



Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India

Johan Bengtsson-Palme¹, Fredrik Boulund², Jerker Fick³, Erik Kristiansson² and D. G. Joakim Larsson^{1*}

¹ Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

² Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

³ Department of Chemistry, Umeå University, Umeå, Sweden

Edited by:

Charles W. Knapp, University of Strathclyde, UK

Reviewed by:

Shaolin Wang, University of Virginia, USA

Andrew C. Singer, Natural Environment Research Council Centre for Ecology & Hydrology, UK
Justin Joseph Donato, University of St. Thomas, USA

*Correspondence:

D. G. Joakim Larsson, Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Guldhedsgatan 10, Göteborg SE-413 46, Sweden
e-mail: joakim.larsson@fysiologi.gu.se

There is increasing evidence for an environmental origin of many antibiotic resistance genes. Consequently, it is important to identify environments of particular risk for selecting and maintaining such resistance factors. In this study, we described the diversity of antibiotic resistance genes in an Indian lake subjected to industrial pollution with fluoroquinolone antibiotics. We also assessed the genetic context of the identified resistance genes, to try to predict their genetic transferability. The lake harbored a wide range of resistance genes (81 identified gene types) against essentially every major class of antibiotics, as well as genes responsible for mobilization of genetic material. Resistance genes were estimated to be 7000 times more abundant than in a Swedish lake included for comparison, where only eight resistance genes were found. The *sul2* and *qnrD* genes were the most common resistance genes in the Indian lake. Twenty-six known and 21 putative novel plasmids were recovered in the Indian lake metagenome, which, together with the genes found, indicate a large potential for horizontal gene transfer through conjugation. Interestingly, the microbial community of the lake still included a wide range of taxa, suggesting that, across most phyla, bacteria has adapted relatively well to this highly polluted environment. Based on the wide range and high abundance of known resistance factors we have detected, it is plausible that yet unrecognized resistance genes are also present in the lake. Thus, we conclude that environments polluted with waste from antibiotic manufacturing could be important reservoirs for mobile antibiotic resistance genes.

Keywords: antibiotic resistance, horizontal gene transfer, lake sediment, metagenomics, mobile genetic elements, pharmaceutical pollution, plasmids

INTRODUCTION

Antibiotics constitute our core line of defense against infectious diseases caused by bacteria and are also fundamental for our ability to perform advanced surgery and treat cancer. However, in later years, an antibiotic resistance crisis has emerged, causing increased mortalities and substantial costs; expenses have recently been estimated to be over 1.5 billion euros every year in Europe alone (Norrby et al., 2009). There is increasing evidence that the resistance we see in pathogens did not initially appear in the clinical setting, but that environmental bacteria have contributed to the resistance gene pool shared among pathogens today (Martinez, 2008; Forsberg et al., 2012; Gaze et al., 2013; Wellington et al., 2013). Resistance genes are commonly encountered on, or associated with, mobile genetic elements (Stokes and Gillings, 2011; Heuer and Smalla, 2012) such as plasmids (Schlüter et al., 2007; Szczepanowski et al., 2008), integrons (Gaze et al., 2011), and transposons (Toleman and Walsh, 2011; Nigro and Hall, 2012). This enables their transfer within and between bacterial cells and species, and their genetic context

should therefore be considered a major factor when assessing the risk of transfer from a source environment, and onwards into clinically relevant bacteria. Even if much evidence points toward an ancient origin for most—if not all—mobile resistance determinants (D'Costa et al., 2011), the picture of the environments in which they are initially selected for, as well as their dissemination routes from there into human pathogens, is much less clear. It is reasonable to assume that such transfer of genes from the environment will occur in the future, and that we can expect pathogens to pick up additional resistance determinants from the environmental resistome. To mitigate this problem, it is important to identify source environments where resistance genes can be selected for, and subsequently transferred into human or animal pathogens. Likely, environments under sufficient selection pressure to maintain a range of known antibiotic resistance genes in relatively large quantities also harbor a fair number of yet undescribed resistance factors, and thereby may serve as a recruitment ground for novel resistance determinants. Environments polluted by direct discharges from manufacturing of antibiotics appear to

provide such conditions (Larsson et al., 2007; Li et al., 2008a,b; Larsson, 2014).

We have for some years studied the Patancheru area in India, housing a range of pharmaceutical industries. Discharges, both from a common effluent treatment plant and illegal dumping of production waste (Greenpeace, 2004; Boralkar et al., 2005; Larsson et al., 2007), has led to unprecedented antibiotic contamination of both surface and ground water (Fick et al., 2009) as well as river sediment (Kristiansson et al., 2011). In this work, we have investigated another environment where resistance genes may flourish; Kazipally lake, located in the same area, and polluted by direct dumping of waste from pharmaceutical production. Fick et al. (2009) reported several pharmaceuticals in the surface water of Kazipally lake, with concentrations of norfloxacin of up to 520,000 ng/L and enoxacin of up to 84,000 ng/L, approximately corresponding to the clinical breakpoint for fluoroquinolone resistance in enterobacteriaceae (European Society of Clinical Microbiology and Infectious Diseases, 2014). To capture a broad diversity of resistance genes, we have here used shotgun sequencing metagenomics to investigate the bacterial community of the lake. In contrast to PCR, this approach allow us to search for in principle any known resistance gene in the same analysis, given that it is present in sufficient abundance to be detected using the sequencing effort of the study. Shotgun metagenomics also offers the opportunity to study the taxonomic and functional content of the bacterial communities in a comprehensive manner. We have previously studied river sediments from the Patancheru area using shotgun pyrosequencing, revealing a high abundance of certain antibiotic resistance genes (Kristiansson et al., 2011). It is clear, however, that the amount of data obtained by this sequencing technique only allows us to scratch the surface of the environmental resistome. To be able to more comprehensively describe the resistome of the polluted lake, we have therefore employed Illumina sequencing, as it delivers considerably more DNA data than does pyrosequencing, given the same sequencing costs.

Using the metagenomic sequencing data we have described the diversity of resistance genes in the highly polluted Kazipally lake, to better understand how the selection pressures here, and possibly in environments similar to this lake, could shape the local resistome. Furthermore, we aimed to define the genetic context of the identified resistance genes, to evaluate their potential to be transferred between bacteria. Several studies have suggested that stressors such as antibiotics could contribute to a general mobilization of genetic material in bacterial communities (Beaber et al., 2004; Prudhomme et al., 2006), and we have therefore specifically looked for genes involved in genetic mobility, and have also identified known and novel plasmids. In addition, strong environmental disturbances have been proposed to reduce the taxonomic diversity of ecological communities (Banks et al., 2013). We have thus explored the diversity of microbial taxa in the lake to understand if such a shift in community structure is likely to have taken place due to the considerable releases of chemicals into the lake.

MATERIALS AND METHODS

SAMPLING AND DNA SEQUENCING

Sediment samples were collected from the Kazipally lake in the Patancheru region, close to Hyderabad in India (N 17°34.45,

E 078°21.38). We did not have access to sediment from any nearby, non-polluted lake that could serve as a local reference. Although there are many environmental metagenomes available for comparisons (see discussion), we collected sediment samples from the Nydalasjön lake, close to Umeå, Sweden (N 63°49.30, E 020°20.85) and analyzed it in a directly comparable way. The latter is a healthy lake with viable populations of fish, insects and other fauna. It is used for recreational swimming activities in the summer, and it does not receive any direct discharges of sewage or industrial waste. Six samples were collected at the Indian site, and 20 samples at the Swedish site. Samples from within the same lake were pooled in order to minimize spatial variability. For each sample, the top 5 cm of sediment were collected in test tubes. Indian samples were refrigerated for approximately 1 week before shipping to Sweden where they were frozen at -80°C , whereas the Swedish samples were frozen immediately. Although an extended time before freezing might have influenced the community structure to some extent, there are no apparent reasons to believe that this would select for bacteria harboring antibiotic resistance genes in any consistent way. DNA was extracted from the samples using the PowerSoil DNA isolation kit (MO BIO, Carlsbad, CA) according to the manufacturer's instructions. Extracting sufficient amounts of high quality DNA from the polluted Indian sediment turned out to be a challenge, likely due to chemicals in the sediment material interfering with the DNA extraction process, in combination with low DNA concentrations. We solved this by amplifying the extracted DNA using the Repli-G mini kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. The same treatment was also applied to the sediment samples from the Swedish lake, to enable better comparisons. It should be acknowledged that Repli-G as well as other amplification procedures might not amplify all sequences to the same extent (Pinard et al., 2006), and thus quantitative estimates should be interpreted with some caution. As the samples from both lakes were amplified using Repli-G, potential biases should affect the DNA libraries in the same way. Paired-end sequencing libraries (2×100 bp) were prepared using the TrueSeq DNA Kit for multiplexing. Sequencing was done at Science for Life Laboratories (Stockholm, Sweden) using the Illumina HiSeq2000 platform. The sequence data have been submitted to the European Nucleotide Archive under project accession number PRJEB6102 (<http://www.ebi.ac.uk/ena/data/view/PRJEB6102>).

SEQUENCE ANALYSIS

The quality of each sequencing library was assessed using FastQC (Andrews, 2010), and reads with lower quality scores than 28 over more than 75% of the read were discarded using a paired-end aware read filterer (Pearf, part of the PETKit; Bengtsson-Palme, 2012). In addition, when more than 5% of the bases in a read fell below a quality score of 28, the read was trimmed at the first low-quality base. Reads trimmed to less than 30 bp were discarded. If one read in a read-pair was discarded, the paired read was discarded as well (Pearf options “-q 28 -f 0.25 -t 0.05 -l 30”).

Quality filtered reads were mapped against the Resqu antibiotic resistance database (version 1.1; <http://www.1928diagnostics.com/resdb>). The database contains 3019 non-redundant protein sequences corresponding to 325 resistance gene families (see Table S1 for the complete list). All data in the database have been

manually extracted from the literature. The mapping was done using Vmatch (Kurtz, 2010), allowing one mismatched amino acid per translated read (options “-showdesc 60 -dnvprot 11 -l 20 -h 1”). The number of matches to each resistance gene were counted, and subsequently normalized to the length of the respective gene. The length-normalized numbers were then further normalized to the number of bacterial 16S rRNA sequences divided by the length of the 16S gene, yielding an approximation of the number of genes per 16S rRNA sequence for each resistance gene, while still avoiding bias due to differential length of the resistance genes.

Read pairs were scanned with Metaxa 2.0 (Bengtsson et al., 2011; Bengtsson-Palme et al., 2014) to extract bacterial 16S rRNA (SSU) sequences (default options). Read pairs identified as bacterial SSUs by Metaxa were assigned to and grafted on their closest matching bacterial rRNA sequences in SILVA (release 108) using Megraft (version 1.0.2; Bengtsson et al., 2012), clustered in Usearch (Edgar, 2010; options “-usersort -id 0.97”), and further subjected to rarefaction analysis in the R package Vegan (Oksanen et al., 2011).

METAGENOMIC ASSEMBLY

Assembly of the quality filtered reads was done using Velvet (Zerbino and Birney, 2008) on the C3SE computer cluster at Chalmers University of Technology. Non-circular contigs were further post-processed in Peacat (part of the PETKit; Bengtsson-Palme, 2012; options: “-overlap 25 -n”) to further merge contigs with overlaps of at least 25 bp (Table S2). Reads were mapped back to contigs using BLAT (Kent, 2002; options “-out=blast8 -q=dna -t=dna -minIdentity=95”). At least 50 matching bp were required for a read to be counted as mapped. Only the best match of each mapped read was used for determination of feature abundance. Annotation was performed using BLAST (Altschul et al., 1997) against plasmids and viruses in GenBank (Benson et al., 2014; blastall options “-p blastn -m 8 -a 16 -e 1e-4 -F F”). Annotation against the NCBI non-redundant protein database was performed using BLAT (options “-out=blast8 -q=prot -t=dnax”). Translated contigs were searched against the Pfam (Finn et al., 2014) and TIGRFAM (Haft et al., 2013) databases and HMM profiles for type 4 secretion systems (Smillie et al., 2010; Guglielmini et al., 2011) using HMMER (Eddy, 2010; hmmsearch options “-noali -cut_ga”). Contigs were annotated for resistance genes using an in-house pipeline for resistance gene and mobile element finding. All annotations were then added to a database containing all annotated features, which was subsequently queried at different, more specific, cutoffs depending on what data that was requested.

Additional assembly of resistance gene contigs were carried out by a custom iterative approach employing Vmatch (Kurtz, 2010) and Trinity (Grabherr et al., 2011). In this approach, reads mapped to a resistance gene were extracted and used as seeds for a Vmatch search against the complete set of read pairs matching with at least 49 bp to any of the seed reads. These reads were then assembled using Trinity. The resulting contigs were then used as seeds for another search using Vmatch to the complete set of quality-filtered reads, as above. All matches (including the previously matching read pairs) were then used for another

Trinity run. This iterative process was then repeated until the total number of assembled nucleotides started to drop rather than increase (40 iterations in this case). The script used has been made available as a software package (Bengtsson-Palme, 2013). The complete set of options used were “-cpu 8 -overlap 49 -contig_min 150.”

ABUNDANCE ANALYSIS

Abundances of Pfam and TIGRFAM families were estimated by calculating the mean number of reads that could be mapped to each specific region matching a Pfam or TIGRFAM family, such that at least 75 bp of the read was mapped to that genetic region, and then subsequently calculating the sum of all those mean values. This number was then divided by the number of reads that passed quality filtering from that sequencing library, yielding a relative abundance estimate. All abundances were then multiplied by 1 million, and can therefore be interpreted as the number of reads mapped to that feature in every million reads generated (Tables S3, S4). Gene Ontology (GO; Ashburner et al., 2000) terms were mapped to the Pfam families using the Pfam2GO mapping supplied by the GO website, and to TIGRFAM families using the mapping supplied by TIGRFAM. GO terms were then mapped back to the third level in the GO hierarchy to generate comparable data for differently annotated biological processes and molecular functions (Tables S5–S9).

IDENTIFICATION OF PLASMIDS

Known plasmids were identified by mapping the quality-filtered reads from the Indian and Swedish lakes to the NCBI plasmid genomes dataset, using BLAT (options “-out=blast8 -q=dna -t=dna -minIdentity=95”). Plasmids with less than 90% coverage were discarded (and thus not counted). Novel plasmids were identified by identifying contigs longer than 3000 bp, with a circular overlap of at least 30 bp, which also had read pairs that mapped to both ends of the contig. Additionally, we also demanded that each circular contig should contain a relaxase or a conjugation system to be classified as a plasmid, to exclude viruses and other non-plasmid circular DNA elements.

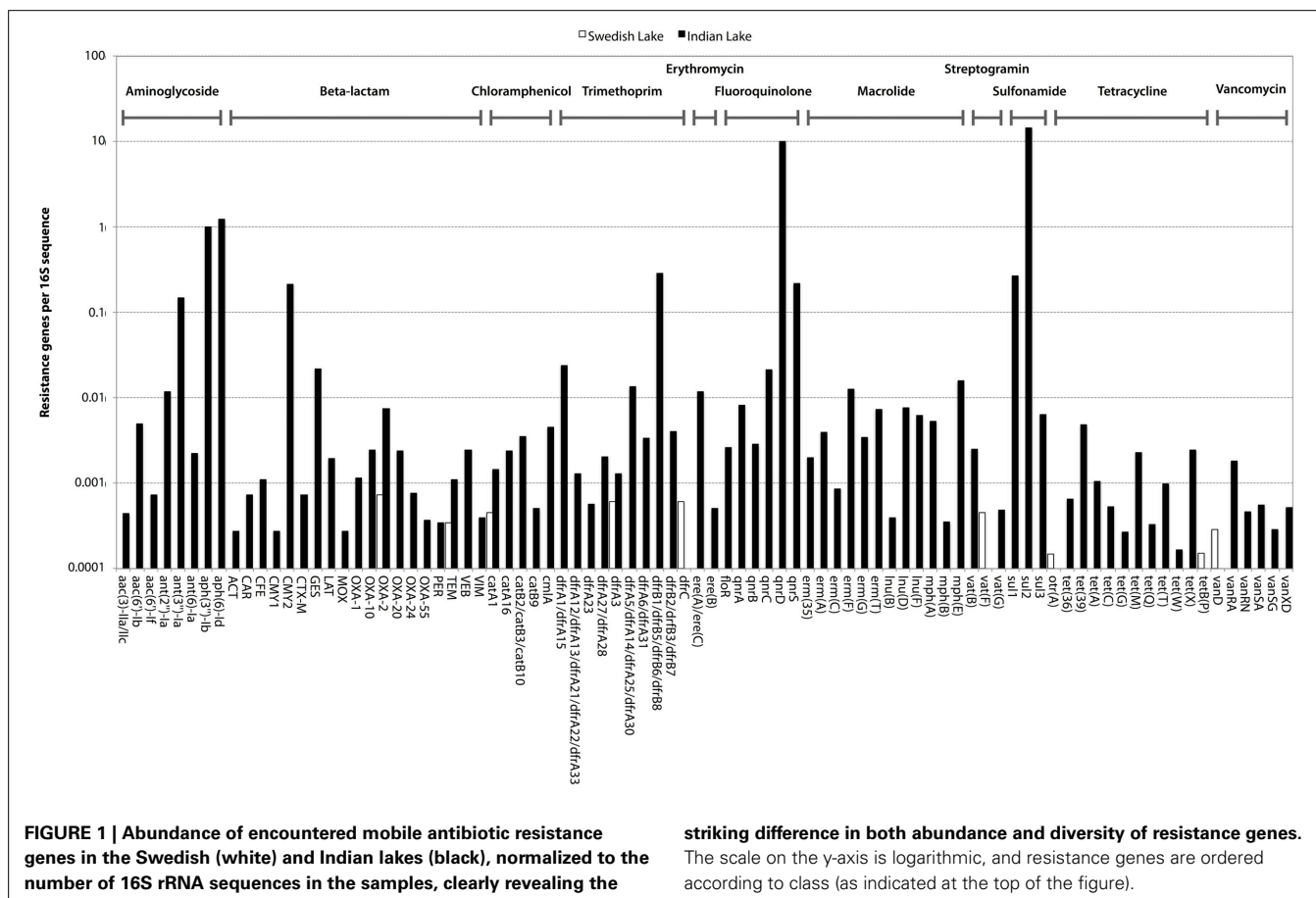
RESULTS

Using Illumina sequencing, we generated 6.69 Gbp of data, corresponding to 21,201,938 paired-end reads passing quality filtering of sediment samples from the Indian lake and 33,422,609 reads from the Swedish lake (Table 1). The reads were assembled separately using Velvet, producing roughly 400,000 contigs for each lake (Table S2).

In the Indian lake, we found 66,960 reads matching resistance genes, while only 10 reads matched in the Swedish lake metagenome. After normalization to gene length and the number of bacterial 16S sequences in each sample, we found about 28.4 resistance genes per 16S sequence in the Indian sample and 0.004 in the Swedish lake (a more than 7000-fold difference). While a large part of the resistance gene content in the polluted lake can be attributed to the sulfonamide resistance gene *sul2* (51.2% of resistance genes detected) and the quinolone resistance gene *qnrD* (35.8%), 81 types of resistance genes could be detected, conferring resistance to in principle all major classes of antibiotics (Figure 1; Table 2).

Table 1 | Summary of the two sequencing libraries used in this study.

Sample	Raw read pairs	Gbp	Passed filtering	% passed	SSU rRNA	% SSU rRNA	Contigs	Total resistance gene reads	% resistance genes
Indian lake sediment	33 466 587	6.69	21 201 938	63.35	7310	0.034	419 550	66 960	0.32
Swedish lake sediment	69 474 306	13.89	33 422 609	48.11	8841	0.026	415 530	10	<0.01



To further assess key functions of the microbial community in the polluted lake, we annotated the assembled contigs against the Pfam, TIGRFAM, and NCBI non-redundant protein databases. We contrasted this with the corresponding analysis of the non-polluted Swedish lake. While only two samples will not allow any proper statistical support for over- and under-representations of protein families, the normalized fold-change between the Indian and Swedish lake can provide us with protein families and features of outstanding interest in these communities. When considering the Pfam protein families with radically higher abundance in the polluted lake, the most striking features are the large proportion of protein families associated with mobility (Figure 2), and a number of protein families indicating higher abundances of phages. Looking at the most over-represented protein families in the Swedish lake, we instead find many proteins related to photosynthesis, which are almost completely absent in the Indian lake.

In addition to resistance genes, we also found a wide range of genes enhancing mobility of genetic material in the Indian samples (Figure 3, Figures S1, S2), such as plasmid conjugation systems, integrases, and transposases. Furthermore, we could detect a large number of *intI* integrons and ISCR sequences. The distribution of integrase sequences was distinctly different in the two lakes, with the Indian lake showing higher overall integrase levels (Figure 4), which could primarily be ascribed to ISCR2, *intI1*, and *intI9* (Figure S2). Several of these mobile elements were located in close proximity to resistance genes (Figure 5). For example, we found a *RelE/StbE* toxin-antitoxin system close to a GES extended spectrum beta-lactamase gene (Weldhagen, 2006; Figure 5A). We also found the fluoroquinolone resistance gene *qnrS*, located close to a gene encoding a bacterial plasmid replicase *RepC* protein (100% identity to *Escherichia coli* protein WP_001672015.1; Figure 5B). An identical *qnrS* gene was also found in a similar context, but with the transposon-associated

Table 2 | Top 20 antibiotic resistance genes in the Indian lake.

Resistance gene	Swedish lake	Indian lake
<i>sul2</i>	0	37332
<i>qnrD</i>	0	20603
<i>aph(6)-Id</i>	0	3221
<i>aph(3')-Ib</i>	0	2595
<i>CMY2</i>	0	778
<i>sul1</i>	0	713
<i>qnrS</i>	0	450
<i>ant(3')-Ia</i>	0	396
<i>dfrB1/dfrB5/dfrB6/dfrB8</i>	0	216
<i>GES</i>	0	60
<i>ere(A)/ere(C)</i>	0	45
<i>qnrC</i>	0	45
<i>mph(E)</i>	0	44
<i>dfrA1/dfrA15</i>	0	36
<i>erm(F)</i>	0	32
<i>ant(2'')-Ia</i>	0	27
<i>dfrA5/dfrA14/dfrA25/dfrA30</i>	0	20
<i>OXA-2</i>	2	19
<i>cmIA</i>	0	18
<i>tet(39)</i>	0	18

Pfam domain DUF772 and a DDE transposase domain located downstream of the *qnrS* gene. This open reading frame had 73% amino acid identity to an *Azoarcus* sp. IS5 transposase (ISAzo41; Siguier et al., 2006), which has not previously been associated with antibiotic resistance genes. However, the *qnrS* gene could be part of this transposable element (Figure 5C). We also found the commonly occurring streptomycin resistance genes *aph(3')-Ib* (*strA*) and *aph(6)-Id* (*strB*) on a contig with 99.9% similarity to a previously sequenced plasmid (AJ431260.1; Figures 5D,E; Tauch et al., 2003). This pair of genes is, however, often found on mobile elements, including transposons and plasmids (Sundin and Bender, 1996; Sundin, 2002), making their relation to that specific plasmid somewhat dubious. The aminoglycoside resistance gene *ant(3'')-Ia* was found together with a gene containing a relaxase domain with a weak hit to *MobP3* (*E*-value 4.5e-11; Smillie et al., 2010), as well as a DNA integrase gene typically found in class 1 integrons (*intI1*; 100% identity to *Klebsiella pneumoniae* protein ACV33249.1; Figure 5F). The contig did not have full-length coverage of any known plasmid, potentially signifying that this constellation of genes here appear in a new context.

The curious absence of any long contigs containing the two most abundant resistance genes (*sul2* and *qnrD*), which instead were fragmented on several smaller contigs with lengths between 74 and 190 bp, suggests that those extremely common resistance genes could be present in multiple contexts and variants, and thereby cause Velvet to perform sub-optimally. We therefore devised a scheme in which Vmatch and Trinity were used to iteratively construct contigs from reads associated with resistance genes. This way, we were able to assemble a 3579 bp contig containing *sul2*, a 2681 bp contig containing *qnrD*, as well as a 2334 bp long contig carrying the *aph(3'')-Ib* (*strA*)

and *aph(6)-Id* (*strB*) genes along with an ISCR2 mobile element (Figure 6).

In the Indian lake, 26 different plasmids present in the RefSeq plasmid genome database were recovered, of which 11 had a mean coverage of more than 100 (Table 3). In contrast, no complete plasmids were detected in the Swedish lake. Two of the recovered plasmids from the Indian lake contained fluoroquinolone resistance genes (*qnrD* and *qnrS2*). In addition, we could assemble 21 new putative small plasmids from the sequences from the Indian lake (Figure 4; Table 4). However, none of those contigs contained any known antibiotic resistance genes.

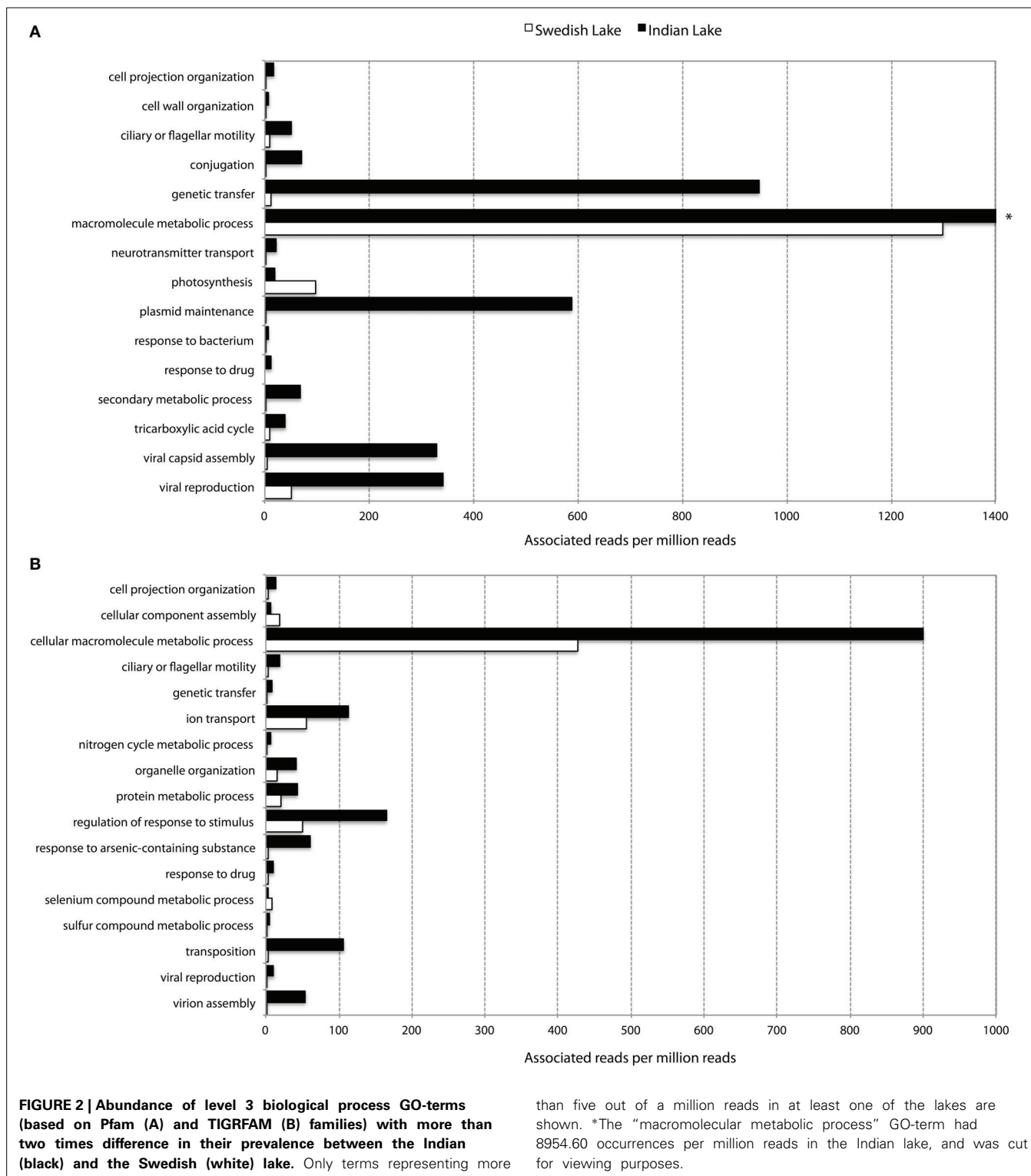
In total, 7310 and 8841 read pairs were identified as SSU sequences in the Indian and Swedish lake, respectively. The taxonomic origin of those read pairs were analyzed at different levels (Figure S3), revealing that the taxonomic composition of the two lakes differed substantially. The Indian lake was dominated by the Clostridia, Deltaproteobacteria, and Gammaproteobacteria classes, while the Swedish lake contained higher relative proportion of bacteria in the Betaproteobacteria and Planctomycetacia classes. The Swedish lake also had higher diversity at all investigated taxonomic levels (consistent with rarefaction analysis results; Figure S4), and also accommodated more unknown and/or unclassified bacterial 16S sequences. However, to us, the difference in diversity between the environments was surprisingly minor, given that the documented pollution of the Indian lake would be likely to pose a strong selection pressure on the community.

At the genus level, the Indian lake was—excluding unknown bacteria—dominated by *Pseudomonas*, *Sulfurimonas*, and *Spirochaeta* species. Additionally, species belonging to the Peptococcaceae, Helicobacteraceae, Desulfomonadaceae, and Desulfobacteraceae families and archaea of the Methanosarcinaceae family were also abundant. The Swedish lake, in contrast, was mostly preoccupied by unknown Proteobacterial genera (Figure S3).

DISCUSSION

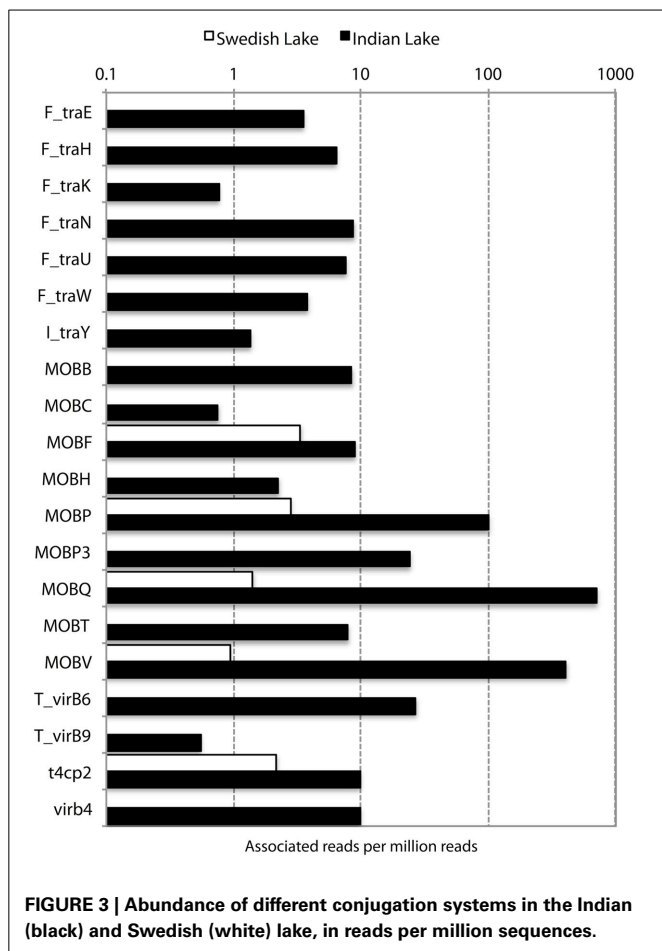
In this work, we have used metagenomic sequence data to describe a high diversity of resistance genes in a highly antibiotic-polluted lake. By assembling the genetic context of the resistance genes, we could conclude that many are located on mobile genetic elements. We could detect high numbers of various genes involved in genetic mobility, in stark contrast with the investigated Swedish lake. Furthermore, we also identified 26 known and 21 novel complete plasmids in the Indian lake. The high abundances of both resistance factors and mobile genetic elements imply that polluted environments could serve as recruitment grounds for antibiotic resistance determinants that ultimately may end up in human pathogens. We also investigated the microbial diversity in the lake, and found that it contained a large number of different microbial taxa. This suggests that the acquisition of mobile resistance genes can contribute to the resilience of microbial communities in polluted environments.

While there are several possible explanations for the exceptional difference in relative abundance of resistance genes between the Indian and Swedish lakes, it should be noted that the total resistance gene abundance in the Indian lake is to our



best knowledge only mirrored by one previously published metagenome, covering antibiotic-contaminated river sediments (Kristiansson et al., 2011). Obviously, there are a myriad of environmental and ecological differences between the two lakes. For example, the Indian lake is known to partially dry out in periods,

while the Swedish lake becomes covered with ice during wintertime. Indeed, the average temperature in the Indian lake is higher throughout the year. The high abundances of several types of resistance genes in the Indian lake are not likely to be a mere reflection of the release of resistant bacteria from human



or animal feces (Czekalski et al., 2012), as common gut bacteria such as *Prevotella*, *Bacteroides*, and *Ruminococcus* are only slightly more abundant in the Indian lake. Additionally, the levels of resistance genes in feces have been reported to be lower than the abundances we detected (Hu et al., 2013). That leaves two plausible explanations for the elevated resistance gene abundance. The first is that pollution of the lake has caused a selection pressure that promotes resistant bacteria on site. As the concentrations of fluoroquinolones reported in the lake water (Fick et al., 2009) approximately correspond to the clinical breakpoint for resistance, this is indeed a plausible explanation. Alternatively, resistant bacteria from antibiotic production have reached the lake via industrial wastewater. Both scenarios would be the result of disposal of waste from pharmaceutical production, but they cannot be distinguished with certainty from each other using the data of this study, and both may contribute to the observed resistance gene abundance.

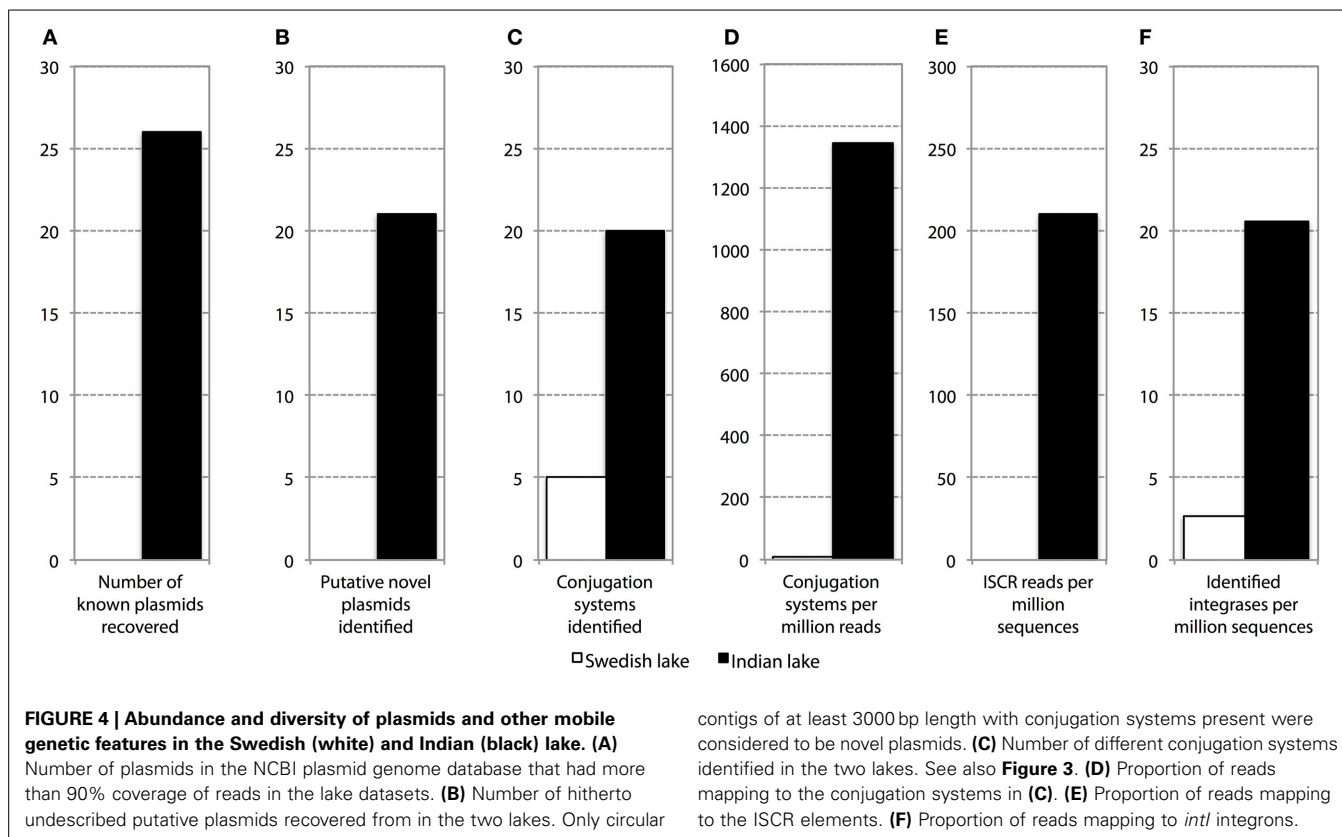
ABUNDANCE AND DIVERSITY OF RESISTANCE GENES

Likely, pollution has had a dramatic impact on the resistance gene content of the Indian lake (Figure 1). The vast range of resistance gene classes detected suggests that some resistance factors could have been co-selected for under the selection pressure from different antibiotics and/or other factors. This could either be due to co-localization of those resistance factors with other resistance

genes on the same mobile genetic element, co-conjugation—i.e., that the resistance factors are located on separate plasmids sharing parts of the conjugation system (Smillie et al., 2010), or caused by a general enrichment of mobile elements induced by strong selective pressure from chemical pollution. Fluoroquinolones are known to often co-select for resistance to several other types of antibiotics (Robicsek et al., 2006), but it is also plausible that other selective agents are present. Resistance genes were found to correspond to 0.316% of the reads in the Indian lake. As the DNA amplification procedure is not completely random, absolute gene abundances should be interpreted with some caution. The total estimated abundance is comparable to the levels found in contaminated river sediments (0.22%; Kristiansson et al., 2011), but far higher than what has been found in activated sludge (0.0054%; Yang et al., 2013), a Puget Sound marina (0.0017%; Port et al., 2012), municipal wastewater treatment plant effluent (0.012%; Port et al., 2012), or in sediment downstream from an effluent treatment plant in Sweden (0.02%; Kristiansson et al., 2011). In comparison, resistance genes in the Swedish lake was found to comprise 0.00003% of the total reads, which is lower than in the Puget Sound marina, but correspond well to data from other environments with little human impact (Port et al., 2012; Muziasari et al., 2014), where resistance genes are generally below the detection limit of previous studies. Consequently, if a lake with even less human activity would have been chosen for comparison, the levels of resistance genes may very well have been below the detection limit also of this study. Regrettably, comparing the number of different resistance gene types encountered between studies is not straightforward due to differences in read length distributions, sequencing errors, annotation strategies and sequencing effort. In this work, the sequencing depth would allow us to detect a resistance gene present in approximately one out of a thousand bacterial cells. Despite this limitation, we have found resistance genes to every major class of antibiotics in the Indian lake, along with a huge abundance of genetic elements promoting mobilization of DNA (Figures 2–4). In principle, this makes the lake a melting pot for antibiotic resistance, where genes that have not yet been mobilized are not unlikely to become in the future.

PUTATIVE NOVEL PLASMIDS IDENTIFIED IN THE POLLUTED LAKE

Given the large differences between the lakes in terms of abundances of plasmids, conjugation systems, and ISCR elements (Figure 4), it is reasonable to assume that the pollution have selected for increased horizontal gene transfer (Gillings, 2014). Similar environments have been shown to select for bacteria resistant to a broad range of antibiotics (Johnning et al., 2013; Marathe et al., 2013), and indeed we were able to assemble several small plasmids from the Indian lake, known as well as unknown (Tables 3, 4). Of the 26 known plasmids recovered, only two contained described resistance genes (*qnrD* and *qnrS2*). While most of these known plasmids have been isolated from human gut bacteria, the abundances of such genera were only slightly higher in the Indian lake compared to the Swedish. Thus, the presence of these plasmids are unlikely to be due to contamination with human feces in the Indian lake, although that may explain a minor portion of their abundance, or provide a possible source of those plasmids. There were furthermore no described resistance



genes among the 21 putative novel plasmids assembled, indicating that the majority of the known resistance genes are located on larger plasmids, or in such a large number of genetic contexts that assigning them to a single genetic element would be nearly impossible by computational means. Sequencing protocols enabling longer read lengths would greatly aid such assemblies. Although the 21 novel plasmids identified did not carry any known resistance genes, they contained several unidentified open read frames (ORFs) that did not match to any Pfam family, even with permissive cutoffs. Such ORFs may encode proteins that are important for survival under the possibly variable and harsh environmental conditions that the Indian lake presents microbes with. Further evaluation of these novel plasmids using experimental methods would therefore be very interesting.

In total numbers, known resistance genes encompassed more than 0.3% of the reads obtained from the Indian lake, with the *sul2* and *qnrD* genes constituting 0.27% of the total reads. Likely, these genes would be present in more than one context; e.g., the recruitment of reads to the pCGB40 plasmid containing a *qnrD* gene, explains less than 10% of the total abundance of *qnrD* in the Indian lake. Since neither *sul2* nor *qnrD* showed up in any of the longer contigs assembled using Velvet, it seems that high abundance might be detrimental to assembly using standard means, rather than advantageous. As shown by our custom iterative assembly approach it can still be possible to assemble high-abundance, but greatly complicated, regions from metagenomic data using specially adapted means (**Figure 6**). This way, we were able to see that about 14% of the *sul2* reads were

associated with a previously unknown genetic element, likely part of a plasmid, of which 79% of the sequence had 99% identity to the *Photobacterium damsela* plasmid pAQU1. Similarly, around 20% of the reads matching to *qnrD* could be explained by a contig assembled using this approach, with 96% identity to a plasmid from *Providencia rettgeri* (pDIJ09-518a). While these two assembled contigs are reassuring, they still only explain a minor portion of all the reads mapping to *sul2* and *qnrD*. Consequently, there seem to be even more variants, which might require even more sophisticated assembly algorithms to be untangled from the vast amount of reads corresponding to these two resistance genes.

INCREASED GENERAL GENETIC MOBILITY

Apart from the massive amount of resistance genes found in the Indian lake compared to the Swedish one, the immense difference in abundance of factors related to horizontal gene transfer between the two lakes was, by far, the most striking feature in terms of functional differences (**Figures 2–4** and **Figure S1**; **Tables S3, S4**). Considering the most abundant GO-annotations in the two lakes, the Indian lake stands out with 2.85% of its reads connected to genetic transfer (**Figure S5**). This process had higher abundance than e.g., carbohydrate and protein metabolic processes, underscoring how extreme the abundance of this gene class is in the Indian lake. These results are clearly echoed among the differentially distributed level 3 GO terms in the samples, both based on Pfam families (**Figure 2A**), and TIGRFAM families (**Figure 2B**). The Indian lake showed large abundances of genes involved in genetic transfer, conjugation, transposition as

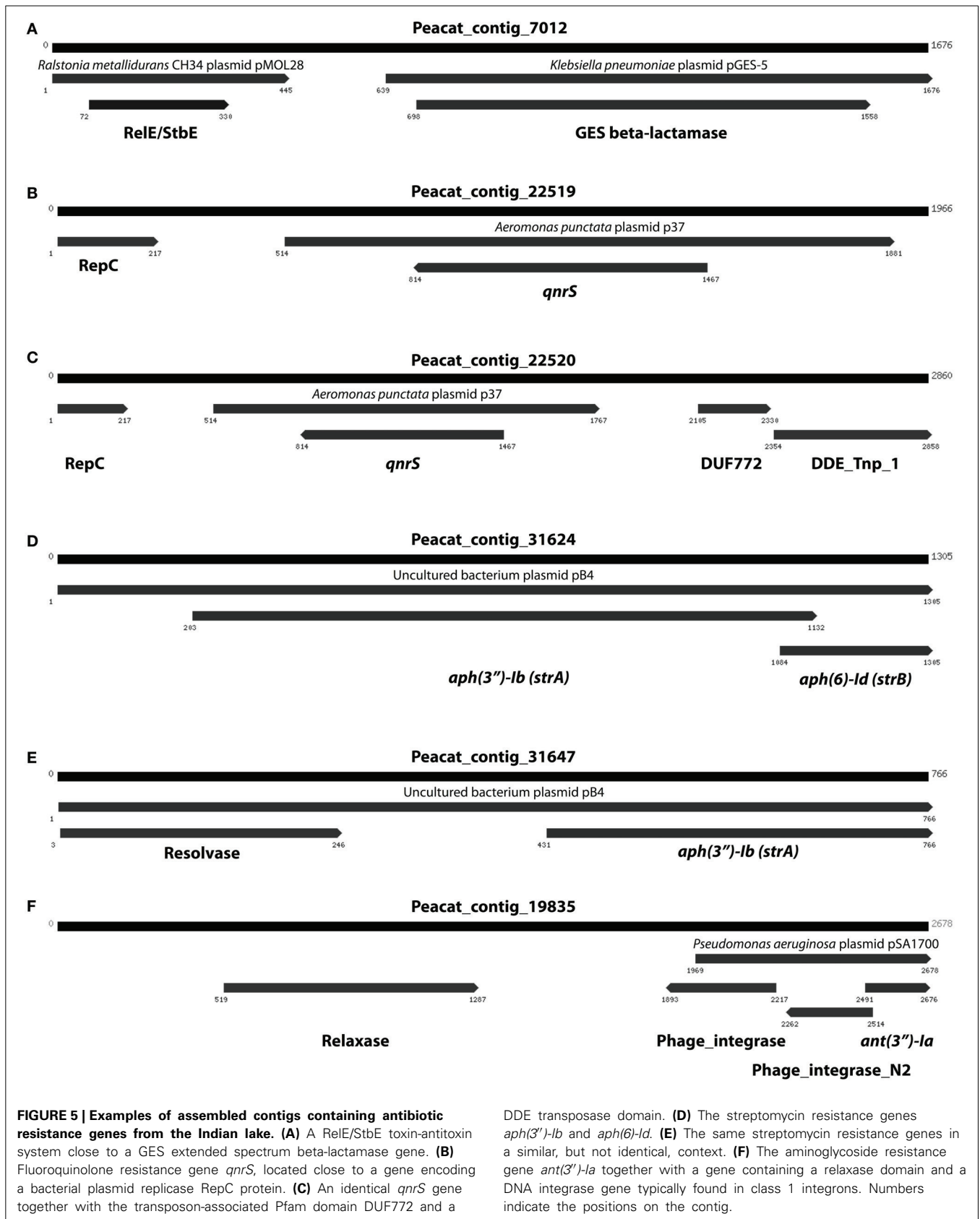
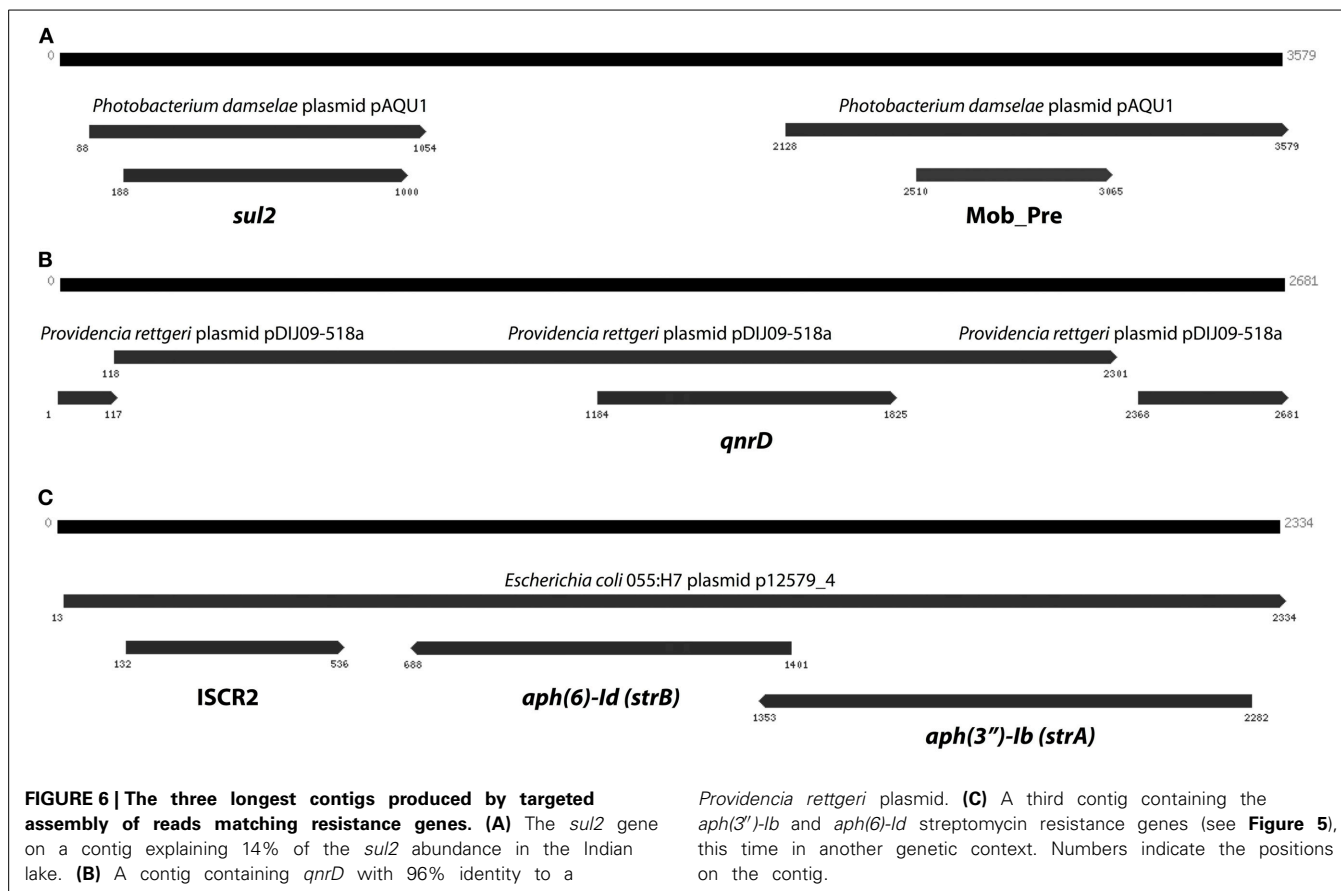


FIGURE 5 | Examples of assembled contigs containing antibiotic resistance genes from the Indian lake. (A) A RelE/StbE toxin-antitoxin system close to a GES extended spectrum beta-lactamase gene. **(B)** Fluoroquinolone resistance gene *qnrS*, located close to a gene encoding a bacterial plasmid replicase RepC protein. **(C)** An identical *qnrS* gene together with the transposon-associated Pfam domain DUF772 and a

DDE transposase domain. **(D)** The streptomycin resistance genes *aph(3'')-Ib* and *aph(6)-Id*. **(E)** The same streptomycin resistance genes in a similar, but not identical, context. **(F)** The aminoglycoside resistance gene *ant(3'')-Ia* together with a gene containing a relaxase domain and a DNA integrase gene typically found in class 1 integrons. Numbers indicate the positions on the contig.



well as plasmid maintenance in both Pfam and TIGRFAM results. It has been suggested that human activities that cause pollution with, e.g., pharmaceuticals, in the long run might influence the evolutionary rate of microorganisms (Gillings and Stokes, 2012). For example, *Streptococcus* species have been shown to become more susceptible to acquiring foreign DNA by transformation under antibiotic exposure (Prudhomme et al., 2006). This may, at least partially, be attributed to the general bacterial SOS response (Beaber et al., 2004). Our observation of a multitude of genes involved in horizontal gene transfer indeed supports this hypothesis.

The contamination of the Indian lake was not only reflected among antibiotic resistance genes, but also in the more general functionality of the bacterial community. Processes such as ion transport, arsenic detoxification, and sulfur metabolism were highly prevalent in the Indian sample, along with more general drug response and the broad-encompassing category of “response to stimulus.” Additionally, macromolecule metabolism also seems to be remarkably abundant in the Indian lake. The general detoxification gene content in the Indian lake is exemplified by findings of the Pfam profiles representing the copper resistance genes *CopB* (3.64 per million reads) and *CopC* (0.71 per million reads), the arsenic resistance gene *ArsC* (34.55 per million reads), the cadmium transporter *Cad* (0.58 per million reads), and the multidrug efflux family Multi_Drug_Res (66.62 per million reads; Table S3), as well as the TIGRFAM family *arsC*

(arsenate reductase; 28.88 per million reads), and selenium and tellurium detoxification proteins belonging to the TMPT_Se_Te family (1.41 per million reads; Table S4). While one should be cautious about drawing conclusions from these data alone, they comprise a suggestion either for co-selection of antibiotic resistance genes and/or mobile elements with a range of detoxification genes; or for a strong general selection pressure by toxicants in the Indian lake. The latter scenario would not be unlikely due to discharge of industrial waste into the lake. Regardless of which, promotion of a broad range of resistance factors toward metals and other toxicants could also promote selection for antibiotic resistance in an environment where both mobile genetic elements and antibiotic resistance genes are prevalent (Pal et al., 2014).

COMMUNITY FUNCTIONS ARE REFLECTED IN TAXONOMY

From a taxonomic perspective, an interesting observation of this study is that the number of eukaryotes seems to differ dramatically between the two lakes. The Swedish lake harbored 2.73% eukaryote 18S rRNA sequences, while we only found 0.01% in the Indian lake. The same pattern was repeated for chloroplast and mitochondrial 16S rRNA (Figure S3). Archaea are often associated with extreme environments (Pikuta et al., 2007), but the overall abundance of archaeal species was lower in the polluted lake, although Euryarchaeota were more common. Moreover, the selection pressures in the Indian lake seem to promote known species and genera of bacteria to a larger extent than in the

Table 3 | Known plasmids recovered from the Indian lake.

Plasmid accession	Plasmid name	Organism	Length (bp)	% GC content	Mean coverage	% covered	Associated resistance genes
gi 514419625 ref NC_021523.1	pCGB40	<i>Escherichia coli</i> strain CGB40	4269	47.1	1890.8	100.0	<i>qnrD</i>
gi 431929441 ref NC_019939.1	pPSEST03	<i>Pseudomonas stutzeri</i> RCH2	2804	61.8	927.8	100.0	
gi 270208707 ref NC_013548.1	pRPSZY	<i>Rhodopseudomonas palustris</i>	2306	48.0	703.0	100.0	
gi 288551640 ref NC_013780.1	pAH3680	<i>Aeromonas hydrophila</i>	3680	55.8	579.4	97.1	
gi 507579660 ref NC_021293.1	pCSA2	<i>Cronobacter sakazakii</i> strain ATCC 29544	5103	55.0	397.3	99.2	
gi 189332428 ref NC_010796.1	pRK10	<i>Serratia marcescens</i>	4241	52.7	294.7	99.2	
gi 449306421 ref NC_020262.1	pSP291-3	<i>Cronobacter sakazakii</i> Sp291	4422	54.0	292.9	94.7	
gi 255929160 ref NC_013090.1	pLK39	<i>Endophytic bacterium</i> LOB-07	4029	55.4	252.6	100.0	
gi 410688106 ref NC_019242.1	pS51B	<i>Enterobacter cloacae</i>	4854	52.4	241.7	99.0	
gi 209947505 ref NC_011404.1	pEC01	<i>Enterobacter cloacae</i>	5002	51.8	167.7	90.5	
gi 410654312 ref NC_019132.1	pSH1148_4.8	<i>Salmonella enterica</i> subsp. <i>enterica</i>	4775	53.3	111.5	96.4	
gi 435855463 ref NC_019988.1	pB1019	<i>Klebsiella pneumoniae</i> strain BB1088	5225	49.7	96.0	100.0	
gi 110666895 ref NC_008246.1	pYAN-1	<i>Sphingobium yanoikuyae</i>	5182	62.0	92.1	91.2	
gi 380083665 ref NC_016979.1	pUUH239.1	<i>Klebsiella pneumoniae</i>	5247	49.2	84.0	100.0	
gi 325168277 ref NC_015182.1	pACMV8	<i>Acidiphilium multivorum</i> AIU301	1729	60.9	57.0	100.0	
gi 10956807 ref NC_002175.1	pMTS1	<i>Methylophaga thalassica</i> strain S1	1995	42.5	54.6	98.1	
gi 410688217 ref NC_019262.1	pAHH04	<i>Aeromonas hydrophila</i>	7191	59.9	45.7	95.3	<i>qnrS2</i>
gi 389842917 ref NC_017935.1	pTHEBA.01	<i>Mesotoga prima</i> MesG1.Ag.4.2	1724	41.6	37.6	100.0	
gi 216996016 ref NC_011640.1	pKpn114	<i>Klebsiella pneumoniae</i>	4211	41.7	36.8	97.7	
gi 410609808 ref NC_019092.1	pLST424C-10	<i>Escherichia coli</i>	8848	49.1	36.1	97.4	
gi 194442198 ref NC_011079.1	pSL254_3	<i>Salmonella enterica</i> subsp. <i>enterica</i>	3605	43.0	30.5	100.0	
gi 410609294 ref NC_019068.1	pKST21	<i>Escherichia coli</i>	1460	51.0	13.2	98.0	
gi 209947511 ref NC_011405.1	pIGRK	<i>Klebsiella pneumoniae</i>	2348	33.4	12.5	100.0	
gi 410108997 ref NC_019014.1	pAAk1	<i>Aeromonas aquariorum</i> AAK1	4161	54.6	7.0	95.1	
gi 387615182 ref NC_017723.1	p58	<i>Escherichia coli</i> ETEC H10407	5800	48.9	6.9	90.4	
gi 189009159 ref NC_010722.1	pMM1	<i>Pseudomonas aeruginosa</i>	2140	45.8	4.4	92.6	

Swedish one, which consistently had higher number of unknown and unclassifiable taxonomic entries. Furthermore, the overall diversity also seems to be slightly lower in the Indian lake (Figure S4), as would be expected as a result of strong selection pressure (Banks et al., 2013). However, the difference was rather small, showing that a vast range of species actually do survive the comparably high levels of antibiotics and other chemicals present in the Indian lake (Fick et al., 2009). This is an interesting finding, as that indicates that a plethora of species are either intrinsically resistant, or have acquired resistance, and are therefore able to survive and reproduce in the contaminated lake.

The overall composition of species in the Indian lake showed high numbers of proteobacteria and firmicutes. Within proteobacteria, a large part of the abundance could be attributed to sulfate reducing bacteria such as Desulfobacteraceae, Desulfomicrobiaceae, and Desulfuromonadaceae. We also found large quantities of iron reducers in the Deferribacteraceae family, suggesting that both sulfur and iron compounds are important in the element cycling in the sediment community of the Indian lake. Additionally, we found large numbers of Clostridia, a class of obligate anaerobes, representing a strong indication that this environment might be very oxygen-poor. Clostridia and many other Firmicutes are able to form endospores, ensuring their survival in harsh conditions, such as drought. Since we found

protein families related to spore formation in the Indian lake, but not in the Swedish lake (Tables S5, S6), the presence of microorganisms capable of such transformations is not surprising. Moreover, the Indian lake is known to be largely dried out in periods, meaning that spore formation might not only be a survival factor related to chemical stressors, but also to physical stress.

MITIGATING RISKS FOR SPREAD AND EMERGENCE OF RESISTANCE FACTORS

Given the broad range and relatively large abundance of known resistance genes we could identify in the Indian lake, it would be surprising not to find as yet undiscovered resistance genes in this environment. Many antibiotic resistance genes encountered in clinical settings are thought to originate from environmental bacteria (Forsberg et al., 2012; Wright, 2012; Finley et al., 2013; Gaze et al., 2013), where they probably have been present for millions of years (D'Costa et al., 2011). It would be astonishing if this gene-flow suddenly stopped, and novel antibiotic substances we discover would not face horizontally-transferred resistance development. However, the more likely scenario is that new resistance genes await us down the line, as has happened with e.g., fluoroquinolones and carbapenems (Lupo et al., 2012). Such novel resistance genes can be hard to discover

Table 4 | Novel plasmids retrieved from the Indian lake.

Plasmid contig	Length (bp)	% GC content	Avg. coverage	% covered
lake_2007_contig_201	3295	58.1	89.7	100.0
lake_2007_contig_1230	4784	59.8	23.3	100.0
lake_2007_contig_30733	3975	48.9	16.1	100.0
lake_2007_contig_6114	5252	58.2	15.9	100.0
Peacat_contig_45232	5164	60.4	15.8	100.0
lake_2007_contig_5550	4525	51.0	14.3	100.0
lake_2007_contig_12425	5024	51.2	14.2	99.7
lake_2007_contig_4528	5356	44.8	11.3	100.0
lake_2007_contig_127348	3660	62.4	9.6	100.0
Peacat_contig_40251	4791	47.1	9.6	100.0
lake_2007_contig_19565	4418	55.6	9.2	99.4
lake_2007_contig_3701	5828	48.1	8.9	100.0
lake_2007_contig_10491	6072	46.8	7.7	99.3
lake_2007_contig_57444	3019	46.6	7.7	100.0
lake_2007_contig_42990	3153	36.6	7.6	100.0
lake_2007_contig_2219	7768	39.6	6.3	99.8
lake_2007_contig_187761	6203	32.2	5.8	100.0
lake_2007_contig_6987	4400	35.8	5.6	100.0
Peacat_contig_24026	3122	64.9	5.5	96.7
lake_2007_contig_12553	4616	44.0	5.5	100.0
lake_2007_contig_22993	3577	40.8	5.5	99.6

using bioinformatic approaches, although a great deal can be achieved by analysis of homologous sequences (Boulund et al., 2012; Flach et al., 2013). Instead, studies of genetic material from environmental communities inserted into laboratory strains subjected to selection pressures, so called functional metagenomics (Handelsman et al., 1998), will be instrumental in detecting such completely novel resistance genes. Functional metagenomics has been previously applied to detect e.g., novel beta-lactamases (Allen et al., 2009) and chloramphenicol resistance genes (Lang et al., 2010). Such large-scale screenings of heavily contaminated environments similar to the Indian lake could provide us with important insights into the range and nature of resistance genes that we might face in the clinic in the future.

An important aspect of research on environmental antibiotic resistance is to track down potential sources of antibiotic contamination, and particularly to assess in which arenas transfer of resistance genes from environmental bacteria into pathogens is more likely to occur (Pruden et al., 2013). Often, milieus such as wastewater treatment plants have been proposed as such potential high-risk environments (Rizzo et al., 2013). However, the findings of this work and previous studies of environments contaminated by pharmaceutical production waste (Kristiansson et al., 2011) suggest that much more effort is needed in unearthing risks related to pollution from production facilities. Particularly, risks are high in settings where humans and/or animals interact with polluted water, soil or other materials. In such situations, there is an obvious risk that resistance genes selected for in a polluted environment might spread into the human or animal gut microflora, and thereafter propagate into pathogens or

pathogenic opportunists. Consequently, more studies are needed on resistance genes and their mobility in highly polluted areas as well as in their neighboring milieu, and further on in the human microbiome of their inhabitants. It is of high importance to identify such pollution sources, and to inform management (Pruden et al., 2013). Furthermore, we also need to understand how resistance plasmids evolve and spread, and which factors can assist in such processes. Metagenomics can play a significant role in identifying settings of particularly high risk, and thereby suggest where to start future mitigation efforts.

AUTHOR CONTRIBUTIONS

D. G. Joakim Larsson, Erik Kristiansson, and Jerker Fick designed the study. Jerker Fick was responsible for collecting the samples. Johan Bengtsson-Palme and Fredrik Boulund performed the bioinformatics analysis. Johan Bengtsson-Palme, Fredrik Boulund, Erik Kristiansson, and D. G. Joakim Larsson interpreted the results. Johan Bengtsson-Palme, Erik Kristiansson, and D. G. Joakim Larsson drafted the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Swedish Research Council (VR), the Swedish Research Council for Environment, Agriculture and Spatial Planning (FORMAS), and the Swedish Foundation for Strategic Environmental Research (MISTRA). The authors would like to thank Dr. Carl-Fredrik Flach for valuable input on the manuscript and Birgitta Weijdegård for assistance with the DNA extractions. The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/Journal/10.3389/fmicb.2014.00648/abstract>

REFERENCES

- Allen, H. K., Moe, L. A., Rodbummer, J., Gaarder, A., and Handelsman, J. (2009). Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J.* 3, 243–251. doi: 10.1038/ismej.2008.86
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Andrews, S. (2010). *FastQC*. Available online at: <http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., et al. (2013). How does ecological disturbance influence genetic diversity? *Trends Ecol. Evol. (Amst.)* 28, 670–679. doi: 10.1016/j.tree.2013.08.005
- Beaber, J. W., Hochhut, B., and Waldor, M. K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427, 72–74. doi: 10.1038/nature02241
- Bengtsson, J., Eriksson, K. M., Hartmann, M., Wang, Z., Shenoy, B. D., Grelet, G.-A., et al. (2011). Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea,

- bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie Van Leeuwenhoek* 100, 471–475. doi: 10.1007/s10482-011-9598-6
- Bengtsson, J., Hartmann, M., Unterseher, M., Vaishampayan, P., Abarenkov, K., Durso, L., et al. (2012). Megraft: a software package to graft ribosomal small subunit (16S/18S) fragments onto full-length sequences for accurate species richness and sequencing depth analysis in pyrosequencing-length metagenomes and similar environmental datasets. *Res. Microbiol.* 163, 407–412. doi: 10.1016/j.resmic.2012.07.001
- Bengtsson-Palme, J. (2012). *PETKit*. Available online at: <http://microbiology.se/software/petkit/>
- Bengtsson-Palme, J. (2013). *TriMetAss*. Available online at: <http://microbiology.se/software/trimetass/>
- Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., et al. (2014). *Metaxa2: Improved Identification and Taxonomic Classification of Small and Large Subunit rRNA in Metagenomic Data*. Available online at: <http://microbiology.se/software/metaxa2/>
- Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2014). GenBank. *Nucleic Acids Res.* 42, D32–D37. doi: 10.1093/nar/gkt1030
- Boralkar, D. B., Alvares, C., Devotta, S., Sharma, P. N., and Thyagarajan, G. (2005). *Report of Visit to Hyderabad (A.P.) 19-20 October 2004*. New Delhi: Supreme Court of India.
- Boulund, F., Johnning, A., Pereira, M. B., Larsson, D. G. J., and Kristiansson, E. (2012). A novel method to discover fluoroquinolone antibiotic resistance (qnr) genes in fragmented nucleotide sequences. *BMC Genomics* 13:695. doi: 10.1186/1471-2164-13-695
- Czekalski, N., Berthold, T., Caucchi, S., Egli, A., and Bürgmann, H. (2012). Increased levels of multiresistant bacteria and resistance genes after wastewater treatment and their dissemination into lake Geneva, Switzerland. *Front. Microbiol.* 3:106. doi: 10.3389/fmicb.2012.00106
- D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., et al. (2011). Antibiotic resistance is ancient. *Nature* 477, 457–461. doi: 10.1038/nature10388
- Eddy, S. (2010). *HMMER*. Available online at: <http://hmmer.janelia.org>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- European Society of Clinical Microbiology and Infectious Diseases. (2014). *EUCAST Clinical Breakpoints Version 4.0*. Available online at: http://www.eucast.org/clinical_breakpoints/
- Fick, J., Söderström, H., Lindberg, R. H., Phan, C., Tysklind, M., and Larsson, D. G. J. (2009). Contamination of surface, ground, and drinking water from pharmaceutical production. *Environ. Toxicol. Chem.* 28, 2522–2527. doi: 10.1897/09-073.1
- Finley, R. L., Collignon, P., Larsson, D. G. J., McEwen, S. A., Li, X.-Z., Gaze, W. H., et al. (2013). The scourge of antibiotic resistance: the important role of the environment. *Clin. Infect. Dis.* 57, 704–710. doi: 10.1093/cid/cit355
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Flach, C.-F., Boulund, F., Kristiansson, E., and Larsson, D. G. J. (2013). Functional verification of computationally predicted qnr genes. *Ann. Clin. Microbiol. Antimicrob.* 12:34. doi: 10.1186/1476-0711-12-34
- Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. A., and Dantas, G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337, 1107–1111. doi: 10.1126/science.1220761
- Gaze, W. H., Krone, S. M., Larsson, D. G. J., Li, X.-Z., Robinson, J. A., Simonet, P., et al. (2013). Influence of humans on evolution and mobilization of environmental antibiotic resistome. *Emerging Infect. Dis.* 19. doi: 10.3201/eid1907.120871
- Gaze, W. H., Zhang, L., Abdoulsalam, N. A., Hawkey, P. M., Calvo-Bado, L., Royle, J., et al. (2011). Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *ISME J.* 5, 1253–1261. doi: 10.1038/ismej.2011.15
- Gillings, M. R. (2014). Integrons: past, present, and future. *Microbiol. Mol. Biol. Rev.* 78, 257–277. doi: 10.1128/MMBR.00056-13
- Gillings, M. R., and Stokes, H. W. (2012). Are humans increasing bacterial evolvability? *Trends Ecol. Evol. (Amst.)* 27, 346–352. doi: 10.1016/j.tree.2012.02.006
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Greenpeace. (2004). *State of Community Health at Medak District*. Bangalore: Greenpeace.
- Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2011). The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.* 7:e1002222. doi: 10.1371/journal.pgen.1002222
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, D387–D395. doi: 10.1093/nar/gks1234
- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–R249. doi: 10.1016/S1074-5521(98)90108-9
- Heuer, H., and Smalla, K. (2012). Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol. Rev.* 36, 1083–1104. doi: 10.1111/j.1574-6976.2012.00337.x
- Hu, Y., Yang, X., Qin, J., Lu, N., Cheng, G., Wu, N., et al. (2013). Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat. Commun.* 4, 2151. doi: 10.1038/ncomms3151
- Johnning, A., Moore, E. R. B., Svensson-Stadler, L., Shouche, Y. S., Larsson, D. G. J., and Kristiansson, E. (2013). Acquired genetic mechanisms of a multiresistant bacterium isolated from a treatment plant receiving wastewater from antibiotic production. *Appl. Environ. Microbiol.* 79, 7256–7263. doi: 10.1128/AEM.02141-13
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202
- Kristiansson, E., Fick, J., Janzon, A., Grabic, R., Rutgersson, C., Weijdegård, B., et al. (2011). Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. *PLoS ONE* 6:e17038. doi: 10.1371/journal.pone.0017038
- Kurtz, S. (2010). *The Vmatch Large Scale Sequence Analysis Software*. Available online at: <http://vmatch.de/>
- Lang, K. S., Anderson, J. M., Schwarz, S., Williamson, L., Handelsman, J., and Singer, R. S. (2010). Novel florfenicol and chloramphenicol resistance gene discovered in Alaskan soil by using functional metagenomics. *Appl. Environ. Microbiol.* 76, 5321–5326. doi: 10.1128/AEM.00323-10
- Larsson, D. G. J. (2014). Pollution from drug manufacturing: review and perspectives. *Phil. Trans. R. Soc. B* 369:20130571. doi: 10.1098/rstb.2013.0571
- Larsson, D. G. J., de Pedro, C., and Paxeus, N. (2007). Effluent from drug manufactures contains extremely high levels of pharmaceuticals. *J. Hazard. Mater.* 148, 751–755. doi: 10.1016/j.jhazmat.2007.07.008
- Li, D., Yang, M., Hu, J., Ren, L., Zhang, Y., and Li, K. (2008a). Determination and fate of oxytetracycline and related compounds in oxytetracycline production wastewater and the receiving river. *Environ. Toxicol. Chem.* 27, 80–86. doi: 10.1897/07-080.1
- Li, D., Yang, M., Hu, J., Zhang, Y., Chang, H., and Jin, F. (2008b). Determination of penicillin G and its degradation products in a penicillin production wastewater treatment plant and the receiving river. *Water Res.* 42, 307–317. doi: 10.1016/j.watres.2007.07.016
- Lupo, A., Coyne, S., and Berendonk, T. U. (2012). Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. *Front. Microbiol.* 3:18. doi: 10.3389/fmicb.2012.00018
- Marathe, N. P., Regina, V. R., Walujkar, S. A., Charan, S. S., Moore, E. R. B., Larsson, D. G. J., et al. (2013). A treatment plant receiving waste water from multiple bulk drug manufacturers is a reservoir for highly multi-drug resistant integron-bearing bacteria. *PLoS ONE* 8:e77310. doi: 10.1371/journal.pone.0077310
- Martinez, J. L. (2008). Antibiotics and antibiotic resistance genes in natural environments. *Science* 321, 365–367. doi: 10.1126/science.1159483
- Muziasari, W. I., Managaki, S., Pärnänen, K., Karkman, A., Lyra, C., Tamminen, M., et al. (2014). Sulphonamide and trimethoprim resistance genes persist in sediments at Baltic Sea aquaculture farms but are not detected in the surrounding environment. *PLoS ONE* 9:e92702. doi: 10.1371/journal.pone.0092702
- Nigro, S. J., and Hall, R. M. (2012). Antibiotic resistance islands in A320 (RUH134), the reference strain for *Acinetobacter baumannii* global clone 2. *J. Antimicrob. Chemother.* 67, 335–338. doi: 10.1093/jac/dkr447

- Norrby, R., Powell, M., Aronsson, B., Monnet, D. L., Lutsar, I., Bocsan, I. S., et al. (2009). *The Bacterial Challenge: Time to React*. Stockholm: European Centre for Disease Prevention and Control.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R. O., O Hara, R. B., et al. (2011). *vegan: Community Ecology Package*. Available online at: <http://CRAN.R-project.org/package=vegan>
- Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. G. J. (2014). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, D737–D743. doi: 10.1093/nar/gkt1252
- Pikuta, E. V., Hoover, R. B., and Tang, J. (2007). Microbial extremophiles at the limits of life. *Crit. Rev. Microbiol.* 33, 183–209. doi: 10.1080/10408410701451948
- Pinar, R., de Winter, A., Sarkis, G. J., Gerstein, M. B., Tartaro, K. R., Plant, R. N., et al. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* 7:216. doi: 10.1186/1471-2164-7-216
- Port, J. A., Wallace, J. C., Griffith, W. C., and Faustman, E. M. (2012). Metagenomic profiling of microbial composition and antibiotic resistance determinants in Puget Sound. *PLoS ONE* 7:e48000. doi: 10.1371/journal.pone.0048000
- Pruden, A., Larsson, D. G. J., Amézquita, A., Collignon, P., Brandt, K. K., Graham, D. W., et al. (2013). Management options for reducing the release of antibiotics and antibiotic resistance genes to the environment. *Environ. Health Perspect.* 121, 878–885. doi: 10.1289/ehp.1206446
- Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B., and Claverys, J.-P. (2006). Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae*. *Science* 313, 89–92. doi: 10.1126/science.1127912
- Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M. C., et al. (2013). Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *Sci. Total Environ.* 447, 345–360. doi: 10.1016/j.scitotenv.2013.01.032
- Robicsek, A., Jacoby, G. A., and Hooper, D. C. (2006). The worldwide emergence of plasmid-mediated quinolone resistance. *Lancet Infect. Dis.* 6, 629–640. doi: 10.1016/S1473-3099(06)70599-0
- Schlüter, A., Szczepanowski, R., Pühler, A., and Top, E. M. (2007). Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol. Rev.* 31, 449–477. doi: 10.1111/j.1574-6976.2007.00074.x
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010). Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* 74, 434–452. doi: 10.1128/MMBR.00020-10
- Stokes, H. W., and Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* 35, 790–819. doi: 10.1111/j.1574-6976.2011.00273.x
- Sundin, G. W. (2002). Distinct recent lineages of the strA-strB streptomycin-resistance genes in clinical and environmental bacteria. *Curr. Microbiol.* 45, 63–69. doi: 10.1007/s00284-001-0100-y
- Sundin, G. W., and Bender, C. L. (1996). Dissemination of the strA-strB streptomycin-resistance genes among commensal and pathogenic bacteria from humans, animals, and plants. *Mol. Ecol.* 5, 133–143. doi: 10.1111/j.1365-294X.1996.tb00299.x
- Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., Kaiser, O., et al. (2008). Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *J. Biotechnol.* 136, 54–64. doi: 10.1016/j.jbiotec.2008.03.020
- Tauch, A., Schlüter, A., Bischoff, N., Goesmann, A., Meyer, F., and Pühler, A. (2003). The 79,370-bp conjugative plasmid pB4 consists of an IncP-1beta backbone loaded with a chromate resistance transposon, the strA-strB streptomycin resistance gene pair, the oxacillinase gene bla(NPS-1), and a tripartite antibiotic efflux system of the resistance-nodulation-division family. *Mol. Genet. Genomics* 268, 570–584. doi: 10.1007/s00438-002-0785-z
- Toleman, M. A., and Walsh, T. R. (2011). Combinatorial events of insertion sequences and ICE in Gram-negative bacteria. *FEMS Microbiol. Rev.* 35, 912–935. doi: 10.1111/j.1574-6976.2011.00294.x
- Weldhagen, G. F. (2006). GES: an emerging family of extended spectrum beta-lactamases. *Clin. Microbiol. Newsl.* 28, 145–149. doi: 10.1016/j.clinmicnews.2006.09.004
- Wellington, E. M., Boxall, A. B., Cross, P., Feil, E. J., Gaze, W. H., Hawkey, P. M., et al. (2013). The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. *Lancet Infect. Dis.* 13, 155–165. doi: 10.1016/S1473-3099(12)70317-1
- Wright, G. D. (2012). The origins of antibiotic resistance. *Handb. Exp. Pharmacol.* 13–30. doi: 10.1007/978-3-642-28951-4_2
- Yang, Y., Li, B., Ju, F., and Zhang, T. (2013). Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. *Environ. Sci. Technol.* 47, 10197–10205. doi: 10.1021/es4017365
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 September 2014; accepted: 07 November 2014; published online: 02 December 2014.

Citation: Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E and Larsson DGJ (2014) Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Front. Microbiol.* 5:648. doi: 10.3389/fmicb.2014.00648

This article was submitted to *Antimicrobials, Resistance and Chemotherapy*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Bengtsson-Palme, Boulund, Fick, Kristiansson and Larsson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.