Characterization of urban wastewater treatment plant effluent from Tokyo using metagenomics and β-lactam-resistant Enterobacteriaceae isolates

Tsuyoshi Sekizuka  
National Institute of Infectious Diseases

Kentaro Itokawa  
National Institute of Infectious Diseases

Rina Tanaka  
National Institute of Infectious Diseases

Masanori Hashino  
National Institute of Infectious Diseases

Koji Yatsu  
National Institute of Infectious Diseases

Makoto Kuroda (✉ makokuro@niid.go.jp)  
National Institute of Infectious Diseases  https://orcid.org/0000-0002-0487-4405

Research

Keywords: urban sewage, effluent, metagenomics, Enterobacteriaceae, ESBL, carbapenemase, heavy-metal resistance

Posted Date: August 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-52275/v1

License: ☇ ☀ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background:** The World Health Organization has endorsed a global action plan on antimicrobial resistance (AMR), which calls upon all nations to adopt mitigation strategies based on the one-health approach. In particular, environmental factors to evaluate their potential risk for promoting the evolution of AMR have not been sufficiently characterized, yet. Instead of the limited nosocomial AMR surveillance, sewage AMR surveillance could highlight a broader picture of the global burden of AMR, including local-specific features, such as urban-, suburban-, industrial-, or agricultural-based, based on the One-Health approach.

**Methods:** We characterized environmental AMR burden in effluents of eight wastewater treatment plants (WWTPs) around the Tama River and Tokyo Bay by metagenomic analysis of AMR genes (ARGs) and heavy metal resistance genes and whole-genome sequencing of AMR bacteria (ARB) isolates.

**Results:** We found that seasonal features of ARGs eluted from WWTPs and sul1 and qacE (or qacEΔ1), which are the main contributors of class 1 integron, were rich genetic factors as baseline ARG pollutants. In addition, aadA6 and msr(E) were predominant ARGs against aminoglycoside and macrolide resistance, respectively. Among the detected ARB isolates, complete genome sequencing demonstrated that *E. coli* O16:H5-ST131-fimH41 producing CTX-M-27 exhibited marked genome clonality with clinical isolates from several countries. Complete plasmid sequences highlighted that ARG distribution was involved in plasmid replicon and its bacterial hosts.

**Conclusions:** Here, we provide a baseline for investigating environmental AMR dissemination and the possibility of using metagenomics for comprehensive AMR surveillance in sewage. This study demonstrated that AMR monitoring could uncover the actual status of the effluents from WWTP and would suggest how we should conduct further studies to reduce the related AMR burden to prevent it from becoming a potential health risk.

Introduction

The World Health Organization has endorsed a global action plan on antimicrobial resistance (AMR), which calls upon all nations to adopt mitigation strategies within two years [1]. However, there is still a need to fully understand the ecology and evolution of AMR based on the one-health approach. In particular, the properties of microbial resistome in ecosystems dominated by humans and how to monitor such environmental factors to evaluate their potential risk for promoting the evolution of AMR have to be sufficiently characterized. Instead of the limited nosocomial AMR surveillance, sewage AMR surveillance could highlight a broader picture of the global burden of AMR, including local-specific features, such as urban-, suburban-, industrial-, or agricultural-based, based on the One-Health approach.

There is growing concern for sludge management due to high levels of contaminants, and the design of current wastewater treatment plants (WWTPs) does not restrict the elimination of emerging contaminants and their metabolites. These contaminants are released into rivers or streams without interruption as sewage effluents. The prevalence of AMR bacteria (ARB) and AMR genes (ARGs) in the rivers would increase downstream of WWTPs because of the effluents [2, 3]. Indeed, the high density of bacteria in WWTPs could provide an optimum environment for horizontal gene transfer (HGT) among environmental bacteria and human pathogens. Moreover, wastewater and active sludge in WWTPs can act as reservoirs and environmental suppliers of antibiotic resistance, implying that they are hotspots for HGT under the selective pressure of antibiotics, disinfectants, and metals, even at low concentrations, enabling the dissemination of antibiotic resistance genes between different bacterial species. ARGs are located in multidrug-resistance (MDR) plasmids, which could have been transferred into broad recipient targets among different Proteobacteria strains, indicating a high possibility of HGT among bacteria in wastewater [4]. Therefore, urban WWTPs are among the main sources of both ARB and ARGs released into the environment [5]. However, the evolution of resistance and the spread of ARGs in WWTPs has still not been widely researched and clearly evidenced.

A novel study by an international team of scientists suggested that analyzing DNA in urban sewage from different countries may tell the full story of global AMR levels and provide a cheaper and easier method of conducting global AMR surveillance. [6]. The Technical University of Denmark (DTU) has globally collaborated for the Global Sewage Surveillance Project (https://www.compare-europe.eu/Library/Global-Sewage-Surveillance-Project) to analyze environmental AMR burden, pathogens, and other potential hazardous agents of regional-specific sewage influents from 79 sites in 60 countries using whole-genome sequencing (WGS) analysis. The DTU team analyzed the first round of collected influents of sewage samples in 2016 and demonstrated that metagenomic analysis of untreated sewage from 60 countries revealed a clear geographic distinction in AMR levels, with countries in Asia, Africa, and South America having more abundance and a higher variety of AMR genes under poor sanitation and public health than those in Europe, North America, and Oceania [6].

To estimate the risk of environmental exposure value of ARB, it would be necessary to investigate quantitative data in extensive sampling spots. qPCR and metagenomics approach could be cost-effective to highlight region-specific contamination. However, such ARG-based detections are not able to reveal the absolute abundance of the ARB of interest in the environmental source, and it may be insufficient to detect...
a specific vector and carrier as the most suspected risk factor. The risks of ARB by colonization or infection to humans from the environment must be assessed based on ARB clone genetics rather than by the prevalence of ARGs [7].

The purpose of this study was to characterize the environmental AMR burden from the effluent of WWTPs around the Tama River and Tokyo Bay area based on metagenomic analysis of ARGs and whole-genome sequencing of ARB isolates. This study demonstrated that AMR monitoring could uncover the actual status of the effluents from WWTP and would suggest how we should conduct further studies to reduce the related AMR burden to prevent it from becoming a potential health risk.

Results

Bacterial proportion in WWTP effluents

Surface water from the sewage effluent samples was collected from eight WWTP sites along the Tama River and around Tokyo Bay (Fig. 1) and the recreational beach area (BEC1) (Fig. 1). Basic information on the WWTPs is summarized in Table 1. The sampling was conducted in summer and winter seasons for two years between August 2017 and February 2019 (Table S1). The collected water samples were filtered through a 0.2-µm pore membrane, and DNA was extracted from the membrane, followed by metagenomic DNA-seq short-read sequencing (Fig. S1).

Taxonomic classification of domain rank (Fig. S2A) suggested that the effluents of fresh water area (WP1–8) were mostly composed of bacterial sequences, whereas those of brackish water (WP9 and BEC1) were rich in eukaryote organisms, such as mussels (Mytilus edulis) and diatoms (Chaetoceros spp.), found in seawater. Further taxonomic classification of family rank (Fig. 1B) suggested that fresh water area (WP1–8) effluents predominantly contained Comamonadaceae, which are aerobic gram-negative β-proteobacteria as reported in sewage active sludge digesters [8], and Mycobacteriaceae, which are rich organisms in sewage active sludge [9]. The effluents of brackish water (WP9 and BEC1) were rich in Rhodobacteriaceae, which mostly habitats aquatic environments.

Non-metric multidimensional scaling (NMDS) showed that plots of the effluents of fresh water area (WP1–8) were closely scattered, but the plots were markedly divided in a season-dependent manner (Fig. 2A). Linear discriminant analysis effect size (LEfSe) analysis revealed that multiple aquatic habitat bacteria were mostly observed in winter, whereas some specific members of the human gut flora (Streptococcus, Escherichia, and Anaerococcus), Cyanobacteria (Dolichospermum), and purple sulfur bacteria (Rheinheimera) increased in summer (Fig. 2B), demonstrating typical seasonal features.

ARG proportion in WWTP effluents

Abovementioned metagenomic DNA-seq short reads were analyzed by Blastn homology search in the ResFinder ARGs database under normalization using reads per kilobase of exon per million mapped reads (RPKM). Prior to the evaluation of ARGs, we assessed whether all metagenomic DNA-Seq short reads were properly obtained to characterize the bacterial population size using RPKM normalization for 16S rRNA genes and rpoB RNA polymerase β-subunit orthologs (Fig. S2B). All samples, except for BEC1, in winter 2019 showed no significant difference with an average of four copies of the 16S rRNA gene per one copy of rpoB, suggesting that RPKM estimation is an effective approach to characterize the relative abundance of ARGs. Resistance genes to quaternary ammonium compounds (QAC), sulfonamide (SA), aminoglycoside (AMG), β-lactam, and macrolide (MAC) were markedly detected in effluents from all WWTPs, except for WP9 and BEC1 (Fig. 3A). Unlike the bacterial proportion shown in Fig. 2A, a seasonal difference was not observed in ARG variations (Fig. 3B), but it appeared to be dependent on sampling year and season.

We speculated that possible occasional trends of ARB dissemination may affect ARG variations at every sampling and WP site. The detected ARGs in each AMR category were further classified into orthologous genes (or gene families) (Fig. 4A and Table S2). The most detected ARG was the SA resistance gene sul1 at 440.5 RPKM (Table S2); sul2 was detected at rather low levels, but sul3 was not detected. The second most detected ARG was the QAC resistance gene qacE at 203.8 RPKM (Table S2), suggesting that RPKM estimation is an effective approach to characterize the relative abundance of ARGs. Resistance genes to quaternary ammonium compounds (QAC), sulfonamide (SA), aminoglycoside (AMG), β-lactam, and macrolide (MAC) were markedly detected in effluents from all WWTPs, except for WP9 and BEC1 (Fig. 3A). Unlike the bacterial proportion shown in Fig. 2A, a seasonal difference was not observed in ARG variations (Fig. 3B), but it appeared to be dependent on sampling year and season.

Regarding clinically important ARGs, multiple AMG- and MAC-resistant genes were identified (Fig. 4A). The seven most predominant AMG resistance genes belonged to the aadA and aph gene families (Table S2), which contribute to streptomycin resistance, implying that streptomycin could be the most potent selective agent among aminoglycosides. The two most predominant MAC resistance genes were msn(E) and mph(E), which are primarily located on the chromosomes of Acinetobacter and Proteus species (The Comprehensive Antibiotic Resistance Database, CARD https://card.mcmaster.ca/ontology/39685).
Regarding β-lactamase gene type, \( \text{bla}_{\text{OXA}} \), \( \text{bla}_{\text{GES}} \), and \( \text{bla}_{\text{MP}} \) were the most markedly detected (Fig. 4A), whereas the \( \text{bla}_{\text{CTX-M}} \) cluster, which is a popular extended-spectrum β-lactamase (ESBL) gene in clinical settings, was in observed in minute abundance at 0.82 RPKM (Table S2), indicating that other β-lactamase type genes, such as \( \text{bla}_{\text{OXA}} \) and \( \text{bla}_{\text{GES}} \), may be present in baseline levels were derived from environmental bacteria but not pathogenic bacteria.

In addition to the distribution of ARG, metagenomic analysis suggested a unique season-dependent ARG detection. For instance, \( \text{qnrS2}, \text{qnrS6}, \) and \( \text{aad(6')-Ib} \) increased markedly in summer (Fig. 4B), whereas \( \text{aadA17} \) was detected in winter (Fig. 4C). These season-dependent ARG detection patterns suggested that, to some extent, ARGs could be affected by seasonal elements in a particular temperature-dependent manner (above 30°C in summer and below 10°C in winter; Table S1) similar to that observed in the growth of ARB in activated sludge of the anaerobic–anoxic–oxic (A2O) water treatment system.

However, another possibility remains that seasonal (or temporal) infectious trends may influence the differences in antimicrobial use in the WP-dependent treated area, leading to differential human gut flora as excreta. Thus, possible WP-dependent ARG differences were elucidated using a fold change \((\log_2 \text{ratio})\) of RPKM (Fig. S3A). In fact, season-dependent ARGs as well as steadily increasing ARGs were observed in WP-specific features (Fig. S3B). WP2 effluents showed that the RPKM of major ARGs \([\text{AMG}, \text{aadA2} \text{ and } \text{aadA4} ; \text{MAC}, \text{msr(E)} \text{ and } \text{mph(E)}; \text{SA, sul1}; \) and \( \text{QAC, qacF} \) increased steadily (Fig. S3B), suggesting that the population of major ARB species may increase predominantly. Likewise, the WP-specific increase in ARGs was identified at other WPs, and metagenomic RPKM value could be a suitable index as one of the evaluation standards to monitor such increasing trend of ARB in WWTP-specific areas.

### Heavy-metal resistance genes (HMRGs) in WWTP effluents

Resistance genes to mercury (Hg), copper (Cu), and arsenate (As) were markedly detected in effluents from all WWTPs (Fig. S4A and Table S3). Unlike the bacterial proportion shown in Fig. 2A, seasonal difference was not observed in heavy-metal resistance (Fig. S4B), but it appeared to be dependent on sampling year and season as well as on ARGs (Fig. S5).

### Genome analysis of ESBL/carbapenemase-producing organisms (EPO and CPO) isolated from WWTP effluents and environment

The metagenomic approach is good for revealing overall ARGs, but it is insufficient to characterize potentially pathogenic ARB. Therefore, we isolated EPO and CPO from fresh 50-mL samples of each effluent (Fig. S6). CHROMagar ESBL selection demonstrated that potential clinically pathogenic Enterobacteriaceae (\( \text{E. coli}, \text{Klebsiella}, \text{and Enterobacter} \) spp.) were observed markedly at approximately 500 colonies from the effluents of WP4, WP5, and WP8 at all tested sampling times (Fig. S6). More than hundred typical chromogenic colonies, which would be sufficient to characterize environmental pollution due to EPO and CPO, were isolated at each sampling time (Table 2). These isolates were subjected to whole-genome sequencing (Fig. S6 and Table S4). CTX-M-positive organisms were isolated as follows: 134 (31%) (Table 2), CTX-M-9 group \((\text{bla}_{\text{CTX-M}})$, \( \text{bla}_{\text{CTX-M-13}}, \text{bla}_{\text{CTX-M-15}}, \text{bla}_{\text{CTX-M-14}}, \text{bla}_{\text{CTX-M-24}, \text{bla}_{\text{CTX-M-27}}, \text{and } \text{bla}_{\text{CTX-M-58}}) \) was the predominant group with 71 isolates (53%), and the second was the CTX-M-1 group \((\text{bla}_{\text{CTX-M-3}}, \text{bla}_{\text{CTX-M-15}}, \text{and } \text{bla}_{\text{CTX-M-55}}) \) with 48 isolates (36%) (Table S4). The EPO and CPO isolates on CHROMagar ESBL and its sequence type (ST) are summarized in Table S5. Eighteen CPO isolates (seven \( \text{bla}_{\text{PER}} \)-positive, eight \( \text{bla}_{\text{KPC}} \)-positive, and three \( \text{bla}_{\text{NDM}} \)-positive) were identified (Table 2).

Among the 71 \( \text{E. coli} \) isolates, nine isolates of predominant ST131 (11 isolates) possessed \( \text{bla}_{\text{CTX-M-27}} \) gene (Table S4). Core-genome single nucleotide variation (SNV) phylogeny (see detailed procedures in Fig. S7) of the ST131 isolates were compared with publicly available \( \text{E. coli} \) ST131 draft or complete genome sequences (a total of 315 strains; see Table S6), suggesting that primarily four subclonal types of ST131 were identified from the WWTP effluents, two isolates of ST131-\( \text{fimH41} \) with \( \text{bla}_{\text{CTX-M-27}}, \) seven isolates of ST131-\( \text{fimH80} \) with \( \text{bla}_{\text{CTX-M-27}}, \) one isolate of ST131-\( \text{fimH80} \) with \( \text{bla}_{\text{CTX-M-55}}, \) and one isolate of ST131-\( \text{fimH80} \) with \( \text{bla}_{\text{CTX-M-15}} \) (Fig. S8).

One isolate (WP5-S18-ESBL-09) of O16:H5-ST131-\( \text{fimH41} \) with \( \text{bla}_{\text{CTX-M-27}} \) exhibited marked clonality, showing less than 12 SNVs with clinical isolates from several countries (Southeastern Asia, EU, and Oceania) (Fig. 5A), and genome recombination and pan-genome analysis suggested that the genome structure of these strains was very stable, except for some ARGs that were transferred with conjugative plasmid (Fig. 5B). Three isolates of ST131-\( \text{fimH80} \) with \( \text{bla}_{\text{CTX-M-27}} \) exhibited marked clonality, showing 8–13 SNVs to clinical isolate SAMD00126441 in Japan (Fig. S8), suggesting that a clinically potential subclone \((\text{O25:H4-ST131-\text{fimH80} with } \text{bla}_{\text{CTX-M-27}}) \) has been detected in WWTP effluent from Japan. In addition to \( \text{E. coli} \) ST131, ST38 (10 isolates), ST10 (five isolates), ST405 (four isolates), ST69 (three isolates), and ST648 (three isolates) were observed in this study (Tables S4 and S5). Core genome SNV phylogeny for each ST isolate, including publicly available strains (Table S6), suggested that CTX-M-14 is one of the major CTX-M variants in \( \text{E. coli} \) ST38, ST69, ST405, and ST648 (Fig. 6). For
instance, two isolates of ST69 (WP7-S17-ESBL-01 and WP8-S17-ESBL-12) harbored \( \text{bla}_{\text{CTX-M-14}} \), but relative clones in other countries harbored \( \text{bla}_{\text{CTX-M-15}}, \text{bla}_{\text{CTX-M-24}}, \) and \( \text{bla}_{\text{CTX-M-27}} \) (Fig. 6).

Among the hundred isolates of *Aeromonas* species (Table 2), 10 isolates of *A. hydrophila*, 10 isolates of *A. caviae*, one isolate of *A. media*, and four isolates of *A. veronii* were identified as CTX-M producers (Table S4). Although four isolates of CTX-M-14 producing *A. caviae* were detected in different WWTP effluents (WP2, 4, and 7), core-genome SNV phylogeny revealed that they were clonal in 21 SNVs with same recombination and pangenome regions (Fig. S9A), suggesting that the *A. caviae* clone may be successfully predominating in the general WWTP environment in Tokyo but not in WWTPs at specific locations.

Among the 51 isolates of *Klebsiella* species (Table 2), eight isolates of *K. pneumoniae*, seven isolates of *K. quasipneumoniae*, two isolates of *K. variicola*, and seven isolates of other *Klebsiella* sp. were identified as CTX-M producers (Table S4). Core-genome SNV phylogeny revealed that four isolates of CTX-M-3-producing *K. quasipneumoniae* ST668 exhibited clonality in 14 SNVs and shared similar recombination and pangenome regions (Fig. S9B), although these isolates were obtained from the same place (WP5) but at different sampling times (summer 2017, summer 2018, and winter 2019), suggesting that the ST668 clone may remain in active sludge at WP5 for at least one and a half years.

### Plasmidome analysis of EPO and CPO isolated from WWTP effluents and environment

ESBL and carbapenemase β-lactamase genes are acquired by conjugative plasmid transfer among variable strains in Proteobacteria. Genetic network analysis among the AMR plasmids could be useful to uncover the mode of relational transfer and to trace dissemination. Of the potential 433 EPO/CPO isolates (Table 2), 85 strains were identified as complete genome sequences in this study (Table S4). Of the 304 complete plasmid sequences, 73 β-lactamase-positive plasmids (Table S7) were subjected to plasmidome network analysis based on orthologous analysis (see details in Fig. S10). To characterize global ARG dissemination through plasmid transfer, 758 publicly available complete plasmid sequences showing clear description of isolation and source were selected from 19,904 complete plasmid sequences (from NCBI in November 2019). A total of 831 complete plasmids (73 in this study and 758 in public domain; see Table S8) were subjected to pangenome analysis using Roary and NMDS (vegan in R package) and visualized; the results revealed that 14 communities were classified and clustered (Fig. 7).

Most network communities (Co) showed notable incompatibility (Inc) replicon-dependent distribution, suggesting that the network analysis was well-performed based on the genetic features of each Inc replicon. The major Co comprised Co10, Co11, and Co12 of the shared 423 plasmids among all tested 831 plasmids. Co10 was primarily IncFIA and IncFIB (AP001918) replicon plasmids harboring \( \text{bla}_{\text{CTX-M}} \) gene in *E. coli* from human sources, whereas Co12 was primarily IncFIA(HI1) and IncFIB(K) replicon plasmids harboring \( \text{bla}_{\text{CTX-M}}, \text{bla}_{\text{KPO}}, \) and \( \text{bla}_{\text{GES}} \) genes in *Klebsiella* from human sources (Fig. 7). This network analysis highlighted the potential mutual role of Co11 plasmids between Co10 and Co12 plasmids because Co11 plasmids are placed around Co10 (rich in *E. coli*) and next to Co12 (rich in *Klebsiella*). Co11 is primarily rich in IncFII, IncFil(pH7A8), IncX1, and IncN replicon plasmids harboring \( \text{bla}_{\text{CTX-M}}, \text{bla}_{\text{KPO}}, \) and \( \text{bla}_{\text{NDM}} \) genes and mainly detected in *E. coli* from human and animal sources. Therefore, Co11 plasmids may contribute to the horizontal ARG transfer between *E. coli* and *Klebsiella*. Regarding other notable Co-exhibiting specific replicon and host bacteria, Co3 comprised IncHI2 and IncHI2A replicon plasmids harboring \( \text{bla}_{\text{CTX-M}} \) and \( \text{bla}_{\text{IMP}} \) genes and was mainly detected in *Enterobacter* from human sources. Co4 comprised IncB/O/K/Z and IncI1-gamma replicon plasmids harboring \( \text{bla}_{\text{CTX-M}} \) gene and were primarily detected in *E. coli* from human, environmental, and animal sources. From the viewpoint of carbapenemase genes, gene-specific distribution was found as follows: \( \text{bla}_{\text{IMP}} \) in Co1 and Co3; \( \text{bla}_{\text{KPC}} \) in Co6, Co8, and Co12; \( \text{bla}_{\text{NDM}} \) in Co5, Co11, and Co13. Such an ARG distribution could be involved in plasmid replicon and its bacterial hosts.

### Discussion

We have demonstrated that urban WWTP effluent could be a potential source of AMR burden by the detection of ARGs using a metagenomic approach and whole genome analysis of the ARB isolates. Furthermore, there is concern that hospital and community effluents include a considerable proportion of ARGs in WWTPs [12]. Hospitals are not the only sources of antibiotic resistance in the environment. According to the European Centre for Disease Prevention and Control, a significant proportion of antibiotics consumed by humans are in the community rather than in healthcare settings [13], suggesting that outpatient therapy could be the most effective factor in increasing the proportion of antimicrobials, selected ARGs, and ARB in WWTP. Therefore, a nation (or regional)-wide survey of ARGs and ARB in WWTP effluents could provide a rapid and efficient method for assessing environmental AMR burden from urban populations. As suggested in this study, ARG profiles in WWTP could then reflect the structure and diversity of resistant bacteria in the gastrointestinal tracts of urban residents within the WWTP catchment area (Figs. 3 and 4). This may be particularly true when the WWTP primarily treats domestic wastewater without significant contributions from agricultural and industrial sources [14]. Through the two-year investigation, we found some bacterial hosts and ARGs (Fig. 4B and C) that increased seasonally (Fig. 2).
The risk assessment of transmission of AMR vectors from the environment to humans cannot be adapted from the model for pathogens because most AMR vectors are supposed to be composed of non- or low-pathogenic bacterial species. Therefore, they can colonize humans without notable symptoms, resulting in a healthy carrier [7]. As colonization may be asymptomatic in most humans, this may cause further dissemination under frequent abuse of antimicrobials in clinical use, leading to an underestimation of the extent of transmission of ARB from the environment to humans as well as to humans in the community. Therefore, such reservoir ARBs should be considered for the overall AMR risk assessment. A major reservoir of antibiotic-resistant gram-negative bacteria is the gut microbiota of humans and animals, particularly apparent after exposure to antibiotics. The most active potential bacteria that play a pivotal role as carriers and vectors appear to be members of the classes γ-proteobacteria and β-proteobacteria and of the phyla Actinobacteria and Firmicutes [15]. Members of the family Enterobacteriaceae and of genera such as Aeromonas, Acinetobacter, Pseudomonas, Enterococcus, and Staphylococcus have been frequently documented as carriers in wastewater samples, and some of them may contribute to the role of vector for HGT. Although nonpathogenic carrier bacteria are not able to colonize and infect humans, their proliferation and subsequent dissemination in the environment would increase the abundance and diversity of ARGs in vectors. Hence, it may increase the risk of transmission of pathogenic ARB, such as hypervirulent Klebsiella, to humans.

Thus far, the healthy carrier rate of EPO has been rising worldwide. In Japan, the detection rate of EPO was reported to be 12.2% in healthy adult volunteers [16] and 15.6% in healthy food handlers [17]. Because the number of ESBL-positive healthy carriers has been increasing around the world, Japan is no exception to the issue of increasing AMR healthy carriers. Indeed, according to the Japan Nosocomial Infections Surveillance (JANIS), clinical reports of antibiotic-resistant gram-negative bacteria are increasing (https://janis.mhlw.go.jp/english/index.asp).

In addition to nosocomial AMR surveillance, the WWTP metagenomics approach will prevent the dissemination of the bacterial flora of predominantly healthy individuals carrying notable ARG/Bs.

Among potential EPO healthy carriers, globally disseminated E. coli ST131 clones should be first investigated. This study suggested that multiple ST131 clones in the WWTP effluents harbor the currently prevalent CTX-M variants (Fig. S8). E. coli O16:H5-ST131-fimH41 with blaCTX-M-27 is a marked clone (Fig. 5). Indeed, the increase in this clonal group harboring blaCTX-M-27 has been reported in China [18], EU [19], and Oceania countries [20], which are one of the marked clonal disseminations associated with ST131 CTX-M-27 producers in Japan [21] because this increase so far has been speculated by the observation that CTX-M-27 producers have a higher MIC of ceftazidime compared with that of CTX-M-14 producers [22]. In contrast to ST131 with blaCTX-M-27, other E. coli STs still harbor blaCTX-M-14, which is the predominant variant since the early 20th century in Japan [21], suggesting that these other STs have not been shifted to acquire more potential CTX-M-27 thus far.

Plasmidome network analysis (Fig. 7) exhibited a good approach to illuminate bacterial communication through plasmid-based HGT. This study highlighted bacterial species-dependent plasmid tropism and its compatibility among distinct species. The blaCTX-M variants were identified in wide variable plasmids, whereas carbapenem genes (blaNDM, blaKPC, and blaIMP) relatively showed a typical plasmid replicon with specific distribution, suggesting that the regional-specific distribution of carbapenemase [23] may be associated with its locality [24]. However, we have reported NDM-5- and CTX-M-55-coproducing E. coli GSH8M-2 isolated from WWTP effluent (WP8) in Tokyo Bay [25]. Such co-producers will be the next clonal dissemination with powerful activity against multiple β-lactams. Therefore, such network analysis will be useful for the prediction of future clonal expansion in local to global dissemination.

The diverse bacterial population from household, hospital, and industrial wastewaters reach the WWTP where conditions for HGT are optimal due to the formation of biofilms in activated sludge and stress caused by various pharmacological and heavy metal contamination [26]. It is now well-demonstrated that even low concentrations of antibiotics can result in the selection of ARGs [27]. Furthermore, the fluoroquinolone antibiotic ciprofloxacin markedly induces reactive oxygen species (ROS), leading to a mutant-generating cell subpopulation and homologous recombination [28]. Such an SOS response in bacteria increases the mutation rate by increasing the expression of error-prone DNA polymerases [29]. Thus, it is very difficult to standardize the upper limit of an antibiotic concentration in wastewater. When the antibiotic concentration is low by dilution and active processing in WWTPs, non-corresponding contaminants (e.g., heavy metals, organic pollutants, and physical and chemical factors) may share or substitute the role of antibiotics and contribute to the propagation of ARGs in the environment.

The main route of ARG dissemination may switch from active transmission, which is dominated by antibiotics, to passive transmission, which is motivated and affected by non-corresponding contaminants [30]. In fact, Azuma et al. reported the concentration of pharmaceutical compounds (PhCs) in the Japanese environment and demonstrated 48 PhCs in the range of 1 ng/L (losartan carboxylic acid) to 228 μg/L (acetaminophen sulfate) in hospital effluent, and the contribution of the pollution load derived from hospital effluent to WWTP influent was estimated as 0.1–15% [31].

Heavy metals are also one of the most important contaminants as selective pressure to ARB [32, 33]. Unlike antibiotics, metals are not subjected to degradation and can subsequently represent prolonged selective pressure. Overall, heavy metals may be one of the dominant factors in estuaries and marine environments, and they can also play an important role in the maintenance and proliferation of ARGs when
antibiotic-selective pressure is weak. In particular, heavy metals are used as feed supplements, followed by accumulation in manures, thus, indicating the potential for co-selection of ARGs [34]. Significant correlations exist between ARGs and heavy metals (Cu, Zn, and As), suggesting that the contamination of ARGs is related to the chemical pollution of the sediment [32]. The presence of heavy metals provides another co-selection pressure for ARG [34, 35]. Such co-selection would generate a co-resistant phenotype located together on the same genetic determinant responsible for resistance to antibiotics and other contaminants. In particular, even short-term copper stress significantly enhances the abundance of ARGs and mobile genetic elements (MGEs) along with increasing Cu concentrations and can significantly change the potential of soil ARGs [36].

Many societies still face water scarcity due to increasing population and land use, and conventional water resources are insufficient in many regions to meet the needs of the growing population. Any water treatment or distribution system should focus on decreasing the discharge of ARB because transmission of ARGs is evidenced by the findings that non-potable wastewater harbors a higher number of ARGs [37] and that these ARGs are distributed across a broad range of microorganisms rather than any specific group [38]. In addition to conventional standard treatment (biological treatment, followed by A₂O and chlorination for disinfection), advanced treatment technologies, such as photocatalysis, membrane filtration, activated carbon adsorption, and advanced oxidation processes (AOPs), should be implemented to remove emerging contaminants from wastewater [39].

Conclusions

Urban WWTP effluent, even with proper treatment, may cause AMR burden with a high frequency of acquired ARGs in the environment. The dissemination of ARB/Gs in the environment can increase the risk of infectious diseases [40], but there is scarce direct evidence about their epidemiological effects. This study revealed the presence of ARB/Gs in WWTP effluents, suggesting that every urban community, including hospitals, healthy carriers, and travelers, can be a potential source of ARB/Gs. WWTP is speculated to be a hotspot where various bacteria can acquire ARGs. Monitoring of ARB/Gs in WWTP effluent is expected to identify the presence of undetected nosocomial infections and healthy carriers at the earliest, leading to the detection of potential ongoing dissemination of ARB/Gs in the overall community.

Materials And Methods

Sample collection

Sewage effluent samples were collected from eight WWTPs (WP1–WP9, and WP6 was treated as missing number due to incorrect sampling locations) along the Tama River and around Tokyo Bay (Table 1 and Fig. 1). As an environmental control sample, surface water from a recreational beach (BEC1) was collected (Fig. 1). Four sampling months during the two years were August 2017, February 2018, August 2018, and February 2019 (Table S1). The water samples in this study were collected on five consecutive days without recent rainfall to exclude the effect of weather. All samples were collected in 500-mL sterilized containers (Coming® Easy-Grip round, plastic storage bottles of 500 mL). Detailed information on these samples is summarized in Table S1.

Metagenomic DNA-seq analysis of water samples

Brief experimental procedures are shown in Fig. S1. To trap and collect organisms with size bigger than that of bacteria, water samples were filtered with TPP Rapid Filtermax Vacuum Filtration (TPP, Trasadingen, Switzerland) in 500-mL bottle with large 49-cm² PES 0.2-µm membrane. The membrane was removed by scraping from the bottle, and the recovered membrane was stored at −30 °C before DNA extraction. One-fourth square of the collected membrane was cut into small pieces and placed into ZR-96 BashingBead™ Lysis Tubes (0.1 and 0.5 mm; Zymo Inc., Irvine, CA, USA). Roche Bacterial lysis buffer (800 µL) was added to the bead tube and frozen at −30 °C and thawed at 23 °C. The tube was subjected to bead-beating (1,500 rpm for 10 min) with GenoGrinder 2010. After brief centrifugation (8,000 × g for 3 min), 400 µL of the upper supernatant was subjected to DNA purification using the Roche MagNa Pure Compact instrument (DNA_Bacteria_v3 protocol; Elution: 50 µL). DNA concentrations and purity were measured using a Qubit DNA HS kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

A metagenomic DNA-seq library was prepared using the QIAsymphony FX DNA library kit (QIAGEN, Hilden, Germany), followed by short-read sequencing using a NexSeq 500 platform (2 × 150-mer paired-end) (Illumina, San Diego, CA, USA). Adapter and low-quality sequences in short-read data were trimmed using Sickle version 1.33 (https://github.com/najoshi/sickle) with the following parameters: average quality threshold, “-q 20” and minimum length threshold, “-l 40”. The metagenomic analysis was performed using KrakenUniq version 0.5.8 [41] with default parameters. For metagenomic database construction, DNA sequences were gathered from the NCBI nucleotide database from KrakenUniq in June 2019 as follows: complete genome sequences of bacteria (n = 13,737), archaea (n = 295), and virus (n = 8,972); human genome sequences (GRCh38.p13); and nucleotide sequences in the nucleotide database of bacteria (n = 7,041,584), archaea (n = 355,027), virus (n = 2,386,054), fungi (n = 4,520,379), and protozoa (n = 1,670,542). Subsequently, the database with these sequences was built using the
KrakenUniq build program with a K-mer length of 31 bp. All metagenomic data were summarized with Pavian version 1.0.0 [42]; rare genera with one sequence count in one sample were excluded. Alpha diversity indices with InvSimpson diversity were calculated with the R package “vegan” version 2.5.6 (https://CRAN.R-project.org/package=vegan). NMDS ordination of the relative abundance was generated at the genus level, and NMDS plots were visualized using method metaMDS from the “vegan” with Bray–Curtis distances. In an attempt to identify metagenomic biomarkers, LEfSe analysis was performed with relative abundance at the genus level in bacteria [43]. In this study, p < 0.05 was considered significant for both statistical methods. ARGs with markedly differential reads per kilobase per million mapped reads (RPKM) numbers were defined as those with a linear discriminant analysis (LDA) score (log_{10}) of > ± 2.0.

**Resistome analysis**

For resistome analysis, database construction of ARGs and HMRGs was performed. A database was also constructed with 16S rRNA and rpoB genes for evaluation of resistome analysis. DNA sequences of ARGs and 16S rRNA genes were obtained from the Bacterial Antimicrobial Resistance Reference Gene Database (NCBI BioProject ID, PRJNA313047), ResFinder database (https://bitbucket.org/genomicepidemiology/resfinder_db/src/master/) [44], and SILVA database (https://www.arb-silva.de/no_cache/download/archive/current/Exports/) [45], followed by database construction using Makeblastdb in BLAST + program. The protein sequences of HMRGs (both confirmed and predicted genes) and RpoB were downloaded from BacMet version 2.0 (http://bacmet.biomedicine.gu.se/download_temporary.html) [46] and the InterPro database (http://www.ebi.ac.uk/interpro/entry/tigrfams/TIGR02013/protein/UniProt#/table) [47], followed by database construction using RAPSearch v2.24 [48]. Operational taxonomic units (OTUs) of ARGs and HMRGs in the database were created by clustering at ≥ 90% sequence identity and ≥ 80% coverage using vsearch version 2.10.4 [49]. These DNA and protein sequences in metagenomic read data were searched using mega-BLAST (e-value threshold, 1E-20; identity threshold, 95%) and RAPSearch (e-value threshold, 1E-10; identity threshold, 90%), respectively. The detected genes were summarized for each OTU of ARGs and HMRGs. For normalization, RPKM was calculated using the following formula: RPKM = number of detected reads against OTUs/average gene length of detected OTUs (bp) × total number of trimmed reads) × 10^9. The NMDS plot for resistome analysis was also visualized as described in the previous section. Statistical analysis was performed by exact Wilcoxon rank–sum test using the R package “exactRankTests” version 0.8.30, followed by multiple comparisons by the Benjamini–Hochberg false discovery rate (FDR) procedure (q-value) using the R package “FSA” version 0.8.30. Differences were considered significant with a p-value of < 0.05 and q-value of < 0.4.

**Whole-genome sequencing of bacterial isolates**

The experimental procedures are shown in Fig. S2. Freshly collected 50 mL of water sample was centrifuged at 7,000 × g for 3 min, and the cell pellets were suspended with residual water (~ 500 μL) spread on CHROMagar ESBL plates (CHROMagar, Paris, France), followed by incubation at 36 °C for 24 h. Marked suspected colonies of *E. coli*, *Klebsiella*, *Enterobacter*, *Aeromonas*, and *Pseudomonas* (Fig. S2) were isolated as single clones, followed by whole-genome sequencing, similar to the metagenomic DNA-seq described in the above section. In addition to gram-negative ESBL-producing bacteria, potential CPO were isolated as described previously [50]. Briefly, 500 mL of effluent was filtered, and the membrane was incubated in 20 mL of Luria–Bertani (LB) broth supplemented with 1 mg/L meropenem at 37 °C for 14 h, after which the culture was spread on CHROMagar ESBL plates (CHROMagar, Paris, France).

For complete genome sequencing, long-read sequencing data of isolated bacterial genomes were produced with more than 100-fold coverage and assembled using Flye version 2.5 [51]. Primary error correction using long reads was performed by Minimap version 0.2-r124 [52] and Racon version 1.1.0 [53], followed by circularization using Circlator version 1.5.3 [54]. The remaining error of tentative complete circular sequences was corrected using Pilon version 1.18 with Illumina short reads [55]. Short plasmids (< approximately 10 kb) were assembled by A5-miseq version 20140604 using unmapped short reads against chromosomal and long plasmid sequences. The draft genome sequencing was assembled using A5-miseq with only Illumina short-read data. Gene annotation was performed using DFAST version 1.2.3 [56] with databases as follows: DFAST default database, ResFinder database [44], Bacterial Antimicrobial Resistance Reference Gene (BARRG) database (PRJNA313047), and Virulence Factors Database [57]. Multilocus sequence typing (MLST) was performed using “mlst” program version 2.16.2 (Seemann T, mlst Github https://github.com/tseemann/mlst) with PubMLST database (https://pubmlst.org/). Plasmid replicon typing was analyzed using ABRicate program version 0.3 (Seemann T, Abricate, Github https://github.com/tseemann/abricate) with the PlasmidFinder database [58].

**SNV phylogenetic analysis and pan-genome analysis**

The flowchart for SNV phylogenetic analysis is summarized in Fig. S7. Assembly data and/or Illumina raw reads of four species (*E. coli*, *A. baumannii*, *A. caviae*, and *K. quasipneumoniae*), were retrieved from the NCBI database as of April 2019. If assembly data were not deposited in the NCBI assembly database, raw sequence data downloaded from the SRA database were assembled using SKESA version 2.3.0 [59]. MinHash sketch for each genome sequence of this study (n = 53) and NCBI data (n = 76,449) were constructed using the sourmash program version 2.0.0a4 [60] with the following parameters: compute -k 21 -s 4000. The MinHash sketch database was built using the sourmash.
program (parameter: index -k 21) with the NCBI dataset, followed by MinHash search against 53 genome sequences of this study. The collected dataset was analyzed for core genome SNV and pan-genome analyses for each species or ST type. In SNV analysis, the longest chromosomal sequences from this study were selected as reference sequences. Repeat and prophage regions of reference sequences were analyzed using NUCmer (MUMmer 3.0) [61] and prophet [62] programs, respectively. If available data were only contig sequences from the NCBI database, the SimSeq program (https://github.com/jstjohn/SimSeq) was used to construct the simulated 150-mer paired-end reads of 200 bp insert length. Read mapping was performed using BWA-MEM [63] with default parameters against the reference chromosomal sequences, followed by variant calling using VarScan version 2.3.4 [64]. SNVs on repeat and predicted prophage regions were removed, and recombination regions were predicted using the Gubbins software [65], followed by filtering SNVs on recombination regions. Core genome SNV phylogenetic analysis was performed by the approximate maximum likelihood phylogenetic method using FastTree v2.1.10 [66], followed by visualization using fandango version 1.3.0 [67] and Interactive Tree Of Life (iTOL) version 3 [68]. SNV network analysis was performed using the median joining network method and TCS network method of PopART (http://popart.otago.ac.nz). Pan-genome analysis of predicted open reading frames (ORFs) was performed using Roary version 3.12.0 with default parameters [69].

**Plasmidome analysis**

The flowchart for plasmidome phylogenetic analysis is summarized in Fig. S10. Complete plasmid sequences were retrieved from the NCBI database as of November 2019, followed by re-annotation and ARG detection using DFAST with ResFinder and BARRG databases. The plasmid database of these nucleotide and protein sequences was constructed using BLAST. For plasmidome analysis, complete plasmids in public database were selected according to the following criteria: (i) ≥ 90% nucleotide sequence identity and ≤ 1E-100 e-value against β-lactamase gene-positive complete plasmids revealed in this study; (ii) sharing ≥ 10 ORFs (≥ 99% identity) against these plasmids; (iii) possession of the ARGs; and (iv) existence of metadata related to isolation source. Pan-genome analysis using the collected plasmids was performed using Roary version 3.12.0 [69] with a parameter of 99% BLASTP percentage identity cutoff, followed by construction of a distance matrix using R package “proxy” with edge weights of plasmids sharing ≥ 10 ORFs. NMDS and community structure analysis were performed by “vegan” and “igraph” of R package, respectively. Plasmidome data were visualized using Cytoscape version 3.7.2.

**Declarations**

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Availability of data and materials**

All complete sequences are available from the DDBJ/EMBL/GenBank database (accession numbers AP021908–AP022304, see Table S4). All raw read sequence files are available from the DRA/SRA database [accession numbers DRR198489–DRR198524 (metagenomic data, see Table S1) and DRR199157–DRR199560 (whole-genome data, see Table S4)].

**Competing interests**

The authors declare that they have no competing interests.

**Funding**
This work was supported by the Research Program on Emerging and Reemerging Infectious Diseases from the Japan Agency for Medical Research and Development, AMED (grant numbers: JP19fk0108048 and JP20fk0108131). This work was also supported by a grant for research on emerging and reemerging infectious diseases and immunization (H30 Shinkogyosei-Ippan-002) from the Ministry of Health, Labour and Welfare, Japan. The funding agencies had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

**Author Contributions**

TS, KI, and MK contributed to collection of water samples.

MK conducted bacterial isolation.

RT and MH performed genome sequencing.

TS and KY performed comparative genome analysis.

MK wrote the manuscript.

**Acknowledgments**

We would like to thank Editage (www.editage.com) for English language editing.

**References**

1. World Health Organization (WHO). Geneva S. *Global Antimicrobial Resistance Surveillance System: Manual for Early Implementation*. 2016: p. 1–4.

2. Amos GC, Hawkey PM, Gaze WH, Wellington EM. Waste water effluent contributes to the dissemination of CTX-M-15 in the natural environment. J Antimicrob Chemother. 2014;69:1785–91.

3. Martí E, Jofre J, Balcazar JL. Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant. PLoS One. 2013;8:e78906.

4. Osinska A, Harnisz M, Korzeniewska E. Prevalence of plasmid-mediated multidrug resistance determinants in fluoroquinolone-resistant bacteria isolated from sewage and surface water. Environ Sci Pollut Res Int. 2016;23:10818–31.

5. Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, Michael I, Fatta-Kassinos D. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. Sci Total Environ. 2013;447:345–60.

6. Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, Roder T, Nieuwenhuijse D, Pedersen SK, Kjeldgaard J, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun. 2019;10:1124.

7. Manaia CM. Assessing the Risk of Antibiotic Resistance Transmission from the Environment to Humans: Non-Direct Proportionality between Abundance and Risk. Trends Microbiol. 2017;25:173–81.

8. Hao L, Bize A, Conteau D, Chapleur O, Courtois S, Kroff P, Desmond-Le Quemener E, Bouchez T, Mazeas L. New insights into the key microbial phylotypes of anaerobic sludge digesters under different operational conditions. Water Res. 2016;102:158–69.

9. Fang H, Cai L, Yu Y, Zhang T. Metagenomic analysis reveals the prevalence of biodegradation genes for organic pollutants in activated sludge. Bioresour Technol. 2013;129:209–18.

10. Gillings MR. Integrons: past, present, and future. Microbiol Mol Biol Rev. 2014;78:257–77.

11. Matsumura Y, Pitout JD, Gorni R, Matsuda T, Noguchi T, Yamamoto M, Peirano G, DeVinney R, Bradford PA, Motyl MR, et al. Global Escherichia coli Sequence Type 131 Clade with blaCTX-M-27 Gene. Emerg Infect Dis. 2016;22:1900–7.

12. Khan FA, Soderquist B, Jass J. Prevalence and Diversity of Antibiotic Resistance Genes in Swedish Aquatic Environments Impacted by Household and Hospital Wastewater. Front Microbiol. 2019;10:688.

13. ECDC: *Antimicrobial consumption in the EU/EEA. Annual epidemiological report*. 2018. https://www.ecdc.europa.eu/sites/default/files/documents/Antimicrobial-consumption-EU-EEA.pdf 2018.
14. Li B, Yang Y, Ma L, Ju F, Guo F, Tiedje JM, Zhang T. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. ISME J. 2015;9:2490–502.

15. Vaz-Moreira I, Nunes OC, Manaia CM. Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. FEMS Microbiol Rev. 2014;38:761–78.

16. Higa S, Sarassari R, Hamamoto K, Yakabi Y, Higa K, Koja Y, Hirai I. Characterization of CTX-M type ESBL-producing Enterobacteriaceae isolated from asymptomatic healthy individuals who live in a community of the Okinawa prefecture, Japan. J Infect Chemother. 2019;25:314–7.

17. Nakane K, Kawamura K, Goto K, Arakawa Y. Long-Term Colonization by bla(CTX-M)-Harboring Escherichia coli in Healthy Japanese People Engaged in Food Handling. Appl Environ Microbiol. 2016;82:1818–27.

18. Zhong YM, Liu WE, Liang XH, Li YM, Jian ZJ, Hawkey PM. Emergence and spread of O16-ST131 and O25b-ST131 clones among faecal CTX-M-producing Enterobacteriaceae in healthy individuals in Hunan Province, China. J Antimicrob Chemother. 2015;70:2223–7.

19. Birgy A, Levy C, Bidet P, Thollot F, Derkx V, Bechet S, Mariani-Kurkdjian P, Cohen R, Bonacorsi S. ESBL-producing Escherichia coli ST131 versus non-ST131: evolution and risk factors of carriage among French children in the community between 2010 and 2015. J Antimicrob Chemother. 2016;71:2949–56.

20. Rogers BA, Ingram PR, Runnegar N, Pitman MC, Freeman JT, Athan E, Havers S, Sidjabat HE, Gunning E, De Almeida M, et al. Sequence type 131 fimH30 and fimH41 subclones amongst Escherichia coli isolates in Australia and New Zealand. Int J Antimicrob Agents. 2015;45:351–8.

21. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M beta-lactamases: temporal and geographical shifts in genotype. J Antimicrob Chemother. 2017;72:2145–55.

22. Kuroda H, Yano H, Hirakata Y, Arai K, Endo S, Kanamori H, Yamamoto H, Ichimura S, Ogawa M, Shimojima M, et al. Molecular characteristics of extended-spectrum beta-lactamase-producing Escherichia coli in Japan: emergence of CTX-M-15-producing E. coli ST131. Diagn Microbiol Infect Dis. 2012;74:201–3.

23. Bonomo RA, Burd EM, Conly J, Limbago BM, Poirel L, Segre JA, Westblade LF. Carbapenemase-Producing Organisms: A Global Scourge. Clin Infect Dis. 2018;66:1290–7.

24. Torimaru Y, Peirano G, Motyl MR, Adams MD, Chen L, Kreiswirth B, DeVinney R, Pitout JD. Global Molecular Epidemiology of IMP-Producing Enterobacteriaceae. Antimicrob Agents Chemother 2017, 61.

25. Sekizuka T, Inamine Y, Segawa T, Kuroda M. Characterization of NDM-5- and CTX-M-55-coproducing Escherichia coli GSH8M-2 isolated from the effluent of a wastewater treatment plant in Tokyo Bay. Infect Drug Resist. 2019;12:2435–9.

26. Karkman A, Do TT, Walsh F, Virta MPJ. Antibiotic-Resistance Genes in Waste Water. Trends Microbiol. 2018;26:220–8.

27. Anderson DI, Hughes D. Microbiological effects of sublethal levels of antibiotics. Nat Rev Microbiol. 2014;12:465–78.

28. Prisby JP, Garcia-Villada L, Zhai Y, Lewin-Epstein O, Wang AZ, Liu J, Xiao J, Mei Q, Fitzgerald DM, Bos J, et al. Gamblers: An Antibiotic-Induced Evolvable Cell Subpopulation Differentiated by Reactive-Oxygen-Induced General Stress Response. Mol Cell. 2019;74:785–800 e787.

29. Qin TT, Kang HQ, Ma P, Li PP, Huang LY, Gu B. SOS response and its regulation on the fluoroquinolone resistance. Ann Transl Med. 2015;3:358.

30. Gao H, Zhang L, Lu Z, He C, Li Q, Na G. Complex migration of antibiotic resistance in natural aquatic environments. Environ Pollut. 2018;232:1–9.

31. Azuma T, Otomo K, Kunitou M, Shimizu M, Hosomaru K, Mikata S, Ishida M, Hisamatsu K, Yunoki A, Mino Y, Hayashi T. Environmental fate of pharmaceutical compounds and antimicrobial-resistant bacteria in hospital effluents, and contributions to pollutant loads in the surface waters in Japan. Sci Total Environ. 2019;657:476–84.

32. Su HC, Pan CG, Ying GG, Zhao J, Zhou JJ, Liu YS, Tao R, Zhang RQ, He LY. Contamination profiles of antibiotic resistance genes in the sediments at a catchment scale. Sci Total Environ. 2014;490:708–14.

33. Wistrand-Yuen E, Knopp M, Bjork K, Koskineni S, Berg OG, Andersson DI. Evolution of high-level resistance during low-level antibiotic exposure. Nat Commun. 2018;9:1599.

34. Zhu YG, Johnson TA, Su JQ, Qiao M, Guo GX, 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.

35. Xu Y, Xu J, Mao D, Luo Y. Effect of the selective pressure of sub-lethal level of heavy metals on the fate and distribution of ARGs in the catchment scale. Environ Pollut. 2017;220:900–8.

36. Kang W, Zhang YJ, Shi X, He JZ, Hu HW. Short-term copper exposure as a selection pressure for antibiotic resistance and metal resistance in an agricultural soil. Environ Sci Pollut Res Int. 2018;25:29314–24.
37. Garner E, Chen C, Xia K, Bowers J, Engelthaler DM, McLain J, Edwards MA, Pruden A. Metagenomic Characterization of Antibiotic Resistance Genes in Full-Scale Reclaimed Water Distribution Systems and Corresponding Potable Systems. Environ Sci Technol. 2018;52:6113–25.

38. Hultman J, Tamminen M, Parmanen 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.

39. Ahmed MB, Zhou JL, Ngo HH, Guo W, Thomaidis NS, Xu J. Progress in the biological and chemical treatment technologies for emerging contaminant removal from wastewater: A critical review. J Hazard Mater. 2017;323:274–98.

40. Maheshwari M, Yaser NH, Naz S, Fatima M, Ahmad I. Emergence of ciprofloxacin-resistant extended-spectrum beta-lactamase-producing enteric bacteria in hospital wastewater and clinical sources. J Glob Antimicrob Resist. 2016;5:22–5.

41. Breitwieser FP, Baker DN, Salzberg SL. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biol. 2018;19:198.

42. Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics. 2020;36:1303–4.

43. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12:R60.

44. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67:2640–4.

45. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.

46. Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DG. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res. 2014;42:D737–43.

47. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang HY, El-Gebali S, Fraser MI, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019;47:D351–60.

48. Ye Y, Choi JH, Tang H. RAPSearch: a fast protein similarity search tool for short reads. BMC Bioinformatics. 2011;12:159.

49. Rogens T, Flouri T, Nichols B, Quince C, Meta H. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2015;4:e2584.

50. Sekizuka T, Yatsu K, Inamine Y, Segawa T, Nishio M, Kishi N, Kuroda M. Complete Genome Sequence of a blaKPC-2-Positive Klebsiella pneumoniae Strain Isolated from the Effluent of an Urban Sewage Treatment Plant in Japan. mSphere 2018, 3.

51. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6.

52. Li H. Minimap and minimap: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2015;31:3872–80.

53. Vasar R, Sovic I, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737–46.

54. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol. 2015;16:294.

55. Walker BJ, Abee T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9:e112963.

56. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics. 2018;34:1037–9.

57. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res. 2012;40:D641–5.

58. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58:3895–903.

59. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 2018;19:153.

60. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. Large-scale sequence comparisons with sourmash. F1000Res. 2019;8:1006.

61. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.

62. Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. ProphET, prophage estimation tool: A stand-alone prophage sequence prediction tool with self-updating reference database. PLoS One. 2019;14:e0223364.

63. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.

64. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25:2283–5.
65. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15.

66. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.

67. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics. 2018;34:292–3.

68. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.

69. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3.

70. Sekizuka T, Inamine Y, Segawa T, Hashino M, Yatsu K, Kuroda M. Potential KPC-2 carbapenemase reservoir of environmental Aeromonas hydrophila and Aeromonas caviae isolates from the effluent of an urban wastewater treatment plant in Japan. Environ Microbiol Rep. 2019;11:589–97.

Tables

Table 1. General information on wastewater treatment plants and sampling sites in this study
| Place ID | Sampling site (GPS) coordinates | Population estimate | Area (ha) | Sewage treatment (m³/day) | A2O process | Sand Filtration/Membrane Filtration | Chlorination | BOD effluent (mg/L)* | COD effluent (mg/L)* |
|----------|---------------------------------|---------------------|----------|---------------------------|-------------|-----------------------------------|-------------|----------------------|---------------------|
| WP1      | 35°41′06.8′N 139°24′37.2′E      | 130,000             | 1,134    | 63,300                    | +           | -                                 | +           | 3                    | N/A                 |
| WP2      | 35°39′50.5′N 139°26′16.0′E      | 230,000             | 2,744    | 48,296                    | +           | -                                 | +           | 3                    | 8                   |
| WP3      | 35°40′06.1′N 139°25′40.0′E      | 263,000             | 3,902    | 78,245                    | +           | -                                 | +           | 3                    | 8                   |
| WP4      | 35°39′02.7′N 139°28′45.6′E      | 360,000             | 5,900    | 111,978                   | +           | -                                 | +           | 3                    | 8                   |
| WP5      | 35°39′13.8′N 139°30′45.2′E      | 489,000             | 5,124    | 200,813                   | +           | -                                 | +           | 5                    | 7                   |
| WP7      | 35°33′42.9′N 139°45′11.0′E      | 2,127,000           | 14,675   | 1,138,446                 | +           | -                                 | +           | 3                    | 7                   |
| WP8      | 35°37′53.8′N 139°44′45.6′E      | 705,000             | 6,440    | 607,162                   | +           | -                                 | +           | 16                   | 12                  |
| WP9      | 35°37′49.9′N 139°47′02.9′E      | N/A                 | N/A      | 15,968                    | +           | -                                 | +           | 1                    | 9                   |
| BEC1     | 35°37′50.0′N 139°46′31.3′E      | N/A                 | N/A      | N/A                       | N/A         | N/A                               | N/A         | N/A                  | N/A                 |

Average value as reported in April 2018. N/A, not available; A2O process, anaerobic–anoxic–oxic process, BOD, biological oxygen demand; COD, chemical oxygen demand.
### Table 2. Whole-genome sequencing of CHROMagar ESBL-positive isolates obtained from WWTPs

|                      | Summer, 2017 | Winter, 2018 | Summer, 2018 | Winter, 2019 | Total |
|----------------------|--------------|--------------|--------------|--------------|-------|
| Total isolates       | 109          | 104          | 114          | 106          | 433   |
| EPO/CPO              |              |              |              |              |       |
| *E. coli*            | 18           | 19           | 18           | 16           | 71    |
| *Klebsiella*         | 9            | 8            | 18           | 16           | 51    |
| *Enterobacter*       | 5            | 15           | 5            | 4            | 29    |
| *Acinetobacter*      | 11           | 4            | 12           | 1            | 28    |
| *Pseudomonas*        | 25           | 14           | 22           | 25           | 86    |
| *Aeromonas*          | 20           | 33           | 26           | 21           | 100   |
| Others               | 21           | 11           | 13           | 23           | 68    |
| **β-lactamase type** |              |              |              |              |       |
| CTX-M                | 32           | 37           | 34           | 31           | 134   |
| IMP                  | 0            | 3            | 2            | 2            | 7     |
| KPC-2                | 0            | 0            | 5<sup>a</sup> | 3            | 8     |
| NDM                  | 0            | 0            | 2<sup>b</sup> | 1            | 3     |
| GES                  | 2            | 6            | 4            | 8            | 20    |

<sup>a</sup> [50, 70]  
<sup>b</sup> [25]

See the information of whole-genome sequence for all isolates in Table S4.

EPO, extended-spectrum β-lactamase-producing organisms  
CPO, carbapenemase-producing organisms

### Description Of Supplementary Files

**Fig. S1.** Experimental procedure for metagenomic DNA analysis of WWTP effluent.

**Fig. S2.** A) Taxonomic classification of domain ranks from effluents of fresh water area (WP1 – 8) based on metagenomic DNA-seq sequencing. B) Bacterial population size using RPKM normalization for 16S rRNA genes and *rpoB* RNA polymerase β-subunit orthologs showed no significant difference with an average of four copies of 16S rRNA gene per one copy of *rpoB*, suggesting that RPKM estimation is an effective approach to characterize the content of ARGs.
Fig. S3. A) Seasonal (or temporal) or B) WP-dependent trends of ARG difference showing a fold change (log₂ ratio) in RPKM values.

Fig. S4. A) HMRGs in effluents from WWTPs (WP) and beach (BEC1). Metagenomic DNA-seq short reads were analyzed by homology search in BacMet2 database (bioinformatics resource of antibacterial biocide- and metal-resistance genes. http://bacmet.biomedicine.gu.se/), followed by normalization with RPKM. See Table S3 for the detail detection of ARGs list. B) NDMS plots of metagenomic sequencing reads classified for heavy-metal resistance in WWTP effluents (WP1–9) and recreational beach (BEC1). Fresh water area in WWTPs (WP1–8, see Fig. 1) were closely scattered (shaded with gray square), but the plots were markedly divided in a sampling time-dependent manner with partial WWTP dependency.

Fig. S5. A) Trends of metal-resistance gene variations, detected metal-resistance genes for each metal at every sampling and WP site, and a unique season-dependent gene detection in B) summer and C) winter.

Fig. S6. Experimental procedure for potential ESBL-producing organisms.

Fig. S7. Flowchart for verification of strain clonality using comparative genomic analysis.

Fig. S8. Core-genome SNV phylogeny (see detail procedures in Fig. S7) of E. coli ST131 isolates using publicly available ST131 draft or complete genome sequences (a total of 315 strains; see Table S6).

Fig. S9. Characterization of other CTX-M-positive organisms from WWTP effluents.

Genome sequence of A) A. caviae and B) K. quasipneumoniae isolates were subjected to SNV network, genome recombination, and pan-genome analysis, and the results suggested a marked clonality among isolates, although the isolates were obtained at different WP sites or time points.

Fig. S10. Flowchart for plasmidome analysis.

Table S1. Metagenomic DNA-seq information for effluent of WWTP and environment water samples in this study.

Table S2. Summary of RPKM values for OTU of antimicrobial resistance.

Table S3. Summary of RPKM value for orthologous cluster of HMRGs.

Table S4. Whole-genome sequence information for the strain growing on CHROMagar ESBL.

Table S5. Summary of isolates on CHROMagar ESBL and its sequence type in this study.
Table S6. Strain information of *E. coli* used in core-genome SNV phylogenetic and pan-genome analyses.

Table S7. Complete plasmid sequence information of the strain growing on CHROMagar ESBL.

Table S8. Complete plasmid sequence list for plasmidome analysis.

Table S9. Summary of absolute count of plasmids against each category in plasmidome analysis.

**Figures**

**Figure 1**

Outline map illustration of water sampling sites along the Tama River and around Tokyo Bay area. Effluents of WWTPs and surface water from recreational beach (BEC) were obtained for further AMR investigation in this study.
Figure 2

Metagenomic analysis for bacteria detection of WWTP effluents. A) NDMS plots of metagenomic sequencing reads classified at the bacterial genus level for WWTP (WP1–9) effluents and recreational beach sample (BEC1). Fresh water areas in WWTPs (WP1–8, see Fig. 1) were closely scattered (shaded with gray square). Plots were markedly divided in a season-dependent manner but were not WWTP dependent. B) The seasonal feature was observed in multiple bacterial genera by LEfSe analysis (score threshold: 2.0).
Figure 3

Metagenomic analysis for ARG detection of WWTP effluents. A) ARG detection in the effluents from WWTPs (WP) and beach (BEC1). Metagenomic DNA-seq short reads were analyzed by homology search in ResFinder ARGs database, followed by normalization with RPKM. See Table S2 for the detailed list of detected ARGs. B) NDMS plots of metagenomic sequencing reads classified for ARGs in effluents of WWTPs (WP1–9) and recreational beach (BEC1). Fresh water area in WWTPs (WP1–8, see Fig. 1) were closely scattered (shaded with gray square), but the plots were markedly divided in a sampling time-dependent manner with partially WWTP-dependency for WP8.
Figure 4

Profiling of ARGs based on AMR categories. A) Trends of ARG variations, detected ARGs in each AMR category at every sampling and WP site, and a unique season-dependent ARG detection in B) summer and C) winter.
Characterization of E. coli ST131 isolates from WWTP effluents. A) Among 11 E. coli ST131 isolates, nine isolates predominantly possessed blaCTX-M-27 genes (Table S4). Core genome SNVs phylogeny (see detail procedures in Fig. S7) suggested that primarily four subclonal types of ST131 were identified from WWTP effluents (Fig. S8). E. coli WP5-S18-ESBL-09 is genotyped as O16:H5-ST131-mH41 with blaCTX-M-27 and exhibited a marked clonality showing <12 SNVs with clinical isolates in several countries (Southeastern Asia, EU, and Oceania). B) SNV network, genome recombination, and pan-genome analysis suggested that the genome structure of these ST131 strains were very stable, except for some ARGs that are transferred with conjugative plasmid.
Figure 6

Characterization of E. coli and other ST isolates from WWTP effluents. E. coli ST38 (10 isolates), ST10 (five isolates), ST405 (four isolates), ST69 (three isolates), and ST648 (three isolates) in this study (Tables S4 and S5) were subjected to core genome SNV phylogeny for each ST isolates (see detail procedures in Fig. S7).
Figure 7
Plasmidome analysis of EPO and CPO isolates from WWTP effluents and environment. Of the potential 433 EPO/CPO isolates (Table 2), 85 strains were identified as complete genome sequences in this study (Table S4). Of the 304 complete plasmid sequences, 73 β-lactamase-positive plasmids (Table S7) were subjected to plasmidome network analysis based on orthologous analysis (see details in Fig. S10). A total of 831 complete plasmids (73 in this study, and 758 from public database; see Table S8) were subjected to pan-genome (Roary) and NMDS (vegan in R package) analysis and visualized, revealing 14 communities (Co).

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- SupplementFigures.pdf
- SupplementaryTableS1.pdf
- SupplementaryTableS2.pdf
- SupplementaryTableS3.pdf
- SupplementaryTableS4.pdf
- SupplementaryTableS5.pdf
- SupplementaryTableS6.pdf
- SupplementaryTableS7.pdf
• SupplementaryTableS8.pdf
• SupplementaryTableS9.pdf