Discovery of the fourth mobile sulfonamide resistance gene

Downloaded from: https://research.chalmers.se, 2021-08-31 20:32 UTC

Citation for the original published paper (version of record):
Razavi, M., Marathe, N., Gillings, M. et al (2017)
Discovery of the fourth mobile sulfonamide resistance gene
Microbiome, 5(1)
http://dx.doi.org/10.1186/s40168-017-0379-y

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology.
It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004.
research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Razavi *et al. Microbiome* (2017) 5:160
DOI 10.1186/s40168-017-0379-y

Microbiome

RESEARCH Open Access

Discovery of the fourth mobile sulfonamide resistance gene

Mohammad Razavi[1,2], Nachiket P. Marathe[1,2], Michael R. Gillings[3], Carl-Fredrik Flach[1,2], Erik Kristiansson[1,4] and D. G. Joakim Larsson[1,2*]

Abstract

Background: Over the past 75 years, human pathogens have acquired antibiotic resistance genes (ARGs), often from environmental bacteria. Integrons play a major role in the acquisition of antibiotic resistance genes. We therefore hypothesized that focused exploration of integron gene cassettes from microbial communities could be an efficient way to find novel mobile resistance genes. DNA from polluted Indian river sediments were amplified using three sets of primers targeting class 1 integrons and sequenced by long- and short-read technologies to maintain both accuracy and context.

Results: Up to 89% of identified open reading frames encode known resistance genes, or variations thereof (> 1000). We identified putative novel ARGs to aminoglycosides, beta-lactams, trimethoprim, rifampicin, and chloramphenicol, including several novel OXA variants, providing reduced susceptibility to carbapenems. One dihydropteroate synthase gene, with less than 34% amino acid identity to the three known mobile sulfonamide resistance genes (*sul1–3*), provided complete resistance when expressed in *Escherichia coli*. The mobilized gene, here named *sul4*, is the first mobile sulfonamide resistance gene discovered since 2003. Analyses of adjacent DNA suggest that *sul4* has been decontextualized from a set of chromosomal genes involved in folate synthesis in its original host, likely within the phylum Chloroflexi. The presence of an insertion sequence common region element could provide mobility to the entire integron. Screening of 6489 metagenomic datasets revealed that *sul4* is already widespread in seven countries across Asia and Europe.

Conclusions: Our findings show that exploring integrons from environmental communities with a history of antibiotic exposure can provide an efficient way to find novel, mobile resistance genes. The mobilization of a fourth sulfonamide resistance gene is likely to provide expanded opportunities for sulfonamide resistance to spread, with potential impacts on both human and animal health.

Keywords: Evolution, Bioprospecting, Resistome, Metagenomics, Pharmaceutical, Environment

Background

Bacterial pathogens can become insensitive to antibiotics due to mutations in pre-existing DNA, or by acquisition of antibiotic resistance genes (ARGs), many of which are likely to originate from environmental bacteria [1]. These genes spread via mobile genetic elements, such as plasmids and transposons, which facilitate the transfer of genetic material between bacterial cells and species [2].

* Correspondence: joakim.larsson@fysiologi.gu.se
[1]Centre for Antibiotic Resistance Research (CARe) at University of Gothenburg, Gothenburg, Sweden
[2]Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
Full list of author information is available at the end of the article

Integrons play a major role in the acquisition and dissemination of ARGs. These genetic elements capture and express genes; they are often associated with transposons and can be carried by conjugative plasmids [3]. Integrons are composed of three key features: an integron integrase gene (*intI*), an integron-associated recombination site (*attI*), and an integron-associated promoter (Pc). The *intI* gene encodes a site-specific tyrosine recombinase, which performs integration and excision of genetic elements, known as gene cassettes, at the recombination site, *attI*. Then, the integrated gene or genes are expressed by a dedicated promoter (Pc) embedded in *intI* or the *attI* site. This mechanism for the integration and excision of new functional modules helps bacteria

BioMed Central

© The Author(s). 2017 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

rapidly adapt to selection pressures, including the acquisition of resistance phenotypes [4].

Integrons have been found in ~ 6% of all sequenced bacterial genomes [5]. These ancient genetic elements can recruit diverse gene cassettes, most often encoding proteins of unknown function [3]. In contrast, integrons carried by pathogens are often resident on mobile elements and carry resistance genes. The relative abundance of mobile integrons in pathogens, particularly the class 1 integron, is now much higher than it is in environmental organisms [6–10]. The success of these clinical class 1 integrons depends partly on their association with transposons of the *Tn*402 and *Tn*21 families. In particular, the *Tn*402 transposon targets the resolution (*res*) site of plasmids, thus inserting the class 1 integron into a wide diversity of conjugative plasmids [3].

Genes encoding resistance to almost all families of antibiotics have been accumulated by class 1 integrons over the last 100 years [11, 12]. The recruitment of novel resistance genes into integrons is thus of considerable concern [13, 14], and their presence in non-pathogenic species or environmental microbial communities implies a risk for future transfer to human pathogens. Given the vast diversity of both bacteria and genes in the external environment [3, 15], environmental gene cassettes are likely to be an important source of novel resistance genes to pathogens.

Early knowledge of genes with a potential to become clinically relevant resistance genes is important, because this helps us to better understand how resistance develops and to prepare surveillance and control measures to reduce their dissemination. Both functional metagenomics and sequence-based metagenomics have been used in the past to identify candidate resistance genes in environmental communities [16, 17]. Functional metagenomics relies on phenotypic screening and is thus hampered by the high abundance of well-characterized resistance genes. This makes functional metagenomics cost- and labor-intensive to find rare, novel resistance genes. Sequence-based metagenomics relies on similarities to known resistance genes, thus easily missing truly novel genes. Additionally, the overwhelming majority of the sequenced DNA has no relevance for resistance, increasing sequencing costs and computation [18]. Neither of these techniques can easily pinpoint mobile genes that have a higher probability of being transferred to pathogens. An approach that specifically targets mobile resistance genes would have the potential to identify genes of concern in a more efficient way. Novel ARGs located in mobile elements, such as clinical class 1 integrons, would be at increased risk of becoming a clinical problem [13].

In addition to focusing on mobile elements, exploring microbial communities with existing selection pressure from antibiotics would probably further increase the chances of finding novel resistance genes. Environments impacted by discharges from antibiotic manufacturing could hence be relevant to investigate. For about a decade, we have studied an Indian treatment plant that receives highly antibiotic-contaminated wastewater from drug manufacturing. As much as 80% of the bacteria isolated from this environment harbor class 1 integrons [19]. Downstream river sediments also contain elevated abundances of class 1 integrons [20]. The antibiotic consumption in India is high and to a large extent uncontrolled [21]. Accordingly, antibiotic resistance in, for example, *Enterobacteriaceae* has become a major problem [22, 23]. Environments where untreated sewage is mixed with "environmental" bacteria could therefore be worthwhile to explore for novel resistance genes.

In this study, we have characterized class 1 integrons to identify novel ARGs and to expand our knowledge of the gene cassettes employed in integrons. Amplicons of integron gene cassettes were generated from Indian river sediments that were heavily contaminated by industrial discharges of antibiotics and by untreated sewage and hospital waste [19, 24].

Results

Sequencing amplicons resulted in 216,807 long PacBio reads (LRs or partial integrons) with an average length of 1.25 kilobases (kb) and 14,184,598 short Illumina reads (SRs) with a maximum length of 250 bases. After filtering low-quality reads, 13,506,840 SRs along with all the LRs were fed into Proovread [25], resulting in 170,257 corrected LRs. Clustering of the LR dataset resulted in 102,550 non-redundant reads (Table 1). A total of 198,436 open reading frames (ORFs) were identified by Prodigal [26]. Clustering of all identified ORFs at 99% amino acid identity led to 19,723 unique ORFs. The ORFs were annotated against the NCBI protein and nucleotide database to identify their putative function. To the best of our knowledge, the numbers of different partial integrons and putative gene cassettes in this study are considerably higher than any previous study that has identified gene cassettes from environmental samples (Table 2).

For both sites, resistance gene cassettes (known and putative) dominated (51 to 89%) (Fig. 1 and Additional file 1: Figure S1). The primer pair HS458-HS459, which was designed to preferentially recover "clinical" integrons, showed the strongest dominance of known and putative resistance gene cassettes (78 to 89%). The primer pair HS464-GCP2, which recovers a slightly more even mix of clinical as well as environmental (pre-clinical) integrons, contained a somewhat lower proportion (52 to 59%), whereas 51 to 73% of the amplicons from the MRG284-MRG285 primer pair included known and putative

Table 1 Total number of unique long reads by sample and primer combinations in the sequenced amplicons

		Primers			
		HS464-GCP2	HS458-HS459	MRG284-MRG285	Primers not detected
Samples	RS_{PETL}	30,780 (1.3 kb)	11,317 (1.2 kb)	6665 (1.1 kb)	594 (1.1 kb)
	RS_{Pune}	25,229 (1.4 kb)	16,182 (1.3 kb)	10,853 (1.1 kb)	930 (1.1 kb)

The average lengths of the long reads are presented in parenthesis

resistance gene cassettes (Additional file 1: Figure S1). The latter pair recovers primarily chromosomal integrons. Additional file 1: Figure S2 illustrates putative functions of the ORFs of gene cassettes in the "clinical" versus "environmental" integrons from the HS464-GCP2 primer pair (see the Methods section). The distribution of putative functions was very similar between the two.

Known and putative ARGs are reported in Additional files 2 and 3. The number of genes providing resistance to different classes of antibiotics is indicated in Table 3. Moreover, to assess if the identified ORFs included genes that were not previously described as integron gene cassettes, we matched them against INTEGRALL [27]. Of the 19,723 non-redundant ORFs identified in this study, 5942 (~ 30%) were previously reported (see the Methods section). However, the rest of the ORFs had lower nucleotide similarity (identity < 95% and coverage < 70) to the sequences deposited in INTEGRALL. Additional file 1: Figure S3 shows the putative function of these recent ORFs based on known homologs in the NCBI protein database. Hypothetical proteins form the largest portion in both samples, followed by ARGs (known as well as putatively novel). The full list of these ORFs, annotated based on NCBI protein database, is presented in Additional file 4.

Nine of the novel genes predicted to provide resistance to aminoglycosides, beta-lactams, rifampicin, chloramphenicol, trimethoprim, and sulfonamides were tested and functionally confirmed by expressing them in *Escherichia coli* (Additional file 1: Table S1). These genes and their contexts are presented in Fig. 2 and Additional file 1: Figure S4. Phylogenetic analysis of all putative OXA-variants showed that their closest relatives were OXA-2, OXA-10, and OXA-46 (Additional file 1: Figure S5). The identified OXA-10 variants did not contain the N143S and G157D substitutions, which are associated with ceftazidime resistance and extended spectrum beta-lactamase resistance (ESBL) characteristics [28]. When expressed in *E. coli*, all provided resistance to ampicillin, whereas the synthesized OXA-10-like gene also conferred resistance to cefotaxime. All the tested OXA variants conferred reduced susceptibility to carbapenems corresponding to a 21- to 63-fold increase in the minimal inhibitory concentration (MIC) for ertapenem and around a two- to three- fold increase for imipenem.

A putative sulfonamide resistance gene with 69% amino acid identity to the closest known dihydropteroate synthase (DHPS) and between 31 and 33% identity to known mobile sulfonamide resistance genes (Table 4) was identified. Experimental validation showed that this gene conferred full resistance to sulfamethoxazole with an MIC > 1024 μg/ml (more than 256-fold increase compared to control). Based on its ability to provide sulfonamide resistance, its mobile character, as demonstrated by its presence in integrons, and the homology to previously known sulfonamide resistance genes, we proposed the name *sul4*(GenBank: MG649393). To date, only three different mobile sulfonamide resistance genes have been identified, whereas for most of the other classes of antibiotics, many more resistance genes are known (e.g., for beta-lactams, aminoglycosides, tetracyclines, and trimethoprim). Since a new mobile sulfonamide resistance gene is a significant observation, we focused on this gene for further characterization.

Fourteen LRs in river sediment collected in Pune (RS_{pune}) and 48 LRs in river sediment collected near PETL (RS_{PETL}) contained *sul4*. These represent two different cassette arrangements. In the first case, *sul4* was the first cassette after the *attI* site, followed by the complete *qacE* gene, and found only in RS_{PETL} (Fig. 2a). In the second case, found in both samples, *sul4* was the second gene cassette following a hypothetical protein, which then was followed by a transposase from the insertion sequence common regions (IS*CR*) family (Fig. 2b). These two arrangements were further

Table 2 Comparison with previous PCR amplicon studies of gene cassettes

Study	Sample type	Location	No. of unique ORFs
[43]	Marine sediment	Halifax Nova Scotia, Canada	1372
[44]	Sludge	Sydney Tar Ponds, Nova Scotia, Canada	708
[70]	Sediment	Minas Gerais state, Brazil	143
[42]	Marine sediment	Suez and Tokyo Bay	146 + 68
Current study	Sediment	Pune and Patancheru, India	19,723

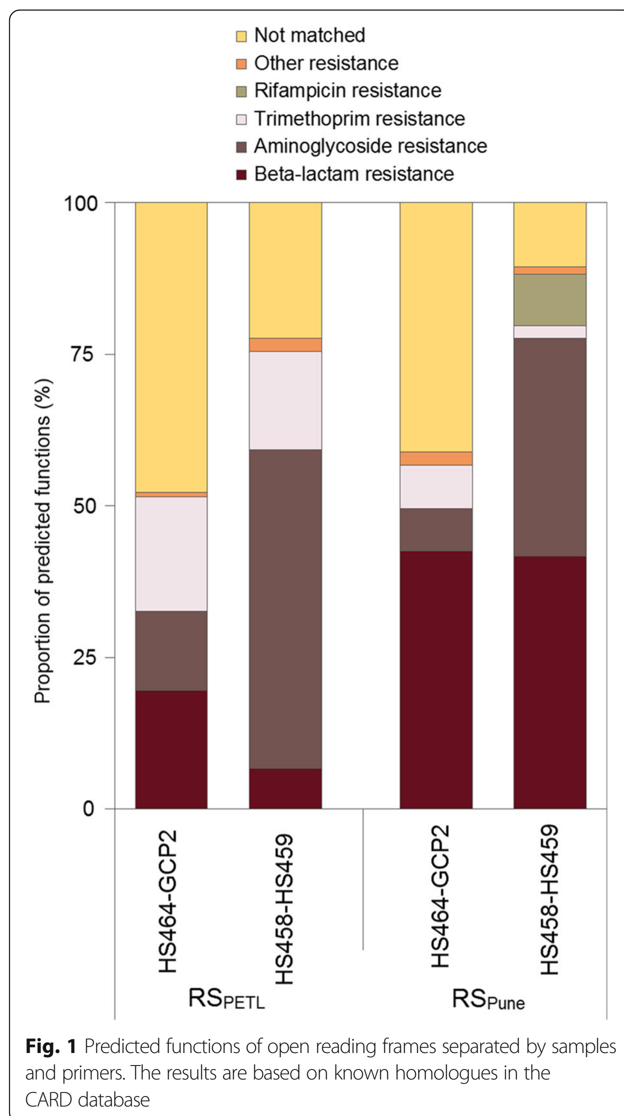

Fig. 1 Predicted functions of open reading frames separated by samples and primers. The results are based on known homologues in the CARD database

Table 3 Number of known and putative novel ARGs to different families of antibiotics

Antibiotic family	Number of known ARG	Number of putative novel ARGs
Aminoglycoside	32	689
Beta-lactam	44	103
Trimethoprim	8	240
Rifampicin	3	2
Chloramphenicol	8	4
Macrolide	3	7
Sulfonamide	2	1
Quinolone	1	0

confirmed by performing PCR on the original unamplified samples using primer pairs targeting *sul4* and *qacE*/IS*CR20* (Fig. 2a, b), both of which generated amplicons of the expected size.

Screening 6489 publicly available metagenomic datasets revealed the presence of *sul4* in seven different countries across Asia and Europe (Table 5). The detailed descriptions of these datasets along with normalized counts of sulfonamide resistance genes are presented in Additional file 5. Figure 2c–h shows the recovered contigs containing *sul4* from some of these datasets. The full list of datasets examined is presented in Additional file 6.

The collapsed phylogenetic tree of Sul4 and 8875 different dihydropteroate synthase enzymes (protein similarity less than 95%) retrieved from the NCBI protein database are presented in Fig. 3. The Sul4 protein was located in a clade with DHPS proteins from members of the phylum Chloroflexi, found in various environments, including wastewater [29].

A structural prediction of Sul4 and the other three sulfonamide resistance proteins is presented in Additional file 1: Figure S6. All the proteins are structurally similar due to the preservation of the distorted cylinder in the center, with α-helices around the inner β-strands and coils. Additional file 1: Figure S7 also presents the alignment and the comparison of the secondary structures between Sul1, Sul2, Sul3, and Sul4 and a sensitive DHPS (with protein data bank (PDB) ID: 1AJ0) [30].

Discussion

We used a targeted PCR of integrons from polluted environmental samples followed by amplicon sequencing using next generation sequencing technologies to greatly extend our knowledge of mobile antibiotic resistance genes found as gene cassettes. Combining the accuracy of short reads from Illumina sequencing with the higher resolution of long reads from PacBio sequencing generated a clear view of the gene cassettes and their immediate context, also providing clues about their evolutionary history. We therefore suggest that a similar approach could be applied to search for ARGs as well as other functional genes in different contexts.

Putative novel ARGs for aminoglycosides, beta-lactams, trimethoprim, rifampicin, chloramphenicol, and sulfonamides were identified. Several novel OXA variants provided reduced susceptibility to carbapenems, providing an additional battery of integron-borne genes that could contribute to resistance against last-line antibiotics. Many of the known and putative ARGs that were found were previously not reported as gene cassettes, thus revealing a potential to be spread via integrons.

A mobile sulfonamide resistance gene with only 31–33% identity to previously known mobile sulfonamide resistance genes was discovered, providing a very high level of

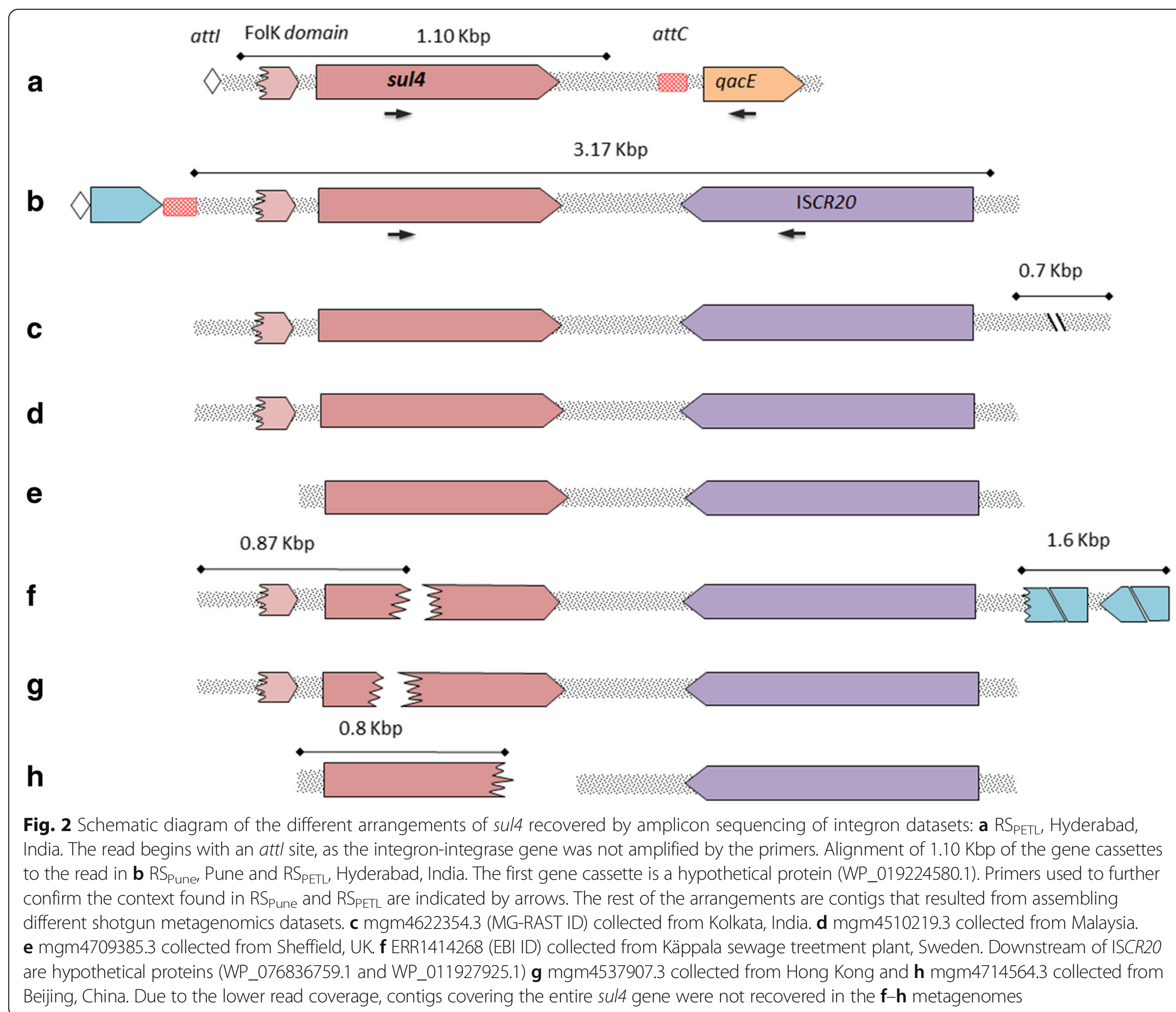

Fig. 2 Schematic diagram of the different arrangements of *sul4* recovered by amplicon sequencing of integron datasets: **a** RS_{PETL}, Hyderabad, India. The read begins with an *attI* site, as the integron-integrase gene was not amplified by the primers. Alignment of 1.10 Kbp of the gene cassettes to the read in **b** RS_{Pune}, Pune and RS_{PETL}, Hyderabad, India. The first gene cassette is a hypothetical protein (WP_019224580.1). Primers used to further confirm the context found in RS_{Pune} and RS_{PETL} are indicated by arrows. The rest of the arrangements are contigs that resulted from assembling different shotgun metagenomics datasets. **c** mgm4622354.3 (MG-RAST ID) collected from Kolkata, India. **d** mgm4510219.3 collected from Malaysia. **e** mgm4709385.3 collected from Sheffield, UK. **f** ERR1414268 (EBI ID) collected from Käppala sewage treetment plant, Sweden. Downstream of IS*CR20* are hypothetical proteins (WP_076836759.1 and WP_011927925.1) **g** mgm4537907.3 collected from Hong Kong and **h** mgm4714564.3 collected from Beijing, China. Due to the lower read coverage, contigs covering the entire *sul4* gene were not recovered in the **f**–**h** metagenomes

resistance when expressed in *E. coli*. Only three mobile sulfonamide resistance genes (*sul1*, *sul2*, and *sul3*) have previously been identified. The *sul4* gene was retrieved by amplifying gene cassettes using class 1 integron-specific primers, the most common type of integron found in human pathogens. The gene cassette contains an ORF for the *sul4* gene, and a partial domain of FolK (COG0801) that is present in the upstream region of *sul4*, similar to the fused *folKP* gene in chlamydia. This domain can be found in dihydroneopterin aldolase (i.e., FolK/SulD) which is involved in the folate biosynthesis pathway. The enzyme encoded by *folK* harbors the activity of EC 4.1.2.25 and produces 6-hydroxymethyl-7,8-dihydropterin diphosphate, which later is used by the dihydropteroate synthase (EC 2.5.1.15) [31]. Downstream of *sul4*, a transposase (IS*CR20*) belonging to the IS*CR* family is identified in one of the two contexts. This family of insertion sequences lacks inverted repeats (IR) and, without the need of another

Table 4 Similarity of the Sul4 protein to the closest known dihydropteroate synthase and the three previously known mobile sulfonamide resistance proteins

Accession in NCBI	Identity (%)	Coverage (%)	Description	Year of discovery
CUS02277.2	69	99	Dihydropteroate synthase [Ardenticatena]	2016
AEJ33969.1	32	92	Sul1	1975
AAL59753.1	33	80	Sul2	1988
ACJ63260.1	31	93	Sul3	2003

Table 5 Summary of 108 shotgun metagenomics datasets containing reads which are mapped to Sul4 with 100% identity, covering more than 20 amino acids (see Additional file 5)

Country	No. dataset	No. mapped reads	Biome	Collection date
India	38	10,803	Aquatic	2008–2016
China	21	249	Air,terrestrial,aquatic	2011–2016
Hong Kong	1	34	Aquatic (sewage)	2013
Malaysia	1	15	Mangrove	2011
Sweden	32	422	Aquatic (sewage)	2012
UK	9	290	Aquatic (sewage)	2016
Switzerland	6	46	Aquatic (sewage)	2014

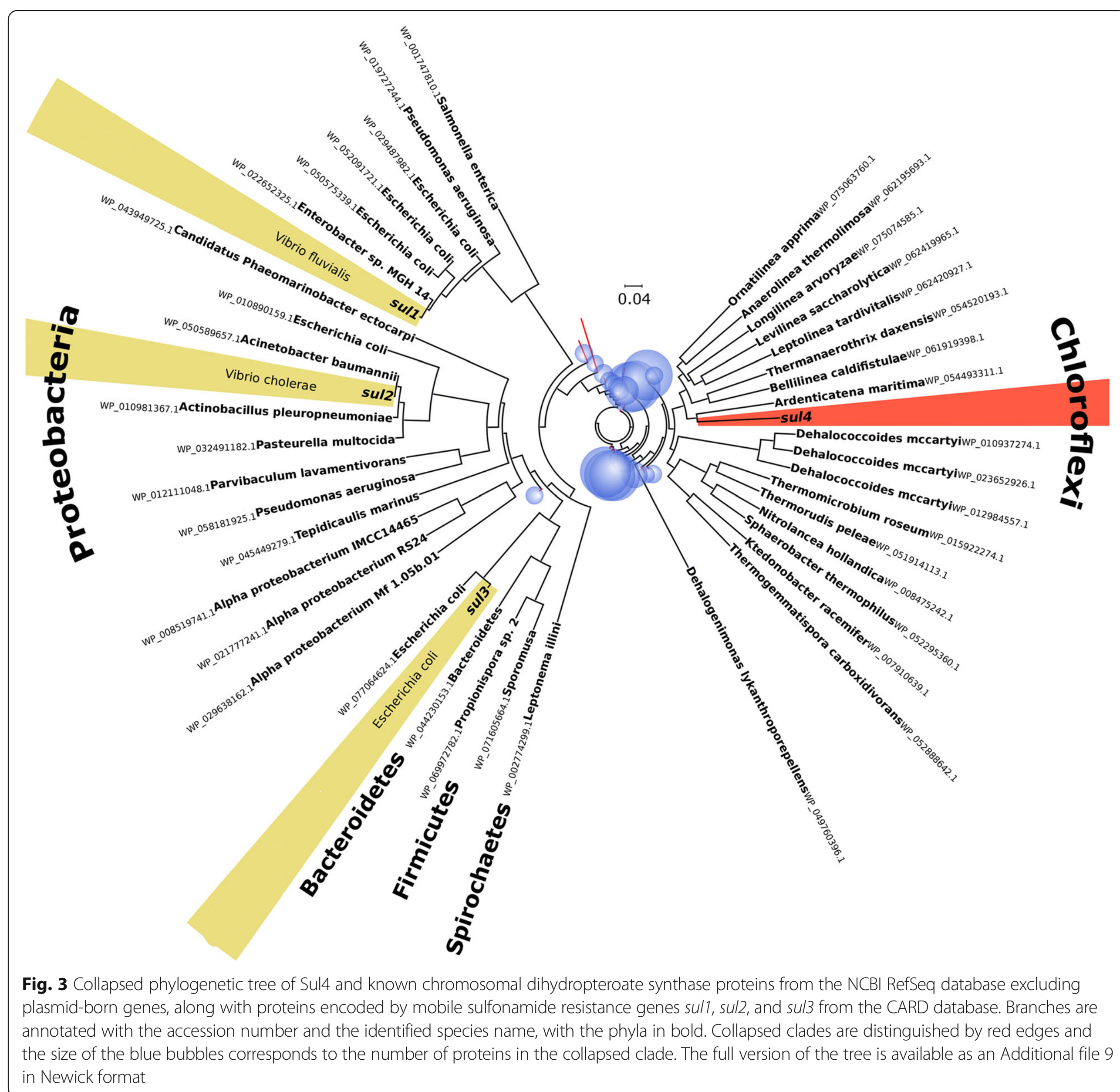

Fig. 3 Collapsed phylogenetic tree of Sul4 and known chromosomal dihydropteroate synthase proteins from the NCBI RefSeq database excluding plasmid-born genes, along with proteins encoded by mobile sulfonamide resistance genes *sul1*, *sul2*, and *sul3* from the CARD database. Branches are annotated with the accession number and the identified species name, with the phyla in bold. Collapsed clades are distinguished by red edges and the size of the blue bubbles corresponds to the number of proteins in the collapsed clade. The full version of the tree is available as an Additional file 9 in Newick format

transposase protein, they can be mobilized along with their adjacent DNA sequence through rolling-circle (RC) transposition [32, 33]. We did not find any other ARGs adjacent to IS*CR20* in the studied samples; however, IS*CR*s, such as IS*CR1*, have been found adjacent to the 3′ conserved segment (3′-CS) of integrons, and with the loss of their *terIS* sites, they can mobilize the entire integron. Moreover, sulfonamide resistance genes are also known to be carried by IS*CR* elements (e.g., association of *sul2* and IS*CR2*, GenBank: KX900483.1) and are reported in complex integrons (e.g., *sul1* in GenBank:AY079169.1). The gene encoding the IS*CR20*-like protein, found in the integron near *sul4*, has been reported earlier to be adjacent to *sul2* in the *Bibersteinia trehalosi* genome (GenBank:CP006956.1$_{(295,195..297771)}$) and as a complex integron in *Enterobacteriaceae* isolates (GenBank:DQ520941.1$_{(1873..3163)}$). Taken together, these findings provide strong support that *sul4* has been decontextualized from the chromosome of its original host. The IS*CR* could potentially provide mobility to the entire integron.

Structural prediction of Sul4 indicates strong overall similarities to Sul1, Sul2, and Sul3. The α/β barrel structure contains the binding sites for 7,8-dihydropterin pyrophosphate (DHPP), para-aminobenzoic acid (pABA), and sulfonamide. After DHPP has bound deep in the cylinder, sulfonamide binds near the surface of the protein. Thus, sulfonamide binding is affected by changes near the surface (e.g., insertion of amino acid in coils after amino acid 190) of DHPS [34]. Most of the α-helices in Sul4 are preserved, but the coils and β-strands have changed considerably from sensitive DHPS, which possibly contributes to reducing the affinity of sulfonamide and the Sul4-DHPP complex structure.

Although extensively used since 1935, the use of sulfonamides in human medicine has become mainly limited to treating gastrointestinal or urinary infections. However, sulfonamides are still broadly used in animals for treatment, growth promotion, and prophylactic purposes. There is a lack of reliable records for the global usage of sulfonamides in animals. Data covering 10 European countries show that sulfonamides and trimethoprim constitute 17% of the sales of veterinary antibacterial agents [35], and in the US, 380,186 kg of sulfonamides was distributed legally during 2015 for food-producing animals [36]. High concentrations of sulfonamide residues in animal manure in China indirectly indicate heavy usage [37, 38]. Hence, further spread of sulfonamide resistance would have severe consequences, particularly for the animal sector.

Fourteen years has passed since the discovery of the third mobile sulfonamide resistance gene. The fact that so few genes have been detected, despite almost 80 years of intense usage of sulfonamides, is intriguing, as there are considerably more types of mobile genes for tetracycline resistance, beta-lactamases, or aminoglycoside acetyltransferases [39]. Our finding of a fourth mobile sulfonamide resistance gene indicates that there are still ongoing forces that introduce, mobilize, and maintain new sulfonamide resistance genes in bacterial communities. We do not yet know the present host-range of the *sul4* gene, nor its context outside of integrons. Our results show, however, that *sul4* provides high-level resistance in *Escherichia coli*. This finding suggests that *sul4* can provide clinical resistance in *Enterobacteriaceae*, similar to the previously discovered sulfonamide resistance genes. Because of founder effects [40], one may question how effectively *sul*4 might spread. It might be that the founder effect is not critical, as is apparent from the spread of beta-lactamases and other types of resistance genes. Moreover, the presence of *sul4* in different samples from different continents suggests that the gene has found a way to spread successfully.

The *sul4* gene was abundant both at the PETL and Pune sites (Fig. 2b). The recovered contigs from Sweden and Kolkata suggest that IS*CR20* has had a role in mobilizing *sul4* and its flanking regions, probably via rolling-circle transposition. In Sweden, the partially recovered *sul4* and the IS*CR20* were located upstream from two hypothetical proteins. It seems that the IS*CR20* has truncated one of the hypothetical proteins, as we could not find the full length ORF (Fig. 2f). In the Kolkata samples, the *sul4* and IS*CR20* were adjacent to an unknown sequence with no detectable ORF. These downstream sequences, which do not appear to follow the structure of a classical integron, suggest the insertion of *sul4* and flanking regions in different regions of the bacterial host genomes. In China, *sul4* was found in Beijing smog in three different samples, and highlights the possible role of aerial transport of this ARG. Unfortunately, these datasets are not sequenced deep enough to assemble the reads and investigate the context of *sul4*.

An association of *sul4* with the phylum Chloroflexi, as suggested by phylogenetic analysis, is further supported by the high abundance of *sul4* in aquatic metagenomes from an algal bloom in Kolkata. No reads of this datasets were mapped to other mobile sulfonamide resistance genes, which are typically markers of anthropogenic pollution. Studies have shown that the phylum Chloroflexi is one of the dominant bacterial phyla in these aquatic ecosystems [41]. We believe that further investigations on Chloroflexi could provide clues about the original host of *sul4* and how it has been decontextualized.

Amplifying integrons from polluted river sediment resulted in identification of a large range of gene cassettes, the majority of which were known or putative ARGs. To our knowledge, such high diversity of ARGs in integron gene cassettes has not been described previously in any

bacterial community [42–44]. Prior selection by antibiotics is the most plausible explanation behind the selection of bacteria with such cassettes, either in the actual sediment and/or in the gut microbiota of humans that contribute fecal residues to the sediment. Selection by antibiotics is likely an important factor in the initial mobilization of such genes, enabling them to shift from a functional role in general metabolism to become mobile resistance genes. The high abundance of resistance gene cassettes both in environmental and clinical integrons indicates an extensive exchange of gene cassettes between them. Close interactions between different types of integrons could facilitate the accumulation of novel resistance determinants and virulence factors into clinical integrons. Moreover, the finding of DNA from human fecal bacteria together with a high abundance of integrons at both sampling sites [20, 24] further highlights the opportunity for such interactions potentially allowing a gene flow of novel resistance determinants to pathogens. Therefore, these results provide part of the necessary ecological connectivity that could contribute to increased resistance in clinics [14, 45].

Conclusions

A targeted amplicon sequencing approach was used to greatly extend our knowledge of integron-born gene cassettes, particularly those with antibiotic resistance function. Combining the accuracy of short reads with the higher resolution of long reads generated a clear view of the gene cassettes and their immediate context, providing some clues about their evolutionary history. A range of novel resistance gene cassettes against different families of antibiotics were identified, including the fourth mobile sulfonamide resistance ever found, namely, *sul4*.

Methods

Sample collection

Sediments from the Mutha River (RS_{Pune}) were collected from within the city of Pune in Maharashtra, India, (referred as Pune river or RS_{Pune}) and pooled into one composite sample (for details, see Additional file 1: Table S2). Pune is the second largest city in the state of Maharashtra, and the river passing through the city is heavily contaminated by untreated sewage [24]. Sediment samples were also collected from the Isakavagu/Nakkavagu River, which flows past an industrial waste water treatment plant (Patancheru Enviro Tech Ltd.; PETL) near Hyderabad, India. The PETL samples, described previously [20], were pooled and are referred to here as RS_{PETL}. The treated waste water and river sediments were contaminated with exceptional levels of fluoroquinolone antibiotics (up to 31 mg/L and up to 0.9 g/kg organic material, respectively) and harbor bacterial communities with very a high abundance of resistance genes and integrons [15, 19, 20, 46, 47].

DNA extraction, PCR, and sequencing

Total genomic DNA was extracted from individual frozen sediment samples using the PowerSoil® DNA isolation kit (MoBio, Carlsbad, CA) according to the manufacturer's instructions (note that unamplified DNA was used, in contrast to repliG-amplification as in [20]). Concentration of the extracted DNA was determined using a dsDNA High Sensitivity (HS) Assay kit on the Qubit® Fluorometer (Invitrogen, USA). The subsamples were pooled, and DNA from each sample was amplified using three sets of previously used primer pairs (Fig. 4) [42, 48, 49]. All PCR reactions were carried out using phusion high-fidelity DNA polymerase (Thermo Scientific, USA). The primers HS458-HS459 amplify entire gene cassette arrays by binding to the 5′ and 3′ conserved segments of clinical class 1 integrons. The primers HS464-GCP2 target the class 1 integrase gene and a conserved region of the *attC* recombination site. The primers MRG284-MRG285 amplify the entire gene cassette array from the *attI* site to a conserved region beyond the cassette array in the pre-clinical class 1 integrons; PCR products were purified using a PCR purification kit (Qiagen, Germany) and quantified using the Qubit® Fluorometer. Amplicons were then sent for single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences) to LRs and shotgun metagenomic sequencing to produce SRs (paired-end 250 bp reads on the Illumina Mi-Seq2000 platform) at Science for Life Laboratories (Uppsala and Stockholm, Sweden). The library construction of LRs was carried out using the SMRTbell Template Prep Kit 1.0 (part number: 100-259-100). The SMRT-bell libraries were sequenced on a PacBio RSII platform with P5-C3 chemistry using two SMRT cells. The metagenomic sequencing data and the corresponding meta-data have been deposited in the NCBI database under the Bio-Project ID: PRJNA400874.

Sequence analysis

The quality of the SR dataset was assessed using FastQC [50]. Reads with low-quality bases were trimmed to reach a score of 21, and those with less than 80 bases in length were filtered using high-throughput quality control (HTQC) [51]. If only one end of the paired-end reads had acceptable quality, we used it as a single read. The resulting paired and single reads were used to correct LRs with Proovread as a hybrid correction pipeline for single-molecule real-time sequencing [25]. Proovread maps SRs to LRs using sequence alignment and then, with the generated short-read consensus, corrects errors in the LRs. Proovread also calculates updated position-

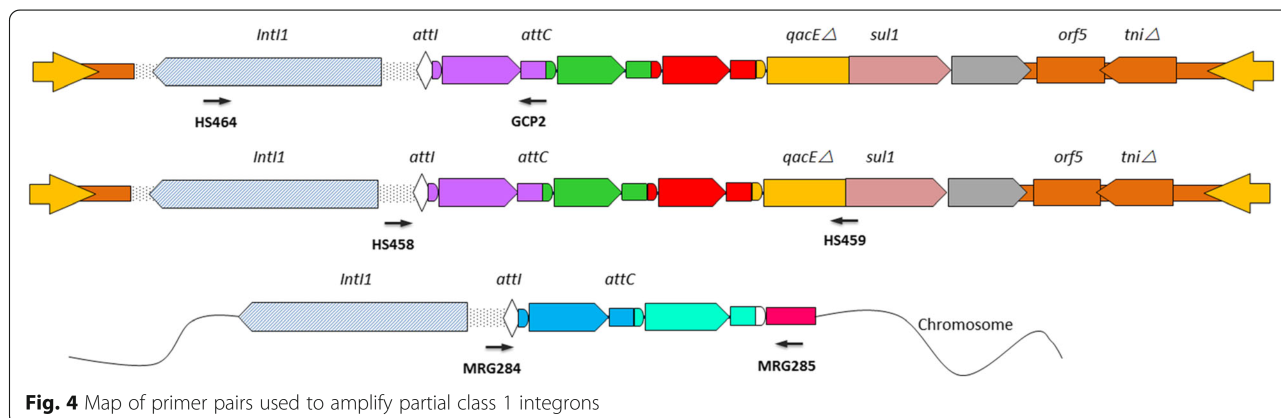

Fig. 4 Map of primer pairs used to amplify partial class 1 integrons

specific quality scores based on the coverage and composition of the consensus.

Redundant LRs were identified by clustering them using CD-HIT (with following parameters: -c 1 -uS 0.05 -S 5 -n 8 -d 0 -r 1) [52]. Blastn in the BLAST+ package (with following parameters: mode blastn-short, word_size 7, gapOpen 5, gapextent 2, reward 1, penalty -3) was employed to find the primers in the 5′ and 3′ ends of the LRs [53]. The LRs were annotated as follows. First, the ORFs were predicted using Prodigal (-p meta) [26]. The functions of the ORFs were identified through similarity searches against the non-redundant nucleotide and protein NCBI databases (update at 07 November 2016). BLAST+ in blastn mode was used for the nucleotide alignments while Diamond was used for the protein alignments. ORFs which were not annotated as integrase or with related terms (i.e., *IntI*, *IntI*1 integrase, XerD domain) or *qacEΔ* or related terms (i.e., qacE delta, partial quaternary ammonium compound resistance protein, partial ethidium bromide resistance protein) and were longer than 75 amino acids were considered putative integron gene cassettes. The LRs amplified by HS464-GCP2 primers were further divided into "clinical" and "environmental" integrons [54]. Blastn in the BLAST+ package with a sequence identity of 100 was used to identify a previously widespread clinical integrase (NCBI accession ID: KC417379$_{(1..1014)}$) from downstream of the HS464 primer sequence. The rest of the LRs were classified as environmental integrons.

Putative novel resistance genes were identified based on their sequence identity and the length of the alignment (coverage) to known homologues in CARD (version 1.1.0) [55] and the NCBI database. We classified ORFs with at least 95% identity to closest homologs in the CARD database as "known resistance genes" and those with identity between 60 and 95% and with coverage greater than 65% as "putative novel resistance genes." The gene cassettes with known function were clustered to remove redundancy using CD-HIT. HattCI was used to identify *attC* sites in the LR [56].

The abundance of mobile sulfonamide resistance genes (*sul1*, *sul2*, *sul3*, and *sul4*) was quantified in 6489 metagenomic datasets as follows. First, shotgun metagenome datasets were collected from the MG-RAST database [57] (sequence type: shotgun metagenome) and our local database from previous studies. The reads were mapped to the Sul proteins using Usearch (with following parameters: -search_global -id 1 -maxaccepts 0 -maxrejects 0) [58], and the best hits with higher sequence identity and longer alignment length were selected. Datasets containing more than five reads mapped to Sul4 were analysed with Metaxa 2.1 [59] to extract the number of bacterial 16S rRNA sequences (SSU). The count data were normalized as was done in a previous publication [15]. To identify the context of *sul4*, the reads in the selected metagenomic datasets were filtered and trimmed using HTQC and assembled using Megahit 1.1.1 [60].

The phylogenetic analysis on the *sul4* and OXA variant gene cassettes was done as follows. Chromosomal proteins annotated with the term "dihydropteroate synthase" were retrieved from the NCBI RefSeq database. All 18,822 proteins along with the Sul1, Sul2, and Sul3 proteins from the CARD database were clustered using CD-hit (with parameters -c 0.95) to remove redundancy. Beta-lactamases classified as OXA were also retrieved from the CARD database. Multiple alignments were done using MAFFT (–auto) [61], which brought efficiency to the pipeline with its parallelism implementation and efficient memory utilization. Then, phylogenetic trees were produced by quicktree [62] using the neighbor-joining algorithm. The Python package ETE3 was used to draw and collapse the phylogenetic trees [63] to better visualize the relationship between the mobile sulfonamide resistance genes and their closest relatives.

To identify genes previously not described in integrons, LRs were searched against the INTEGRALL database [27]. All the accession numbers ($n = 8471$ November

2016) in INTEGRALL were retrieved, and their sequences were downloaded from the NCBI GenBank database. The collected sequences were utilized as a reference database for nucleotide comparison between the LRs and recorded ORFs, using blastn in the BLAST+ package. Novel ORFs in the integrons were identified based on the sequence identity and the length of the alignment. We classified hits with an identity greater than 95% and coverage greater than 70% as previously reported integron-associated ORFs.

Functional verification of candidate novel resistance genes

Putative novel resistance genes were grouped according to the classes of antibiotics against which they were likely to confer resistance. Nine candidate novel genes/gene variants with high correction scores and low identity to the closest known resistance gene in each class were selected for functional verification. The candidate novel genes were synthesized at ThermoFisher Scientific, Germany, using their GeneArt Gene Synthesis service and subcloned into the expression vector pZE21-MCS1 using *Kpn*1 and *Bam*H1, as described previously [64]. The recombinant plasmids containing novel resistance gene candidates were then transformed into *E. coli* C600Z1 (Expressys, Germany) by electroporation. The MICs of the antibiotics for the strains containing the candidate novel resistance genes were determined using E-tests on Mueller-Hinton Agar plates (BioMérieux, France) with the addition of 100 ng/μl anhydrotetracycline (aTC), which acts as an expression inducer for the pZE21-MCS1 gene inserts [65]. The strain containing empty vector was used as a negative control. The protein sequences of the synthesized genes are presented in Additional file 7.

Additional files

Additional file 1: Figure S1. Predicted functions of open reading frames recovered by the chromosomal integron primer pair MRG284-MRG285 separated by samples. The results are based on known homologues in the CARD database. **Figure S2.** Predicted functions of open reading frames of the "clinical" and "environmental" integrons from the HS464-GCP2 amplicons separated by samples. The results are based on known homologues in the CARD database. **Figure S3.** Functional annotation of the open reading frames not previously reported in integrons. The results are based on known homologues in the NCBI protein database. Putative resistance genes are determined based on annotation in the NCBI database. **Figure S4.** Genetic arrangements of functionally verified resistance gene cassettes as identified by PCR amplification of the integrons. The synthesized gene cassettes are distinguished by thicker borders. Both synthesized OXA-2-like gene cassettes have the same arrangement. **Figure S5.** Collapsed phylogenetic tree of the identified OXA-variant gene cassettes and 289 known OXA-variants retrieved from the CARD database. The identified genes are described by Id numbers and located adjacent to OXA-10, OXA-2 and OXA-46 clades, which are highlighted in the tree. The collapsed clades are based on [28, 66] and distinguished by red edges, and the size of the bubbles correspond to the number of proteins in the collapsed clade. The full version of the tree is available in Additional file 8 in Newick format. **Figure S6.** Prediction of the tertiary structures of sulfonamide resistance proteins using I-TASSER server [67]. Color spectrum, from blue to red, corresponds to the detected secondary structure of the proteins based on the order of the amino acids. C-score scales the confidence of each predicted structure between −5 to 2. Sul1: c-score = 0.86, Sul2: c-score = 1.20, Sul3: c-score = 1.25, Sul4: c-score = 1.07. **Figure S7.** Sequence Alignments of sulfonamide resistance proteins and a sensitive DHPS with a crystal structure stored in the Protein Data Bank (PDB). The alignment was performed in UCSF Chimera [68] using the Muscle algorithm [69]. α-Helixes and β-strands are marked with yellow and green colors, respectively. α-Helixes are more preserved than the β-strands and coils. **Table S1.** Functional verification of the synthesized putative novel resistance genes. **Table S2.** Sampling site coordinates for RS_{Pune}. (DOCX 1985 kb)

Additional file 2: List of known ARGs, categorized by different families of antibiotics, identified as gene cassettes in both samples. (XLSX 20 kb)

Additional file 3: List of putative novel ARGs. (XLSX 110 kb)

Additional file 4: List of previously not reported gene cassettes. (XLSX 325 kb)

Additional file 5: Relative abundance of mobile sulfonamide resistance genes (*sul1–4*) in metagenomic samples containing *sul4*. (XLSX 28 kb)

Additional file 6: List of metagenomic samples searched for *sul4* genes. (XLSX 376 kb)

Additional file 7: Protein sequences of the synthesized genes in fasta format. (TXT 1 kb)

Additional file 8: Phylogenetic tree of the identified OXA-variant gene cassettes and 289 known OXA-variants retrieved from the CARD database. (TXT 9 kb)

Additional file 9: Phylogenetic tree of DHPS proteins encoded by chromosomal genes and mobile sulfonamide resistance genes. (TXT 283 kb)

Abbreviations

ARG: Antibiotic resistance gene; DHPP: 7,8-Dihydropterin pyrophosphate; DHPS: Dihydropteroate synthase enzyme; ESBL: Extended spectrum beta-lactamase resistance; IR: Inverted repeat; ISCR: Insertion sequence common region; LR: Long read; MIC: Minimal inhibitory concentration; ORF: Open reading frame; pABA: Para-aminoBenzoic acid; PETL: Patancheru Enviro Tech Ltd., an industrial waste water treatment plant; RC: Rolling circle transposition; RS_{PETL}: River sediment collected near PETL; RS_{Pune}: River sediment collected in Pune; SMRT: Single-molecule real-time; SR: Short read

Acknowledgements

Not applicable.

Funding

This work was funded by the Swedish Research Council for Environment, Agriculture and Spatial Planning (FORMAS), the Swedish Research Council (VR), and the Centre for Antibiotic Resistance Research (CARe) at the University of Gothenburg to DGJL.

Availability of data and materials

The raw sequencing data of the sediment samples have been deposited in the NCBI Sequence Read Archive (SRA) under the bio-project PRJNA400874. Long reads that contian synthesized genes were deposited in GenBank with the following accession numbers: MG649394, MG649395, MG649396, MG649397, MG649398, MG649399, MG649400, MG649401, MG649402 and MG649403.

Authors' contributions

MR preformed the bioinformatic analyses and drafted the manuscript; NPM collected and prepared samples for sequencing, did functional verification of novel genes, and edited the manuscript; MRG contributed to conceiving the study and edited the manuscript; CF gave advice to the design of the study and edited the manuscript; EK assisted with bioinformatic analyses and edited the manuscript; DGJL conceived and designed the study and provided significant input in the writing of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
No ethical approval is needed/applicable nor is consent from any participant, since the study did not involve sampling from humans or animals.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

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

Author details
[1]Centre for Antibiotic Resistance Research (CARe) at University of Gothenburg, Gothenburg, Sweden. [2]Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. [3]Department of Biological Sciences, Genes to Geoscience Research Centre, Macquarie University, Sydney, New South Wales, Australia. [4]Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden.

Received: 26 September 2017 Accepted: 29 November 2017
Published online: 15 December 2017

References
1. Finley RL, Collignon P, Larsson DGJ, McEwen SA, Li X-Z, Gaze WH, Reid-Smith R, Timinouni M, Graham DW, Topp E. The scourge of antibiotic resistance: the important role of the environment. Clin Infect Dis. 2013; 57:704–10.
2. Gaze WH, Krone SM, Larsson DGJ, Li X-Z, Robinson JA, Simonet P, Smalla K, Timinouni M, Topp E, Wellington EM. Influence of humans on evolution and mobilization of environmental antibiotic resistome. Emerg Infect Dis. 2013; 19:e120871.
3. Gillings MR. Integrons: past, present, and future. Microbiol Mol Biol Rev. 2014;78:257–77.
4. Escudero JA, Loot C, Nivina A, Mazel D. The integron: adaptation on demand. Microbiol Spectr. 2015;3:MDNA3-0019-2014.
5. Cury J, Jové T, Touchon M, Néron B, Rocha EP. Identification and analysis of integrons and cassette arrays in bacterial genomes. Nucleic Acids Res. 2016; 44:4539–50.
6. Rao AN, Barlow M, Clark LA, Boring J, Tenover FC, McGowan J. Class 1 integrons in resistant Escherichia coli and Klebsiella spp., US hospitals. Emerg Infect Dis. 2006;12:1011–4.
7. van Essen-Zandbergen A, Smith H, Veldman K, Mevius D. Occurrence and characteristics of class 1, 2 and 3 integrons in Escherichia coli, Salmonella and Campylobacter spp. in the Netherlands. J Antimicrob Chemother. 2007; 59:746–50.
8. Marchant M, Vinué L, Torres C, Moreno MA. Change of integrons over time in Escherichia coli isolates recovered from healthy pigs and chickens. Vet Microbiol. 2013;163:124–32.
9. Hardwick SA, Stokes H, Findlay S, Taylor M, Gillings MR. Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. FEMS Microbiol Lett. 2008;278:207–12.
10. Gillings MR. Class 1 integrons as invasive species. Curr Opin Microbiol. 2017;38:10–5.
11. Stalder T, Barraud O, Casellas M, Dagot C, Ploy M-C. Integron involvement in environmental spread of antibiotic resistance. Front Microbiol. 2012;3:119.
12. Partridge SR, Tsafnat G, Coiera E, Iredell JR. Gene cassettes and cassette arrays in mobile resistance integrons. FEMS Microbiol Rev. 2009;33:757–84.
13. Martínez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol. 2015;13:116–23.
14. Bengtsson-Palme J, Larsson DGJ. Antibiotic resistance genes in the environment: prioritizing risks. Nat Rev Microbiol. 2015;13:396.
15. Pal C, Bengtsson-Palme J, Kristiansson E, Larsson DGJ. The structure and diversity of human, animal and environmental resistomes. Microbiome. 2016;4:54.
16. Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E, Larsson DGJ. Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. Front Microbiol. 2014;5:648.
17. Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. Functional metagenomics reveals diverse β-lactamases in a remote Alaskan soil. ISME J. 2009;3:243–51.
18. Schmieder R, Edwards R. Insights into antibiotic resistance through metagenomic approaches. Future Microbiol. 2012;7:73–89.
19. Marathe NP, Regina VR, Walujkar SA, Charan SS, Moore ER, Larsson DGJ, Shouche YS. A treatment plant receiving waste water from multiple bulk drug manufacturers is a reservoir for highly multi-drug resistant integron-bearing bacteria. PLoS One. 2013;8:e77310.
20. Kristiansson E, Fick J, Janzon A, Grabic R, Rutgersson C, Weijdegård B, Söderström H, Larsson DGJ. Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS One. 2011;6:e17038.
21. Group GARP-IW. Rationalizing antibiotic use to limit antibiotic resistance in India. Indian J Med Res. 2011;134:281–94.
22. Khajuria A, Praharaj AK, Kumar M, Grover N. Carbapenem resistance among Enterobacter species in a tertiary care hospital in central India. Indian society of Crit Care Med 2014;18:750-753.
23. Alagesan M, Gopalakrishnan R, Panchatcharam SN, Dorairajan S, Ananth TM, Venkatasubramanian R. A decade of change in susceptibility patterns of Gram-negative blood culture isolates: a single center study. Germs. 2015;5:65.
24. Marathe NP, Chandan P, Gaikwad SS, Jonsson V, Kristiansson E, Larsson DGJ. Untreated urban waste contaminates Indian river sediments with resistance genes to last resort antibiotics. Water Res. 2017;124:388–97.
25. Hackl T, Hedrich R, Schultz J, Förster F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 2014;30:3004–11.
26. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:1.
27. Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics. 2009;25:1096–8.
28. Evans BA, Amyes SG. OXA β-lactamases. Clin Microbiol Rev. 2014;27:241–63.
29. Björnsson L, Hugenholtz P, Tyson GW, Blackall LL. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removalc. Microbiology. 2002;148:2309–18.
30. Achari A, Champness JN, Bryant PK, Rosemond J, Stammers DK. Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. Nat Struct Mol Biol. 1997;4:490–7.
31. Rébeillé F, Macherel D, Mouillon JM, Garin J, Douce R. Folate biosynthesis in higher plants: purification and molecular cloning of a bifunctional 6-hydroxymethyl-7, 8-dihydropterin pyrophosphokinase/7, 8-dihydropteroate synthase localized in mitochondria. EMBO J. 1997;16:947–57.
32. Toleman MA, Bennett PM, Walsh TR. ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev. 2006;70:296–316.
33. Harmer CJ, Hall RM. IS26-mediated formation of transposons carrying antibiotic resistance genes. mSphere. 2016;1:e00038–16.
34. Levy C, Minnis D, Derrick JP. Dihydropteroate synthase from Streptococcus pneumoniae: structure, ligand recognition and mechanism of sulfonamide resistance. Biochem J. 2008;412:379–88.
35. Grave K, Torren-Edo J, Mackay D. Comparison of the sales of veterinary antibacterial agents between 10 European countries. J Antimicrob Chemother. 2010;65:2037–40.
36. FDA: Summary report on antimicrobials sold or distributed for use in food-producing animals. 2015. https://www.fda.gov/AnimalVeterinary/NewsEvents/CVMUpdates/ucm476256.htm. Accessed 5 Oct 2017.
37. Zhu Y-G, Johnson TA, J-Q S, Qiao M, Guo G-X, Stedtfeld RD, Hashsham SA, Tiedje JM. Diverse and abundant antibiotic resistance genes in Chinese swine farms. Proc Natl Acad Sci U S A. 2013;110:3435–40.
38. Park JY, Ruidisch M, Huwe B. Transport of sulfonamide antibiotics in crop fields during monsoon season. Environ Sci Pollut Res Int. 2016;23: 22980–92.
39. Bush K, Jacoby GA. Updated functional classification of β-lactamases. Antimicrob Agents Chemother. 2010;54:969–76.
40. Gillings MR, Paulsen IT, Tetu SG. Ecology and evolution of the human microbiota: fire, farming and antibiotics. Genes. 2015;6:841–57.
41. Ramanan R, Kim B-H, Cho D-H, H-M O, Kim H-S. Algae–bacteria interactions: evolution, ecology and emerging applications. Biotechnol Adv. 2016;34:14–29.
42. Elsaied H, Stokes HW, Kitamura K, Kurusu Y, Kamagata Y, Maruyama A. Marine integrons containing novel integrase genes, attachment sites, attI,

and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. ISME J. 2011;5:1162–77.
43. Koenig J, Boucher Y, Charlebois RL, Nesbø C, Zhaxybayeva O, Bapteste E, Spencer M, Joss MJ, Stokes H, Doolittle W. Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. Environ Microbiol. 2008;10:1024–38.
44. Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y, Doolittle WF. Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds. PLoS One. 2009;4:e5276.
45. Baquero F, Alvarez-Ortega C, Martinez J. Ecology and evolution of antibiotic resistance. Environ Microbiol Rep. 2009;1:469–76.
46. Larsson DGJ, de Pedro C, Paxeus N. Effluent from drug manufactures contains extremely high levels of pharmaceuticals. J Hazard Mater. 2007;148:751–5.
47. Fick J, Söderström H, Lindberg RH, Phan C, Tysklind M, Larsson DGJ. Contamination of surface, ground, and drinking water from pharmaceutical production. Environ Toxicol Chem. 2009;28:2522–7.
48. Holmes AJ, Gillings MR, Nield BS, Mabbutt BC, Nevalainen K, Stokes H. The gene cassette metagenome is a basic resource for bacterial genome evolution. Environ Microbiol. 2003;5:383–94.
49. Gillings MR, Xuejun D, Hardwick SA, Holley MP, Stokes H. Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? ISME J. 2009;3:209–15.
50. Andrews S: FastQC: a quality control tool for high throughput sequence data. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 5 Oct 2017.
51. Yang X, Liu D, Liu F, Wu J, Zou J, Xiao X, Zhao F, Zhu B. HTQC: a fast quality control toolkit for Illumina sequencing data. BMC Bioinform. 2013;14:1.
52. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinform. 2009;10:1.
54. Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM, Zhu Y-G. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. ISME J. 2015;9:1269–79.
55. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013;57:3348–57.
56. Pereira MB, Wallroth M, Kristiansson E, Axelson-Fisk M. HattCI: fast and accurate attC site identification using hidden Markov models. J Comput Biol. 2016;23:891–902.
57. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinform. 2008;9:386.
58. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.
59. Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ, Nilsson RH. METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. Mol Ecol Resour. 2015;15:1403–14.
60. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.
61. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
62. Howe K, Bateman A, Durbin R. QuickTree: building huge neighbour-joining trees of protein sequences. Bioinformatics. 2002;18:1546–7.
63. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 2016;33:1635–8.
64. Flach C-F, Boulund F, Kristiansson E, Larsson DGJ. Functional verification of computationally predicted qnr genes. Ann Clin Microbiol Antimicrob. 2013;12:34.
65. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. 1997;25:1203–10.
66. Poirel L, Naas T, Nordmann P. Diversity, epidemiology, and genetics of class D β-lactamases. Antimicrob Agents Chemother. 2010;54:24–38.
67. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinform. 2008;9:40.
68. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605–12.
69. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
70. Oliveira-Pinto C, Costa PS, Reis MP, Chartone-Souza E, Nascimento AM. Diversity of gene cassettes and the abundance of the class 1 integron-integrase gene in sediment polluted by metals. Extremophiles. 2016;20:283–9.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit