Limited dissemination of the wastewater treatment plant core resistome

Munck, Christian; Albertsen, Mads; Telke, Amar; Ellabaan, Mostafa M Hashim; Nielsen, Per Halkjær; Sommer, Morten Otto Alexander

Published in:
Nature Communications

Link to article, DOI:
10.1038/ncomms9452

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Limited dissemination of the wastewater treatment plant core resistome

Christian Munck1,*, Mads Albertsen2,*, Amar Telke3,*, Mostafa Ellabaan1, Per Halkjær Nielsen2 & Morten O.A. Sommer1

Horizontal gene transfer is a major contributor to the evolution of bacterial genomes and can facilitate the dissemination of antibiotic resistance genes between environmental reservoirs and potential pathogens. Wastewater treatment plants (WWTPs) are believed to play a central role in the dissemination of antibiotic resistance genes. However, the contribution of the dominant members of the WWTP resistome to resistance in human pathogens remains poorly understood. Here we use a combination of metagenomic functional selections and comprehensive metagenomic sequencing to uncover the dominant genes of the WWTP resistome. We find that this core resistome is unique to the WWTP environment, with <10% of the resistance genes found outside the WWTP environment. Our data highlight that, despite an abundance of functional resistance genes within WWTPs, only few genes are found in other environments, suggesting that the overall dissemination of the WWTP resistome is comparable to that of the soil resistome.
Investigations of environmental sources of antibiotic resistance genes have shown that many clinically relevant resistance genes such as the cephalosporin resistance gene family _ctx-m_, the quinolone resistance gene family _qnr_ and the vancomycin resistance gene family _van_ have close homologues on the chromosomes of environmental non-pathogenic bacterial species\(^1\)\(^2\)\(^3\)\(^4\). These findings highlight that resistance genes can spread from environmental reservoirs to human pathogens. This has led to a growing interest in uncovering which environmental microbial communities participate most in antibiotic resistance gene exchange with the aim of establishing the most impactful policies that limit the dissemination of antibiotic resistance\(^5\)\(^6\)\(^7\)\(^8\).

To assess the resistance gene load of a given environmental microbial community, researchers have used targeted PCR or metagenomic read mapping to detect the presence of well-characterized resistance genes\(^9\)\(^10\). These studies have shown that representatives of clinically relevant resistance genes can be found in virtually all environments. Although such studies are important and relevant, they rarely consider the abundance of the targeted resistance genes relative to the complete resistome of the studied environment. This is largely due to our incomplete annotation and knowledge of antibiotic resistance genes\(^10\). Indeed, less biased methods such as metagenomic functional selections have revealed that <1% of the resistance genes found in soil have high identity (>99% nucleotide identity) hits in NCBI database\(^8\)\(^11\). Accordingly, the majority of the soil resistome has not been found outside the soil environment. On the other hand similar approaches have been used to show that a subset of the human gut microbiome may exchange resistance genes with human pathogens\(^12\). So far, metagenomic functional selections have largely been applied to study the resistomes of soil and gut microbial communities; yet, other communities are believed to contribute substantially to the dissemination of resistance genes. Hence, to make informed policy decisions addressing the challenges associated with antibiotic resistance, the abundance and dissemination of antibiotic resistance genes in these communities must be characterized.

Wastewater treatment plants (WWTPs) are generally considered to be important hubs for horizontal gene transfer of antibiotic resistance genes\(^13\)\(^14\)\(^15\). In these facilities, wastewater from diverse sources such as hospitals, households and animal production farms is mixed, often containing both pathogenic bacteria and traces of antibiotics\(^16\)\(^17\)\(^18\). Accordingly, the dense microbial communities of WWTPs should provide ideal conditions for horizontal exchange of antibiotic resistance genes between a wide range of bacterial species, including human pathogens.

Several PCR-based studies have shown that clinically relevant resistance genes, including _tem1_, _ampC_, _ndm-1_, _vanA_ and _ermB_, can be found in WWTPs\(^19\)\(^20\)\(^21\). Similarly, metagenome sequencing studies characterizing either complete or closed circular metagenomes (plasmids and some phages) have used sequence read mapping to resistance gene databases, to identify several different resistance genes in WWTPs\(^22\)\(^23\)\(^24\)\(^25\)\(^26\). Furthermore, traditional cultivation-based studies have also shown that vancomycin-resistant enterococci, methicillin-resistant staphylococci and cephalosporin-resistant Enterobacteriaceae can be isolated from WWTPs\(^20\)\(^27\). However, owing to the specificity of PCR-based screening and database mapping, these approaches can only identify known resistance genes and their close homologues. Consequently, such methods cannot uncover the full spectrum of functional resistance genes within an environment. Likewise, cultivation-based approaches will only recover a small proportion of bacterial species, with estimates indicating that 85%-99% of the bacteria from WWTPs are not readily cultured using current protocols\(^28\). Here we combine functional selection and metagenomic sequencing to characterize and trace the most abundant antibiotic resistance genes in the WWTP environment. We find that only a small subset of the core WWTP resistome is found outside the WWTP environment.

### Table 1 | Selection conditions.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Concentration (µg ml(^{-1}))</th>
<th>Code</th>
<th>Class</th>
<th>Number of colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>16</td>
<td>AMP</td>
<td>β-Lactam/penicillin</td>
<td>660</td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>16</td>
<td>AMX</td>
<td>β-Lactam/penicillin</td>
<td>500</td>
</tr>
<tr>
<td>Carbenicillin</td>
<td>500</td>
<td>AMX</td>
<td>β-Lactam/penicillin</td>
<td>1,100</td>
</tr>
<tr>
<td>Piperacillin</td>
<td>500</td>
<td>PIP</td>
<td>β-Lactam/penicillin</td>
<td>800</td>
</tr>
<tr>
<td>Cefazodime</td>
<td>30</td>
<td>CTZ</td>
<td>β-Lactam/cephalosporin</td>
<td>350</td>
</tr>
<tr>
<td>Amikacin</td>
<td>22</td>
<td>AMK</td>
<td>Aminoglycoside</td>
<td>40</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>40</td>
<td>GEN</td>
<td>Aminoglycoside</td>
<td>40</td>
</tr>
<tr>
<td>Spectinomycin</td>
<td>33</td>
<td>SPC</td>
<td>Aminoglycoside</td>
<td>330</td>
</tr>
<tr>
<td>Azithromycin</td>
<td>120</td>
<td>AZI</td>
<td>Macrolide</td>
<td>120</td>
</tr>
<tr>
<td>Erythromycin</td>
<td>200</td>
<td>ERY</td>
<td>Macrolide</td>
<td>200</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>100</td>
<td>TET</td>
<td>Tetracycline</td>
<td>100</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>120</td>
<td>CAM</td>
<td>Phenicol</td>
<td>30</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>200</td>
<td>RIF</td>
<td>Rifamycin</td>
<td>210</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>2,000</td>
<td>TMP</td>
<td>Benzylpyrimidine</td>
<td>2,000</td>
</tr>
<tr>
<td>Sulfamethoxazole/trimethoprim</td>
<td>1,500</td>
<td>SXT</td>
<td>Sulfonamide/dihydrofolate reductase inhibitor</td>
<td>1,500</td>
</tr>
</tbody>
</table>

**Table 1** | Selection conditions.

AMK, amikacin; AMP, ampicillin; AMX, amoxicillin; AZI, azithromycin; CAM, chloramphenicol; CAR, carbenicillin; CTZ, cefazodime; ERY, erythromycin; GEN, gentamicin; PIP, piperacillin; RIF, rifampicin; SPC, spectinomycin; SXT, sulfamethoxazole/trimethoprim; TET, tetracycline; TMP, trimethoprim.

**Results**

**Metagenomic functional selection.** Using metagenomic functional selection, we identified the dominant members of the resistome in a large modern WWTP, Aalborg West (AAW), sampled in the third quarter of 2012. This facility is located in Denmark and receives both hospital and household wastewater, and has been in operation since 1979 (Supplementary Table 1 and Methods). A total of 800 Mb of DNA was cloned in an *Escherichia coli* expression library (Supplementary Methods) and screened on 15 different antibiotics representing 8 chemical classes: β-lactam, aminoglycoside, macrolide, tetracycline, phenicol, rifamycin, sulphonamide and dihydrofolate reductase inhibitor (Table 1). In total 8,540 resistant clones were identified with an average of 534 colonies per antibiotic ranging from 30 clones for chloramphenicol to more than 2,000 clones for...
trimethoprim (Table 1). Of the 8,540 identified clones, the inserts of 749 clones were selected proportionally among the classes of antibiotics and sequenced, resulting in the identification and annotation of 79 unique functionally validated resistance genes (Supplementary Table 2). Rarefaction curve analysis showed that we sampled the majority of the resistance genes in the library (Supplementary Fig. 1). Accordingly, these genes are expected to represent the most abundant genes in the WWTP resistome and we refer to them as the core WWTP resistome.

**WWTP community stability.** Considering the constant flow of material through the AAW WWTP, up to 50,000 m$^3$ per day, it could be expected that the microbial community varied greatly over time. However, using a modified version of the Bray–Curtis dissimilarity index ($(1 – Bray–Curtis dissimilarity index) \times 100$), we calculated the similarity between all pairwise combinations of samples collected from AAW over a 7-year period ($n = 24$) (Fig. 1a)$^{29}$. Sample similarity was indexed from 0 to 100, with 0 indicating completely disjoint sample and 100 indicating identical samples. The comparison revealed that the bacterial communities were remarkably stable (Fig. 1a). Notably, samples collected 7 years apart had an average similarity index of 48 (Fig. 1a) and samples collected <1 year apart had an average similarity index of 68, highlighting that the WWTP microbial community is highly stable and suggesting that variation in wastewater inflow has a limited impact on the core bacterial communities.

Comparison of the relative abundances of the ten most abundant bacterial orders in the WWTP community across the 7 years of sampling showed that the WWTP community is highly stable and unique to this environment (Fig. 1b). As recently reported, the WWTP environment is dominated by the phyla Actinobacteria (Actinomycetales and Acidimicrobiales), Bacteroidetes (Saprospirales) and Proteobacteria (Rhizobiales, Burkholderiales and Rhodobacterales), highlighting that WWTPs contain distinct microbial communities$^{30}$. Still, owing to horizontal gene transfer, a stable microbial community at the 16S ribosomal RNA sequence level does not necessarily imply a stable resistome.

**WWTP core resistome stability.** To investigate the long-term stability of the WWTP core resistome, we sequenced seven metagenomes spanning 2 years from the same WWTP used for the metagenomic functional selections (AAW; Supplementary Table 3). Approximately 230 Gbp of sequence data corresponding to more than 1.5 billion raw reads were sequenced. By mapping the metagenomic sequencing reads to the functional selected resistance genes, we could trace the WWTP core resistome dynamics in the WWTP over the 2-year period (Fig. 2a). Extrapolation of the rarefaction curves predicted that the samples collected over the 2-year sampling period on average contained 70% (s.d. = 13%) of the whole resistance gene set (Fig. 2b). This highlights that despite vast amounts of wastewater passing through the treatment plant the resistome remains stable and we consider this set of genes part of the WWTP core resistome.

To evaluate the impact of sequencing depth, we exhaustively sequenced sample AAW.5.2010, which was collected 2 years before the sample used for the functional selection. In this metagenome, consisting of more than 1.1 billion reads, we identified 67% of the functional selected resistance genes and extrapolations of the rarefaction curve estimated that up to 82% genes could be found in this sample (Fig. 2a,b). This further underscores that the core resistome was largely stable during the 2-year sampling period.

**Core resistome hosts.** To identify possible hosts carrying the functional selected resistance genes, we assembled contigs from the deeply sequenced metagenome (AAW.5.2010). The relative low coverage of the functionally selected resistance genes challenges their assembly. In addition, the complexity of the WWTP community also challenges the assembly. A total of 43% of all metagenomic reads could be mapped back to the assembled contigs. Seven contigs contained functional selected resistance genes (>95% identity and >90% gene coverage). The contigs were blasted against the genbank nt database, to identify host organisms. In general, the query coverage was low, indicating that the contigs assembled from the metagenomes are located on hitherto uncharacterized genomes, which is largely a result of the limited availability of WWTP-specific bacterial genomes$^{31}$. However, a small contig containing CAR_05 had high identity and query coverage to the tem-1 β-lactamase gene on cloning vectors (Supplementary Table 4). The challenges of assembling the low coverage genes from the metagenome highlight the power of combining functional selection and metagenomic sequencing, to trace functionality through different metagenomes.

**The WWTP core resistome is stable across multiple facilities.** To explore whether the AAW WWTP core resistome was shared with other WWTP facilities, we sequenced additional eight metagenomes from four different Danish full-scale WWTPs: Egaa, Ejby, Hjoerring and Aalborg East. Two of the plants (Egaa and Ejby) were sampled repeatedly over a 1-year period and all plants, except Aalborg East, received household as well as hospital wastewater (Supplementary Table 3). In total, these additional WWTP metagenomes consisted of >124 Gbp corresponding to >800 million reads. Mapping of the metagenomic reads to the AAW core resistome revealed that the majority of the resistance genes were present in the different metagenomic samples (Fig. 2a,c). Extrapolations of the rarefaction curves estimate that
on average the non-AAW WWTPs contained 66% (s.d. = 11%) of the functionally selected resistance genes from AAW (Fig. 2c,d).

The stability of the WWTP microbial community and core resistome could support the hypothesis that resistance gene exchange within WWTPs is a rare event as was recently reported for soil microbial communities 11. However, as activated sludge in WWTPs is continuously mixed and exposed to varying levels of antibiotics, one would expect that horizontal gene transfer of resistance genes would be more frequent in WWTPs compared with soil 16–18. To investigate whether the WWTP core resistomes correlate with the microbial composition, we used procrustes analysis to correlate resistomes and 16S rRNA sequences from each of the 15 WWTP metagenome samples 11. Our analysis shows that WWTP resistome and phylogeny correlated (P < 0.001, based on 999 permutations), indicating that the WWTP core resistome is linked to the microbial community (Supplementary Fig. 2). This supports the hypothesis that the WWTP resistome is only exchanged to a limited extent within the WWTP environment.

Dissemination of the WWTP core resistome. Recent studies based on metagenomic read mapping showed that the metagenomic sequence data from WWTPs mapped to a distinct set of resistance genes compared with other environments 23. However, this finding could arise from database biases resulting in an inability to identify shared resistance genes or it could suggest that the WWTP resistome is only disseminated to a limited extent to other environments.

To investigate whether the WWTP core resistome is involved in gene exchange networks comprising other environments, as well as human pathogens, we used BLAST to perform a homology search of the WWTP core resistome to genes in the NCBI nucleotide database. We found that the WWTP resistance genes had an average sequence identity of 62% to the best hit in the database, and that just six genes had an identity > 95%, demonstrating that the vast majority of the functional selected resistance genes represent novel sequences not captured by the current antibiotic resistance gene databases (Fig. 3a).
To further investigate whether the WWTP core resistome was disseminated outside the treatment plant facilities, we mapped reads from online available metagenomes to the functionally selected WWTP resistome. These auxiliary metagenomes contained sequences from the human gut, cow rumen, permafrost and aquifer, in total more than two billions reads 37-40. Notably, only six (8%) of the WWTP core resistance genes were found in these non-WWTP metagenomes, confirming that the WWTP core resistome is distinct from the resistomes of other environments (Fig. 4a). Rarefaction curve analysis of the non-WWTP metagenomes showed that the low identification frequency of the WWTP resistance genes was not due to undersampling (Supplementary Fig. 3). Four of the six genes found in the non-WWTP metagenomes had a nucleotide identity >95% to clinically relevant resistance genes, highlighting that a PCR-based approach would probably have successfully identified these genes. However, reliance solely on PCR or metagenomic read mapping to antibiotic resistance gene databases would have neglected the vast majority of the WWTP core resistome, which seems confined to the WWTP environment.

To explore whether the WWTP metagenomes contained important resistance genes, we compiled a list of 27 resistance genes commonly found in Gram-negative and Gram-positive pathogens41-43 (Supplementary Table 5). This list includes commonly encountered representatives of the ctx-m and erm gene families, as well as several important tetracycline resistance genes. A total of 578 sequence reads from the deeply sequenced AAW.5.2010 metagenome mapped to 10 of the resistance genes (Supplementary Fig. 4a). Nearly half of the reads (48%) mapped to the tetA resistance gene, whereas no reads mapped to the ctx-m genes. In contrast, a significantly higher number of 9,401 reads (P = 0.008, Mann–Whitney U-test) mapped to a total of 53 of the 79 functionally selected resistance genes (Supplementary Fig. 4b). These mapping results confirm that clinically important resistance genes are indeed present in the WWTP environment, yet again they also highlight how reliance on targeted methodologies can be misleading when assessing the contribution of the WWTP core resistome to the dissemination of resistance genes outside of the WWTP environment.

As our study is focused on dissemination of the WWTP core resistome, we cannot rule out the possibility that low abundance species engage in important resistance gene exchange in WWTPs. Indeed, mobile plasmids carrying known resistance genes have been isolated from WWTPs26,44. However, our results show that only a small fraction of the WWTP core resistome participate in gene exchange, suggesting that WWTP microbiomes appear as probably as soil microbiomes to be sources for resistance gene dissemination. This conclusion should be cautioned by the fact that we only sampled WWTPs in one country. It is possible that these findings may be different in other parts of the world that deploy simpler WWTPs at higher temperatures and higher load of antibiotics.

Figure 3 | Genbank nucleotide database hit identity. Functionally selected resistance genes from (a) WWTP AAW (n = 79), (b) human gut (n = 94)32,33 and (c) soil (n = 101)33-36 were blasted against the Genbank nt database and the top hit was aligned to the query. The panels show the nucleotide identity distributions of the pairwise alignment for each environment (Methods).
An explanation for the limited dissemination of the WWTP core resistome could be that the resistance genes are not selected for or mobilized. However, studies have shown that even very low concentrations of antibiotics, probably present in WWTPs, can select for resistance determinants\(^{45,46}\). Therefore, we analysed the functionally selected inserts for the presence of canonical indicators of mobilization such as transposases and phage elements. We found that three (4\%) of the functionally selected inserts carried canonical indicators of mobility including integrases associated with class I integrons, transposes and resolvases (Fig. 4b). Two of the resistance genes carried on these inserts (SPC_01 and AZI_01) had a high identity hit (95\% identity) to the NCBI nucleotide database. Furthermore, the resistance genes located on the AZI_01 and ERM_04 inserts could be found in the auxiliary metagenomes (Fig. 4a). This supports the hypothesis that mobilization is a major limiting factor for the dissemination of antibiotic resistance genes; for example, once a resistance gene is mobilized, it easily becomes disseminated across ecosystems. It should be noted that, owing to the relative short insert size (~2 kb), some mobility elements may be overlooked. However, previous functional selections from human faecal samples using the same insert size found mobilization elements in 14\% of the inserts, which is significantly higher than what is observed in this study (\(P = 0.027, \chi^2\)-test)\(^{12}\).

Of particular interest was the spectinomycin resistance gene \(aadA\) in the SPC_01 insert. This gene was located in a class I integron and mapping of reads from the deeply sequenced metagenome to the insert containing the \(aadA\) gene showed that the integrase gene was roughly 50 times more abundant than the \(aadA\) gene, highlighting that the integron was not always associated with the \(aadA\) gene (Fig. 4c). This gene is part of the WWTP core resistome and has previously been found to be associated with wastewater environments including a French hospital, pig manure sampled in Denmark and on a plasmid isolated from a Danish soil/manure sample\(^{47–51}\) (Supplementary Fig. 5). Interestingly, the gene had also been found in an \(E. coli\) isolate from a bloodstream infection (99.9\% identity). This confirms that once mobilized, members of the core WWTP resistome can be found outside this environment and even in clinical isolates (Supplementary Fig. 5). Yet, only a small fraction of the core WWTP resistome has crossed this mobilization barrier.

**Discussion**

Integrating metagenomic functional selection with deep metagenomic sequencing represents a powerful approach to tracing functionally validated antibiotic resistance genes across different environments. Using functional selections, we identified a core WWTP resistome consisting of mostly novel resistance genes that confer resistance to all antibiotics tested, highlighting that
WWTPs is a reservoir of previously uncharacterized resistance genes. 165 rRNA sequencing of longitudinal samples collected over a 7-year period from the same WWTP revealed that the microbial community within the WWTP is remarkably stable. In addition, longitudinal metagenomic sequencing from the same WWTP demonstrated that the majority of the genes within the WWTP core resistome were also stably maintained within the WWTP. This indicates that the resistome is linked to the community, which was confirmed with procrustes analysis. Integration of metagenomic data sets from other WWTPs showed that a majority of the functionally verified resistance genes are also maintained in those microbial communities. These data point towards the existence of a shared WWTP core resistome harboured by the dominant members of the WWTP microbial community.

Notably, <10% of the WWTP core resistome sampled in this study is found in other metagenomic data sets or in sequences deposited at NCBI. Although database bias towards human pathogens and commensals could explain these findings, this also support the idea that members of the WWTP core resistome rarely take part in gene exchange networks with human pathogens. Instead, the WWTP microbial community can be considered a unique and stable microbial community harbouring a rich resistome that is difficult to access by human pathogens or other bacteria. This conclusion is consistent with a previous metagenomic study based on read mapping to existing antibiotic resistance gene databases, showing that the WWTP resistance gene profiles are distinct from other environments.

Our findings do not preclude the existence of clinically important antibiotic gene exchange in the WWTP. In fact, we do detect a subset of clinically important resistance genes in the deeply sequenced metagenome. Consequently, species found in low abundance in WWTPs, including human pathogens, may exchange resistance genes within this environment contributing to clinical problems of antibiotic resistance. However, our findings are consistent with the hypothesis that the abundant WWTP core resistome is only disseminated to a very limited extent to other microbial communities. The few resistance genes identified in this study to be disseminated to other microbial communities are associated with canonical mobility elements, suggesting that mobilization, rather than functionality, is the main barrier preventing spread of resistance genes.

Although our study demonstrates that the WWTP core resistome harbours an unappreciated diversity of resistance genes that are unique to this environment, the genetic diversity captured using functional selection is still limited, compared with the total genetic diversity within the WWTP environment. Consequently, the functional selection only allows us to analyse the more abundant resistance genes. Larger studies are needed to characterize the full WWTP resistome and its relation to clinical problems of antibiotic resistance. However, application of metagenomic methods such as those presented in this study should improve the risk classification of various environments, as it allows relating the number of clinically relevant resistance genes to the overall number of functional resistance genes. This will provide a useful tool for establishing control policies for environments that contribute substantially in the dissemination of antibiotic resistance genes.

**Methods**

**DNA extraction.** A total of 15 activated sludge samples were collected over a period of 3 years from five different Danish WWTPs with biological nitrogen and phosphorus removal. DNA was extracted using the FastDNA spin kit for soil (MP Biomedicals) according to the manufacturer’s instructions, except an initial incubation for 3 min in 50°C phenol. Briefly, 1 ml sample aliquots were centrifuged at 13,000 r.p.m. (17,000g) for 5 min and the supernatant discarded. The pellets were redissolved in 250 μl phosphate buffer (included in the FastDNA kit) and transferred to the FastDNA bead beating tubes. To each tube, 750 μl preheated phenol (Biological grade, pH 8, 0.1 M EDTA, Sigma-Aldrich) was added and the samples incubated for 3 min at 90°C with occasional shaking. The samples were then bead beaten using the manufacturer’s instructions and afterwards centrifuged at 13,000 r.p.m. for 10 min, to separate the phenol phase from the top liquid phase containing the DNA. The top phase was extracted and the DNA further purified using the FastDNA kit according to the manufacturer’s instructions. DNA concentrations were measured using Qubit (Life Technologies) and DNA integrity evaluated using gel electrophoresis.

**Library for functional selection.** DNA from WWTP AAW sampled in the third quarter of 2012 was used for functional selection. The DNA was sheared by Covaris E220 instrument (USA). Two hundred microlitres of DNA solution was sheared in a Covaris mini-tube according to Covaris shearing protocol for 2 kbp DNA fragments. Sheared DNA was end repaired using the End-It end repair kit (Epicentre) with the following protocol:

1. Combine:
   a. 68 μl of sheared DNA (90 ng·μl⁻¹)
   b. 10 μl 10 × end repair buffer
   c. 10 μl 2.5 mM dNTP mix
   d. 10 μl 10 mM ATP
   e. 2 μl of End It enzyme mix

2. Incubate at room temperature for 45 min.
3. Stop reaction by heating at 70°C for 20 min.

End-repaired DNA was size selected by electrophoresis through a 1% low melting point agarose gel in 0.5 × Tris-Borate-EDTA (TBE) buffer. A gel slice corresponding to 1,500–3,000 bp was excised from the gel and DNA was extracted using a QIAquick Gel Extraction Kit (Qiagen). DNA was ligated into pZ2E1 at the HincII site using the Fast Link ligation kit (Epicenter) with the following ligation protocol:

1. Combine:
   a. 1.0 μl 10 × fast ligation buffer
   b. 0.5 μl 10 mM ATP
   c. 1.0 μl HincII-cut pZ2E1 MCS 1 vector (65 ng·μl⁻¹)
   d. 2.50 μl sheared, end-repaired, gel-purified DNA insert (154 ng·μl⁻¹)
   e. 1.0 μl Fast-Link DNA ligase (2 units per μl)

2. Incubate at room temperature for 16 h.
3. Heat inactivate at 70°C for 20 min.

Then, 3 μl of the fresh ligation mixture was used for electro-transformation (Becton Aeroprotector) into 50 μl of electrocompetent E. coli TOP10 cells (Invitrogen). After transformation using standard protocols for a 2-mm electroporation cuvette (Biorad), cells were recovered in 1 ml super optimal broth with catabolite repression (SOC) medium for 1 h at 37°C. Libraries were titred by plating out 1, 0.1 and 0.01 μl of recovered cells onto lysogeny broth (LB) agar plates containing 50 μg·ml⁻¹ kanamycin. For each library, insert size distribution was estimated by gel electrophoresis of colony PCR products obtained by amplifying the insert using primers flanking the HincII site of the multiple cloning site of the pZ2E1 MCS1 vector (which contains a selectable marker for kanamycin resistance). The average insert size for all libraries was found to be 2.07 kb. The total size of metagenomic library was 8.28 × 10⁶ bp as determined by multiplying average PCR-based insert size with the number of colony forming units. The rest of the recovered cells were inoculated into 10 ml of LB containing 50 μg·ml⁻¹ kanamycin and grown overnight. The overnight culture was frozen down with 15% glycerol and kept at −70°C for subsequent sequencing.

**Functional selections of antibiotic-resistant clones from metagenomic library.** 100 μl of the insert library freezer stock corresponding to 5 × 10⁷ colony forming units, were plated out on LB agar plates containing binary combinations of kanamycin (50 μg·ml⁻¹) and 1 of 15 different antibiotics (Table 1), and incubated at 37°C for 16 h. Assay drug concentrations were chosen to prevent growth of overnight control libraries with the empty vector. Based on the original library titres as well as titres of the freezer stocks, the average amplification of a given library can be estimated. On average, each unique clone in the library was plated out with 10× coverage. To minimize the redundancy of clones in subsequent analysis, on average, the number of colonies picked for sequencing corresponded to approximately the number of resistant clones on an agar plate divided by the estimated average library amplification.

Each of the clones picked was inoculated into 96 deep-well plates containing liquid LB medium supplemented with kanamycin (50 μg·ml⁻¹) and the antibiotic
to which resistance had been selected (Table 1), and grown overnight to verify resistance phenotype.

**Sequencing and analysis of metagenomic inserts.** Functional selected metagenomic inserts were sequenced using Sanger sequencing. A total of 749 clones were sequenced bidirectionally using the following primers:

- **Forward primer:** pZE21_81_104_57C: 5’-GAA TTC ATT AAA GAG GAG AAA GGT-3’
- **Reverse primer:** pZE21_151_174rc_58C: 5’-TTT CGT TTG ATT TGA TGC CTC TAG-3’

All reads (≥500 bp) obtained after sequencing were assembled into contigs using CLC Main Workbench 6. The sequence from each contig corresponds to single clone containing unique insert. The full insert sequence for each unique insert was obtained by merging the forward primer sequence reads and reverse primer sequence reads using EMBL-Merger (Merger: www.ncbi.nlm.nih.gov/blast/spacer; accessed: 20 March 2013). Open reading frames were identified and annotated using ORFinder (ref. 55) with the settings SLIDINGWINDOW:1:3 and MINLEN:275. All high-quality PE reads were screened for phiX contamination using bowtie2 v. 2.1.0 (ref. 56), with standard settings and all matching reads removed. The potential phiX contamination is due to the use of an un-indexed phiX as a quality control, which can result in index carryover from nearby clusters with indexes.

Forward and reverse reads were merged using FLASH v. 1.2.7 (ref. 57), with the setting -r 300 -f 475 -s 50 -m 25 -M 200 and afterwards merged reads smaller than 425 bp or larger than 525 bp were discarded.

The merged reads were de-replicated and formatted for use in the sparse workflow (ref. 58). The merged reads were clustered using the usearch v. 7.0.1090-cluster_otelus with default settings. Operation taxonomic units (OTU) abundance was estimated using the usearch v. 7.0.1090–usearch_global with –id 0.97. Taxonomy was assigned using the RDP classifier (ref. 59) as implemented in the parallel_assign_taxonomy.qp.py script in QIIME (ref. 60) using the Greengenes database version 13_5 (ref. 60).

**16S rRNA data processing.** All sequenced sample libraries were subsampled to 50,000 raw reads and low-quality reads removed using Trimomatic v. 0.32 (ref. 55) with the settings SLIDINGWINDOW:1:3 and MINLEN:275. All high-quality PE reads were screened for phiX contamination using bowtie2 v. 2.1.0 (ref. 56), with standard settings and all matching reads removed. The potential phiX contamination is due to the use of an un-indexed phiX as a quality control, which can result in index carryover from nearby clusters with indexes.

Forward and reverse reads were merged using FLASH v. 1.2.7 (ref. 57), with the setting -r 300 -f 475 -s 50 -m 25 -M 200 and afterwards merged reads smaller than 425 bp or larger than 525 bp were discarded.

The merged reads were de-replicated and formatted for use in the sparse workflow (ref. 58). The merged reads were clustered using the usearch v. 7.0.1090-cluster_otelus with default settings. Operation taxonomic units (OTU) abundance was estimated using the usearch v. 7.0.1090–usearch_global with –id 0.97. Taxonomy was assigned using the RDP classifier (ref. 59) as implemented in the parallel_assign_taxonomy.qp.py script in QIIME (ref. 60) using the Greengenes database version 13_5 (ref. 60).

**16S rRNA data analysis and visualization.** All data analysis and visualizations were conducted using R (ref. 62) through the Rstudio IDE (http://www.rstudio.com/). Briefly, OTU counts and associated taxonomic assignments were imported and merged to a phyloseq object (ref. 61). All samples were rarefied to 10,000 reads and subsequently OTUs with a maximum count of <5 in the 24 samples were removed.

Community stability within WWTP AAW as a function of time was calculated by comparing Bray–Curtis β-diversity between all samples.

**Rarefaction curve analysis.** The rarefaction curves are based on the number of reads that map (>95% identity and >95% coverage) to the functional selected resistance genes. The curves are computed using the R package Vegan (step size = 1 x 10⁵) and ggplot2 (ref. 62,63). Abundance estimates are calculated using the Vegan implementation of the species accumulation model described by Chiu et al. (ref. 64).

**BLAST of functionally selected resistance genes.** Functionally selected resistance genes from WWTP AAW, human gut (ref. 12,13) and soil (ref. 36–38) were blasted against NCBI genbank nt database, excluding self-hits. The top hit was aligned to the functionally selected resistance gene using EMBoss Matcher and the identity was calculated (ref. 50).

**Mapping to clinically relevant genes.** The trimmed metagenome reads were mapped to reference database (Supplementary Table 1) using CLC Genomics Workbench v. 7.03, using a minimum of 95% similarity over the full read length. The number of reads hitting the putative antibiotic gene were counted and further data analysis and visualization was conducted using R (ref. 62).

**Procrustes analysis.** Square-root-transformed metagenome abundance counts of the antibiotic resistance genes were used for a principal component analysis (PCA) analysis in R through the vegan package (ref. 52). For comparison, matching samples (for a few samples, the closest time point had to be used as matching samples were not available) from the larger 16S rRNA data set were extracted and square-root-transformed OTU abundance counts used for a PCA analysis. The two PCAs were compared using the procrustes function implemented in vegan and significance tested using the protest function with 999 permutations.

**References.**

12. Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of
11. Forsberg, K. J. et al.
16. Renew, J. E. & Huang, C.-H. Simultaneous determination of fluoroquinolone,
8. Walsh, T. R., Weeks, J., Livermore, D. M. & Toleman, M. A. Dissemination of
0. Luo, Y. et al. Proliferation of multidrug-resistant New Delhi metallo-

Acknowledgements
This research was primarily funded by the Lundbeck Foundation, EU FP7-Health Program Evotar (282004) and the Danish Council for Independent Research (Technology and Production Sciences). M.O.A.S. acknowledges additional funding from The Danish Free Research Councils, ERC and the Novo Nordisk Foundation.

Author contributions
P.H.N. and M.O.A.S. designed the study. A.T. performed functional selection. M.A. collected samples and performed metagenomic sequencing. C.M., M.A., A.T. and M.E. performed data analysis. C.M., M.A., P.H.N. and M.O.A.S. wrote the manuscript.

Additional information
Accession codes: Nucleotide sequences of the functionally selected resistance genes have been deposited in the Genbank database with accession codes KT387137 to KT387215. The raw metagenome reads have been deposited in the European Nucleotide Archive under study PRJEB8087. The assembled metagenome scaffolds from AAW.5.2010 have been deposited in the MG-RAST database with accession code MG-RAST4611649.3. The raw V13 16S rRNA amplicon sequences are part of the MiDAS data set (www,midasfieldguide.org) and have been deposited in the European Nucleotide Archive under study PRJEB8105. The accession codes for the specific samples used are: ERS634988, ERS634989, ERS634990, ERS634991, ERS635004, ERS635036, ERS635037, ERS635038, ERS635362, ERS635433, ERS635452, ERS635453, ERS635454, ERS635455, ERS635477, ERS635478, ERS635521, ERS635529, ERS635539, ERS635540, ERS635541, ERS635549 and ERS635550.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/