



**HAL**  
open science

## Microarray (phylochip) analysis of freshwater pathogens at several sites along the Northern German coast transecting both estuarine and freshwaters

Julia Baudart, Delphine Guillebault, Erik Mielke, Thomas Meyer, Neeraj Tandon, Sabine Fischer, Wilfried Weigel, Linda Medlin

### ► To cite this version:

Julia Baudart, Delphine Guillebault, Erik Mielke, Thomas Meyer, Neeraj Tandon, et al.. Microarray (phylochip) analysis of freshwater pathogens at several sites along the Northern German coast transecting both estuarine and freshwaters. *Applied Microbiology and Biotechnology*, 2017, 101 (2), pp.871-886. 10.1007/s00253-016-7937-2 . hal-03941743

**HAL Id: hal-03941743**

**<https://hal.science/hal-03941743>**

Submitted on 16 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Microarray (phylochip) analysis of freshwater pathogens at several sites along the Northern German coast transecting both estuarine and freshwaters

Julia Baudart<sup>1</sup> · Delphine Guillebault<sup>2</sup> · Erik Mielke<sup>3</sup> · Thomas Meyer<sup>3</sup> · Neeraj Tandon<sup>4</sup> · Sabine Fischer<sup>4</sup> · Wilfried Weigel<sup>4</sup> · Linda K. Medlin<sup>5</sup>

Received: 20 July 2016 / Revised: 2 October 2016 / Accepted: 10 October 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Monitoring the quality of drinking water is an important issue for public health. Two of the main objectives of the European Project  $\mu$ AQUA were (i) the development of specific probes to detect and quantify pathogens in drinking water and (ii) the design of standardized sampling programs of water from different sources in Europe in order to obtain sufficient material for downstream analysis. Our phylochip contains barcodes that specifically identify freshwater pathogens for enabling the detection of organisms that can be risks for human health. Monitoring for organisms with molecular tools is rapid, more accurate and more reliable than traditional methods. Rapid detection means that mitigation strategies come into play faster with less harm to the community and to humans. Samples were collected from several waters in France, Germany, Ireland, Italy and Turkey over 2 years. We present microarray results for the presence of freshwater pathogens from brackish and freshwater sites in Northern Germany, and cyanobacterial cell numbers inferred from these sites. In a companion study from the same samples, cyanobacterial toxins were analyzed using two methods and those sites with highest

toxin values also had highest cell numbers as inferred from this microarray study.

**Keywords** Phylochips · Microarray · Freshwater pathogens · Cyanobacteria · Molecular barcodes

## Introduction

Pathogenic organisms occurring in lakes and rivers used as drinking water reservoirs represent a particularly serious health hazard and cause no less than 170,000 cases of water-related diseases annually worldwide. Moreover, global climate changes and massive migration fluxes from Africa and Asia are expected to perturb the existing ecological balance determining major changes in the type, abundance and distribution of pathogenic microbes and likely causing (re)-emergence of water-related pathogens. This has created a great deal of concern at the World Health Organization (WHO) and at the UNECE (United Nations Economic Commission for Europe), who have issued a protocol (<http://www.unece.org/index.php?id=35086>) asking governments to take appropriate actions to cope with the present alarming situation, which will deteriorate further if no direct and strong measures are taken. Reliable and safe water supplies are lacking for approximately 100 million European inhabitants, despite a large proportion of the European drinking water being obtained from surface waters. *Escherichia coli* has been chosen as faecal indicator bacteria (FIB) of water safety and is part of drinking water regulations (EU Directive 98/83/EC for the quality of water intended for human consumption) and is also used as a faecal pollution indicator for recreational bathing waters in Europe (Directive 2006/7/EC for the management of bathing water quality, which replaces Directive 76/160/EEC). This FIB is mainly used to monitor the degradation of water quality caused by human or animal faecal wastes.

✉ Linda K. Medlin  
lkm@mba.ac.uk

<sup>1</sup> Sorbonne Universités, UPMC University Paris 06, CNRS, Laboratoire de Biodiversité et Biotechnologies Microbiennes, (LBBM), Observatoire Océanologique, F-66650 Banyuls/Mer, France

<sup>2</sup> Microbia Environnement, Avenue de Fontaule, 66550 Banyuls sur Mer, France

<sup>3</sup> MariLim Aquatic Research, GmbH Heinrich-Wöhlk-Str.14, 24232 Kiel, Germany

<sup>4</sup> Scienion AG, Volmerstr. 7b, 12489 Berlin, Germany

<sup>5</sup> Marine Biological Association of the UK, The Citadel, Plymouth PL1 2PB, UK

Although, widely used in monitoring contamination levels, *E. coli* alone can lead to misleading information about pathogen levels. Indeed, diverse waterborne pathogens (including bacteria and protozoans) are natural inhabitants of water and have not an oral-faecal transmission pathway (e.g. *Legionella* spp., *Aeromonas* spp., *Mycobacterium avium* complex, some species of *Vibrio* spp. and *Naegleria* spp). Their presence in water cannot be correlated to the faecal pollution assessed by *E. coli* or other faecal indicators. The emerging of non-faecal waterborne outbreaks lead to the urgent monitoring of a larger diversity of pathogens in water resources having different sources. Our microarray for universal freshwater pathogens, both eukaryotic and prokaryotic, fills this need.

Molecular tools are an effective means of monitoring for very diverse biological targets (Ramírez-Castillo et al. 2015). The EU  $\mu$ AQUA project was conceived to implement a reliable system for the early and sensitive detection of waterborne pathogens and toxins based on the phylochip microarray detection method for monitoring. A microarray is an efficient, sensitive, robust, rapid and inexpensive test for the presence of any organisms, and they have been applied to detect and enumerate toxic algae, pathogenic bacteria, parasites and viruses, and to detect the presence of toxins and of potential toxin producers in European freshwaters (see review in Kegel et al. 2016).  $\mu$ AQUA developed an easy to use advanced molecular tool consisting of a universal microarray chip for the (1) high-throughput detection of known and emerging water-borne pathogens (bacteria, viruses, protozoa and cyanobacteria), (2) monitoring the presence of select species of diatoms, whose presence is reliable bioindicators to assess overall water quality and (3) detection the cyanobacterial toxins from gene expression. This universal microarray is designed to replace traditional methods for monitoring, which are laborious, technically demanding and time-consuming and detects RNA rather than amplified gene products and as such is more quantitative and other microarrays that are PCR based and detect only one group of pathogens (see Discussion). The method now has an ISO number (ISO 16578:2013(en), molecular biomarker analysis—general definitions and requirements of specific nucleic acid sequences, ([http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=57,185](http://www.iso.org/iso/catalogue_detail.htm?csnumber=57,185)) and, thus, is a recognized and validated method to detect nucleic acids in environmental samples. PCR-independent methods using high density microarrays (phylochips) provide the best evidence for analyzing changes in relative abundance in ribosomal ribonucleic acid (rRNA) transcript and gene concentration levels because of its high specificity, sensitivity and repeatability (DeAngelis et al. 2011) and reveals higher diversity than clone library studies (DeSantis et al. 2007).

The  $\mu$ AQUA microarray was field tested in six countries. Microarray results for one of the monitoring sites, the Tiber River in Italy, have been published (Marcheggiani et al. 2015)

as has the PCR and rtPCR results from a Bulgarian site (Panaiotov et al. 2015), and in this paper, we present microarray results for bacteria, cyanobacterial and protozoa for selected sites in Germany monitored regularly by Marilim for 3 years from 2012 until 2014.

## Material and methods

### Sampling site and water sampling

Fifty litres of water were collected at nine freshwater and six brackish water sites (Fig. 1, Tables 1 and 2). The estuarine stations Schlei 1/Kappeln (BW2) and Kieler Förde 1/Neumühlen-Dietrichsdorf (BW1) are located inner coastal waters, which are influenced by steady freshwater influx. Several rivers (Füsinger Au, Koseler Au (near Bohnert), the Hüttener Au and the Osterbyer Au (near Fleckeby) as well as some unnamed brooks flow into the Schlei. The station Kieler Förde 1/Neumühlen-Dietrichsdorf (BW1) is located close to the river Schwentine. Only small salinity fluctuations exist at station Fehmarnsund (BW6). Three stations (Schlei 1, BW2; Kieler Förde 1, BW1 and Eckernförder Bucht, BW5) are close to harbour facilities, which could receive higher charges of abiotic pollution. BW3 (Kieler Förde 2) and BW5 (Eckernförder Bucht) have the highest salinities nearly 15 psu = practical salinity units.

The freshwater stations are Wittensee/Sande (FW8), Schwentine/Wellingdorf (FW3), Schwentine/Rosen See (FW1 and FW2), Selenter See/Fargau (FW4 and FW5), Neukirchener See (FW6 and FW7) and Keller See/Malente (FW9). The latter two are interconnected by creeks and channels.

At every sampling site, 50 l of water were collected approximately 20 cm below the water surface using two 20 and one 10 l sterilized carboys and transported to Marilim in a thermo box (between 0.5 to 1.5 h transport time). All abiotic measurements were done in situ. Oxygen and temperature were measured with a WTW Multi 3420 Set G with FDO 925–3 Probe and salinity/conductivity and temperature were measured with a WTW LF 325-A Set. (WTW GmbH, Weilheim, Germany). Five turbidity measurements were made from every sample using infrared spectrometry (AquaLytic Turbidimeter AL250T-IR, Dortmund, Germany). Wind speed and direction were taken from local weather stations. To maintain light levels approximately constant, all samples were collected between 7:30 am and 10:15 am. From the six freshwater sites, two of them (Keller See and Neukirchener See) were repeatedly sampled from 2012 to 2014.

### Sample preparation

In the laboratory, the water sample (50 l) was concentrated to 1 l using the Hemoflow kidney dialysis hollow fibre



**Fig. 1** Location of freshwater (FW) and estuarine (BW) sampling sites. The sampling sites and their numbers can be cross-referenced to Tables 1 and 2 for more details

filter (Hemoflow HF80S, Fresenius, Bad Homburg, Germany) and back-flashed with 1 l of elution buffer (0.01 M phosphate buffered saline, 0.5 M Tween 80, 0.01 % sodium hexametaphosphate, 10 % antifoam B

**Table 1** Freshwater sites, sampling dates and abiotic on-site measurement data. (–) no data. The sample collected in Schwentine in April 2012 was not analyzed further because of poor RNA quality after extraction (xxx)

Location	Sample site code	Date	Time	Air temperature [°C]	Water temperature [°C]	Air moisture [%]	Turbidity [NTU]	Conductivity [ $\mu\text{S}/\text{cm}^2$ ]
Schwentine	xxx	04.04.12	07:30	5.8	6.9	71.8	4.0	458
Seelenter See (Fargau)	FW4	21.08.12	09:15	19.5	21	75.9	1.11	301
Rosensee (Schwentine)	FW1	12.09.12	10:00	17.1	17.3	63.6	4.31	415
Wittensee (Sande)	FW8	12.06.13	09:00	21.9	17.8	–	1.13	396
Kellersee 1	FW9	13.08.13	09:20	19.2	20.0	–	3.25	410
Neukirchener See 1	FW6	22.08.13	08:40	20.5	19.9	–	42.62	292
Seelenter See (Fargau)	FW5	27.08.13	08:50	20.2	19.0	–	1.92	327
Rosensee (Schwentine)	FW2	29.08.13	08:44	18.6	19.1	48.9	3.67	437
Schwentine	FW3	14.10.13	09:40	15.0	11.2	59.5	1.59	466
Neukirchener See 2	FW7	27.03.14	08:00	6.2	8.1	76.0	5.43	430

**Table 2** Estuarine sites with mesohaline conditions, sampling dates and abiotic on-site measurement data. (–) = no data. The sample collected in Orther Bucht in July 2013 was not analyzed further because of poor RNA quality after extraction (xxx)

Location	Sampling site code	Date	Time	Air temperature [°C]	Water temperature [°C]	Turbidity [NTU]	Salinity [psu]
Schlei 1 (Kappeln)	BW2	05.06.13	09:00	21.7	16.9	7.04	9.5
Kieler Förde 1 (Neumühlen Dietrichsdorf)	BW1	26.06.13	09:05	13.9	15.0	2.0	12.1
Orther Bucht (Fehmarn)	xxx	30.07.13	09:53	19.3	21.3	1.16	11.0
Kieler Förde 2 (Heikendorf)	BW3	06.08.13	10:00	20.9	21.3	1.42	14.9
Schlei 2 (Maasholmer Breite)	BW4	15.08.13	10:15	19.8	17.8	2.08	13.4
Eckernförder Bucht	BW5	20.08.13	09:53	19.0	18.2	0.67	14.8
Fehmarnsund	BW6	09.09.13	09:55	14.5	16.7	1.48	12.7

emulsion) to yield a 1-l concentrate. During experimental pre-testing, the eluate was examined to see if any cells were lysed. No lysed cells were noted, and if they had been, the cut-off of the filter is 70 Da; the size of urea and any DNA/RNA from lysed cells would have been retained by the filter.

Two hundred millilitre of the concentrated eluate was sequentially filtered through the eight different filters on the same day as its collection of decreasing pore size (20, 10, 5, 2, 0.8, 0.45, 0.1 and 0.025 µm filters, Millipore, Billerica, Mass, USA). Each filter was placed into 1 ml of Tri-reagent (Sigma Aldrich, St Louis, MO, USA) (Chomczynski and Sacchi 1987), then stored frozen at –80 °C.

### RNA extractions

Before RNA extraction, all filters corresponded to a same water sample were pooled, and total RNA was extracted with Tri-Reagent following the protocol in Lewis et al. (2012) as modified in Kegel et al. (2013). From the total RNA extractions, one freshwater and two brackish water samples were not analyzed further because their RNA extractions were of poor quality (Tables 1 and 2).

### Probe design

The protocols used in the EU project µAQUA were adopted from the EU MIDTAL project for toxic algae (Lewis et al. 2012). Nevertheless, the challenge of µAQUA was to provide a single microarray able to analyze in the same water sample, many members of four phylogenetically different taxa of pathogenic bacteria, toxic cyanobacteria, pathogenic protozoa and diatoms as indicator species of water quality. All probes were designed to improve the specificity of the detection and to allow the hybridization with the targets under unique condition (buffer and temperature of hybridization).

Briefly probes representing species, genera, classes or phyla of the four target groups were either collected from the literature and extended to 25 nts or newly designed with the ARB

program (<http://www.arb-silva.de>) from the ribosomal rRNA genes following a hierarchical fashion. Thus, for a species or genus to be present, all of the taxonomic hierarchy leading to that taxon also had to be present (Table 3). This hierarchical approach to positive target signals prevents the detection of false positives. All probes were checked in silico for the specific recognition of their targets using the nucleotide Basic Local Alignment Search Tool (BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) and Silva (<http://www.arb-silva.de>) databases. We only accepted probes that had a 100 % match to their target species, with no variations among strains of the intended target as some phylochips allow (Phylochip, DeSantis et al. 2007). In a couple of instances, our probes had a single base mismatch, and for those we designed, a competitor probe to bind to the single base mismatches, allow true targets to bind to the probe barcode. Their biophysical properties were analyzed using the Oligonucleotide Properties Calculator software (<http://www.basic.northwestern.edu/biotools/oligocalc.html>). Positive control probes and higher taxonomic probes targeting kingdom and phylum level came from the MIDTAL microarray (Lewis et al. 2012), which is commercially available from Microbia Environnement (Banyuls-sur-Mer, France).

Probes, whose results are presented here, were designed for these bacterial families, genera and species: *Salmonella*, *Shigella*, *Campylobacter*, *Campylobacter coli* and *Campylobacter jejuni*, *Enterobacteriaceae*, *E. coli*, *Legionella*, *Legionella pneumophila*, *Clostridium perfringens*, *Listeria*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Yersinia*, *Yersinia enterocolitica*, *Vibrionaceae*, *Aeromonas*, *Bacillus cereus*, *Pseudomonas* and *Mycobacterium*; for these cyanobacterial genera and species: *Microcystis aeruginosa*, *Planktothrix*, *Nodularia spumigena*, *Anabaena*, *Aphanizomenon*, *Aphanizomenon flos-aquae* and *Cylindrospermopsis*; and for these protozoans: *Cryptosporidium parvum* and *Cryptosporidium hominis*, *Giardia*, *Entamoeba*, *Naegleria fowleri* and *Naegleria lovensis*, and various higher group level probes were designed ranging from family, order, class, phylum to domain depending on target availability. This hierarchy is presented in Table 3.

**Table 3** Summary of hierarchical probes for *Cyanobacteria*, *Bacteria* and *Protozoa*. The hierarchy should be read from left to right across each row, beginning with either a target species (if a species probe was designed) or a target genus and ending with the kingdom level probes. Where multiple order levels are noted, then the first order probe comes under the second order and so on

Species	Genus or multiple genera	Order	2nd level of multiple orders <sup>a</sup>	3rd level of multiple orders <sup>a</sup>	Phylum/division	Kingdom
A. Hierarchy for <i>Cyanobacteria</i> (has multiple levels of order probes)						
<i>Microcystis</i>	<i>Microcystis</i>	Chlorococcales	Chroococcales/Nostocales		Cyanobacteria	Eubacteria
<i>Anabaena/Aphanizomenon/Cylindrospermopsis</i>	<i>Anabaena/Aphanizomenon/Cylindrospermopsis</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Anabaena/Aphanizomenon</i>	<i>Anabaena/Aphanizomenon</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Anabaena/Aphanizomenon/Nodularia</i>	<i>Anabaena/Aphanizomenon/Nodularia</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Aphanizomenon</i>	<i>Aphanizomenon</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Cylindrospermopsis</i>	<i>Cylindrospermopsis</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Nodularia/Anabaena</i>	<i>Nodularia/Anabaena</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Nodularia/Aphanizomenon</i>	<i>Nodularia/Aphanizomenon</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Nodularia</i>	<i>Nodularia</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Nodularia</i>	<i>Nodularia</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Planktothrix</i>	<i>Planktothrix</i>	Oscillatoriales	Nostocales/Oscillatoriales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
B. Hierarchy for <i>Bacteria</i>						
<i>Nodularia spumigena</i>	<i>Nodularia</i>	Nostocales	Oscillatoriales/Nostocales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
<i>Planktothrix agaritii</i>	<i>Planktothrix</i>	Oscillatoriales	Nostocales/Oscillatoriales	Chroococcales/Nostocales	Cyanobacteria	Eubacteria
Species	Genus	Family/class	Phylum/division	Kingdom		
<i>Yersinia enterocolitica</i>	<i>E. coli/Shigella</i>	Enterobacteriaceae	Gammaaproteobacteria	Eubacteria		
<i>Legionella pneumophila</i>	<i>E. coli/Shigella/Salmonella</i>	Enterobacteriaceae	Gammaaproteobacteria	Eubacteria		
<i>Campylobacter coli/jejuni</i>	<i>Salmonella</i>	Enterobacteriaceae	Gammaaproteobacteria	Eubacteria		
<i>Staphylococcus aureus</i>	<i>Yersinia</i>	Enterobacteriaceae	Gammaaproteobacteria	Eubacteria		
<i>Clostridium perfringens</i>	<i>Vibrio</i>	Enterobacteriaceae	Gammaaproteobacteria	Eubacteria		
	<i>Legionella</i>					
	<i>Pseudomonas</i>					
	<i>Campylobacter</i>					
	<i>Staphylococcus</i>					
	<i>Listeria</i>					
	<i>Bacillus</i>					
	<i>Mycobacterium</i>					
C. Hierarchy for <i>Protozoa</i>						
Species	Genus	Family	Order	Class	Kingdom	
<i>Cryptosporidium hominis</i>	<i>Cryptosporidium</i>	Cryptosporidae	Emeriorina	Conoidasida	Eukaryote	
<i>Cryptosporidium parvum/hominis</i>	<i>Cryptosporidium</i>	Cryptosporidae	Emeriorina	Conoidasida	Eukaryote	
<i>Cryptosporidium parvum/hominis</i>	<i>Cryptosporidium</i>	Cryptosporidae	Emeriorina	Conoidasida	Eukaryote	
<i>Cryptosporidium parvum</i>	<i>Cryptosporidium</i>	Cryptosporidae	Emeriorina	Conoidasida	Eukaryote	
<i>Giardia intestinalis</i>	<i>Giardia</i>				Eukaryote	
<i>Naegleria fowleri</i>	<i>Naegleria</i>				Eukaryote	
<i>Naegleria lovaniensis</i>	<i>Naegleria</i>				Eukaryote	
<i>Entamoeba histolytica</i>	<i>Entamoeba</i>				Eukaryote	

<sup>a</sup> The second name is the stronger probe and the probe in the hierarchy

## Microarray construction and hybridization

A total of 246 probes (taxon and controls) were spotted by Scienion AG (Berlin, Germany) on sciCHIP Epoxy using a sciFLEXARRAYER S11 as follows: each microarray slide contained two arrays with eight replicates for each probe. Hybridizations of each sample were performed on different slides, thus producing a pseudo-replicate. Considering two arrays per sample, each probe is therefore represented by 16 spots and the signal for the 16 spots was averaged. One millilitre of the mixture obtained from the pooled filters previously stored in TRIzol<sup>®</sup> Reagent at  $-80^{\circ}\text{C}$  plus an internal extraction quality control (500,000 cells of *Dunaliella tertiolecta*) was processed for total RNA extraction using TRIzol<sup>®</sup> Reagent according to the patented MIDTAL procedure (patent WO2015008011 A1). RNA quality and purity (260/280 ratio 1.8–2.2 and 260/230 ratio 1.8–2.3) were measured by NanoDrop<sup>®</sup> Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The integrity and size distribution of total RNA were checked with a 2100 Bioanalyzer (Agilent Technologies Inc. Santa Clara, CA, USA). Clean-up was performed using the RNeasy Plus Mini kit (Qiagen, Düsseldorf, Germany). An extra clean-up round for samples that did not meet the total RNA purity (260/230 and 260/280 ratios) criteria using the PowerClean<sup>®</sup> Pro RNA Clean-Up Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA). The PowerClean<sup>®</sup> Pro RNA Clean-Up Kit will remove this brown colour as well as any PCR-inhibiting substances, such as heme, polysaccharides, polyphenols fulvic acids and dyes. This kit was validated with RNA isolated from a variety of problematic soils and also with RNA samples spiked with commercial humic acids (data not shown). One microgram of total RNA extracted of field samples was labelled and purified using a Platinum Bright 647 Infrared Nucleic Acid kit (Leica Biosystem, Nussloch GmbH, Germany) according to the manufacturer's instructions. The degree of labelling (DoL) was determined by measuring concentration and incorporation of the dye using a NanoDrop<sup>®</sup> Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Samples with DoL values between 1.0 and 3.0 were processed to hybridisation. Labelled RNAs were fragmented by adding 1/10 volume of fragmentation buffer (salt buffer) (100 mM  $\text{ZnCl}_2$  in 100 mM Tris-HCl, pH 7.0) (Lewis et al. 2012) and incubated for 15 min at  $70^{\circ}\text{C}$  and immediately chilled down on ice to room temperature. The reaction was stopped by adding 1/10 volume of 0.5 M EDTA, pH = 8 to the sample. Microarray hybridizations were performed following optimized procedures based on protocols published in Kegel et al. (2013) and Medlin (2013). Briefly, labelled field samples (1  $\mu\text{g}$  RNA) were mixed with 5X (2X) hybridization buffer containing 3  $\mu\text{L}$  Poly-dA (1  $\mu\text{M}$ ) and 10 ng TATA box protein (TBP) control made up to a final volume of 60  $\mu\text{L}$  (30  $\mu\text{L}$ ). Poly-dA is added to block the poly-T spacer on the probe, and

TBP is the TATA box gene fragment added as the positive hybridization control (Lewis et al. 2012). The labelled RNA was then denatured for 5 min at  $95^{\circ}\text{C}$ . After denaturation, the samples were placed on ice and 15  $\mu\text{L}$  of 4X KREA block (background blocker from Leica Biosystems, Nussloch GmbH, Germany) were added. The hybridization mixture was equally distributed to each array, which were pre-covered with coverslips cleaned with ethanol (LifterSlips, Erie Scientific, VWR International, Radnor, PA, USA). Slides were placed into SCIENION's sciHYBCHAMBER (Scienion, Berlin, Germany) that maintained a moist environment to avoid evaporation, and hybridizations were carried out for 1 h at  $65^{\circ}\text{C}$  using a water bath. In order to remove unhybridised RNA, the slides were successively washed with three washing steps with increasing buffer stringency under agitation and in the dark to protect the fluorophore. The first buffer (2X SSC/10 mM EDTA/0.05 % SDS) and the second buffer (0.5X SSC/10 mM EDTA) washings were done at room temperature for 10 min. Finally, a third most stringent wash (0.2X SSC/10 mM EDTA) was performed at  $50^{\circ}\text{C}$ .

## Microarray scanning

Fluorescence scanning of slides was performed using the LS Reloaded confocal fluorescence scanner (Tecan Group Ltd., Männedorf, Switzerland) with a resolution of 6  $\mu\text{m}$  and an excitation wavelength of 635 nm.

## Data analysis

The scanned images were analyzed with GenePix 5.1 software (Molecular Devices, Sunnyvale, Ca) to align spots with the corresponding gal files. The signal to noise ratios of the spots were calculated from the median spot intensities and the local background fluorescence. The mean values for the spot intensities for the eight replicates were calculated from the median spot intensities. The hybridization results were analyzed using the hierarchy file and the GPR-Analyzer v1.28 software (Dittami and Edvardsen 2013), and signals for the lowest taxon possible were eliminated from analysis if the hierarchy was broken. Multiple probes at each hierarchy level also helped to ensure that we did not have any false negatives. During the testing of probe specificity, we eliminated all probes that did not display clear, strong signals with pure cultures of each target. The cut-off to accept a positive hit of a set of probes was a normalized signal  $>0.2$  and a signal to noise ratio  $>2$  according the MIDTAL protocol (Lewis et al. 2012). Positive controls were represented by the TBP and eukaryotic probes (POSITIVE\_25\_dT, EUK 328, EUK 1209) and the *Dunaliella* probes (DunGS05\_25\_dT, DunGS02\_25\_dT). To compare different hybridization experiments, fluorescent signals were normalized using the internal positive control TBP (Positive\_25\_DT). This probe was spotted in five different

concentrations, and the concentration giving the lowest signal was used for the normalization. All microarray results were uploaded to the  $\mu$ AQUA database at <http://www.microaqua.eu>.

We analysed the microarray signal of each probes and for each biological target under PermutMatrix developed for microarray analyses (Caraux and Pinloche 2005). PermutMatrix is a free software that allows a heat map representation of the microarray data (<http://www.atgc-montpellier.fr/permutmatrix/>).

### Calibration curves for cyanobacteria quantification using microarray

To perform the calibration curves, cyanobacteria were quantified by cell counts using the Utermöhl method (Utermöhl 1958). RNA was extracted from a known cell number from pure cultures of *Microcystis*, *Planktothrix*, *Nodularia*, *Aphanizomenon* and *Anabaena*. Microarray analysis were performed with four different concentrations of RNA equivalent to 10,000, 50,000, 100,000 and 500,000 cells to construct concentration curves and to convert microarray signals to cell numbers. From these concentration curves, we can extrapolate to numbers of cells per litre in the field samples.

## Results

Air temperature ranged from 5.8 °C (Schwentine April 2012) to 21.9 °C (Wittensee June 2012). Water temperatures ranged between 6.9 °C (Schwentine April 2012) and 21.3 °C (Orther Bucht in July 2013 and Kieler Förde in August 2013) (Tables 1 and 2). Rising temperatures from April to July/August enhanced the growth rate of *Aphanizomenon*, *Anabaena* and *Nodularia*. Each sample was visually inspected, and the dominant species were noted along with their relative abundance but not enumerated. *Aphanizomenon*, *Anabaena* and *Nodularia* started to bloom at the end of June 2013, whereas *Cylindrospermopsis* reproduced mainly at lower temperatures (highest reproduction at 16.9 °C—Schlei, June 2013). At the end of the *Aphanizomenon*, *Anabaena* and *Nodularia* blooms, the blooming of *Cylindrospermopsis* decreased, then increased again for short time. Thus, at temperatures higher than 18 °C, *Cylindrospermopsis* does not compete with *Aphanizomenon*, *Anabaena* and *Nodularia*.

Turbidity values did not much differ between months and locations, except for the Neukirchener See (42.62 NTU) and the Eckernförder Bucht (0.67 NTU) in August 2013. Long-term data are unavailable. We have no evidence that turbidity affected cyanobacterial growth, but cyanobacteria can colonize successfully low light areas by using phycoerythrin in addition to chlorophyll, which increases the efficiency of light utilization for photosynthesis.

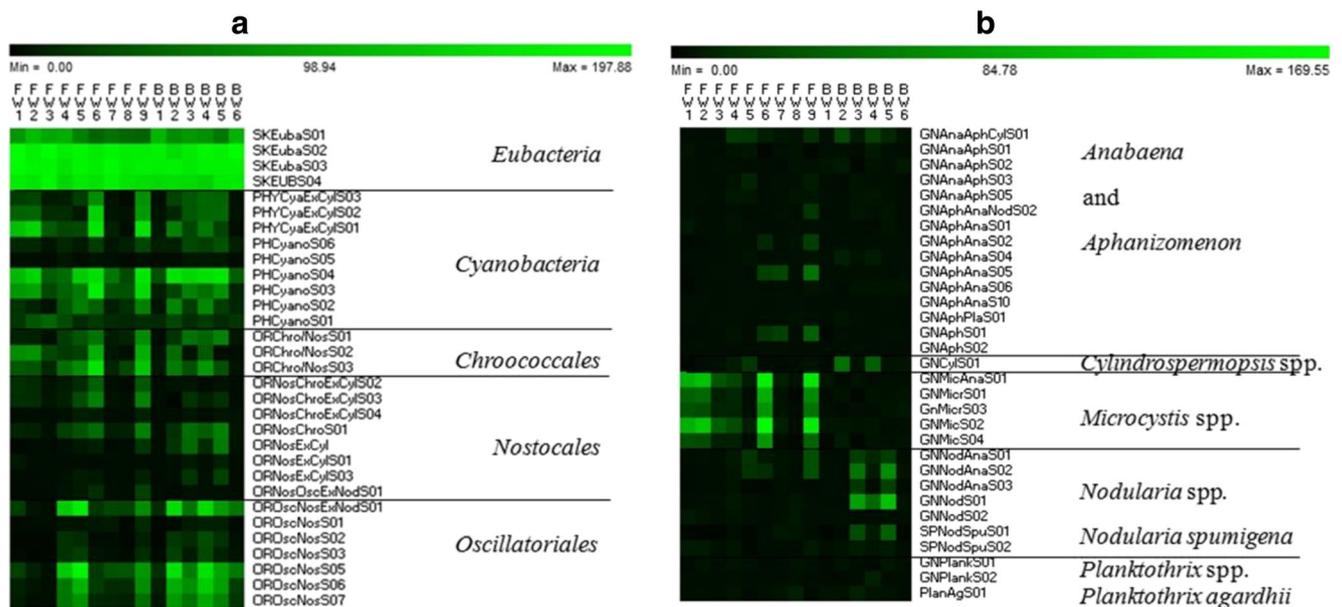
The hollow fibre ultrafiltration method provided a concentrate that was used for downstream RNA extraction for the

microarray. We did not encounter any problems with the filtration but only with the RNA extraction of two samples, which were not used. Hollow fibre filtration has been shown to be a very effective means of concentrating microorganisms from large volumes with little or no bias for cell retention and high recovery of spiked cells in experimental setups to test filtration efficiency (Hill et al. 2005). Cells in the concentrate are not lysed because this is a kidney dialysis filter, and all blood cells must remain intact for effective dialysis. Thus, the results presented below likely reflect an accurate representation of the microbial community present at the time of sampling, and relative abundances of the cyanobacteria were inferred from calibration curves made from actively growing cultures.

## Cyanobacteria

All eubacterial probes were highlighted validating bacterial presence in all samples (Fig. 2a). Cyanobacterial phylum level probes produced a high fluorescent signal in all samples except for FW3, FW7, FW8, BW1, and BW6 sites where some fluorescent signals were weaker. One phylum level probe (PhCynaoS05) did not work in all samples. At the order level, fluorescent signal distribution among these probes showed that cyanobacterial populations were different between freshwater and brackish water. The freshwater samplings had stronger signals for the orders *Chroococcales* and *Oscillatoriales*, whereas the brackish ones had higher signals for the order *Oscillatoriales* (Fig. 2a). *Nostocales* signals were not very strong among samples irrespective of their origin and were more or less absent from the estuarine site, BW1 site Kieler Förde 1/Neumühlen Dietrichsdorf and from the freshwater site, FW3 Schwentine. Schwentine has the highest conductivity of these freshwater sites (Table 1), and this may likely influence the presence of the cyanobacteria. Genus and species cyanobacterial probe signals are presented in Fig. 2b with highest values at two close brackish sites: Kieler Förde/Heikendorf, (BW3) and Eckernförder Bucht/Eckemförde (BW5) and at two freshwater samplings (Neukircher See, FW6 and Kellersee, FW9).

**Estuarine sites** The *Anabaena/Aphanizomenon* complex, *Cylindrospermopsis* and *Microcystis* signals were weak at these sites, even absent at BW1, BW3, BW5 and BW6. *Nodularia* genus level probes were well highlighted at different brackish water sites (Kieler Förde/Heikendorf, BW3; Schlei2/Massholmer Breite, BW4 and Eckernförder Bucht/Eckemförde (BW5), and the species *N. spumigena* was detected in both BW3 and BW5 samplings. *Planktothrix* signals were very weak at all sites, and the positive signal of the species *Planktothrix agardhii* at BW3 is likely a false positive because the genus level probe is not highlighted. *Cylindrospermopsis* signals were strongest at BW2 and BW4.



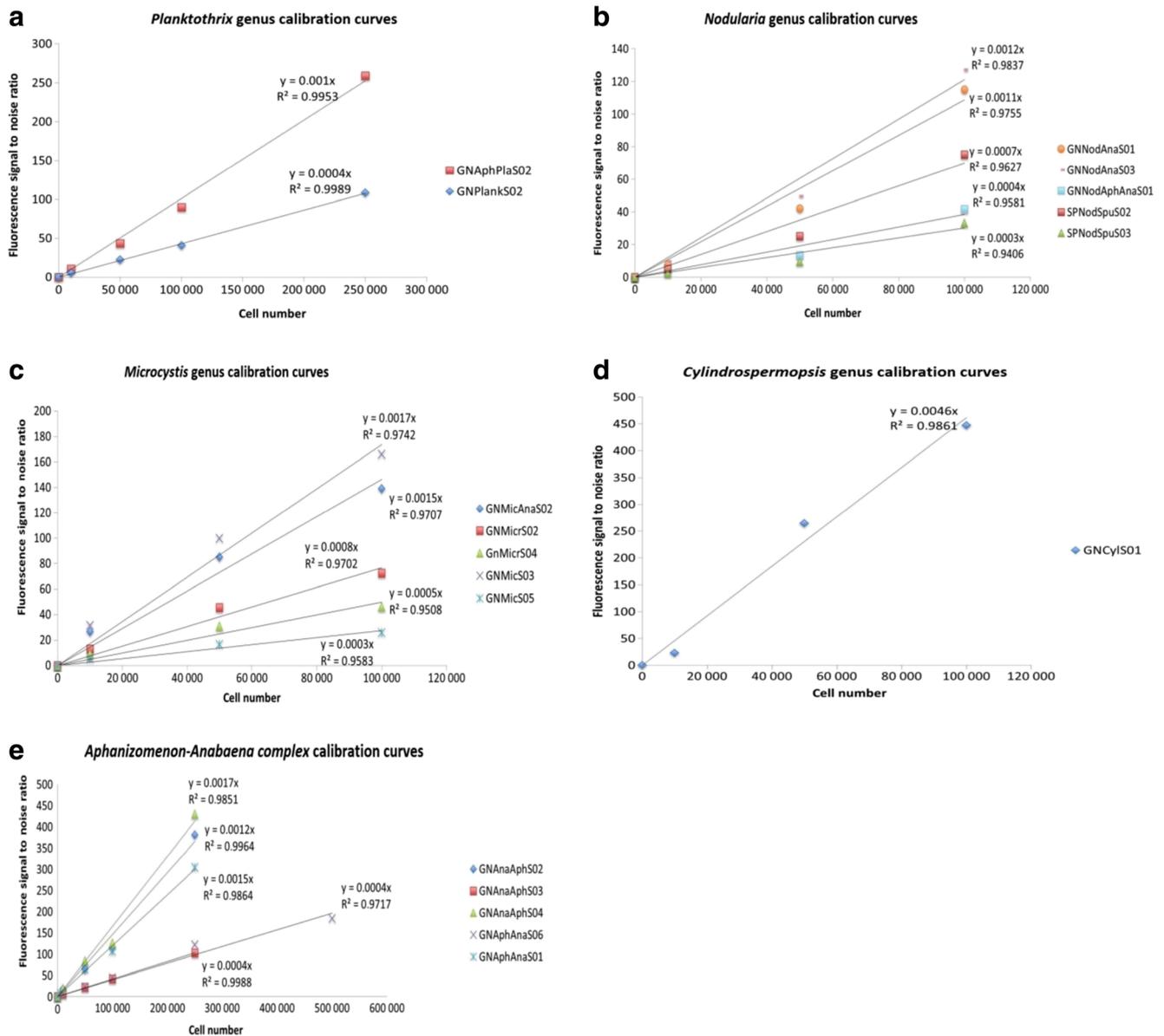
**Fig. 2** Hierarchical probes for taxonomic levels in *Cyanobacteria*. Normalized fluorescent signal measured for higher hierarchy level (a) and for genus and species of cyanobacteria (b) at both fresh and brackish water sites

**Freshwater sites** The *Anabaena/Aphanizomenon* complex is found in the freshwater sites of Neukircher See, FW6 and FW7, and Kellersee/Malente FW9 sites. Strong variation of the *Anabaena/Aphanizomenon* probes suggests that different genera and species were present. In this complex, the probes labelled with multiple generic names, the order of the names indicates the strength of the probe signal when presented with similar amounts of RNA. Thus, a stronger Aph/Ana probe indicates a stronger signal for *Aphanizomenon*, and a stronger Ana/Aph signal indicates a stronger signal for *Anabaena*. In general, *Aphanizomenon* presented higher probe signals and were notably weak or absent at freshwater samplings FW1–3 (see Table 1 and Fig. 1 for names, dates and locations). *Cylindrospermopsis* was detected in only one site (Seelter See/Fargau FW5), and *Nodularia* showed a high fluorescent signal especially at sites FW9, with no apparent signal for the *spumigena* species level probe. *Planktothrix* was absent or presented a low signal at all samplings. *Microcystis* was detected at all samplings except in Seelter See (Fargau) and Wittensee/Sande (FW5 and FW8). The fluorescent signals were especially strong at four samplings (FW1, FW2, FW3, FW6 and FW9) (Fig. 2b). These probes achieved the highest signal in the freshwater samplings/sites of all the cyanobacterial probes, especially at FW6 and 9, both of which were taken in August. The outflow from Neukirchener See (FW6) feeds directly into Kellersee (FW9) by canals and creeks. FW samplings 1, 2 and 3 are taken along the Schwentine River at two different sites with Rosensee (Schwentine) being sampled in August and September. The signals are slightly weaker at the downstream site (FW3) suggesting that the population is being concentrated or increasing as it nears the mouth of the river.

**Inferred cell numbers from microarray signal** Using the calibration curves constructed using RNA concentrations from known numbers of cells from pure cultures (Fig. 3), we inferred the most likely cell number for five genera and one species (Table 4). We had more than one probe for each of these taxa, except the species level probe for *Nodularia spumigena*. Thus, we averaged the cell numbers to obtain a relative number based on multiple probes. Numbers for *Microcystis*, *Planktothrix* and *Nodularia* ranged greatly among the different sites and across dates (Table 4). In most cases, the signal intensities for the various probes for the same taxon did not vary greatly, meaning that cell numbers inferred from the signal were more or less equal with the following exceptions. GNMicS04, a genus level probe for *Microcystis*, was usually two to three times higher than the three other genus level probes, and GNMicS02 was generally half the values of GNMicS01 and S03. GNNodS02 was an order of magnitude greater than GNNodS01, thus cell numbers were very different using the two probes.

Cell numbers of *Microcystis* were higher in the freshwater samples than the brackish water samples, and the highest cell numbers at FW6 likely contributed to the highest turbidity measured at this sampling time/site (Table 1). The highest cell numbers were inferred for *Microcystis* at three freshwater sites (FW2 on 29.08.13 sampling at Schwentine/Rosen See, FW5 on 27.08.13 at Seelter See/Fargau and FW9 on 13.08.13 at Kellersee 1).

The cell numbers inferred from genus level probe GNPlankS02, for *Planktothrix*, were an order of magnitude higher than the other genus level probe. Cell numbers for this genus were more or less the same in both the freshwater and the brackish water sites.



**Fig. 3** Calibration curves for Cyanobacteria. Regression of microarray signal to cell numbers. Each panel (a–e) represents a different species

The genus level probes for *Anabaena* and *Aphanizomenon* recognize both genera but to varying degrees of specificity. Those labelled AnaAmp preferentially recognize *Anabaena*, and those labelled AmpAna preferentially recognize *Aphanizomenon*, with two exceptions, GNPhAnaS01 = ANA major and GNPhAnaS06 = ANA major. There is one genus level probe that recognizes only *Aphanizomenon*, and it occasionally produced a higher signal than the others. GNPhAnaS01 was the weakest probe of all the probes of these genera. *Aphanizomenon* cell numbers were highest at three freshwater samplings (FW6, 7, 9). *Anabaena* was not particularly abundant at any of the samplings.

*Nodularia* had maximum abundances at two brackish sites and one freshwater water site, all collected within the same

2 weeks in August 2013 (Tables 1 and 2). *N. spumigenis* was more abundant in brackish water reaching its highest numbers at the same two sites as the genus level probes.

*Cylindrospermopsis* was present in low numbers in both habitats with slightly higher cell numbers at BW2 in June 2013 (Table 2).

### Pathogenic bacteria

A high eubacterial signal was obtained at both the freshwater and estuarine stations throughout the sampling period (Fig. 2a). Signal intensities at the bacterial phylum, and class levels followed the general trend seen at the domain level (Fig. 2a for eubacterial kingdom level probes and Fig. 4a for

**Table 4** Summary of inferred cell numbers from calibration curves generated from the regression of known amounts of RNA or cell numbers against microarray signal. Where multiple probes for the same taxon exist, cell numbers were averaged and the three highest cell concentrations are in bold and in red. The vertical separates the freshwater sites (left) from the brackish ones (right)

Probes	$R^2$	Cell number per 0.5 $\mu\text{g}$ RNA	Freshwater sites									Brackish sites					
			FW1	FW2	FW3	FW4	FW5	FW6	FW7	FW8	FW9	BW1	BW2	BW3	BW4	BW5	BW6
GNMierS01	1	0.000398	95006	147890	44410	14184	11446	192546	16416	3838	178679	6421	3427	2415	3232	1168	6351
GnMierS03	1	0.000333	49483	73758	12972	8181	5468	136724	9206	1810	99090	2957	8293	1120	4953	1728	10754
GNMicS02	1	0.000838	15434	19500	6513	2773	2648	42792	4089	1464	30647	1976	1736	875	1293	760	5534
GNMicS04	1	0.000158	239868	391641	126610	40654	31721	549264	43642	7587	478955	16711	22562	5569	11710	2880	15489
Average			99948	<b>158197</b>	47627	16448	12821	<b>230332</b>	18338	3675	<b>196843</b>	7016	9005	2495	5297	1634	9532
GNcylS01	1	0.004620	173	378	416	375	2381	350	408	256	463	255	6135	188	3241	139	890
GNPlankS02	1	0.000313	3258	7573	4836	3044	6535	3570	7477	5473	5734	3640	7689	2094	3248	2768	8284
GNaphPlaS01=PLKTX	1	0.000800	1433	1899	1873	901	2180	2000	2957	1648	2006	1751	1954	696	4350	566	2461
Average			2346	<b>4736</b>	3355	1972	4357	2785	<b>5217</b>	3561	3870	2695	4821	1395	3799	1667	<b>5373</b>
GNNodS01	0.93809300	0.002823	459	692	547	493	3806	999	1024	454	5615	605	599	1595	1680	1234	867
GNNodS02	0.94721651	0.000182	5965	9867	8574	5914	54436	38304	17010	5875	86087	9176	9424	58010	15968	56526	18453
SPNodSpuS02	0.94966041	0.000699	1461	2546	2444	1376	1589	1430	3702	1708	1696	2354	2648	5443	1746	5182	2141
Average			3212	5280	4560	3204	<b>29121</b>	19651	9017	3165	<b>45851</b>	4890	5011	<b>29802</b>	8824	<b>28880</b>	9660
GNaphS01	1	0.000390	3089	4853	4506	3889	5932	3679	6084	5175	5051	3723	4884	41902	42742	42232	9221
GNanaAphS01=ANA major	0.98557875	0.001875	1562	1421	891	500	1177	643	1329	1166	936	976	1084	320	650	292	1301
GNanaAphS02=ANA major	0.93309842	0.001464	433	340	421	423	604	1155	2630	1328	1080	314	818	4027	2046	4158	826
GNanaAphS03=ANA major	0.99744199	0.001867	1597	2054	969	610	633	743	1179	553	1920	973	1316	835	1066	622	1255
GNaphAnaS01=ANA major	0.91431614	0.002032	581	890	720	2236	5254	2050	2022	839	3843	798	7715	779	4356	923	1757
Average			1453	1911	1501	1532	2720	1654	2649	1812	2566	1357	3163	<b>9573</b>	<b>10172</b>	<b>9645</b>	2872
GNaphAnaS06=ANA major	0.99066886	0.000753	1922	2372	2087	1303	8105	1585	3369	1781	5289	2034	1994	1121	1770	839	2878
GNaphAnaS04=APHA major	0.98935478	0.000392	3935	3828	4719	1834	4292	3541	3886	4291	4856	6244	10349	968	4168	1461	13092
Average			2928	3100	3403	1568	<b>6198</b>	2563	3627	3036	<b>5073</b>	4139	<b>6172</b>	1044	2969	1150	7985

bacterial phylum and class probes). Gammaproteobacteria were detected with a high signal in all samples except for Wittensee/Sande (FW8), Keller See/Malente (FW9) and Eckernförder Bucht/Eckemförde (BW5). One probe of the Gammaproteobacteria (CLGamS13) did not seem to work in any sample, and likely this might have been a spotting error for this probe, hence the benefit of having multiple probes for the same taxon. Globally, *Firmicutes* probes show higher intensity in freshwater samples. *Actinobacteria* were widely detected in both freshwater and estuarine water samples, and signal intensities were globally higher for freshwater sites. Family level probes designed for *Enterobacteriaceae* and *Vibrionaceae* showed a widely distribution in both freshwater and estuarine sites with weaker signals as compared with the *Gammaproteobacteria* probes. Thus, at no site was a particular phylum or class on our array over represented as a higher percentage of the total bacterial signal, with a possible exception of *Actinobacteria* at the freshwater samplings and *Gammaproteobacteria* at all samplings.

**Estuarine sites** *Gammaproteobacteria* and *Actinobacteria* were detected in all estuarine sites. At the family level, *Enterobacteriaceae* were present at all of the water sites, with a complete hierarchy recorded (Fig. 4a) but globally signals

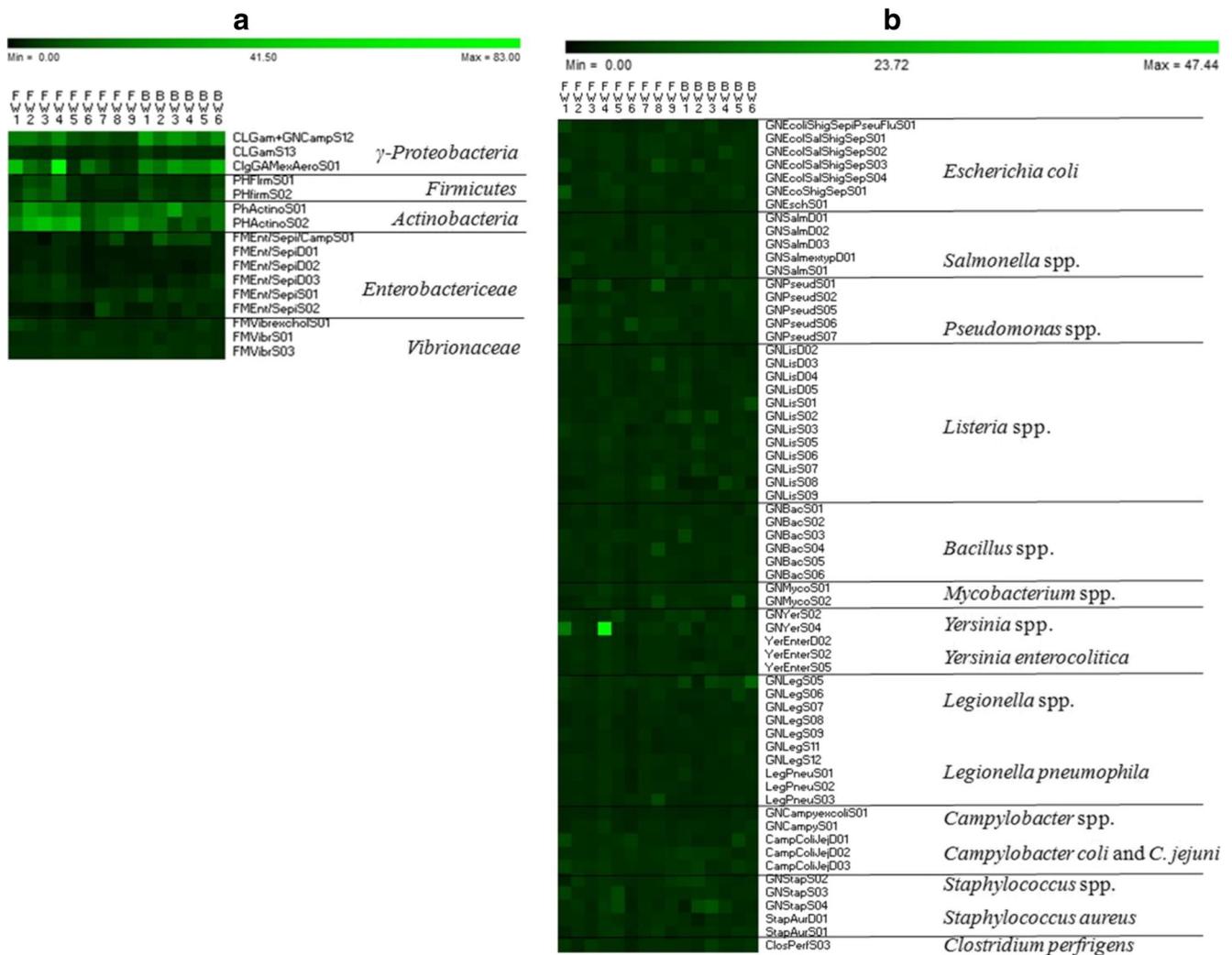
were still weak. All generic probes for *Enterobacteriaceae* family members (*E. coli*, *Salmonella* and *Yersinia* including *Y. enterocolitica*) were positive.

All of the other pathogens were detected at the estuarine sites over the sampling period with a complete probe hierarchy, and their presence could not be rejected because of an incomplete or broken hierarchy.

**Freshwater sites** All pathogens targeted with genera and species level probes in the study were present at all the freshwater sites over the sampling period and showed globally slightly higher signals than those obtained at the estuarine sites (Fig. 4b). Signal intensities for one *Yersinia* spp. probe were higher than the rest of the probes for two freshwater sites (FW1 and FW4), but very low signals for *Y. enterocolitica* were detected at these samplings sites.

### Protozoa

**Estuarine sites** *C. parvum* and *hominis* were present at all of the brackish water sites, with a complete hierarchy recorded (Figs 5). Lower levels of *Giardia*, *Entamoeba histolytica* and *N. fowleri* and *N. lovensis* were also present (Fig. 5b).



**Fig. 4** Heat map of the relative abundance of (a) the bacterial hierarchical probes from family to kingdom and (b) from genus to species. Vertical lines separate sampling dates

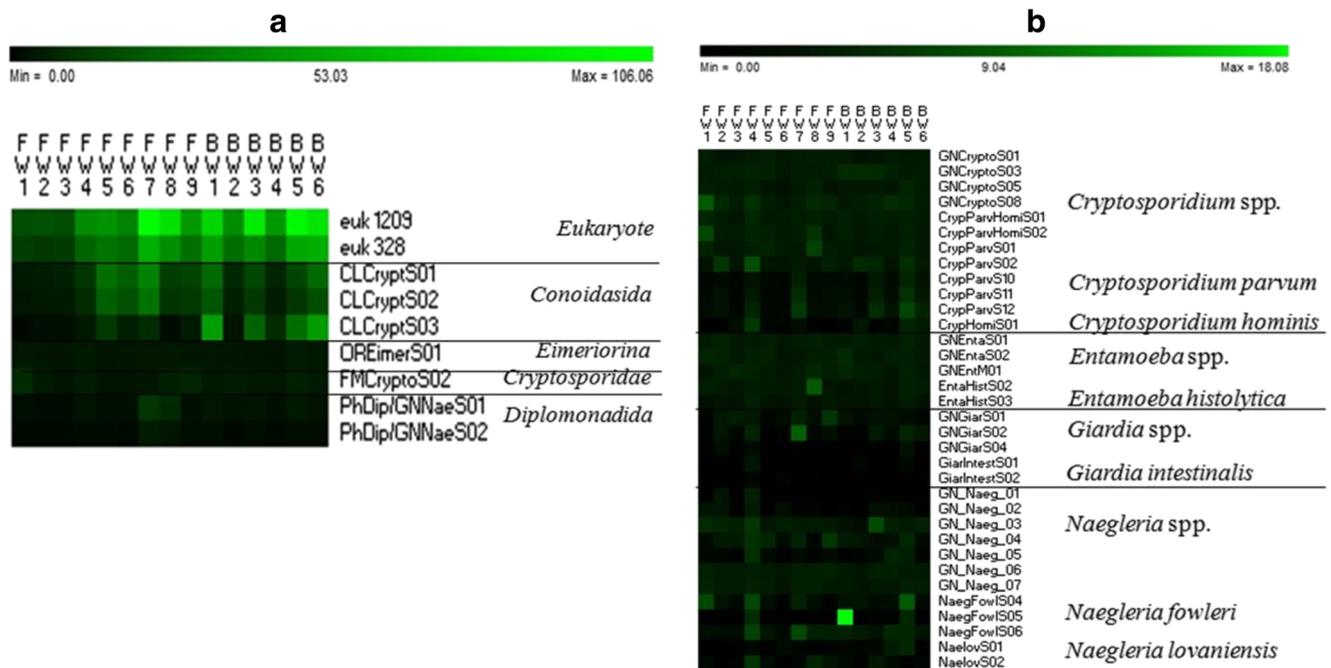
**Freshwater sites** *C. parvum* and *C. hominis* were also present at all of the freshwater sites throughout the sampling periods, with slightly higher signals than at the brackish water sites, especially for the class level probes, likely indicating the presence of other parasitic species/genera than we have probes for on our microarray (Fig. 5b). Lower levels of *E. histolytica*, *N. fowleri* and *Giardia* were also present (Fig. 5b).

## Discussion

### Differences among sites

At the freshwater sites, nine samples were analyzed during the period of the spring bloom (1× in March, April and June and during that of the summer bloom (4× in August, 1× in October) from 2012 to 2014. Three stations were positioned in rivers and five stations were in standing water. The

conductivity was between 292 and 466  $\mu\text{S}/\text{cm}^2$  and is in the expected range. Seven samples were taken at stations that were brackish (ranging from 9.5 psu at Schlei and Kappeln to 14.9 psu at Kieler Förde/Heikendorf and were regularly examined in June (2×), July (1×), August (3×) to September (1×) for 2 years (2013 and 2014), but one of these was not analyzed further because of poor rRNA extraction (Table 2). The salinity of the sample locations represents the salinity gradient typical of the German Baltic Coast. The low salinity at Schlei/Kappeln is caused by higher freshwater inflow and decreasing influence of salt water of Baltic Sea. The stations Schlei (Kappeln and Maasholmer Breite) have a low exposure to wind, Kieler Förde 1/Neumühlen-Dietrichsdorf, and Orther Bucht is moderately exposed, and Kieler Förde/Heikendorf and Fehmarnsund is heavily exposed, which resulted in frequent water column turnover. High nutrient values at the stations Schlei 1/Kappeln (BW2) and Kieler Förde 1/Neumühlen-Dietrichsdorf (BW1) are possibly caused by



**Fig. 5** Heat map of the relative abundance of (a) the bacterial hierarchical probes from family to kingdom and (b) from genus to species. Vertical lines separate sampling dates

runoff from agricultural areas close to these stations. Lakes in Northern Germany are generally considered eutrophic, although some improvement in nutrient status has been seen in recent years and are in general influenced by intense agricultural practices (Schernewski 2003).

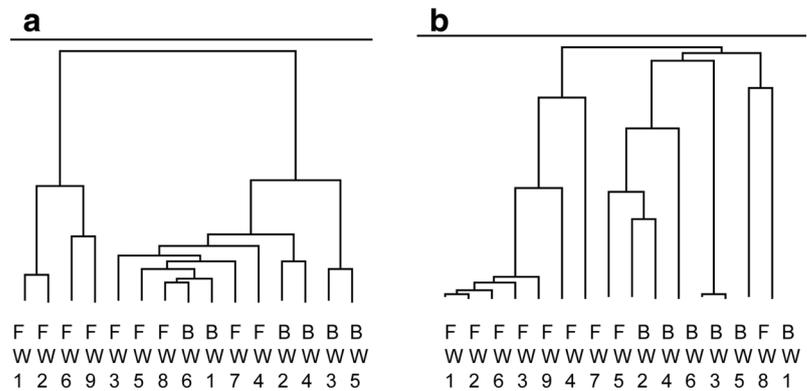
The distance between the samples of the total community as measured by the signal intensity, and the probes that reacted was analyzed using two different distance methods (Euclidian distance and Pearson's correlation) within the PermutMatrix program. The Euclidian distance can be used if the actual difference in values is important, which in the case of the microarray, a higher signal would represent more cells and more of a health concern. The Pearson's correlation would indicate trends in the data. Both distances were displayed using a complete linkage method in which the distance between two clusters is the distance between the furthest points in those clusters, and both methods show that there is a clear difference between some fresh and brackish water sites (Fig. 6) with the following exceptions: Using Euclidian distance (Fig. 6a), FW1 and FW2 (=temporal differences same site) and FW6 and FW9 (spatial and temporal similarity) were in a separate clade, and all remaining sites, remaining fresh and brackish water sites, were in a second clade. The four freshwater sites are grouped together primarily because of the large population of *Microcystis* that occurs at these sites, although FW7 which is a winter/spring sampling of FW6, which was sampled in the summer and more closely related to FW9, a nearby lake sampled also in August. The strong signal for *Nodularia* is responsible for the clustering of BW2 and BW4, which are very close geographically and BW3 and BW5, which have the highest salinity values

(Table 2). Using the Pearson's correlation (Fig. 6b), there is a stronger separation between the FW and BW sites with only two FW sites (FW5 and FW8) being correlated with the brackish water sites. The close relationship between FW8 and BW1 is likely related to the overall lower signals at these two sites, whereas the association of FW5 with BW2 and BW4 may be linked because they shared similar cyanobacterial populations. Overall, the cyanobacterial populations seemed to have had more effect on the overall community structure analyses performed here than did the bacterial or protozoan signals.

### Cyanobacteria

Cyanobacteria were present in all sites and seemed to show a preference for either the freshwater or the brackish water sites, with the exception of *Planktothrix*, which was common to both habitat types. This observation is in agreement with Lopes and Vasconcelos (2011) who showed the presence of this genus in brackish waters, although the fluorescent signals tend to show that *Planktothrix* was present or active at a low level. *Microcystis* was mainly found in freshwater sites, which is consistent with its natural habitat occupation, being not common in brackish water, whereas microcystins were found with the same two toxin detection methods, with the highest amount at Neukirchener See on (22.08.13), and this corresponded with the lowest signal of *Microcystis* at this location at this time, which would imply that this population of *Microcystis* was particularly toxic (Rodriguez et al. 2016). *Cylindrospermopsis* was not common at the freshwater sites but was found at the brackish water sites according to the high

**Fig. 6** Heat map showing the relative abundance of the protozoan hierarchical probes from the species level to kingdom. Vertical lines separate sampling dates



fluorescent signals. Several studies reported the ability of *Cylindrospermopsis* to proliferate in low salinity ecosystem similar to our brackish sites for which salinity is very low ranging between 9.5 and 14.9 psu, especially if there are elevated nutrient conditions (Chapman and Schelske 1997; Padišák 1997, Briand et al. 2004; Antunes et al. 2015). *Anabaena/Aphanizomenon* were preferentially found at the freshwater sites. Our set of probes does not allow a good distinction between *Aphanizomenon* and *Anabaena* whose generic classification was recently revised (Rajaniemi et al. 2005), redistributing many *Aphanizomenon* and *Anabaena* species, emphasizing the difficulty to design specific sequences targeting either one or the other genus. The definition of these two genera needs a better refining to be able to design an efficient genetic barcode to identify these cyanobacteria genera (Rajaniemi et al. 2005). In a separate study of the toxins as detected by two different methods (Rodriguez et al. 2016), anatoxins were found in all but one site (FW8, Wittensee/Sande, 12.06.13) with the microsphere-based assay and all but two sites (FW4, Selenter See/Fargau (21.08.12) and BW4, Schlei/Maasholmer Breite, 15.08.13) with the UPLC-MS/MS method. The toxin measurements would imply that the strains present were toxic. Microcystins were found with the same two toxin detection methods, with the highest amount at FW6, Neukirchener See on (22.08.13), and this corresponded with the highest cell numbers of *Microcystis* (average 230,332 cells/l) at this location at this time. Such findings indicate the importance of employing both a cell identification and toxin identification method for cyanobacteria to have an effective monitoring program. *Nodularia*'s presence in brackish sites was in accordance with previous monitoring of this cyanobacteria occurrence in Baltic Sea. The cell numbers inferred from the microarray signals were more or less the same for multiple probes. At present, we do not have inferred cell numbers validated by cell counts. Thus, for those probes whose signals are markedly different for multiple probes for the same taxon, we do not know which of the multiple probes can be validly used to infer cell numbers. Such validation will be obtained in other EU projects (MICROCOKIT), in which microarray signals are being

validated by FISH cell counts for cyanobacteria. Without such validation, we can only average the signals for the multiple probes to generate an average cell number. If we find that after validation, the probes that produce lower signals infer too low a cell number, then we will eliminate that probe from the microarray for commercialization. More extended and detailed investigations are necessary to find out what are the main effects on growth of different cyanobacterial species at the sampling sites listed here.

### Bacteria

For bacteria, we detected each target with at least one or several probes, which were correctly detected with its higher hierarchy level. All of the targeted bacteria species and genera were present in the freshwater and estuarine water samples with a complete hierarchy. The signals measured for bacteria were weaker as compared to those measured with the cyanobacterial set of probes. This can be explained by a lower abundance of pathogenic bacteria in these water samples because most of pathogenic bacteria investigated in this study have a human-associated waste source, such as urban or domestic wastewater inputs that are highly diluted into receiving waters and thus would present a lower signal. Human-associated sewage is a well-known source of pathogenic bacteria, but these facilities were not in the vicinity of our samplings. Molecular-based methods have shown that the members of the members *Enterobacteriaceae* family (*E. coli*, *Yersinia* spp., *Salmonella* spp.), *C. perfringens* but also the members of the Family *Vibrionaceae*, *Pseudomonas*, *Campylobacter*, *Listeria*, *Staphylococcus* and *Mycobacterium* are common pathogenic bacteria found in urban-domestic sewage waters (Campbell et al. 2015; Wang et al. 2014). Nevertheless, in urban sewage, their relative abundance in the total bacterial community still low (<1 %), and their abundances can decrease depending on the stages of wastewater treatment (Shannon et al. 2007). By consequence, the abundances in receiving sewage areas of the pathogenic bacteria are very low and their concentrations in environment seem to be close to the detection limit of the microarray

method. Similar results were obtained from an investigation performed in Bulgarian sites samples, where large volumes of water samples (50 l) were analyzed by using the same ultra-filtration system and by PCR-based methods (Panaiotov et al. 2015). Our results obtained with the microarray confirm the very low concentration of pathogenic bacteria in surface freshwaters and estuarine waters sampled here. In contrast, the high nutrient load likely present in the water from intense agricultural practices in the area could contribute to the high cyanobacterial signals we recovered, some of which were from toxic strains (Rodriguez et al. 2016).

The results obtained by the microarray over 2 years showed the presence of a very high diversity of pathogenic bacteria detected but with a weak signal in the freshwater sites but also at the estuarine water sites. The salinity of the estuarine sites were low (ranging from 9.5 to 14.9 psu), which highlights that these brackish sites could possible receive freshwater input from rivers including several wastewater treatment plants and discharges from rivers or agricultural runoff. The continental inputs could explain the presence of bacterial pathogens at the brackish sites. A storm event can contribute to street runoff or overflow from wastewater treatment plant, increase significantly the counts of bacteria associated to human sewage in receiving water (Berndtsson and Paul 2015). Nevertheless, during the investigated period, no rain events or storm events were noted.

### Protozoa

All of the targeted protozoan species were present in the samples, with slightly higher signals in the estuarine sites than in the freshwater sites. *Cryptosporidium* had the highest signals among all of the protozoans present on the array. The extent to which parasites from these groups live in saline conditions is unknown except that it is known that *Cryptosporidium* and *Naegleria* will survive exposure to saline conditions (Bradford and Schijven 2002; Zbikowska et al. 2013). It is also unknown if the cells targeted by our probes are parasitic to humans because the targets highlighted in these studies could be parasites of fish, livestock, etc.

Our microarray has successfully detected a wide range of pathogens in fresh and brackish water in Northern Germany. Microarrays have been successfully used to estimate the diversity and abundance from a variety to organisms. Our microarrays used RNA, whereas nearly all others have a PCR step and hybridize the amplicons to the microarray. The problems of using PCR to obtain products for the hybridization is that it tends to favour the most abundant species present so that the results are not truly quantitative and there is always the possibility of having natural inhibitors to the PCR (Maynard et al. 2005; DeAngelis et al. 2011). Gushin et al. (1997) suggested, from their study of nitrifying bacteria, that combining RNA and PCR-

amplified DNA, each labelled with different dyes, could provide an excellent assessment of the active components and total diversity of any community, respectively. Amplified products of freshwater pathogens, many of those the same as on our microarray, have been detected using a phylochip using both 16S rRNA and/or protein coding genes (Wilson et al. 2002; Wang et al. 2004; Maynard et al. 2005; Lee et al. 2006, 2008, 2010), and in the Lee studies, array results from PCR products compared favourably with qPCR results. Brinkmann et al. (2013) designed a similar microarray to ours targeting most of the same protozoa and several bacteria. Their microarray successfully used extracted DNA with no PCR amplification instead of RNA but only it was only tested with lab-based cultures and no field testing.

Now that the microarray method being recognized as a valid method for detecting nucleic acids in environmental samples, microarrays, especially those based on RNA detection, such as the  $\mu$ AQUA microarray, can provide a more reliable method of detecting actively growing pathogens in the environment prior to becoming a health issue. The uniqueness of our microarray is the diversity of biological targets belonging to four widely different target groups ranging from prokaryotes to eukaryotes, which includes re-emerging and emerging pathogens that are the target of much active water quality research. Another innovative aspect is our capability of detecting actively growing cells because hybridization with RNA is representative only of active and not dead cells and thus gives an indication about the risk potential of the detected and identified cells. This makes our microarray the obvious tool for an early warning system. The results presented in the paper prove that both pathogenic eukaryotes and prokaryotes even at a low level can be detected and identified and that the use of RNA as the target molecule gives significant signals and useful information about the cell's state and the actively growing community.

Their presence as detected by the microarray is not affected by PCR or qPCR inhibitors, which could yield false-negative results, and because known targets are on the array, turn-around time for results is hours instead of months as is the case with next generation sequencing (NGS). NGS yields the total microbial community composition rather than just pathogenic targets and is independent of culture requirements for selected pathogens. Our microarray is the only array available that not only uses RNA detection with no PCR amplification but also combines all of the major pathogenic species into one chip using SSU and LSU genes for pro- and eukaryotic organisms at the same time. Our results also underscore the importance of employing both a cell identification and toxin identification method for cyanobacteria to have an effective monitoring program and the need for more frequent monitoring as afforded by molecular techniques to serve as early warning systems for harmful events.

**Acknowledgments** This work was supported by the EU  $\mu$ AQUA project (FP7-KBBE-2010-4, 265409).

**Compliance with ethical standards** LKM declares on behalf of all of the authors that there are no conflicts of interest, no research on humans or animals and no informed consents in this study.

## References

- Antunes JT, Leão PN, Vasconcelos VM (2015) *Cylindrospermopsis raciborskii*: review of the distribution, phylogeography, and eco-physiology of a global invasive species. *Front Microbiol* 6:473. doi:10.3389/fmicb.2015.00473
- Bradford SA, Schijven J (2002) Release of *Cryptosporidium* and *Giardia* from dairy calf manure: impact of solution salinity. *Env Sci Tech* 36:3916–3923
- Briand JF, Leboulanger C, Humbert JF, Bernard C, Dufour P (2004) *Cylindrospermopsis raciborskii* (Cyanobacteria) invasion at mid-latitudes: selection, wide physiological tolerance, or global warming? *J Phycol* 40:231–238. doi:10.1111/j.1529-8817.2004.03118.x
- Brinkmann NE, Francisco R, Nichols TL, Robinson D, Scaeffere FW II, Schaudies RP, Villega EN (2013) Detection of multiple waterborne pathogens using microsequencing arrays. *J App Microbiol* 114:564–573
- Campbell AM, Fleisher J, Sinigalliano C, White JR, Lopez JV (2015) Dynamics of marine bacterial community diversity of the coastal waters of the reefs, inlets, and wastewater outfalls of southeast Florida. *Microbiol* 4:390–408
- Caraux G, Pinloche S (2005) PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 21:1280–1281
- Chapman AD, Schelske CL (1997) Recent appearance of *Cylindrospermopsis* (Cyanobacteria) in five hypereutrophic Florida lakes. *J Phycol* 33:191–195. doi:10.1111/j.0022-3646.1997.00191.x
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156–159
- DeAngelis KM, Wu CH, Beller HR, Brodie EL, Chakraborty R, DeSantis TZ, Fortney JL, Hazen TC, Osman SR, Singer ME, Tom LM, Anderson GL (2011) PCR amplification-independent methods for detection of microbial communities by the high-density microarray phylochip. *Appl Environ Microbiol* 77:6313–6322
- DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* 53:371–383
- Dittami SM, Edvardsen B (2013) GPR-analyzer: a simple tool for quantitative analysis of hierarchical multispecies microarrays. *Environ Sci Pollut Res Int* 20:6808–6815
- Gushin DY, Mobarry BK, Proudnikov D, Stahl DA, Rittmann BE, Mizrabekov AD (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. *Appl Environ Microbiol* 63:2397–2402
- Hill VR, Polaczyk AL, Hahn D, Narayanan J, Cromeans TL, Roberts JM, Amburgey JE (2005) Development of a rapid method for simultaneous recovery of diverse microbes in drinking water by ultrafiltration with sodium polyphosphate and surfactants. *Appl Environ Microbiol* 71:6878–6884
- Kegel JU, Del Amo Y, Medlin LK (2013) Introduction to project MIDTAL: its methods and samples from Arcachon Bay, France. *Env Sci Poll Res* 20:6690. doi:10.1007/s11356-012-1299-9
- Kegel JU, Guillebault D, Medlin LK (2016) Application of microarrays (phylochips) for analysis of community diversity by species identification. *Persp Phycol* 3:93–106.
- Lee DY, Shannon K, Beaudette LA (2006) Detection of bacterial pathogens in municipal wastewater using an oligonucleotide microarray and real-time quantitative PCR. *J Microbiol Meth* 65:453–467
- Lee DY, Lauder H, Cruwys H, Falletta P, Beaudette LA (2008) Development and application of an oligonucleotide microarray and real-time quantitative PCR for detection of wastewater bacterial pathogens. *Sci Total Environ* 398:203–211
- Lee DY, Seto P, Korczak R (2010) DNA microarray-based detection and identification of winterbourne protozoan pathogens. *J Microbiol Meth* 80:129–133
- Lewis J, Medlin LK, Raine R (2012) MIDTAL (microarrays for the detection of toxic algae): a protocol for a successful microarray hybridisation and analysis. Koeltz, Koenigstein
- Lopes VR, Vasconcelos VM (2011) Planktonic and benthic cyanobacteria of European brackish waters: a perspective on estuaries and brackish seas. *Eur J Phycol* 46:292–304
- Marcheggiani S, D'Ugo E, Puccinelli C, Giuseppetti R, D'Angelo AM, Gualerzi CO, Spurio R, Medlin LK, Guillebault D, Baudart-Lenfant J, Weigel W, Helmi K, Mancini L (2015) Detection of emerging and re-emerging pathogens in surface waters close to an urban area. *Int J Environ Res Public Health* 12:5505–5527
- Maynard C, Berthiaume F, Lemarchand K, Harel J, Payment P, Bayardelle P, Masson L (2005) Waterborne pathogen detection by use of oligonucleotide-based microarrays. *Appl Environ Microbiol* 71:8548–8557
- Medlin LK (2013) Note: steps taken to optimise probe specificity and signal intensity prior to field validation of the MIDTAL microarray for the detection of toxic algae. *Env Sci Poll Res* 20:6683. doi:10.1007/s11356-012-1195-3
- Padisák J (1997) *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. *Arch Hydrobiol Suppl Monogr Stud* 107:563–593
- Panaiotov S, Simeonovski I, Levterova V, Karamfilov V, Brankova N, Tankova K, Campbell K, Jacob P, Helmi K, Boots B, D'Ugo E, Marcheggiani S, Mancini L, Breitenbach U, Mielke E, Kantardjiev T (2015) Two-year monitoring of water samples from Dam of Iskar and the Black Sea, Bulgaria, by molecular analysis: focus on *Mycobacterium* spp. *Int J Environ Res Public Health* 12:7430–7443
- Rajaniemi P, Hrouzek P, Kastovska K, Raphae KK, Willame RI, Rantala A, Hoffmann L, Komarek J, Sivonen K (2005) Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria). *Int J Sys Evol Microbiol* 55:1–26
- Ramírez-Castillo FY, Loera-Muro A, Jacques M, Garneau P, Avelar-González FJ, Harel J, Guerrero-Barrera AL (2015) Waterborne pathogens: detection methods and challenges. *Pathogens* 4:307–334. doi:10.3390/4020307
- Rodríguez I, Fraga M, Alfonso A, Guillebault D, Medlin LK, Baudart J, Jacob P, Helmi K, Meyer M, Breitenbach U, Holten NM, Boots B, Supiro R, Cimarelli L, Mancini L, Marcheggiani S, Albay M, Akcaalan R, Köker L, Botana LM (2016) Monitoring of freshwater toxins in European environmental waters by using novel multi-detection methods. *Water Res*, in press
- Schernewski G (2003) Nutrient budgets, dynamics and storm effects in a eutrophic, stratified Baltic Lake. *Acta Hydrochim Hydrobiol* 31:1–10
- Shannon KE, Lee DY, Trevors JT, Beaudette LA (2007) Application of real-time quantitative PCR for the detection of selected bacterial pathogens during municipal wastewater treatment. *Sci Total Environ* 382:121–129
- Utermöhl H (1958) Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. *Mitt Int Ver Theor Angew Limnol* 9:1–38

- Wang Z, Vora GJ, Stenger DA (2004) Detection and genotyping of *Entamoeba histolytica*, *Entamoeba dispar*, *Giardia lamblia*, and *Cryptosporidium parvum* by oligonucleotide microarray. *J Clin Microbiol* 42:3262–3271
- Wang ZH, Yang JQ, Zhang DJ, Zhou J, Zhang CD, Su XR, Li TW (2014) Composition and structure of microbial communities associated with different domestic sewage outfalls. *Genet Mol Res* 13:7542–7552
- Wilson WJ, Strout CL, DeSantis TZ, Stilwell JL, Carrano AV, Andersen GL (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes* 16:119–127
- Zbikowska E, Walczak M, Krawiec A (2013) Distribution of *Legionella pneumophila* bacteria and *Naegleria* and *Hartmannella amoebae* in thermal saline baths used in balneotherapy. *Paras Res* 112: 77–83