAAGGCAGCAGGCG	CACCAGACTTGCCCTCYAAT	49	KF297603.1	150
Acc_ppk1	phosphorous pathway	GATGACCCAGTTCCTGCTCG	CGGCACGAACTTCAGATCG	50,51	KP738079.1	133
ampC	antibiotic resistance	TGAGTTAGGTTCCGGTCAGCA	AGTATTTTGTGCGGGATCG	52	CP002729.1	98
AOA_8	nitrogen pathway	CTATTCATAGTTGTAGTTGCTGTAA	ATGTAGTCTCCTCGGTTGAT	38	KX181620.1	80
AOB_1	nitrogen pathway	GTCTCCATGCTCATGTTTC	GGAAAGCCTTCTTCGCC	38	KX683107.1	134
ciaB	pathogen	GCGTTTGTGAAAAAGATGAAGATAG	GGTGATTTTACTTTCATCCAAGC	53	HE978252.1	80
dsrA	sulphur pathway	ACSCACTGGAAGCAGC	GGTGAGCCGTCGATGTT	54	KM241895.1	222
ftsZ	pathogen	CTGGTGACCAATAAGCAGGTT	CATCCCATGCTGCTGGTAG	53	CP018103.1	60
mecA	antibiotic resistance	CTGATGGTATGCAACAAGTCG	TGAGTTCGTCAGTACCGGATT	52	CP000255.1	96
napA_14	nitrogen pathway	ATGTGGGTGGAGAAGGA	TGAAGCGCTTGAGAATT	38	LN901633.1	130
napA_3	nitrogen pathway	CCCAATGCTCGCCACTG	CATGTTKAGACCCACAG	55	CP014774.1	129
narG_1	nitrogen pathway	GACTTCCGCATGTCRAC	TTYTCGTACCAGGTGGC	55	DQ481141.1	68
narG	nitrogen pathway	TAYGTSGGCAGGARAAACTG	CGTAGAAGAAGCTGGTGCTGTT	56	HM104383.1	110
nifD_33	nitrogen pathway	TGCCGTTCCGCCAGATGCA	AGATGGCGAAGCCGTCATAGC	38	CP010523.2	69
nifH_32	nitrogen pathway	GGCGTCATCACCTCGATCA	GCATAGAGCGCCATCATCTC	38	AP017605.1	176
nirK	nitrogen pathway	ATYGCCGGVCAYGGCGA	GCCTCGATCAGRTTRTGGTT	57	KX036332.1	164
nirS_1-3	nitrogen pathway	CCTAYTGCCGCCRCART	GCCGCCGTCRTGVAGGAA	58	AB164133.1	257
nirS_ef	nitrogen pathway	CACCCGGAGTTCATCKTC	ACCTTGTGGACTGGTGGG	55	JN257972.1	173
norB_79	nitrogen pathway	GAATACTGGCGTTGGT	ATACTCAAAGAAGCCTTC	38	CP012027.1	55
nosZ_2	nitrogen pathway	CGCRACGGCAASAAGTSMSSGT	CAKRTGCAKSGCRTGGCAGAA	59	KJ137778.1	269
nrfA_2	nitrogen pathway	CACGACAGCAAGACTGCCG	CCGGCACTTTCGAGCCC	55	AM408255.1	68
phoD	phosphorous pathway	GCCATTTATGCCGACACT	TCCGATAGGCAGGCACATT	60	XM_002179950.1	130
phoNC	phosphorous pathway	CGGCTCTATCCGTCGGG	CAACATCGCTTGGCAGTG	61	AE008922.1	155
phsA	sulphur pathway	CGACCAGGACCTCATGCC	CTACCTGACCCTGCTTTGC	62	CP002297.1	106
plc	pathogen	CATCAGTTGGAAAAGATGTAAGAAGAAC	TGATTCCAAAATACATGATGTCATCTG	53	DQ184163.1	96
rubisco_bact_cbbL	carbon pathway	AAGGAYGACGAGAACAT	TGCAGSATCATGTCRTT	63,64	omitted	—
rubisco_diatom	carbon pathway	GATGAYGARAACATAACTC	TAAGAACCCTTAACYTCACC	65	EF143298.1	113
rubisco_haptophyte	carbon pathway	GAGAGCGTTTCTATTCTC	CACGTGCGTACATTTCTTC	65	KF536380.1	117
rubisco_synecho	carbon pathway	CATCAAGCTGTCCGAG	TGTTGGCYGTGAAGCC	65	JN692350.1	161
sir	sulphur pathway	CCGTGTACTCTCAACAAGATG	CCAATTCTGCCATGTAAGGAC	66	XM_002289791.1	101
stx2	pathogen	TCTGGCGTTAATGGAGTTYAG	GTGACAGTGACAAAACGCAGA	53	AB046175.1	78

**Table 3.** MFQPCR primers targeting nutrient cycling, pathogens and antibiotic resistance genes.

lanes. All sequencing was done at the Ramaciotti Centre for Genomics (UNSW Sydney). All raw sequence data are publicly available through the National Center for Biotechnology Information (NCBI)<sup>34</sup> under SRA study accession SRP224901. The SRA data record includes 1,596 experiments derived from 1,476 samples.

**Microfluidic quantitative polymerase chain reaction (MFQPCR).** The absolute abundance of nutrient (carbon, nitrogen, phosphorus and sulfur) cycling genes and genes associated with pathogens and antibiotic resistance were determined using microfluidic quantitative polymerase chain reaction (MFQPCR) on the Fluidigm platform. Like traditional qPCR, this method enables the determination of gene abundances through the measurement of fluorescence after each PCR cycle, but eliminates much of the manual pipetting from the sensitive qPCR reaction. Its use in ecological research in various environments has significantly increased in recent years, with a focus on fast and reliable detection of pathogens<sup>35–37</sup>.

Gene abundances were measured in two (2) out of three (3) replicates collected for microbial analyses from both the field surveys and WET tests. Suitable primers, as defined by<sup>38</sup>, were selected from the literature. Five (5) different gBlock gene fragments (Integrated DNA Technologies<sup>39</sup>) were designed as standards for the assays using targeted sequences for each primer pair sourced from NCBI (refer to Table 3) using Primer-BLAST<sup>40</sup>. Standard curves were generated using a dilution series of 10<sup>1</sup> to 10<sup>7</sup> copies/μL. Exact DNA concentrations of the samples were determined spectrophotometrically using the PicoGreen double-strand DNA kit (Life Technologies) on the ClarioSTAR<sup>®</sup> microplate reader (BMG Labtech) and samples were diluted to a final DNA concentration of approximately 7–8 ng/μL. To alleviate sample-specific inhibition<sup>41</sup>, DNA extracts were mixed with T4 gene 32 protein (New England Biolabs<sup>®</sup>), to a final concentration of 80 ng/μL. Specific target amplification (14 cycles) and MFQPCR were conducted at the UNSW Sydney Ramaciotti Centre for Genomics, as per<sup>38</sup> using Evagreen<sup>®</sup> chemistry and the 96.96 Fluidigm Dynamic Array<sup>™</sup> Integrated Fluidic Circuit. Thermocycling consisted of 95 °C for 1 min, followed by 35 cycles of 96 °C for 5 s and 60 °C for 20 s or 25 s, followed by melt curve analysis for 60–95 °C at a ramp rate of 1 °C/3 s. Reactions were conducted in triplicate in 6.7 nL volumes with a final primer concentration of 700 nM.

Data Type	Data File Tab	File Format
Design Table	Mapping	Col 1/2 (A/B)–Unique sample ID/name Col 3 (C)–Dataset Col 4 (D)–Date Col 5/6 (E/F)–Location/ID Col 7/8 (G/H)–Latitude/Longitude Col 9 (I)–WWTP Col 10 (J)–Distance from WWTP Col 11/12 (K/L)–Outfall location/ID Col 13 (M)–Interaction Col 14/15 (N/O)–Type/ID Col 16 (P)–Time point ID Col 17 (Q)–Replicate Col 18 (R)–Bioproject accession Col 19 (S)–Biosample accession Col 20 (T)–16s/18s Sequence ID
MFQPCR	MFQPCR	Col 1–20 (A–T)–Design table Col 21 (U)–Sample volume (mL) Col 22 (V)–Original sample DNA conc. (ng/uL) Col 23 (W)–Dilution factor Col 24 (X)–Submitted sample DNA conc. (ng/uL) Col 25–39 (Y–AM)–Gene abundance (copies/mL) Col 40–54 (AN–BB)–Gene abundance (copies/ng DNA) Col 55 (BC)–X/End of data record
Water Quality	WQ_nutrients	Col 1–20 (A–T)–Design table Col 21 (U)–NH <sub>3</sub> (mg/L) Col 22 (V)–BOD (mg/L) Col 23 (W)–Chl-a (mg/L) Col 24 (X)–NO <sub>x</sub> (mg/L) Col 25 (Y)–Sulphides (mg/L) Col 26 (Z)–TSS (mg/L) Col 27 (AA)–TKN (mg/L) Col 28 (AB)–TN (mg/L) Col 29 (AC)–TOC (mg/L) Col 30 (AD)–TP (mg/L) Col 31 (AE)–Temperature (°C) Col 32 (AF)–pH Col 33 (AG)–Turbidity (NTU) Col 34 (AH)–DO (mg/L) Col 35 (AI)–DO (% saturation) Col 36 (AJ)–TDS (g/L) Col 37 (AK)–Salinity (ppt) Col 38 (AL)–X/End of data record

**Table 4.** Dataset record – Field Microbial Survey.

Data Type	Data File Tab	File Format
Design Table	Mapping	Col 1/2 (A/B)–Unique sample ID/name Col 3 (C)–Dataset Col 4/5 (D/E)–Species/ID Col 6 (F)–Date Col 7/8 (G/H)–WWTP/ID Col 9/10 (I/J)–Latitude/Longitude Col 11/12/13 (K/L/M)–Wastewater concentration/code/ID Col 14/15/16 (N/O/P)–Water/code/ID Col 17/18 (Q/R)–Time point/ID Col 19 (S)–Replicate
Cell Counts	FCM_single-species	Col 1–19 (A–S)–Design table Col 20 (T)–Chlorophyll-a (Fresh) (cells/mL) Col 21 (U)–X/End of data record
Water Quality	WQ_nutrients	Col 1–19 (A–S)–Design table Col 20 (T)–pH Col 21 (U)–DO sat% Col 22 (V)–NH <sub>3</sub> (mg/L) Col 23 (W)–NO <sub>x</sub> (mg/L) Col 24 (X)–TKN (mg/L) Col 25 (Y)–TN (mg/L) Col 26 (Z)–TP (mg/L) Col 27 (AA)–X/End of data record

**Table 5.** Dataset record – Ecotox Single-Species Algae.

## Data Records

The data format and metadata for the full archived datasets<sup>34,42</sup> is provided in Tables 4–6. The design table in each dataset provides the lookup reference for each sample analysed for raw 16S and 18S rRNA gene sequences uploaded to the NCBI SRA database.



Data Type	Data File Tab	File Format
Design Table	Mapping	Col 1/2 (A/B)–Unique sample ID/name Col 3 (C)–Dataset Col 4/5 (D/E)–Species/ID Col 6 (F)–Date Col 7/8 (G/H)–WWTP/ID Col 9/10 (I/J)–Latitude/Longitude Col 11/12/13 (K/L/M)–Wastewater concentration/code/ID Col 14/15/16 (N/O/P)–Water/code/ID Col 17/18 (Q/R)–Time point/ID Col 19 (S)–Replicate Col 20 (T)–Bioproject accession Col 21 (U)–16s Sequence ID Col 22 (V)–16s Biosample accession Col 23 (W)–18s Sequence ID Col 24 (X)–18s Biosample accession
MFQPCR	MFQPCR	Col 1–19 (A–S)–Design table Col 20 (T)–Sample volume (mL) Col 21 (U)–Original sample DNA conc. (ng/uL) Col 22 (V)–Dilution factor Col 23 (W)–Submitted sample DNA conc. (ng/uL) Col 24–35 (X–AI)–Gene abundance (copies/mL) Col 36–47 (AJ–AU)–Gene abundance (copies/ng DNA) Col 48 (AV)–X/End of data record
Cell Counts	FCM_microbes	Col 1–19 (A–S)–Design table Col 20 (T)–Bacteria (Frozen) (cells/mL) Col 21 (U)–Chlorophyll-a (Fresh) (cells/mL) Col 22 (V)–Chlorophyll-a (Frozen) (cells/mL) Col 23 (W)–X/End of data record
Water Quality	WQ_nutrients	Col 1–19 (A–S)–Design table Col 20 (T)–pH Col 21 (U)–H <sub>2</sub> S (mg/L) Col 22 (V)–DO sat% Col 23 (W)–BOD (μmol/L) Col 24 (X)–Temperature (°C) Col 25 (Y)–NH <sub>3</sub> (mg/L) Col 26 (Z)–NO <sub>x</sub> (mg/L) Col 27 (AA)–TKN (mg/L) Col 28 (AB)–TN (mg/L) Col 29 (AC)–TP (mg/L) Col 30 (AD)–X/End of data record
Water Quality	BOD	Col 1–19 (A–S)–Design table Col 20 (T)–pH Col 21 (U)–H <sub>2</sub> S (mg/L) Col 22 (V)–DO sat% Col 23 (W)–BOD (μmol/L) Col 24 (X)–Temperature (°C) Col 25 (Y)–NH <sub>3</sub> (mg/L) Col 26 (Z)–NO <sub>x</sub> (mg/L) Col 27 (AA)–TKN (mg/L) Col 28 (AB)–TN (mg/L) Col 29 (AC)–TP (mg/L) Col 30 (AD)–X/End of data record

**Table 6.** Dataset record – Ecotox Microbial Community.

## Technical Validation

**Single-species WET tests.** The single-species WET tests fulfilled the standard validity criteria<sup>43</sup>, including less than 20% coefficient of variation; greater than one (1) doublings per day of algal cells in control treatments; and less than 1-unit change in pH over the test duration. Note a reference test using a range of concentrations of copper sulfide (Cu<sub>2</sub>S: 2 μg/L, 4 μg/L, 8 μg/L, 16 μg/L, 32 μg/L and 64 μg/L) was also completed to ensure normal reactivity of the microalgae, and to provide a base for potential comparisons with tests conducted at different facilities. These concentrations were achieved by adding aliquots of a CuSO<sub>4</sub> stock solution (2 mg/L) to soft-water of moderate hardness (100–120 mg CaCO<sub>3</sub>/L). The reference test using Cu<sub>2</sub>S showed a standard dose response of the algae, thus confirming the validity of the test conditions.

**Flow cytometry.** Between samples, fluidics lines were cleared with a wash volume of 500 μL undiluted phosphate-buffered saline (PBS) solution. Additional wash steps using wells containing bleach and then milli-Q water were included between effluent types on each microplate and at the end of each microplate to reduce sample carryover. Positive (stained) and negative (unstained) control samples containing aliquots of river water were also analysed at the end of each microplate for protocol validation.

**DNA extraction.** Extracted DNA was quantified using a NanoDrop Spectrophotometer (ThermoFisher Scientific) with concentrations greater than 5–10 ng/ul. To test if the 16S rRNA gene in the different samples could be amplified, PCRs were conducted using the same primer set that was later used for amplicon sequencing (27 f/519r). The PCR reaction mixture (50 μL) contained 10 μL of 5-fold MyTaq Reaction Buffer (Bioline, Alexandria, NSW), 0.4 μM of each primer, 1 U of MyTaq DNA Polymerase (Bioline), and approximately 10 ng of DNA as a template. The following thermal cycling scheme was used: initial denaturation at 95 °C for 2 min, 28

cycles of denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, followed by extension at 72 °C for 1 min. The final extension was carried out at 72 °C for 2 min. Negative controls were performed by using the reaction mixture without template. Genomic DNA from *Escherichia coli* DH5 $\alpha$  was used in the positive control. PCR products were checked with agarose gel electrophoresis. All samples amplified and amplicon size matched the length of the amplified 16S rRNA gene (approximately 500 bp).

**MFQPCR.** The PCR efficiencies of both standards and samples were calculated for each gene using the program LinRegPCR (v2017.1). Only genes with a PCR efficiency of greater than 80% for both standards and samples were further analysed. Further, only those genes with similar PCR efficiencies for the standard and samples ( $\pm 10\%$ ) were used. Afterwards, MFQPCR results were analysed using the Fluidigm Real-Time PCR Analysis program (v4.3.1) using recommended settings (peak sensitivity 7, peak ratio threshold 0.7, quality threshold 0.65). The first analysis step included a manual linear derivative baseline correction. This correction was developed by Fluidigm to correct for baseline drift in real-time PCR data. Ranges of desired melt curve peaks for each primer set were determined manually based on the melt curves of the gBlock standards. Genes with multiple distinguishable melt curve peaks, indicating multiple PCR products due to non-specificity of the primer pair, were excluded from further analyses. Moreover, only samples within the desired melt curve peaks for each primer set were used for further analysis.

Samples were run across two separate 96.96 IFCs; Survey and Ecotox, and quality thresholds were applied separately to each IFC run. A total of 15 primer pairs for genes involved in carbon fixation (rubisco\_synecho), nitrogen cycling (napA\_3, narG\_1, narG, nifD\_33, nifH\_32 and nirK), sulphur cycling (dsrA, phsA and sir), phosphorous uptake (Acc\_ppk1) and pathogen identification (ciaB, ftsZ and stx2) passed the above standard validity tests for the Survey IFC. In addition, the primer pair for the 18S rRNA gene (18S) passed the validity criteria and was included in the analyses as a proxy for the abundance of eukaryotic cells in the samples. However, for genes ciaB and stx2, only single data points (sample-primer combinations) passed the validity tests, thus excluding these two (2) pathogen-related genes from further analyses. For the Ecotox IFC, 12 primer pairs passed the standard validity tests, encompassing the nitrogen cycle (AOA\_8, AOB\_1, napA\_14, napA\_3, narG\_1, narG, nirS\_ef, nrfA\_2), sulphur cycling (dsrA, phsA), phosphorous cycle (phoD) and eukaryotes (18s).

## Usage Notes

In this study, bacterial cell densities were quantified in the microbial WET tests from frozen samples several months after the initial flow cytometry was done using fresh samples. Previous studies have reported decreases in bacterial<sup>44</sup> cell densities between natural and frozen samples. Therefore, these data are recommended for assessment of relative differences between samples since they may reflect a slight underestimation of the bacterial relative to algal cell chlorophyll-a contributions.

While several studies have found that MFQPCR produces copy number estimates that are directly comparable to those produced with traditional qPCR<sup>45,46</sup>, variations in reaction efficiency are common between samples from different sites<sup>47</sup>, and in MFQPCR these differences in efficiency may be substantial<sup>38</sup>. Therefore, whilst copy number tables presented here are suitable for intra-study comparisons and modelling, it is recommended that raw data files are utilised in studies intending to combine inter-study datasets, and uniform efficiency cut-offs be applied prior to further analysis.

Received: 7 February 2020; Accepted: 20 April 2020;

Published online: 27 May 2020

## References

1. Corcoran, E. C. *et al.* *Sick Water? The Central Role of Wastewater Management In Sustainable Development. A Rapid Response Assessment.* (United Nations Environment Programme, UN-Habitat, GRID-Arendal, 2010).
2. UN-Water. *Wastewater Management - Analytical Brief.* (United Nations World Water Assessment Programme, 2015).
3. UN-WWAP. *The United Nations World Water Development Report - Wastewater: The Untapped Resource.*, (United Nations World Water Assessment Programme, Paris, 2017).
4. UN-WWAP. *Water For People, Water For Life - The United Nations World Water Development Report.* (United Nations Educational, Scientific and Cultural Organisation (UNESCO) and Berghahn Books, 2003).
5. Oelsner, G. P. & Stets, E. G. Recent trends in nutrient and sediment loading to coastal areas of the conterminous U.S.: Insights and global context. *Sci. Total Environ.* **654**, 1225–1240, <https://doi.org/10.1016/j.scitotenv.2018.10.437> (2019).
6. Wurtsbaugh, W. A., Paerl, H. W. & Dodds, W. K. Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdiscip. Rev.: Water* **6**, e1373, <https://doi.org/10.1002/wat2.1373> (2019).
7. Drury, B., Rosi-Marshall, E. & Kelly, J. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Appl. Environ. Microbiol.* **79**, 1897–1905, <https://doi.org/10.1128/aem.03527-12> (2013).
8. Koutsou, O. P., Gatidou, G. & Stasinakis, A. S. Domestic wastewater management in Greece: Greenhouse gas emissions estimation at country scale. *J. Cleaner Prod.* **188**, 851–859, <https://doi.org/10.1016/j.jclepro.2018.04.039> (2018).
9. Nguyen, T. *et al.* Insight into greenhouse gases emissions from the two popular treatment technologies in municipal wastewater treatment processes. *Sci. Total Environ.* **671**, 1302–1313, <https://doi.org/10.1016/j.scitotenv.2019.03.386> (2019).
10. Zimmer, A., Winkler, I. T. & De Albuquerque, C. Governing wastewater, curbing pollution, and improving water quality for the realization of human rights. *Waterlines* **33**, 337–356, <https://doi.org/10.3362/2046-1887.2014.034> (2014).
11. Sato, T. *et al.* Global, regional, and country level need for data on wastewater generation, treatment, and use. *Agric. Water Manage.* **130**, 1–13, <https://doi.org/10.1016/j.agwat.2013.08.007> (2013).
12. Small, C. & Nicholls, R. J. A Global Analysis of Human Settlement in Coastal Zones. *J. Coast. Res.* **19**, <https://doi.org/10.2307/4299200> (2003).
13. UN-DESA. *World Population Prospects 2019 - Highlights.* (United Nations, Department of Economic and Social Affairs, Population Division, 2019).
14. Clark, G. & Johnston, E. *Australia State of the Environment 2016: Coasts, Independent Report to the Australian Government Minister for Environment and Energy.* (Australian Government Department of the Environment and Energy, 2017).

15. Schindler, D. W. Recent advances in the understanding and management of eutrophication. *Limnol. Oceanogr.* **51**, 356–363, [https://doi.org/10.4319/lo.2006.51.1\\_part\\_2.0356](https://doi.org/10.4319/lo.2006.51.1_part_2.0356) (2006).
16. Mara, D. *Domestic Wastewater Treatment In Developing Countries* (2013).
17. Palaniappan, M. *et al.* *Clearing The Waters - A Focus On Water Quality Solutions*. (United Nations Environment Programme (UNEP), 2010).
18. Hassett, B. A. *et al.* Pulling apart the urbanization axis: patterns of physiochemical degradation and biological response across stream ecosystems. *Freshw. Sci.* **37**, 653–672, <https://doi.org/10.1086/699387> (2018).
19. Kieft, B. *et al.* Microbial community structure-function relationships in Yaquina Bay estuary reveal spatially distinct carbon and nitrogen cycling capacities. *Front. Microbiol.* **9**, 1282–1282, <https://doi.org/10.3389/fmicb.2018.01282> (2018).
20. Jeffries, T. C. *et al.* Bacterioplankton dynamics within a large anthropogenically impacted urban estuary. *Front. Microbiol.* **6**, <https://doi.org/10.3389/fmicb.2015.01438> (2016).
21. Birrer, S. C., Dafforn, K. A. & Johnston, E. L. Microbial Community Responses to Contaminants and the Use of Molecular Techniques in *Microbial Ecotoxicology* (eds. Cravo-Laureau, C., Cagnon, C., Lauga, B. & Duran, R.), 165–183 (Springer International Publishing, 2017).
22. Carney, R. L. *et al.* River bacterioplankton community responses to a high inflow event. *Aquat. Microb. Ecol.* **75**, 187–205, <https://doi.org/10.3354/ame01758> (2015).
23. Roy, P. S. *et al.* Structure and function of south-east Australian estuaries. *Estuarine Coastal Shelf Sci.* **53**, 351–384, <https://doi.org/10.1006/ecss.2001.0796> (2001).
24. Ferguson, A., Potts, J., Dawson, G. & Mitrovic, S. M. Biogeochemical Function Of The Hunter River Tidal System - An Empirically Based Conceptual Model. (NSW Government, Office of Environment and Heritage (OEH), 2018).
25. Russell, K., Erskine, J. & Glamore, W. C. Tomago Wetland Rehabilitation Project: Integrated, Innovative Approaches. *NSW Coastal Conference* (Kiama, NSW, 2012).
26. US EPA. *Short-Term Methods for Estimating Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms*. 4th edn (United States Environment Protection Agency, 2002).
27. Environment Canada. *Biological Test Method: Growth Inhibition Test Using A Freshwater Alga*. 2nd edn (2007).
28. Brown, M. R. *et al.* A flow cytometry method for bacterial quantification and biomass estimates in activated sludge. *J. Microbiol. Methods* **160**, 73–83, <https://doi.org/10.1016/j.mimet.2019.03.022> (2019).
29. Duhamel, S. & Jacquet, S. Flow cytometric analysis of bacteria- and virus-like particles in lake sediments. *J. Microbiol. Methods* **64**, 316–332, <https://doi.org/10.1016/j.mimet.2005.05.008> (2006).
30. Marie, D., Partensky, F., Vaulot, D. & Brussaard, C. Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Curr. Protoc. Cytom.* **10**, <https://doi.org/10.1002/0471142956.cy1111s10> (1999).
31. Lane, D. J. 16S/23S rRNA sequencing in *Nucleic acid techniques in bacterial systematics* (eds. Stackebrandt, E. & Goodfellow, M.), 115–175 (John Wiley and Sons, New York, NY, 1991).
32. Lane, D. *et al.* Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. USA* **82**, 6955–6959, <https://doi.org/10.1073/pnas.82.20.6955> (1985).
33. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS one* **4**, <https://doi.org/10.1371/journal.pone.0006372> (2009).
34. Ruprecht, J. E. *et al.* Next-generation sequencing of 16S and 18S rRNA genes in Hunter River estuary. *NCBI Sequence Read Archive*, <https://identifiers.org/insdc.sra:SRP224901> (2019).
35. Ishii, S. *et al.* Water quality monitoring and risk assessment by simultaneous multipathogen quantification. *Environ. Sci. Technol.* **48**, 4744–4749, <https://doi.org/10.1021/es500578s> (2014).
36. Ishii, S., Segawa, T. & Okabe, S. Simultaneous quantification of multiple food- and waterborne pathogens by use of microfluidic quantitative PCR. *Appl. Environ. Microbiol.* **79**, 2891–2898, <https://doi.org/10.1128/AEM.00205-13> (2013).
37. Zhang, Q. *et al.* Correlations between pathogen concentration and fecal indicator marker genes in beach environments. *Sci. Total Environ.* **573**, 826–830, <https://doi.org/10.1016/j.scitotenv.2016.08.122> (2016).
38. Crane, S. L. *et al.* Microfluidic qPCR enables high throughput quantification of microbial functional genes but requires strict curation of primers. *Front. Environ. Sci.* **6**, <https://doi.org/10.3389/fenvs.2018.00145> (2018).
39. Sadik, N. *et al.* Quantification of multiple waterborne pathogens in drinking water, drainage channels, and surface water in Kampala, Uganda, during seasonal variation. *GeoHealth* **1**, 258–269, <https://doi.org/10.1002/2017GH000081> (2017).
40. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinf.* **13**, 134, <https://doi.org/10.1186/1471-2105-13-134> (2012).
41. Dandie, C. E. *et al.* Nitric oxide reductase-targeted real-time PCR quantification of denitrifier populations in soil. *Appl. Environ. Microbiol.* **73**, 4250–4258, <https://doi.org/10.1128/AEM.00081-07> (2007).
42. Ruprecht, J. E. *et al.* A novel real-world ecotoxicological dataset of pelagic microbial community responses to wastewater. *Dryad*, <https://doi.org/10.5061/dryad.2fqz612km> (2020).
43. OECD. *Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test*, <https://doi.org/10.1787/9789264069923-en> (2011).
44. Kamiya, E. *et al.* Effects of fixation and storage on flow cytometric analysis of marine bacteria. *J. Oceanogr.* **63**, 101–112, <https://doi.org/10.1007/s10872-007-0008-7> (2007).
45. Ishii, S. *et al.* Microfluidic quantitative PCR for simultaneous quantification of multiple viruses in environmental water samples. *Appl. Environ. Microbiol.* **80**, 7505–7511, <https://doi.org/10.1128/AEM.02578-14> (2014).
46. Byappanahalli, M. N., Nevers, M. B., Whitman, R. L. & Ishii, S. Application of a microfluidic quantitative polymerase chain reaction technique to monitor bacterial pathogens in beach water and complex environmental matrices. *Environ. Sci. Tech. Lett.* **2**, 347–351, <https://doi.org/10.1021/acs.estlett.5b00251> (2015).
47. Brankatschk, R., Bodenhausen, N., Zeyer, J. & Bürgmann, H. Efficiency of real-time qPCR depends on the template: a simple absolute quantification method correcting for qPCR efficiency variations in microbial community samples. *Appl. Environ. Microbiol.*, 07878–07811, <https://doi.org/10.1128/AEM.07878-11> (2012).
48. Maeda, H. *et al.* Quantitative real-time PCR using TaqMan and SYBR Green for *Actinobacillus Actinomycetemcomitans*, *Porphyromonas Gingivalis*, *Prevotella Intermedia*, *tetQ* gene and total bacteria. *FEMS Immunol. Med. Microbiol.* **39**, 81–86, [https://doi.org/10.1016/S0928-8244\(03\)00224-4](https://doi.org/10.1016/S0928-8244(03)00224-4) (2003).
49. Zhu, F. *et al.* Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* **52**, 79–92, <https://doi.org/10.1016/j.femsec.2004.10.006> (2005).
50. He, S., Gall, D. L. & McMahon, K. D. “*Candidatus Accumulibacter*” population structure in enhanced biological phosphorus removal sludges as revealed by polyphosphate kinase genes. *Appl. Environ. Microbiol.* **73**, 5865–5874, <https://doi.org/10.1128/AEM.01207-07> (2007).
51. Zeng, W. *et al.* Community structures and population dynamics of “*Candidatus Accumulibacter*” in activated sludges of wastewater treatment plants using *ppk1* as phylogenetic marker. *J. Environ. Sci.* **67**, 237–248, <https://doi.org/10.1016/j.jes.2017.09.001> (2018).
52. Fernando, D. M. *et al.* Detection of antibiotic resistance genes in source and drinking water samples from a first nations community in Canada. *Appl. Environ. Microbiol.* **82**, 4767–4775, <https://doi.org/10.1128/AEM.00798-16> (2016).
53. Ishii, S., Segawa, T. & Okabe, S. Simultaneous quantification of multiple food and waterborne pathogens by use of microfluidic quantitative PCR. *Appl. Environ. Microbiol.* **9**, 2891–2898, <https://doi.org/10.1128/AEM.00205-13> (2013).

54. Ben-Dov, E., Brenner, A. & Kushmaro, A. Quantification of sulfate-reducing bacteria in industrial wastewater by real-time polymerase chain reaction (PCR) using *dsrA* and *apsA* Genes. *Microb. Ecol.* **54**, 439–451, <https://doi.org/10.1007/s00248-007-9233-2> (2007).
55. Smith, C. J., Nedwell, D. B., Dong, L. F. & Osborn, A. M. Diversity and abundance of nitrate reductase genes (*narG* and *napA*), nitrite reductase genes (*nirS* and *nrfA*), and their transcripts in estuarine sediments. *Appl. Environ. Microbiol.* **73**, 3612–3622, <https://doi.org/10.1128/AEM.02894-06> (2007).
56. López-Gutiérrez, J. C. *et al.* Quantification of a novel group of nitrate-reducing bacteria in the environment by real-time PCR. *J. Microbiol. Meth.* **57**, 399–407, <https://doi.org/10.1016/j.mimet.2004.02.009> (2004).
57. Henry, S. *et al.* Quantification of denitrifying bacteria in soils by *nirK* gene targeted real-time PCR. *J. Microbiol. Meth.* **59**, 327–335, <https://doi.org/10.1016/j.mimet.2004.07.002> (2004).
58. Braker, G., Fesefeldt, A. & Witzel, K. P. Development of PCR primer systems for amplification of nitrite reductase genes (*nirK* and *nirS*) to detect denitrifying bacteria in environmental samples. *Appl. Environ. Microbiol.* **64**, 3769–3775 (1998).
59. Henry, S. *et al.* Quantitative detection of the *nosZ* gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, *narG*, *nirK*, and *nosZ* genes in soils. *Appl. Environ. Microbiol.* **72**, 5181–5189, <https://doi.org/10.1128/AEM.00231-06> (2006).
60. Feng, T. *et al.* Examination of metabolic responses to phosphorus limitation via proteomic analyses in the marine diatom *Phaeodactylum tricornutum*. *Sci. Rep.* **5**, 10373, <https://doi.org/10.1038/srep10373> (2015).
61. Fraser, T. D. *et al.* Quantification of bacterial non-specific acid (*phoC*) and alkaline (*phoD*) phosphatase genes in bulk and rhizosphere soil from organically managed soybean fields. *Appl. Soil Ecol.* **111**, 48–56, <https://doi.org/10.1016/j.apsoil.2016.11.013> (2017).
62. Christensen, G. A. The Role Of Rex In Regulation Of Sulfate Reduction In *Desulfovibrio Vulgaris* Hildenborough, University of Missouri–Columbia (2014).
63. Selesi, D. *et al.* Quantification of bacterial RubisCO genes in soils by *cbbL* targeted real-time PCR. *J. Microbiol. Meth.* **69**, 497–503, <https://doi.org/10.1016/j.mimet.2007.03.002> (2007).
64. Selesi, D., Schmid, M. & Hartmann, A. Diversity of green-like and red-like ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes (*cbbL*) in differently managed agricultural soils. *Appl. Environ. Microbiol.* **71**, 175–184, <https://doi.org/10.1128/AEM.71.1.175-184.2005> (2005).
65. John, D. E. *et al.* Phytoplankton carbon fixation gene (*RuBisCO*) transcripts and air-sea CO<sub>2</sub> flux in the Mississippi River plume. *The ISME Journal* **1**, 517–531, <https://doi.org/10.1038/ismej.2007.70> (2007).
66. Kettles, N., Kopriva, S. & Malin, G. Insights into the regulation of DMSP synthesis in the diatom *Thalassiosira Pseudonana* through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen. *PLoS one* **9**, e94795, <https://doi.org/10.1371/journal.pone.0094795> (2014).

## Acknowledgements

This dataset was collected as part of a larger project on the Hunter River estuary funded by Hunter Water Corporation. J.E.R. was supported by an Australian Government Research Training Program Scholarship. The authors would like to thank all staff and scientists who helped with the ecotoxicology experiments at the Sydney Institute of Marine Science. Special thanks also go to Chris Brownlee and Dr Emma Johansson Beves at the Mark Wainwright Analytical Centre (Flow Cytometry Facility) at UNSW Sydney for their assistance with collection of the cell count data.

## Author contributions

J.E.R., W.C.G., S.C.B., K.A.D. and E.L.J. were involved with the conceptualisation and design of the data collection program and experiments. J.E.R., S.C.B., S.L.C., J.V.D. and F.W. were involved with data acquisition and analysis. J.E.R. lead the writing of the manuscript. All authors were involved with the interpretation of the data and review of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.E.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020