Uncovering bacterial hosts of class 1 integrons in an urban coastal aquatic environment with a single-cell fusion-PCR technology

Qin Qi*‡, Timothy M Ghaly‡, Anahit Penesyan, Vaheesan Rajabal, Jeremy AC Stacey, Sasha G Tetu, Michael R Gillings

* Correspondence: qin.qi@mq.edu.au

‡ These authors contributed equally

AUTHOR ADDRESS

School of Natural Sciences, 14 Eastern Road, Macquarie University, NSW, Australia

Abstract

Horizontal gene transfer (HGT) drives bacterial evolution via transmission of genetic materials across taxa. Class 1 integrons are mobile genetic elements that correlate strongly with anthropogenic pollution and contribute to the spread of antimicrobial resistance (AMR) genes via HGT. Despite their significance to human health, there is a shortage of robust, culture-free surveillance technologies for identifying uncultivated environmental taxa that

harbour class 1 integrons. We developed a modified version of the epicPCR (emulsion, paired isolation and concatenation PCR) technology that links class 1 integrons amplified from single bacterial cells to taxonomic markers from the same cells in emulsified aqueous droplets. Using this single-cell genomic approach and Nanopore long-read sequencing, we successfully assigned class 1 integron gene cassette arrays containing mostly AMR genes to their hosts in coastal water samples that were affected by urban pollution. Our work presents the first application of epicPCR for targeting variable, multi-gene loci of interest. We also identified the *Rhizobacter* genus as novel hosts of class 1 integrons. These findings establish epicPCR as a powerful tool for linking bacterial taxa to class 1 integrons in environmental bacterial communities and offer the potential to direct mitigation efforts towards hotspots of class 1 integron-mediated dissemination of AMR.

Synopsis

We present a novel single-cell genomic surveillance technology for identifying environmental bacterial hosts of a class of mobile genetic elements that are linked to anthropogenic pollution and contribute to the dissemination of antimicrobial resistance.

Keywords

Integrons, antimicrobial resistance (AMR), anthropogenic pollution, class 1 integron gene cassettes, horizontal gene transfer (HGT), mobile genetic elements, environmental bacterial hosts, single-cell genomics

Introduction

Horizontal gene transfer (HGT) drives prokaryotic evolution by contributing to the spread of novel and ecologically significant genes in microbial communities [1]. This process generates genomic diversity upon which selection can act, and often underlies adaptation of microbial populations to environmental changes [2, 3]. HGT most commonly occurs among closely related taxa or between more distantly related taxa found in similar environments [4, 5]. A key example is the dissemination of antimicrobial resistance (AMR) genes across diverse bacterial phyla via HGT events such as conjugation, bacteriophage transduction, and natural transformation of bacterial cells by extracellular DNA [6-9].

To understand the dynamics of HGT, we need to identify the sources of horizontally transferred genes and the range of taxa they subsequently inhabit [10-12]. This is challenging given the sheer diversity of microbial taxa and the difficulty of culturing many environmental microbes under laboratory conditions [13-16]. Metagenomics allows high-throughput profiling of microbial taxonomy and horizontally transferred genes without the need for prior culturing of microbes [17, 18]. However, mobile genetic elements are often extensively associated with repetitive DNA sequences and exhibit complex interactions with genome rearrangements [19, 20]. These limit our ability to attribute laterally transferred genes to host genomes when analysing metagenomic data due to assembly and/or taxonomic binning difficulties.

epicPCR (emulsion, paired isolation and concatenation PCR), also known as OIL-PCR (one-step isolation and lysis PCR), is a fusion PCR-based technique that overcomes these problems by co-amplifying bacterial taxonomic markers along with functional genes of interest and doing so from single uncultured bacterial cells. Single cells are trapped in droplets generated by emulsifying an aqueous phase suspension of environmental bacteria in an emulsion oil [21-26]. This compartmentalises the fusion PCR-type amplification process into

parallel, single cell-level reactions. Within each droplet, chimeric amplicons that combine genes of interest with taxonomic markers corresponding to diverse phylogenetic groups are generated for high-throughput sequencing.

To date, epicPCR has been successfully applied to target functional genes including the bacterial dissimilatory sulfite reductase *dsrB* gene [21, 24], the bacteriophage ribonucleotide reductase *RNR* gene [25], antibiotic resistance genes [22, 23, 26, 27], and clinical class 1 integron genes *intI1*, *sul1*, and Δ*qacE* [23, 27, 28]. In all these studies, epicPCR amplicons comprised short fragments of genes of interest linked to the V4 hypervariable region of 16S rRNA markers. Illumina sequencing of recovered amplicons demonstrated the presence of the genes of interest in a variety of bacterial taxa. However, no studies have so far applied epicPCR to target variable multi-gene loci of interest, such as complete cassette arrays of class 1 integrons.

Class 1 integrons are key players in the dissemination of AMR genes via HGT in natural ecosystems and clinical settings [29, 30]. Integrons are genetic elements that comprise an integron-integrase gene (*intI*), an *attI* recombination site, and a promoter (P_C) that drives the expression of variable gene cassettes arranged in a gene cassette array [31-33]. *intI* encodes a site-specific tyrosine recombinase that catalyses the integration and excision of gene cassettes, which are small mobile DNA elements that usually carry single open reading frames (ORFs) and *attC* sites that can be recombined with *attI* or other *attC* sites [33-36]. All class 1 integrons contain a highly conserved integron-integrase (*intI1*) and *attI1* sequence (Figure 1A). Clinical class 1 integrons are strongly associated with AMR gene cassettes that confer resistance to antibiotics, disinfectants, or heavy metals [29, 31, 37] and are often flanked by characteristic 5'- and 3'-conserved segments (CS) that are not found in non-clinical class 1 integrons. Mobile genetic elements such as broad host-range plasmids and transposons have acquired class 1 integrons, resulting in the proliferation of class 1 integrons across diverse

bacterial hosts via HGT [31, 38]. Recently, it has been reported that phage-plasmids can carry and disseminate class 1 integrons through infection and lysogenic conversion, which represents a previously under-explored route of class 1 integron transmission [39].

Bacterial hosts of class 1 integrons have previously been isolated, identified, and characterised through colony PCR screening and fosmid library preparation procedures [29, 40-43]. The necessity to pre-isolate environmental bacteria severely biases host identification towards the cultivated faction and hinders our ability to attribute relative abundances of class 1 integrons to individual taxa. To overcome these limitations, we developed a modified experimental and bioinformatic pipeline for epicPCR that associates class 1 integrons with their bacterial hosts. Class 1 integron gene cassette arrays are co-amplified with the V4 hypervariable region of the bacterial 16S rRNA marker genes from individual bacterial cells (Figure 1B-C). The fused amplicons are sequenced using Nanopore technology to obtain longer read lengths compared to Illumina sequencing. We applied stringent bioinformatic filtering to ensure that the associations between the class 1 integrons and the taxonomic markers were *bona fide.*

We successfully tested our epicPCR pipeline on coastal water samples and linked six unique class 1 integrons with variable gene cassettes to seven distinct environmental bacterial taxa. While most of the gene cassettes encode AMR genes, we also linked a gene cassette encoding a cupin-domain containing protein of unknown functions to two *Alphaproteobacteria* hosts. Further, our results demonstrated that *Rhizobacter* is a novel host of class 1 integrons. Overall, our study shows that epicPCR is a powerful tool for determining the environmental hosts of class 1 integrons, with the potential to access previously untapped carriers of these genetic elements. This paves the way for future studies that seek to pinpoint the “sources” and “sinks” of class 1 integrons in bacterial communities from diverse natural and human-modified ecosystems.

Materials and Methods

Isolation of bacteria from an urban coastal location

50 mL coastal water samples were collected in three biological replicates after rain from the rockpools downstream from a stormwater outlet at Shark Point (location coordinates: -33.91, 151.27), which is a rocky intertidal zone near Sydney, NSW, Australia (Supplementary Figure S1). The samples were collected in sterile 50 ml Falcon tubes, placed on ice, and transported to the laboratory for immediate processing. Unless otherwise stated, all the chemicals used in this study were supplied by Sigma-Aldrich (United States). The collected water samples were individually filtered using 40 µm EASYstrainer cell strainers (Greiner AG, Germany) to remove debris and was centrifuged at 3220 *g* for 15 minutes. Cell pellets were resuspended in 2 ml of 0.01M phosphate buffered saline (PBS) solution (pH 7.4). Resuspended samples (0.7 ml) were mixed with 0.3 ml sterile glycerol and stored at -80˚C.

Bacterial cell counts by fluorescence microscopy

Approximately 10^5-10^6 bacterial cells were added to 1 mL PBS containing 4% formaldehyde and 2 µg/mL DAPI (4′,6-diamidino-2-phenylindole) for cell fixation and DNA staining respectively. Cell straining was carried out for 30 min in the dark with rotation at room temperature. The cells were then washed, pelleted, and re-suspended in 0.01M PBS solution. 10 µL of the re-suspended bacterial cells were transferred to a haemocytometer and viewed under the x40 lens of an Olympus BX63 fluorescence microscope. Differential interference contrast (DIC) and DAPI fluorescence channel images were acquired. The overlaid images were analysed using ImageJ software. Bacterial cell densities were calculated based on the average number of cells observed in five 0.2 mm x 0.2 mm square grids of the haemocytometer counting chamber (Supplementary Figure 1).

Emulsion, paired-isolation and concatenation PCR (epicPCR)

Our hybrid epicPCR procedures were adapted from the procedures described by Sakowski *et al* [25] and Diebold *et al* [26]. Briefly, epicPCR was performed in three independent technical replicates for the three biological replicates of coastal water samples. Approximately 150,000 bacterial cells from the glycerol stocks were pelleted, re-suspended and vortexed in a 75 µL PCR reagent mix containing the following components (at the specified final concentrations shown in parentheses): GC buffer (1x), dNTP mix (0.4 mM), Phusion Hot-Start II polymerase (0.05 U/µL) (Thermo Scientific, United States), bovine serum albumin solution (1 µg/µL) (Promega, United States), Lucigen Ready-Lyse lysozyme (500 U/µL) (LGC Biosearch Technologies, United States) and three oligonucleotide primers: R926 (2 µM), *intI1*_outer (1 µM), and R519-*qacE* bridging primer (0.04 µM). The primer sequences are shown in Supplementary Table S2. To demonstrate the absence of false associations between 16S rRNA markers and class 1 integrons originating from different cells, we spiked one set of technical replicates with cells of a class 1 integron-free *Escherichia coli* MG1655 strain at a population frequency of 10%.

ABIL emulsion oil for generating water-in-oil emulsion of single bacterial cells was prepared by supplementing mineral oil with 4% ABIL EM 90 (Redox, Australia) and 0.05% Triton X-100 (Promega, United States). 425 µL ABIL oil was added to the PCR mix, which was emulsified at 4 ms^{-1} speed for 45 s using the FastPrep-24 bead beating system (MP Biomedicals, United States). The emulsified PCR mix was aliquoted into 8 x 60 µL portions in PCR tubes. The following thermocycling conditions were then applied: 37˚C for 10 min (lysozyme digest), 98˚C for 5 minutes (lysis of bacterial cells and initial denaturation of DNA) and 38 cycles of 98°C for 10 s (DNA denaturation), 59°C for 20 s (primer annealing), and 72°C for 2 min (extension). A final extension at 72°C was carried out for 5 min. During this stage of epicPCR, the *intI1*_outer primer and the R519-*qacE* bridging primer target *intI1*

in the 5'-conserved segment (CS) and *qacE* in the3'-CS of class 1 integrons respectively (Figure 1B). An 18 bp overhang introduced by the bridging primer to the *qacE* end allows the intermediate PCR products to act as a long primer and amplify the V4-V5 hypervariable region of the 16S rRNA marker gene in the presence of the R926 reverse primer.

Aliquots of emulsified PCR mix for each sample were pooled, vortexed with 900 µL isobutanol and 200 µL of 5M NaCl, and centrifuged for 1 min at 20,817 *g*. The aqueous phase at the bottom of the tubes was retained for downstream DNA purification. A repeat of the previous step using the remaining emulsion was carried out to increase the DNA yield. The PCR products were purified using the NEB Monarch PCR & DNA Cleanup Kit (New England Biolabs, United States) and eluted with 30 µL nuclease-free water. 21 µL of Sera-Mag Select magnetic beads (Cytiva, United States) was mixed with the eluted DNA, pelleted on a magnetic stand, and washed twice with 80% ethanol to deplete <300 bp DNA fragments. The purified and size-selected DNA was eluted in nuclease-free water, which was then added to 100 µL Phusion PCR mix containing GC buffer (1x), dNTP mix (0.4 mM), Phusion Hot-Start II polymerase (0.04 U/µL) and AP27_short primer (0.8 µM), *intI1*_nested primer (0.4 µM), as well as forward and reverse blocking primers with 3' C3 spacer modification (0.32 µM each). AP27_short contains the bases of the R806 primer that binds to the 16S rRNA gene [44] and introduces a non-degenerate tail sequence (in blue).

The PCR mix was aliquoted into 8 PCR tubes with 12 µL in each tube and subjected to the following thermocycling conditions: 98˚C for 30 s, 35 cycles of 98°C for 10 s, 59°C for 20 s and 72°C for 1 min 50 s, followed by a final extension step of 72°C for 5 min. The nested primers used in this step further amplified the chimeric amplicons. Sera-Mag Select magnetic beads were added at 45% of the volume of the pooled epicPCR products to deplete <500 bp DNA fragments. 80% ethanol washes and DNA elution with nuclease-free water were performed as described above.

The non-degenerate tail introduced by the AP27_short nested primer can be targeted by the AP28_short primer (Figure 1C). To confirm that epicPCR amplicons comprising class 1 integrons and 16S rRNA gene was successfully generated, 2 ng of each purified epicPCR product was amplified using *qacE*_F (0.4 µM) and AP28_short (0.4 µM) primers in GoTaq polymerase master mix (Promega, United States) under the following thermocycling conditions: 95˚C for 30 s, 30 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, with a final extension step of 72°C for 5 min. The GoTaq PCR products were checked on 1.5% agarose gel by electrophoresis. Successful amplification of the *qacE*-16S rRNA gene (V4) chimeric region should produce gel bands of approximately 390 bp in size.

epicPCR verification of class 1 integrons in the *Rhizobacter* genus

To experimentally confirm that class 1 integrons could be found in the *Rhizobacter* genus detected in our coastal water samples, we pooled approximately 150,000 cells from the three biological replicates and performed our modified epicPCR procedures in three technical replicates with minor modifications. The R926 outer primer and AP27_short nested primer for targeting the 16S rRNA gene were replaced with *Rhizobacter*-specific primers *Rhizobacter*_16S_ R926 and *Rhizobacter*_16S_ R806 respectively. During the nested epicPCR step, the *intI1*_nested primer was replaced with *qacE*_F. The DNA sequence of the 389 bp chimeric epicPCR product was verified by Sanger sequencing (Marcogen Inc, South Korea).

Oxford Nanopore long-read sequencing

Purified epicPCR amplicons were treated with the NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs, United States) to obtain 5' phosphorylated and 3' dA-tailed (end-repaired) DNA, which was subsequently purified with JetSeq Clean magnetic beads

(Meridian Bioscience, United States) and eluted using nuclease-free water. Native barcodes from the Native Barcoding Expansion kit (ONT, United Kingdom) were ligated to the end-repaired DNA using the Blunt/TA Ligase Master Mix (New England Biolabs, United States). After purification by magnetic beads and elution in nuclease-free water, individual barcoded DNA samples were pooled and ligated to sequencing adaptor molecules in Adapter Mix II using the NEBNext Quick T4 DNA Ligase (New England Biolabs, United States). The ligation products were washed twice using the Short Fragmentation Buffer (SFB) and eluted with the Elution Buffer (EB). The multiplexed barcoded samples were prepared as sequencing libraries and loaded into a FLO-MIN106D flow cell R9.4.1 and a MinION Mk1B sequencer (ONT, United Kingdom) for Nanopore sequencing. The minimum threshold for read length was set to 1 kB on the MinKNOW operating software to minimise the acquisition of cassette-less class 1 integrons. The median number of ONT reads acquired per native barcoded sample was 232,402 after the quality control and demultiplexing procedures.

Analysis of Nanopore sequencing data of epicPCR amplicons

The basecalling of FAST5 raw sequencing data was performed with the high accuracy option using the Nanopore Guppy software v6.1.2 (ONT, United Kingdom). The native barcoded samples were demultiplexed into individual FASTQ files using default parameters. All Nanopore sequence filtering and processing steps for epicPCR amplicon reads were carried out using an in-house pipeline (https://github.com/timghaly/Int1-epicPCR). First, the pipeline quality-filters Nanopore reads using NanoFilt v2.8.0 [45], removing those with an average read quality of less than 7 and read length less than 670 bp, which represents the minimum length of an epicPCR product with a cassette-less integron. Quality-filtered reads are then oriented and trimmed with the final nested epicPCR forward and reverse primer sequences using Pychopper v2.7.0 (https://github.com/epi2me-labs/pychopper). Pychopper identifies both primers in each read using edlib v1.2.3 [46], and orients the reads based on the forward and

reverse primer sequences. Reads that do not contain both primers in the correct orientation are discarded. The pipeline then clusters the primer-oriented reads into amplicon-specific clusters using isONclust v0.0.6.1 [47], and performs error correction on each cluster using isONcorrect v0.0.8 [48]. Importantly, isONcorrect can jointly use all cassette arrangements of the same integron that occur in different clusters, allowing efficient error correction even for amplicons with low sequencing depths. After error correction, a consensus sequence is generated for each cluster using spoa v4.0.7 [49]. All consensus sequences are then pooled, while removing any redundancies, including reverse complement redundancies, using the dedupe programme from the BBTools v35 software package (https://github.com/kbaseapps/BBTools). Next, the consensus sequences are screened for both the sequence of the R519-*qacE* bridging primer, and then the 157 bp region that includes the 5'-end of *intI1* and all of *attI1* using blastn from the BLAST v2.2.31 package. Retained sequences must contain both the bridging primer and the *intI1*/*attI1* region. Any sequences containing more than one hit to either of these sequences are considered unintended chimeras and discarded. Finally, sequences are screened for the correctly fused 16S rRNA gene fragment using Metaxa2 v2.2.3 [50]. The final output of the pipeline for each sample is a set of full-length, primer-oriented amplicon consensus sequences that contain a complete *attI1* sequence, R519-*qacE* bridging primer, and the V4 hypervariable region of the 16S rRNA gene.

Cd-hit v4.8.1 [51] was used to cluster chimeric epicPCR products comprising class 1 integron gene cassette arrays and V4 hypervariable regions of 16S rRNA gene with ≥99% pairwise identity in nucleotide sequences in at least 3 replicates of epicPCR. For epicPCR products that were found in fewer than 3 replicates, the NCBI Genome Workbench was used to perform blastn searches of these sequences against a local database created using all the Nanopore reads obtained in this study. Reads with ≥98% pairwise identity in nucleotide

sequences and ≥98% coverage were aligned using the Geneious bioinformatic software (Biomatters, New Zealand), and the consensus sequences were generated from these alignments within each set of epicPCR replicates. Consensus sequences that could be found in at least 3 epicPCR replicates and showed ≥99% pairwise identity in nucleotide sequences and ≥99% identical sites were added to the list of epicPCR products that have already been shortlisted by the Cd-hit algorithm. IntegronFinder 2.0 [52] [parameters: --local-max --gbk --promoter-attI --calin-threshold 1] was used to predict the ORFs and *attC* sites in the epicPCR products, and the annotations were adjusted according to the NCBI reference class 1 integron gene cassettes shown in Table 1.

Taxonomic classification for bacterial hosts of class 1 integrons

The sequences of V4 hypervariable regions of 16S rRNA gene sequences obtained in this study were searched against the SILVA 16S rRNA gene database website [53] using the Alignment, Classification and Tree (ACT) tool with a cut-off threshold of 0.80. The V4 region sequences were aligned using the Geneious multiple alignment tool using default parameters and cost matrix = 93% similarity. A phylogenetic tree was constructed with PhyML (v3.0) using the maximum likelihood method based on the Hasegawa-Kishino-Yano (HKY85) model with bootstrapping (number of replicates = 100) [54]. To verify genus-level taxonomic classifications, the V4 hypervariable regions for the *Rhizobacter* and *Aquabacterium*-like hosts of class 1 integrons found in this study were searched against the NCBI Nucleotide Collection database by blastn with the exclusion of "uncultured/environmental sample sequences". The 16S rRNA sequences of all matches with 100% pairwise nucleotide identity and 100% coverage were searched against the SILVA 16S rRNA gene database. All the reference genomes of *Rhizobacter* and *Aquabacterium* species were downloaded from the NCBI RefSeq database (last accessed on 11 November 2022). Nucleotide alignment was performed using the sequences of V4 hypervariable regions

against the full 16S rRNA gene sequences from the RefSeq genomes using the Geneious bioinformatic software.

Results and Discussion

Coastal water samples were collected in three biological replicates from the rockpools downstream of a stormwater outlet at Shark Point, Sydney, New South Wales, Australia (Supplementary Figures S1A and S1B). We anticipated that class 1 integron-associated environmental bacteria would be enriched at this location, because surface runoff was expected to carry pollutants from the surrounding urban areas. A preponderance of cells displayed unicellular and planktonic morphologies (Supplementary Figures S1C). We applied modified epicPCR procedures to amplify class 1 integron gene cassette arrays from single bacterial cells encapsulated by PCR reagent droplets homogenously dispersed in an oil emulsion and successfully generated chimeric epicPCR products containing class 1 integrons and 16S rRNA gene sequences (Figure 1). These epicPCR products were Nanopore sequenced, which revealed features conserved in class 1 integrons. After stringent bioinformatic filtering, we identified six class 1 integrons in seven distinct bacterial hosts (Figure 2). All seven combinations of epicPCR products were observed in at least three independent replicates ($n \geq 3$, ≥99% pairwise nucleotide identity and ≥99% identical sites relative to the consensus sequence for each epicPCR product). Taxonomic classifications were obtained using the SILVA ribosomal RNA gene database (Supplementary Table S1) [53]. *Alphaproteobacteria* (n=2) and *Gammaproteobacteria* (n=5) were the two major classes of hosts represented in our dataset, while *Burkholderiales* (n=4) was the most common bacterial order. These observations are broadly consistent with previous findings that class 1 integrons are associated with *Gammaproteobacteria* in urban, semi-urban, and natural ecosystems [29, 42,

[43]. The finding of integrons in two *Alphaproteobacteria* is a notable observation, because *Alphaproteobacteria* are rarely known to be associated with integrons [55]. Currently, there are very few examples of *Alphaproteobacteria* species that are vectors of class 1 integrons in literature [56-58]. This underscores the potential of epicPCR approaches to discover novel host-function associations.

Six different gene cassettes were detected using epicPCR, five of which encoded well-characterised AMR genes (Table 1). The nucleotide sequences of all the gene cassettes reported in this study are summarised in Supplementary Table S3. We compared the nucleotide sequences of all six gene cassettes against annotated examples of reference class 1 integron gene cassettes from the NCBI database [34, 59]. The coding and non-coding regions of these sequenced gene cassettes were highly conserved relative to the reference sequences. In our dataset, there are six different *attC* recombination sites with 100% nucleotide sequence identity to those in the respective reference gene cassettes in Table 1. In addition, we also aligned all the nucleotide sequences in the 5'-CS and 3'-CS regions of all the class 1 integrons from this study and found that both the 157 bp *intI1/attI1* and 190 bp *attC*/*qacE* regions in the chimeric amplicons were identical in all but one exception (Supplementary Figures 2A and 2B). Overall, these class 1 integron structures are typical of those recovered from clinical bacterial isolates [29, 60, 61].

Class 1 gene cassettes detected by epicPCR

Cupins belong to a functionally diverse superfamily of proteins with a beta-barrel fold, and were originally found as highly conserved domains in germin and germin-like proteins of higher plants [62]. We found a gene cassette with an open-reading frame (ORF) encoding a cupin barrel domain of unknown functions in two *Alphaproteobacteria* hosts. One host was unclassified at the order level (Figure 2A), while the other belonged to the *Rhodobacteraceae*

family (Figure 2B). This cupin-domain-encoding gene was first observed in a class 1 integron gene cassette from an unknown host in a previous study [59]. Based on identical nucleotide sequence matches from the NCBI nucleotide database, this gene cassette has been detected in class 1 integrons, non-class 1 integrons and CALINs (clusters of *attC*s lacking an associated integron-integrase) in the chromosomes or plasmids of several *Gammaproteobacteria* hosts from the *Comamonadaceae*, *Enterobacteriaceae* and *Pseudomonadaceae* families (Supplementary Table S4). This suggests that there is selective pressure for maintaining this cupin-domain-encoding gene cassette in both class 1 and non-class 1 integron genetic contexts. To the best of our knowledge, its association with an *Alphaproteobacteria* host has not been reported previously. It is hoped that pinpointing the bacterial hosts and their ecological niches will generate more clues regarding possible functions of this cupin-domain-encoding gene.

The *dfrA5* gene encodes a dihydrofolate reductase that confers a high level of resistance to trimethoprim in Gram-negative bacteria [63]. It is among the most frequently observed trimethoprim resistance genes found in association with class 1 integrons [64]. The *dfrA5* gene cassette has been detected in a wide range of *Enterobacteriaceae* hosts in clinical, agricultural, and environmental contexts [65-69]. We found that the *dfrA5* gene cassette is carried by a *Comamonadaceae*-family host (Figure 2C). The $blaOXA_{10}$ gene encodes a class D beta-lactamase that hydrolyses oxacillin, which is a narrow-spectrum beta-lactam derived from penicillin [70, 71]. Previously, $blaOXA_{10}$ was referred to as *pse*-2 (*Pseudomonas*-specific enzymes) due to its initial discovery in *P. aeruginosa* plasmids. Class 1 integron-associated $blaOXA_{10}$ gene cassettes have been found in a variety of *Gammaproteobacteria* pathogens, including those from the *Acinetobacter*, *Pseudomonas*, *Gallibacterium*, *Providencia*, and *Stutzerimonas* genera [71-75]. Through epicPCR, we associated the $blaOXA_{10}$ gene cassette with a *Moraxellaceae*-family host (Figure 2D). Both the *dfrA5* and the $blaOXA_{10}$ integron gene

cassettes in our dataset show 100% nucleotide identity to the respective reference gene cassettes shown in Table 1.

Genes *qacF* and *qacG2* encode small multidrug resistance (SMR) efflux pumps that confer resistance to quaternary ammonium compounds (QACs) present in a wide range of disinfectants and detergents [31, 76]. We found a *qacG2* gene cassette in an *Aquabacterium*-like host (Figure 2D) and a *qacF* gene cassette at the proximal position in the gene cassette array in a *Rhizobacter* host (Figure 2F). Also found in the class 1 integron of the *Rhizobacter* host is an *aadA2* gene cassette. The *aadA* family of aminoglycoside (3'') adenylyltransferase genes confer resistance to the aminoglycoside antibiotics streptomycin and spectinomycin. Diverse *aadA* genes are some of the most frequently detected gene cassettes in class 1 integrons in environmental and clinical bacterial hosts [34, 77-80]. In the *Rhodocyclaceae*-family host, we found the same *aadA2* gene cassette with one SNP in the non-coding region and one non-synonymous mutation (V257A) in the coding region (Figure 2G and Supplementary Table S3).

Discovery of *Rhizobacter* as novel carriers of class 1 integrons

Rhizobacter bacteria are soil microbes that are often associated with the rhizosphere. The presence of the *Rhizobacter* genus in our coastal water samples was likely due to surface runoff. *Rhizobacter* species have been successfully isolated and characterised from the rhizospheres of plant species from the *Panax* [81, 82] and *Bergenia* genera [83], as well as from freshwater sediment [84]. There are still many questions regarding their potential ecological roles. For example, the *Rhizobacter gummiphilus* species naturally degrades rubber [85-87], which implies beneficial roles in promoting nutrient recycling and bioremediation in soil. On the other hand, *Rhizobacter daucus* is thought to be an agricultural plant pathogen that causes galls in carrot root tissues [88].

An association between the *Rhizobacter* genus and class 1 integrons has not been reported previously. In all the NCBI RefSeq genomes for the *Rhizobacter* genus, we detected various non-class 1 integrons but did not find any bioinformatic evidence of class 1 integrons. We verified the genus-level classification of the *Rhizobacter* class 1 integron host in our epicPCR results by searching for identical matches to the V4 hypervariable region of the 16S rRNA gene (i.e. 100% pairwise nucleotide identity and 100% coverage) in the NCBI nucleotide database and found two *Rhizobacter* isolates (NCBI accession no. AB835065.1 and LC734193.1) whose taxonomic classification at the genus level we confirmed using the SILVA 16S rRNA gene database [89, 90]. To experimentally verify that the association between the *Rhizobacter* and class 1 integron was *bona fide*, we designed *Rhizobacter*-specific 16S rRNA gene primers that target the same binding sequences in the 16S rRNA gene as the degenerate primers R926 and R806. By replacing the *intI1*_nested primer with *qacE*_F (the primer positions are indicated in Figure 1C) in our epicPCR experimental pipeline, we successfully obtained the expected chimeric epicPCR product featuring a *qacE* fragment and the *Rhizobacter* phylogenetic marker. This provides the first experimental evidence of an association between class 1 integrons and the *Rhizobacter* genus.

Unknown antibiotic resistance determinants in environmental bacteria have the potential to transfer to human pathogens and pose significant risks to human health when they proliferate as novel antibiotic resistance genes [91, 92]. Class 1 integrons, and integrons in general, can be vectors of uncharacterised genes with unknown functions. At least some of these uncharacterised genes are expected to have the potential to become novel antibiotic resistance genes in the future. An example of a novel class 1 integron-associated antibiotic resistance gene is the garosamine-specific aminoglycoside resistance (*gar*) gene recently discovered in wastewater metagenomes [93]. Identifying the transmission paths of uncharacterised class 1 integron gene cassettes can provide vital clues concerning the environmental conditions that

contribute to the emergence of novel antibiotic resistance traits [92]. A recent study applied a modelling approach to understand the emergence and spread of novel antimicrobial resistance genes [94]. The dispersal of novel antibiotic resistance genes from environmental bacteria to humans has been highlighted as a major knowledge gap that contributes to uncertainties in these models. Acquiring more experimental data for these gene mobilisation and HGT events can improve future quantitative risk assessment models for antibiotic resistance and our ability to curb transfer of antibiotic resistance genes between bacteria. An epicPCR approach, which links integron gene cassettes to their bacterial hosts, has the potential to contribute to this endeavour given its strengths in identifying environmental microbes that cannot be cultivated under laboratory conditions.

Conclusions

In this study, we established epicPCR as a valuable surveillance platform for uncovering novel associations between class 1 integrons with diverse cassette arrays and their bacterial hosts in environmental water samples. Using this single-cell genomic approach, we successfully identified *Alphaproteobacteria* as novel hosts of a gene cassette encoding a cupin-domain protein of unknown functions, and the *Rhizobacter* genus as novel carriers of class 1 integrons. It is envisaged that this technology can be deployed to analyse water samples from sewage treatment plants, as well as effluents from industrial, clinical, and agricultural settings where antimicrobial usage is high. These surveillance efforts will hopefully lead to a timelier detection of hotspots of class 1 integron mediated AMR transmission and more targeted environmental mitigation measures. epicPCR also paves the way for future studies that seek to uncover the “sources” and “sinks” of integron gene cassettes in bacterial communities, as well as possible barriers to integron-mediated HGT that

may also exist [95]. With greater diversity in combinations of gene cassette and bacterial hosts that are likely to be captured through future epicPCR studies, we will be able to address outstanding questions and gain more insights into possible origins of novel gene cassettes, which has so far remained the greatest mystery in the field of integron research [32, 35, 96].

Figure Caption

Figure 1. Class 1 integrons and their co-amplification with 16S rRNA marker genes via epicPCR.

(A) The class 1 integron carried by the IncW plasmid R388 has the canonical structure of a clinical class 1 integron. The *intI1* genes and *attI1* sequences are high conserved in the 5'-conserved segment (5'-CS). The 3'-conserved segment (3'-CS) comprise a truncated quaternary ammonium compound (QAC) resistance gene (Δ*qacE*) and a sulfonamide resistance gene (*sul1*). Between the 5'-CS and 3'-CS is the integron gene cassette array that contains *attC* recombination sites and gene cassettes. Each gene cassette is typically associated with a single open-reading frame (ORF).

(B) During the first stage of epicPCR, the *intI1*_outer and R519_*qacE* bridging primers target *intI1* and *qacE* sequences in the 5'-CS and 3'-CS respectively and amplify the class 1 integron. A short overhang introduced by the bridging primer to the *qacE* fragment end allows the intermediate PCR product to act as a long primer and co-amplify the V4-V5 hypervariable regions of the 16S rRNA marker gene in the presence of the R926 reverse primer.

(C) During the second stage of epicPCR, the *intI1*_nested and AP27_short nested primers further amplify the chimeric amplicons pooled from the previous step. AP27_short introduces a non-degenerate tail sequence (in blue) to the 3'-end of the amplicons. During standard PCR verification of chimeric amplicon formation, this tail sequence is targeted by the AP28_short non-degenerate primer, which is used in conjunction with the *qacE*_F forward primer. The final chimeric epicPCR amplicons contain highly conserved features of class 1 integrons, including a 157 bp *intI1-attI1* region in the 5'-CS and the *qacE*/Δ*qacE* fragment in the 3'-CS.

Figure 2: Class 1 integrons identified through epicPCR alongside the taxonomic classification of their respective bacterial hosts. The *intI1* gene fragment is displayed in blue, the *attI1* site in purple, and all ORFs in green. The six different *attC* sites are shown in red.

Figure 3: Phylogenetic tree constructed using the maximum likelihood method for the V4 hypervariable regions of 16S rRNA genes for the seven hosts of class 1 integrons identified in this study. The integron gene cassettes associated with *Alphaproteobacteria* (in blue) and *Gammaproteobacteria* (in red) are grouped accordingly. Gene cassettes that were found in *Gammaproteobacteria* hosts of the *Burkholderiales* order are shown in yellow.

Tables

Table 1. Comparison between the class 1 integron gene cassettes sequenced in this work and annotated reference class 1 integron gene cassettes.

ORF in gene cassette	Predicted function of ORF gene product	NCBI reference gene cassette	% Nucleotide identity to NCBI reference gene cassette sequence
Cupin-domain-encoding gene	Unknown	FJ172388	99.8% nucleotide identity; 1 synonymous mutation in the coding region
dfrA5	Trimethoprim-resistant dihydrofolate reductase	X12868	100% nucleotide identity
blaOXA$_{10}$	Oxacillin-hydrolysing beta-lactamase	XXU37105	
qacG2	Quaternary ammonium compound efflux pump	AF327731	99.6% nucleotide identity; 2 non-synonymous mutations (S13F and A57G) in the coding region
qacF	Quaternary ammonium compound efflux pump	AF034958	99.6% nucleotide identity; 1 non-synonymous mutation (S10A) in the coding region
aadA2	Aminoglycoside (3'') adenylyltransferase conferring resistance to streptomycin and spectinomycin	X68227	In the *Rhizobacter* host: 100% nucleotide identity
			In the *Rhodocyclaceae* host: 99.6% nucleotide identity; 1 SNP in the non-coding region, 1 non-synonymous mutation (V257A) in the coding region

Associated Content

Supporting Information (PDF): Tables S1, S2, S3 and S4.

Supplementary Figure 1 (PDF): **(A)** Geographic location of the collection point for the coastal water samples at Shark Point, NSW, Australia **(B)** Photograph of rockpools downstream from a stormwater outlet at the sample collection site. **(C)** DAPI-stained cells (in green) from the glycerol stocks of coastal water samples were imaged against a 0.2 mm x 0.2 mm grid on a haemocytometer by fluorescence microscopy.

Supplementary Figure 2 (PDF): Nucleotide sequence alignments of **(A)** the 157 bp *intI1-attI1* region in the 5'-CS region of class 1 integrons and **(B)** the 190 bp final *attC-qacE* region in the 3'-CS region of class 1 integrons that were sequenced in this study.

Author Contributions

Conceptualisation of study: TMG, QQ, SGT, MRG

Experimental investigation and validation: QQ, AP, JACS

Bioinformatic analysis: TMG, QQ, VR

Project management and supervision: MRG, SGT

Writing the original draft of the manuscript: QQ

Reviewing and editing the manuscript: All authors

Funding Sources

This work was funded by the Australian Research Council Discovery Project DP200101874 grant to MRG and SGT.

Acknowledgement

The authors would like to thank Professor Marko Virta and members his research group at the Department of Microbiology, University of Helsinki for sharing their expertise on epicPCR. We would like to acknowledge the Macquarie University Faculty of Science and Engineering Microscope Facility for access to its instrumentation and staff. We also thank Mr Richard Miller (IT Systems and HPC, Macquarie University) for his technical advice and support.

References

1. Boto, L., Horizontal gene transfer in evolution: facts and challenges. *Proc Biol Sci* **2010,** *277* (1683), 819-27.

2. Jain, R.; Rivera, M. C.; Moore, J. E.; Lake, J. A., Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol* **2003,** *20* (10), 1598-602.

3. Wiedenbeck, J.; Cohan, F. M., Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **2011,** *35* (5), 957-76.

4. Beiko, R. G.; Harlow, T. J.; Ragan, M. A., Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci U S A* **2005,** *102* (40), 14332-7.

5. Gillings, M. R.; Holley, M. P.; Stokes, H. W., Evidence for dynamic exchange of *qac* gene cassettes between class 1 integrons and other integrons in freshwater biofilms. *FEMS Microbiol Lett* **2009,** *296* (2), 282-8.

6. Sørensen, S. J.; Bailey, M.; Hansen, L. H.; Kroer, N.; Wuertz, S., Studying plasmid horizontal transfer *in situ*: a critical review. *Nat Rev Microbiol* **2005,** *3* (9), 700-10.

7. Gillings, M. R., Lateral gene transfer, bacterial genome evolution, and the Anthropocene. *Ann N Y Acad Sci* **2017,** *1389* (1), 20-36.

8. Brown-Jaque, M.; Calero-Cáceres, W.; Muniesa, M., Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid* **2015,** *79*, 1-7.

9. Lerminiaux, N. A.; Cameron, A. D. S., Horizontal transfer of antibiotic resistance genes in clinical environments. *Can J Microbiol* **2019,** *65* (1), 34-44.

10. Hall, J. P.; Wood, A. J.; Harrison, E.; Brockhurst, M. A., Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. *Proc Natl Acad Sci U S A* **2016,** *113* (29), 8260-5.

11. Zhou, H.; Beltrán, J. F.; Brito, I. L., Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Sci Adv* **2021,** *7* (43), eabj5056.

12. Abe, K.; Nomura, N.; Suzuki, S., Biofilms: hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. *FEMS Microbiol Ecol* **2020,** *96* (5).

13. Cárdenas, J. P.; Quatrini, R.; Holmes, D. S., Genomic and metagenomic challenges and opportunities for bioleaching: a mini-review. *Res Microbiol* **2016,** *167* (7), 529-38.

14. Thompson, L. R.; Sanders, J. G.; McDonald, D.; Amir, A.; Ladau, J.; Locey, K. J.; Prill, R. J.; Tripathi, A.; Gibbons, S. M.; Ackermann, G.; Navas-Molina, J. A.; Janssen, S.; Kopylova, E.; Vázquez-Baeza, Y.; González, A.; Morton, J. T.; Mirarab, S.; Zech Xu, Z.; Jiang, L.; Haroon, M. F.; Kanbar, J.; Zhu, Q.; Jin Song, S.; Kosciolek, T.; Bokulich, N. A.; Lefler, J.; Brislawn, C. J.; Humphrey, G.; Owens, S. M.; Hampton-Marcell, J.; Berg-Lyons, D.; McKenzie, V.; Fierer, N.; Fuhrman, J. A.; Clauset, A.; Stevens, R. L.; Shade, A.; Pollard, K. S.; Goodwin, K. D.; Jansson, J. K.; Gilbert, J. A.; Knight, R.; Consortium, E. M. P., A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **2017,** *551* (7681), 457-463.

15. Bahram, M.; Hildebrand, F.; Forslund, S. K.; Anderson, J. L.; Soudzilovskaia, N. A.; Bodegom, P. M.; Bengtsson-Palme, J.; Anslan, S.; Coelho, L. P.; Harend, H.; Huerta-Cepas, J.; Medema, M. H.; Maltz, M. R.; Mundra, S.; Olsson, P. A.; Pent, M.; Põlme, S.; Sunagawa, S.; Ryberg, M.; Tedersoo, L.; Bork, P., Structure and function of the global topsoil microbiome. *Nature* **2018,** *560* (7717), 233-237.

16. Stewart, E. J., Growing unculturable bacteria. *J Bacteriol* **2012,** *194* (16), 4151-60.

17. Douglas, G. M.; Langille, M. G. I., Current and promising approaches to identify horizontal gene transfer events in metagenomes. *Genome Biol Evol* **2019,** *11* (10), 2750-2766.

18. Seiler, E.; Trappe, K.; Renard, B. Y., Where did you come from, where did you go: Refining metagenomic analysis tools for horizontal gene transfer characterisation. *PLoS Comput Biol* **2019,** *15* (7), e1007208.

19. Ghurye, J. S.; Cepeda-Espinoza, V.; Pop, M., Metagenomic assembly: Overview, challenges and applications. *Yale J Biol Med* **2016,** *89* (3), 353-362.

20. Darmon, E.; Leach, D. R., Bacterial genome instability. *Microbiology and molecular biology reviews : MMBR* **2014,** *78* (1), 1-39.

21. Spencer, S. J.; Tamminen, M. V.; Preheim, S. P.; Guo, M. T.; Briggs, A. W.; Brito, I. L.; D, A. W.; Pitkanen, L. K.; Vigneault, F.; Juhani Virta, M. P.; Alm, E. J., Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J* **2016,** *10* (2), 427-36.

22. Roman, V. L.; Merlin, C.; Virta, M. P. J.; Bellanger, X., EpicPCR 2.0: Technical and methodological improvement of a cutting-edge single-cell genomic approach. *Microorganisms* **2021,** *9* (8).

23. Hultman, J.; Tamminen, M.; Parnanen, K.; Cairns, J.; Karkman, A.; Virta, M., Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. *FEMS Microbiol Ecol* **2018,** *94* (4).

24. Qin, H.; Wang, S.; Feng, K.; He, Z.; Virta, M. P. J.; Hou, W.; Dong, H.; Deng, Y., Unraveling the diversity of sedimentary sulfate-reducing prokaryotes (SRP) across Tibetan saline lakes using epicPCR. *Microbiome* **2019,** *7* (1), 71.

25. Sakowski, E. G.; Arora-Williams, K.; Tian, F.; Zayed, A. A.; Zablocki, O.; Sullivan, M. B.; Preheim, S. P., Interaction dynamics and virus-host range for estuarine actinophages captured by epicPCR. *Nat Microbiol* **2021,** *6* (5), 630-642.

26. Diebold, P. J.; New, F. N.; Hovan, M.; Satlin, M. J.; Brito, I. L., Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. *Elife* **2021,** *10*.

27. Sarekoski, A. K., The use of epicPCR to study antimicrobial resistance genes and their bacterial hosts in human impacted environments in West Africa. University of Helsinki: 2020.

28. Wei, Z.; Feng, K.; Wang, Z.; Zhang, Y.; Yang, M.; Zhu, Y. G.; Virta, M. P. J.; Deng, Y., High-throughput single-cell technology reveals the contribution of horizontal gene transfer to typical antibiotic resistance gene dissemination in wastewater treatment plants. *Environ Sci Technol* **2021,** *55* (17), 11824-11834.

29. Gillings, M.; Boucher, Y.; Labbate, M.; Holmes, A.; Krishnan, S.; Holley, M.; Stokes, H. W., The evolution of class 1 integrons and the rise of antibiotic resistance. *Journal of Bacteriology* **2008,** *190* (14), 5095-5100.

30. Domingues, S.; Da Silva, G. J.; Nielsen, K. M., Global dissemination patterns of common gene cassette arrays in class 1 integrons. *Microbiology* **2015,** *161* (7), 1313-1337.

31. Gillings, M. R.; Gaze, W. H.; Pruden, A.; Smalla, K.; Tiedje, J. M.; Zhu, Y. G., Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* **2015,** *9* (6), 1269-79.

32. Ghaly, T. M.; Geoghegan, J. L.; Tetu, S. G.; Gillings, M. R., The peril and promise of integrons: beyond antibiotic resistance. *Trends in Microbiology* **2020,** *28* (6), 455-464.

33. Escudero, J. A.; Loot, C.; Mazel, D., Integrons as adaptive devices. In *Molecular Mechanisms of Microbial Evolution*, Rampelotto, P. H., Ed. Springer International Publishing: 2018; pp 199-239.

34. Partridge, S. R.; Tsafnat, G.; Coiera, E.; Iredell, J. R., Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiology Reviews* **2009,** *33* (4), 757-784.

35. Escudero, J. A.; Loot, C.; Nivina, A.; Mazel, D., The integron: adaptation on demand. *Microbiol Spectr* **2015,** *3* (2), MDNA3-0019-2014.

36. Recchia, G. D.; Stokes, H. W.; Hall, R. M., Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. *Nucleic Acids Res* **1994,** *22* (11), 2071-8.

37. Zheng, W.; Huyan, J.; Tian, Z.; Zhang, Y.; Wen, X., Clinical class 1 integron-integrase gene - A promising indicator to monitor the abundance and elimination of antibiotic resistance genes in an urban wastewater treatment plant. *Environ Int* **2020,** *135*, 105372.

38. Jechalke, S.; Schreiter, S.; Wolters, B.; Dealtry, S.; Heuer, H.; Smalla, K., Widespread dissemination of class 1 integron components in soils and related ecosystems as revealed by cultivation-independent analysis. *Front Microbiol* **2013,** *4*, 420.

39. Pfeifer, E.; Bonnin, R. A.; Rocha, E. P. C., Phage-plasmids spread antibiotic resistance genes through infection and lysogenic conversion. *mBio* **2022,** *13* (5), e0185122.

40. Waldron, L. S.; Gillings, M. R., Screening foodstuffs for class 1 integrons and gene cassettes. *J Vis Exp* **2015,** (100), e52889.

41. Dawes, F. E.; Kuzevski, A.; Bettelheim, K. A.; Hornitzky, M. A.; Djordjevic, S. P.; Walker, M. J., Distribution of class 1 integrons with IS26-mediated deletions in their 3'-conserved segments in *Escherichia coli* of human and animal origin. *PloS One* **2010,** *5* (9), e12754.

42. Stokes, H. W.; Nesbø, C. L.; Holley, M.; Bahl, M. I.; Gillings, M. R.; Boucher, Y., Class 1 integrons potentially predating the association with Tn*402*-like transposition genes are present in a sediment microbial community. *J Bacteriol* **2006,** *188* (16), 5722-30.

43. Rosewarne, C. P.; Pettigrove, V.; Stokes, H. W.; Parsons, Y. M., Class 1 integrons in benthic bacterial communities: abundance, association with Tn*402*-like transposition modules and evidence for coselection with heavy-metal resistance. *FEMS Microbiol Ecol* **2010,** *72* (1), 35-46.

44. Kozich, J. J.; Westcott, S. L.; Baxter, N. T.; Highlander, S. K.; Schloss, P. D., Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **2013,** *79* (17), 5112-20.

45. De Coster, W.; D'Hert, S.; Schultz, D. T.; Cruts, M.; Van Broeckhoven, C., NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **2018,** *34* (15), 2666-2669.

46. Šošic, M.; Šikic, M., Edlib: a C/C++ library for fast, exact sequence alignment using edit distance. *Bioinformatics* **2017,** *33* (9), 1394-1395.

47. Sahlin, K.; Medvedev, P., *De novo* clustering of long-read transcriptome data using a greedy, quality value-based algorithm. *J Comput Biol* **2020,** *27* (4), 472-484.

48. Sahlin, K.; Medvedev, P., Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nat Commun* **2021,** *12* (1), 2.

49. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M., Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* **2017,** *27* (5), 737-746.

50. Bengtsson-Palme, J.; Hartmann, M.; Eriksson, K. M.; Pal, C.; Thorell, K.; Larsson, D. G.; Nilsson, R. H., METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* **2015,** *15* (6), 1403-14.

51. Li, W.; Godzik, A., Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006,** *22* (13), 1658-9.

52. Neron, B.; Littner, E.; Haudiquet, M.; Perrin, A.; Cury, J.; Rocha, E. P. C., IntegronFinder 2.0: Identification and analysis of integrons across bacteria, with a focus on antibiotic resistance in *Klebsiella*. *Microorganisms* **2022,** *10* (4).

53. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F. O., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **2013,** *41* (Database issue), D590-6.

54. Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **2010,** *59* (3), 307-21.

55. Cury, J.; Jove, T.; Touchon, M.; Neron, B.; Rocha, E. P., Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* **2016,** *44* (10), 4539-50.

56. Zhang, Y.; Lu, J.; Wu, J.; Wang, J.; Luo, Y., Potential risks of microplastics combined with superbugs: Enrichment of antibiotic resistant bacteria on the surface of microplastics in mariculture system. *Ecotoxicol Environ Saf* **2020,** *187*, 109852.

57. Zhang, Y.; Yang, X.; Huang, L.; Zhang, R., Whole-genome analysis and description of an IMP-8-producing *Ochrobactrum anthropi*. *J Glob Antimicrob Resist* **2022,** *29*, 275-277.

58. Martini, M. C.; Quiroga, M. P.; Pistorio, M.; Lagares, A.; Centrón, D.; Del Papa, M. F., Novel environmental class 1 integrons and cassette arrays recovered from an on-farm bio-purification plant. *FEMS Microbiol Ecol* **2018,** *94* (3).

59. Gillings, M. R.; Xuejun, D.; Hardwick, S. A.; Holley, M. P.; Stokes, H. W., Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? *ISME J* **2009,** *3* (2), 209-15.

60. Martinez-Freijo, P.; Fluit, A. C.; Schmitz, F. J.; Grek, V. S.; Verhoef, J.; Jones, M. E., Class 1 integrons in Gram-negative isolates from different European hospitals and association with decreased susceptibility to multiple antibiotic compounds. *The Journal of antimicrobial chemotherapy* **1998,** *42* (6), 689-96.

61. 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. *The Journal of antimicrobial chemotherapy* **2007,** *59* (4), 746-50.

62. Dunwell, J. M.; Culham, A.; Carter, C. E.; Sosa-Aguirre, C. R.; Goodenough, P. W., Evolution of functional diversity in the cupin superfamily. *Trends Biochem Sci* **2001,** *26* (12), 740-6.

63. Krucinska, J.; Lombardo, M. N.; Erlandsen, H.; Estrada, A.; Si, D.; Viswanathan, K.; Wright, D. L., Structure-guided functional studies of plasmid-encoded dihydrofolate reductases reveal a common mechanism of trimethoprim resistance in Gram-negative pathogens. *Commun Biol* **2022,** *5* (1), 459.

64. Grape, M.; Motakefi, A.; Pavuluri, S.; Kahlmeter, G., Standard and real-time multiplex PCR methods for detection of trimethoprim resistance *dfr* genes in large collections of bacteria. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **2007,** *13* (11), 1112-8.

65. Stokes, H. W.; Gillings, M. R., Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol Rev* **2011,** *35* (5), 790-819.

66. Heir, E.; Lindstedt, B. A.; Leegaard, T. M.; Gjernes, E.; Kapperud, G., Prevalence and characterization of integrons in blood culture *Enterobacteriaceae* and gastrointestinal *Escherichia coli* in Norway and reporting of a novel class 1 integron-located lincosamide resistance gene. *Ann Clin Microbiol Antimicrob* **2004,** *3*, 12.

67. Lee, J. C.; Oh, J. Y.; Cho, J. W.; Park, J. C.; Kim, J. M.; Seol, S. Y.; Cho, D. T., The prevalence of trimethoprim-resistance-conferring dihydrofolate reductase genes in urinary isolates of *Escherichia coli* in Korea. *The Journal of antimicrobial chemotherapy* **2001,** *47* (5), 599-604.

68. Herrero-Fresno, A.; Larsen, I.; Olsen, J. E., Genetic relatedness of commensal *Escherichia coli* from nursery pigs in intensive pig production in Denmark and molecular characterization of genetically different strains. *J Appl Microbiol* **2015,** *119* (2), 342-53.

69. Mukherjee, S.; Chakraborty, R., Incidence of class 1 integrons in multiple antibiotic-resistant Gram-negative copiotrophic bacteria from the river Torsa in India. *Res Microbiol* **2006,** *157* (3), 220-6.

70. Hall, L. M.; Livermore, D. M.; Gur, D.; Akova, M.; Akalin, H. E., OXA-11, an extended-spectrum variant of OXA-10 (PSE-2) beta-lactamase from *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **1993,** *37* (8), 1637-44.

71. Evans, B. A.; Amyes, S. G., OXA β-lactamases. *Clin Microbiol Rev* **2014,** *27* (2), 241-63.

72. Huang, C.; Long, Q.; Qian, K.; Fu, T.; Zhang, Z.; Liao, P.; Xie, J., Resistance and integron characterization of *Acinetobacter baumannii* in a teaching hospital in Chongqing, China. *New Microbes New Infect* **2015,** *8*, 103-8.

73. Peng, Z. F.; Gao, D. S.; Yang, X.; Liu, H. Y.; Huangfu, H. P.; Wang, C. Q., Bla_{OXA-10} and PSE-1 genes located on class 1 integrons in *Gallibacterium anatis. Curr Microbiol* **2019,** *76* (8), 959-961.

74. Aibinu, I. E.; Pfeifer, Y.; Ogunsola, F.; Odugbemi, T.; Koenig, W.; Ghebremedhin, B., Emergence of β-lactamases OXA-10, VEB-1 and CMY in *Providencia* spp. from Nigeria. *The Journal of antimicrobial chemotherapy* **2011,** *66* (8), 1931-2.

75. Liu, S.; Xu, H.; Guo, X.; Li, S.; Wang, Q.; Li, Y.; Liu, R.; Gou, J., Emergence and genetic characterization of plasmid-encoded VIM-2-producing *Pseudomonas stutzeri* with novel integron In*1998* isolated from cerebrospinal fluid. *Infect Drug Resist* **2021,** *14*, 3415-3424.

76. Gaze, W. H.; Zhang, L.; Abdouslam, N. A.; Hawkey, P. M.; Calvo-Bado, L.; Royle, J.; Brown, H.; Davis, S.; Kay, P.; Boxall, A. B.; Wellington, E. M., Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *ISME J* **2011,** *5* (8), 1253-61.

77. Antunes, P.; Machado, J.; Peixe, L., Dissemination of sul3-containing elements linked to class 1 integrons with an unusual 3' conserved sequence region among *Salmonella* isolates. *Antimicrob Agents Chemother* **2007,** *51* (4), 1545-8.

78. Chang, C. Y.; Chang, L. L.; Chang, Y. H.; Lee, T. M.; Chang, S. F., Characterisation of drug resistance gene cassettes associated with class 1 integrons in clinical isolates of *Escherichia coli* from Taiwan, ROC. *J Med Microbiol* **2000,** *49* (12), 1097-1102.

79. Clark, N. C.; Olsvik, O.; Swenson, J. M.; Spiegel, C. A.; Tenover, F. C., Detection of a streptomycin/spectinomycin adenylyltransferase gene (*aadA*) in *Enterococcus faecalis*. *Antimicrob Agents Chemother* **1999,** *43* (1), 157-60.

80. Binh, C. T.; Heuer, H.; Kaupenjohann, M.; Smalla, K., Diverse *aadA* gene cassettes on class 1 integrons introduced into soil via spread manure. *Res Microbiol* **2009,** *160* (6), 427-33.

81. Yoon, M. H.; Ten, L. N.; Im, W. T.; Lee, S. T., *Methylibium fulvum* sp. nov., a member of the *Betaproteobacteria* isolated from ginseng field soil, and emended description of the genus *Methylibium*. *Int J Syst Evol Microbiol* **2007,** *57* (Pt 9), 2062-2066.

82. Stackebrandt, E.; Verbarg, S.; Frühling, A.; Busse, H. J.; Tindall, B. J., Dissection of the genus *Methylibium*: reclassification of *Methylibium fulvum* as *Rhizobacter fulvus* comb. nov., *Methylibium aquaticum* as *Piscinibacter aquaticus* gen. nov., comb. nov. and *Methylibium subsaxonicum* as *Rivibacter subsaxonicus* gen. nov., comb. nov. and emended descriptions of the genera *Rhizobacter* and *Methylibium*. *Int J Syst Evol Microbiol* **2009,** *59* (Pt 10), 2552-60.

83. Wei, L.; Si, M.; Long, M.; Zhu, L.; Li, C.; Shen, X.; Wang, Y.; Zhao, L.; Zhang, L., *Rhizobacter bergeniae* sp. nov., isolated from the root of *Bergenia scopulosa*. *Int J Syst Evol Microbiol* **2015,** *65* (Pt 2), 479-484.

84. Jin, L.; Ko, S. R.; Ahn, C. Y.; Lee, H. G.; Oh, H. M., *Rhizobacter profundi* sp. nov., isolated from freshwater sediment. *Int J Syst Evol Microbiol* **2016,** *66* (5), 1926-1931.

85. Imai, S.; Yoshida, R.; Endo, Y.; Fukunaga, Y.; Yamazoe, A.; Kasai, D.; Masai, E.; Fukuda, M., *Rhizobacter gummiphilus* sp. nov., a rubber-degrading bacterium isolated from the soil of a botanical garden in Japan. *J Gen Appl Microbiol* **2013,** *59* (3), 199-205.

86. Kasai, D.; Imai, S.; Asano, S.; Tabata, M.; Iijima, S.; Kamimura, N.; Masai, E.; Fukuda, M., Identification of natural rubber degradation gene in *Rhizobacter gummiphilus* NS21. *Biosci Biotechnol Biochem* **2017,** *81* (3), 614-620.

87. Linh, D. V.; Gibu, N.; Tabata, M.; Imai, S.; Hosoyama, A.; Yamazoe, A.; Kasai, D.; Fukuda, M., Complete genome sequence of natural rubber-degrading, Gram-negative bacterium, *Rhizobacter gummiphilus* strain NS21^{T}. *Biotechnol Rep (Amst)* **2019,** *22*, e00332.

88. Goto, M.; Kuwata, H., *Rhizobacter daucus* gen. nov., sp. nov., the causal agent of carrot bacterial gall. *International Journal of Systemic and Evolutionary Microbiology* **1988,** *38* (3).

89. Watanabe, M.; Igarashi, K.; Kato, S.; Kamagata, Y.; Kitagawa, W., Critical effect of H_2O_2 in the agar plate on the growth of laboratory and environmental strains. *Microbiol Spectr* **2022**, e0333622.

90. Tanaka, T.; Kawasaki, K.; Daimon, S.; Kitagawa, W.; Yamamoto, K.; Tamaki, H.; Tanaka, M.; Nakatsu, C. H.; Kamagata, Y., A hidden pitfall in the preparation of agar media undermines microorganism cultivability. *Appl Environ Microbiol* **2014,** *80* (24), 7659-66.

91. Bengtsson-Palme, J.; Larsson, D. G., Antibiotic resistance genes in the environment: Prioritizing risks. *Nat Rev Microbiol* **2015,** *13* (6), 396.

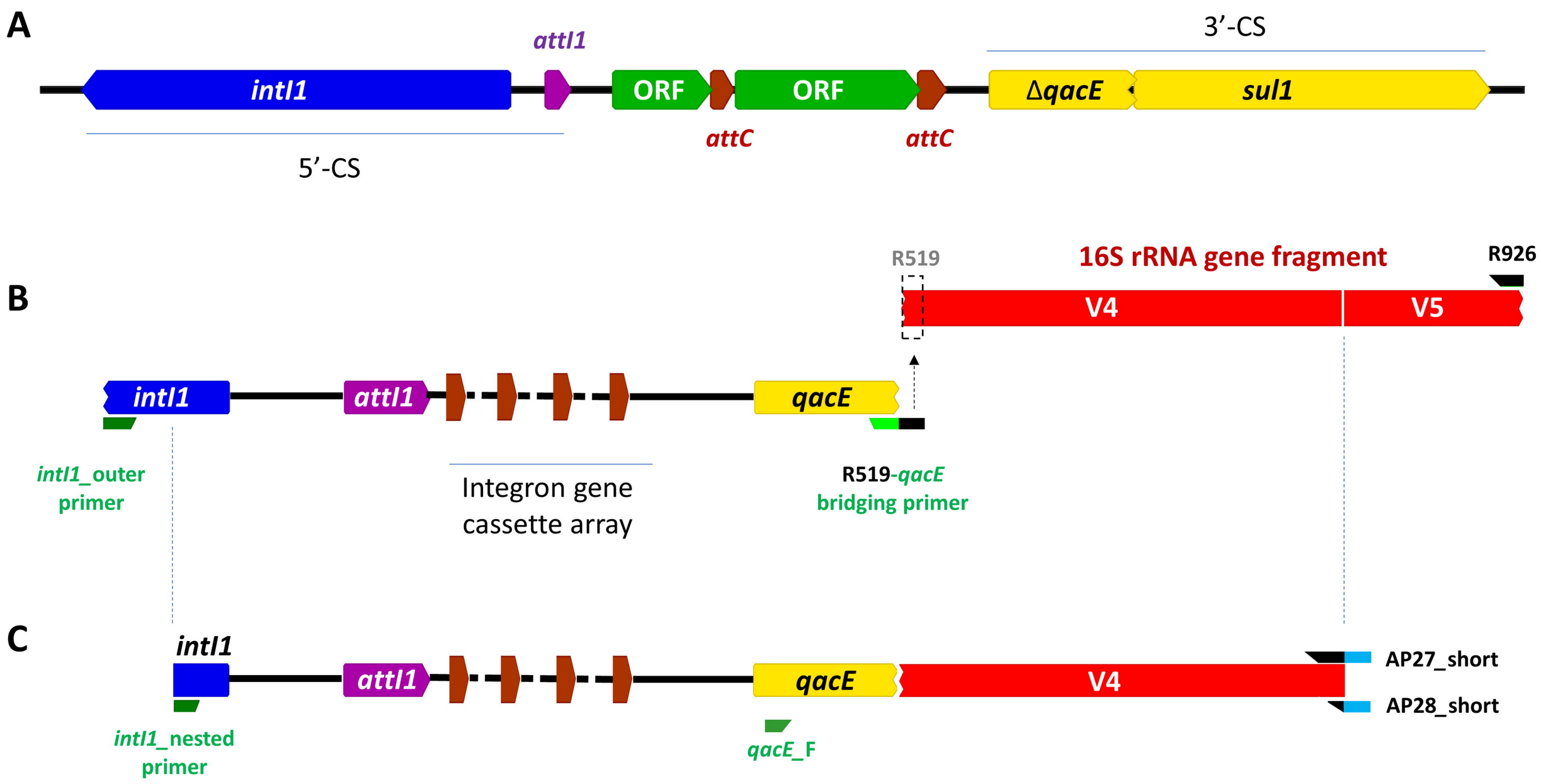
92. Ebmeyer, S.; Kristiansson, E.; Larsson, D. G. J., A framework for identifying the recent origins of mobile antibiotic resistance genes. *Commun Biol* **2021,** *4* (1), 8.

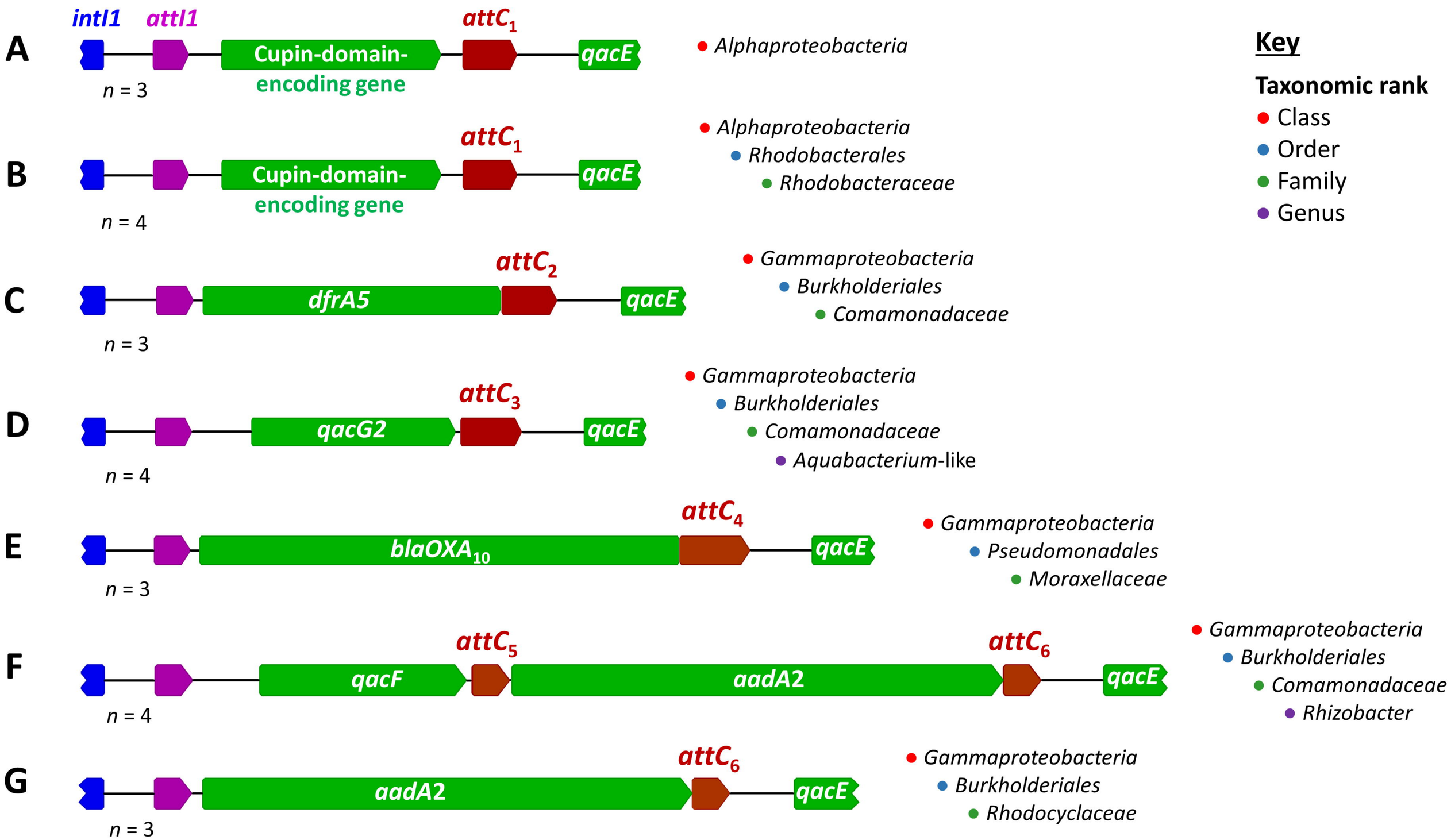
93. Böhm, M.-E.; Razavi, M.; Marathe, N. P.; Flach, C.-F.; Larsson, D. J., Discovery of a novel integron-borne aminoglycoside resistance gene present in clinical pathogens by screening environmental bacterial communities. *Microbiome* **2020,** *8* (1), 114.

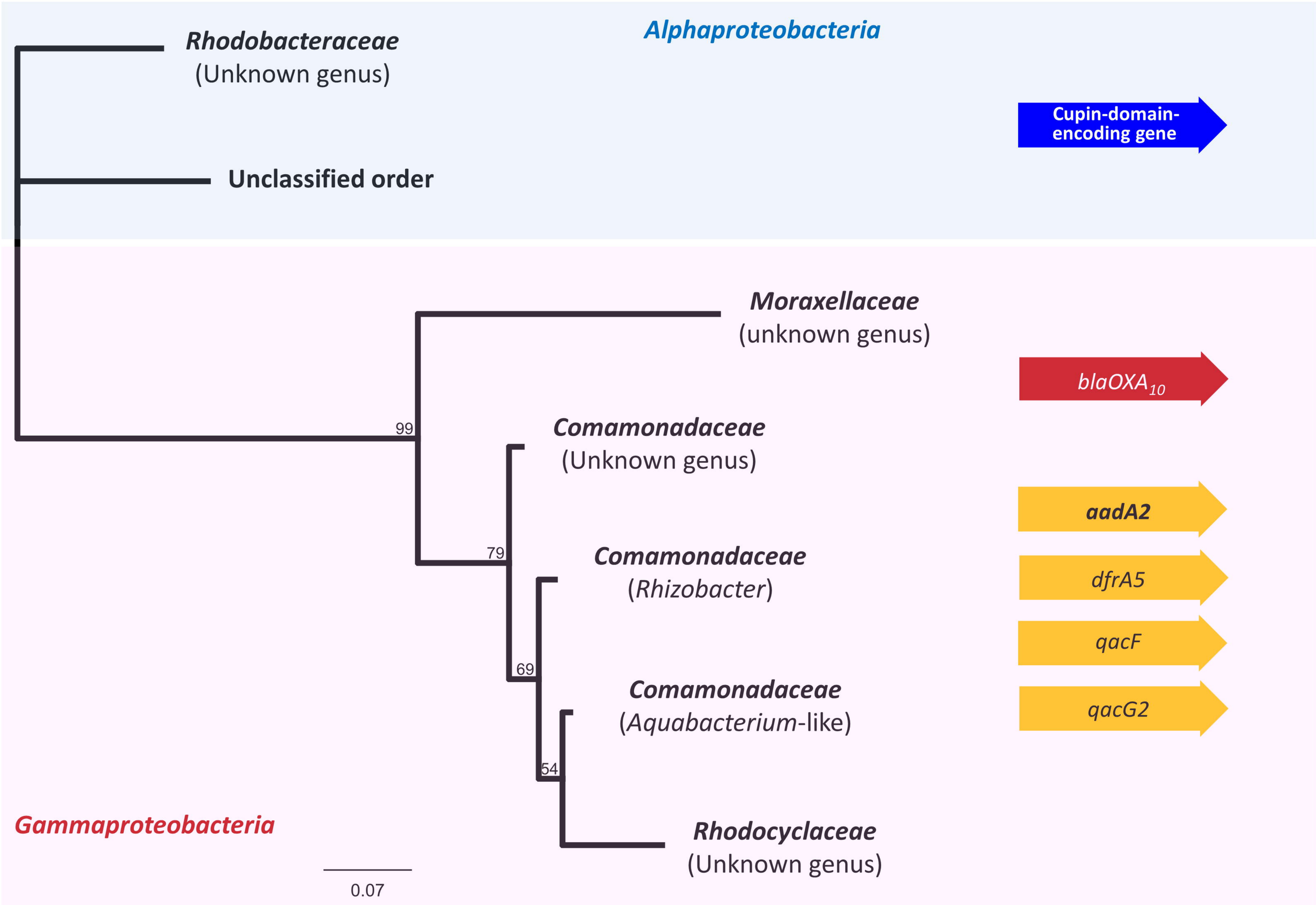
94. Bengtsson-Palme, J.; Jonsson, V.; Hess, S., What is the role of the environment in the emergence of novel antibiotic resistance genes? A modeling approach. *Environ Sci Technol* **2021,** *55* (23), 15734-15743.

95. Kurland, C. G.; Canback, B.; Berg, O. G., Horizontal gene transfer: A critical view. *Proc Natl Acad Sci U S A* **2003,** *100* (17), 9658-62.

96. Ghaly, T. M.; Gillings, M. R.; Penesyan, A.; Qi, Q.; Rajabal, V.; Tetu, S. G., The natural history of integrons. *Microorganisms* **2021,** *9* (11).


A
attI1
3'-CS
intI1
ORF
ORF
ΔqacE
sul1
attC
attC
5'-CS
B
R519
16S rRNA gene fragment
R926
V4
V5
intI1
attI1
qacE
intI1_outer primer
Integron gene cassette array
R519-qacE bridging primer
C
intI1
attI1
qacE
V4
AP27_short
AP28_short
intI1_nested primer
qacE_F


A
intI1
attI1
attC1
Cupin-domain-encoding gene
qacE
n = 3
Alphaproteobacteria
B
attC1
Cupin-domain-encoding gene
qacE
n = 4
Alphaproteobacteria
Rhodobacterales
Rhodobacteraceae
C
attC2
dfrA5
qacE
n = 3
Gammaproteobacteria
Burkholderiales
Comamonadaceae
D
attC3
qacG2
qacE
n = 4
Gammaproteobacteria
Burkholderiales
Comamonadaceae
Aquabacterium-like
E
attC4
blaOXA10
qacE
n = 3
Gammaproteobacteria
Pseudomonadales
Moraxellaceae
F
attC5
attC6
qacF
aadA2
qacE
n = 4
Gammaproteobacteria
Burkholderiales
Comamonadaceae
Rhizobacter
G
attC6
aadA2
qacE
n = 3
Gammaproteobacteria
Burkholderiales
Rhodocyclaceae
Key
Taxonomic rank
Class
Order
Family
Genus


Alphaproteobacteria
Rhodobacteraceae
(Unknown genus)
Unclassified order
Cupin-domain-encoding gene
Moraxellaceae
(unknown genus)
blaOXA10
99
Comamonadaceae
(Unknown genus)
aadA2
79
Comamonadaceae
(Rhizobacter)
dfrA5
qacF
69
Comamonadaceae
(Aquabacterium-like)
qacG2
54
Gammaproteobacteria
Rhodocyclaceae
(Unknown genus)
0.07