Classification of hospital and urban wastewater resistome and microbiota over time and their relationship to the eco-exposome.

Elena Buelow1*, Andreu Rico2, Margaux Gaschet1, José Lourenço3, Sean P. Kennedy4, Laure Wiest5, Marie-Cecile Ploy1#, Christophe Dagot1#

1University Limoges, INSERM, CHU Limoges, RESINFIT, U1092, F-87000 Limoges, France;

2IMDEA Water Institute, Science and Technology Campus of the University of Alcalá, Avenida Punto Com 2, 28805, Alcalá de Henares, Madrid, Spain;

3Department of Zoology, University of Oxford, Oxford, UK;

4Biomics Pole, CITECH, Institut Pasteur, Paris 75015, France;

5Univ Lyon, CNRS, Université Claude Bernard Lyon, Institut des Sciences Analytiques, UMR 5280, 5 Rue de la Doua, F-69100, Villeurbanne, France;

#These authors participated equally to the work.

*Corresponding author: elena.buelow@gmail.com

UMR Inserm 1092, Agents Anti-Microbiens, CBRS, Faculté de médecine, 1 rue du Pr Bernard Descottes, 87000 Limoges

Phone number: +33 55 5 19 56 42 62

Abstract

Wastewaters (WW) are important sources for the dissemination of antimicrobial resistance (AMR) into the environment. Hospital WW (HWW) contain higher loads of micro-pollutants and AMR markers than urban WW (UWW). Little is known about the long-term dynamics of H and U WW and the impact of their joined treatment on the general burden of AMR. Here, we characterized the resistome, microbiota and eco-exposome signature of 126 H and U WW samples treated separately for three years, and then mixed, over one year. Multi-variate analysis and machine learning revealed a robust signature for each WW with no significant variation over time before mixing, and once mixed, both WW closely resembled U signatures. We demonstrated a significant impact of pharmaceuticals and surfactants on the resistome and microbiota of H and U WW. Our results present considerable targets for AMR related risk assessment of WW.

41 The worldwide spread of multidrug-resistant bacteria is an important public health issue with a high
42 health and economic burden[1–3]. A global "One Health" approach is urgently needed to combat the
43 dissemination of antibiotic resistant bacteria (ARB) and antimicrobial resistance genes (ARGs) from
44 humans and livestock to the environment and vice versa, as well as to identify key drivers contributing to
45 the selection, dissemination and persistence of ARB and/or ARGs[4,5]. The natural environment and its
46 biodiversity serve as a wide reservoir of genetic determinants implicated in resistance to antimicrobial
47 compounds[6,7]. Human activity has a significant impact on the terrestrial and aquatic microbial ecosystems
48 through chemical pollutants that are spread via urban, agricultural and industrial waste and which pose
49 an important selective pressure for AMR[8]. For instance, urban and hospital wastewaters (UWW and
50 HWW) contain a high diversity of ARGs and chemicals[9–12]. It is therefore generally accepted that the
51 implementation of efficient wastewater treatment plants (WWTP) is essential in order to reduce the
52 amounts of chemicals, ARGs and ARB that reach the environment[4,9,13]. The treated WW are re-introduced
53 into the aquatic environment and the produced sludge often re-used in agricultural lands[11,14]. However,
54 despite a global reduction of ARGs through treatment, effluents from urban, hospital and industrial
55 wastewater still contain ARGs, antibiotics (ABs) and moderate levels of other pollutants affecting
56 microorganisms (e.g. biocides, heavy metals)[11,15,16]. HWWs have been reported to contain particularly high
57 amounts of ARGs, ARBs and ABs[9,12,17]. It has been debated whether HWW contributes significantly to the
58 load of ARGs in the UWW systems, and whether separate treatment for HWWs should be applied[9,13,18].
59 Recent work has shown that HWW has limited impact on the relative levels of ARGs and mobile genetic
60 elements (MGEs) associated with ARGs like integrons, in hospital receiving urban wastewater (WW)[9,15,19].
61 However, most studies usually analyzed a limited number of samples and yet, longitudinal studies that
62 monitor WW dynamics are so far lacking, which limits the possibility of assessing the risk for AMR
63 mediated through WW.
64 We thus studied the dynamics of the resistome, microbiota and of some chemical-pollutants present in
65 WW (the "eco-exposome") of 126 WW samples (UWW, HWW and mixed WW) in a French city during a
66 period of approximately four years: 34 months with separate treatments for H and U WW and 11 months
67 with H and U WW mixed. We studied H and U WW in the course of their passage through i) two
68 independent wastewater treatment systems applying the conventional secondary (activated sludge)
69 treatment process and ii) then mixed 1:2 (HWW:UWW) into one system. We investigated the relationship
70 of the respective resistome and microbiota with the measured eco-exposome (pharmaceuticals, mainly
71 antibiotics; surfactants and heavy metals). Multi-variate analysis and machine learning, reveal a robust
72 signature of the resistome, microbiota and eco-exposome of HWW compared to UWW with no significant

variation over time. We also showed that, when mixed, both WW closely resemble urban signatures. Furthermore, we demonstrated that pharmaceuticals and surfactants had a large influence on the variability of the monitored resistome and microbiota of H and U WW.

Results:

HWW and UWW have distinct resistome and microbiota signatures. We evaluated the resistome and microbiota of monthly WW (N=126) and river (N= 12) samples. For the resistome, we targeted 78 genes by high-throughput qPCR[9,20] conferring resistance to antibiotics, quaternary ammonium compounds, or heavy metals, grouped into 16 resistance gene classes. The genes targeted include ARGs that are most commonly detected in the gut microbiota of healthy individuals[21,22], clinically relevant ARGs (including genes encoding extended spectrum β-lactamases (ESBLs), carbapenemases, and vancomycin resistance), and heavy metal and quaternary ammonium compound resistance genes suggested to favor cross and co – selection for ARGs in the environment[23,24]. We also targeted genetic elements as important transposase gene families[25] and class 1, 2 and 3 integron integrase genes, that are important vectors for ARGs in the clinics and often used as proxy for anthropogenic pollution[26] (gene targets and grouping of genes into gene classes/according to function are detailed in Supplementary Tables 5 and 6).

H and U WW samples were treated separately through 2012 and 2014, and were mixed at a ratio of 1:2 (HWW:UWW) throughout the year 2015 (Figure 1). H and U WW samples exhibited a distinct signature with respect to the proportional makeup of their resistome (Figure 2a) and microbiota (Figure 2b). Analyzing the data with a Random Forest machine learning approach showed that the distinct H and U WW signatures resulted in a high prediction accuracy. When using the resistome as predictor (on the level of gene classes), 93.5% and 96.7% of untreated HWW and UWW samples respectively, could be correctly classified (Supplementary Figure 1a). We further analyzed the data on the individual gene level to increase resolution of the machine learning approach. Using individual genes resulted in similarly high predictions (93.5% prediction for untreated HWW and 100% prediction for untreated UWW) (Supplementary Figure 1b). Similarly, when using the microbiota, 96.8% of untreated HWW and 89% of untreated UWW samples were correctly classified (Supplementary Figure 1c).

For the treated H and U WW the machine learning prediction accuracy was lower compared to the untreated WW sources but still considerably high for all predictor levels and in particular for the microbiota (≥80%) (Supplementary Figure 2).

For the MWW overall classification success was lower (Supplementary Figure 1 and 2), given a much lower sample size and likely due to the mixing of the two wastewater sources hampering clear signatures.

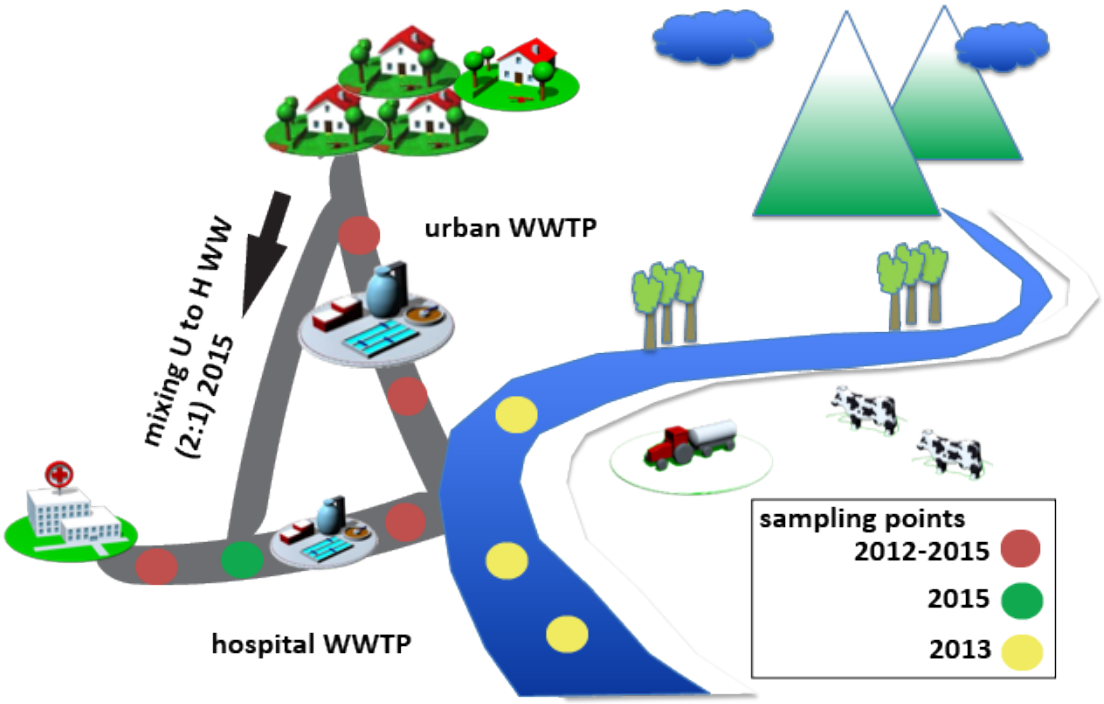


Figure 1: **Sampling site**. Samples were collected in monthly intervals (untreated and treated samples) by flow proportional sampling, from March 2012 through November 2015. From March 2012 to December 2014, UWW and HWW were treated by separate wastewater treatment plants (WWTPs). During the period from January 2015 through November 2015, UWW was mixed into the HWW (1:2 ratio HWW:UWW) and added to the separate HWW treatment line resulting in mixed WW (MWW). In addition, 12 water samples of the effluent receiving river up (river upstream) and downstream (river downstream sampling point 1 and 2) of the effluent release pipes were collected during the winter months of 2013 (January, February, November, and December).

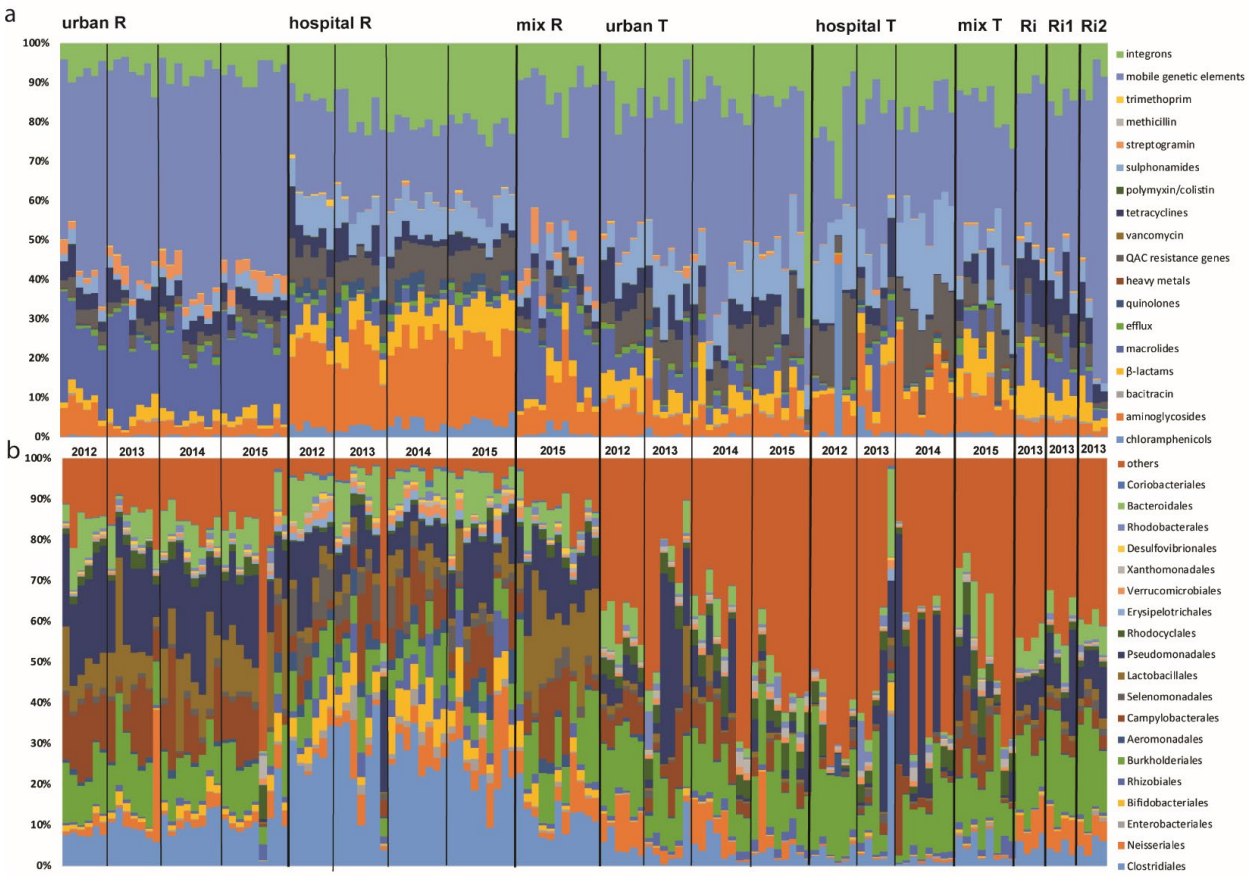


Figure 2: **Proportional abundance of the resistome and microbiota in untreated (R) and treated (T) HWW, UWW and MWW, as well as river samples up (Ri) and downstream (Ri1 and Ri2) of the waste water release pipes throughout the sampling period (2012-2015)**. **a:** proportional abundance of resistome (ARG classes, heavy metals integron integrase genes and mobile genetic elements) for all samples. **b**: proportional abundance of the microbiota (displaying the 20 most abundant bacteria at the order level for all samples, where "others" represents the percentage of the remaining taxa).

131 **Temporal dynamics of H and U WW:** Redundancy analysis was performed to assess putative significant
132 influences of time on the variability of resistome and microbiota compositions using sampling year or
133 season as independent variables (Monte Carlo (MC) permutations (n=499)). Analysis was carried out for
134 all different sample groups together (untreated U, H, M WW and treated U, H, M WW) to potentially
135 identify general patterns influencing all groups at the same time. The same analysis was also performed
136 per individual sample group (untreated and treated H, U and M WW alone, respectively).

137 For untreated HWW and UWW, neither year (p=0.6 resistome and p=0.4 microbiota respectively)
138 nor season (p=0.9 and p=0.3) had a significant impact on the resistome (resistance gene classes) and
139 microbiota composition of the WW over the first 3 years before mixing (Supplementary Table 1a).
140 However, by analyzing the sample groups separately, a yearly and seasonal impact on the level of
141 individual resistance genes in HWW was demonstrated (p=0.004 year, p=0.032 season; Supplementary
142 Table 1b). Redundancy analysis revealed the relationship between individual genes and gene classes with
143 seasons, pointing towards a correlation of increased normalized abundance of individual genes and gene
144 classes detected during summer season (Supplementary Figure 3a, Supplementary Figure 4). The fact that
145 the hospital was only installed in February 2012[27] may explain why the year 2012 is particular for HWW
146 resistome, with overall lower normalized abundance of the resistome and no obvious trend towards the
147 summer season in 2012. After HWW and UWW mixing (in 2015), no significant variation for the resistome
148 and microbiota composition throughout the seasons could be observed considering all sample groups
149 (untreated MWW, HWW and UWW; p=0.6 resistome and p=0.07 microbiota respectively).

150 For treated WW, the analysis exhibited more variation of the resistome and microbiota
151 composition compared to the untreated sources over the years (2012-2014 for H and U WW; Figure 2,
152 Supplementary Table 1c). Redundancy analysis revealed that in particular the microbiota of treated H and
153 U WW effluents varied between the years (p=0.016 for all groups, p=0.021 for treated HWW and p=0.006
154 for treated UWW, Supplementary Table 1d) while no significant seasonal impact could be observed. The
155 resistome varied between the years for treated HWW on both the gene class and the individual gene
156 levels (p=0.004 and p=0.012, Supplementary Table 1d), whereas for the treated UWW this variation was
157 only significant on the individual gene level (p=0.028). For the treated mixed WW (MWW), significant
158 seasonal variation was observed for the microbiota composition (p=0.02), while for the resistome no
159 significant variation could be observed.

160 **The untreated HWW resistome is significantly diluted by UWW in mixed WW.** The bacterial biomass
161 (absolute copy numbers of 16S rRNA genes per liter of water) was comparable for the untreated WW
162 sources (HWW, UWW and MWW) (Supplementary Figure 5). Exemplary, here ratios of HWW over UWW

163 and HWW over MWW were calculated based on the averaged cumulative abundance of the resistome in
164 untreated HWW and UWW (Table 1).
165 The untreated HWW contained significantly more gene classes compared to the untreated UWW,
166 between 3 (transposase genes) and 161-fold (*qnr* genes encoding quinolone resistance) higher (p values
167 ≤ 0.004). When WW were mixed at the experimental ratio of 1:2 (HWW:UWW), the untreated MWW
168 contained significantly less resistance gene classes compared to untreated HWW, between 3 and 22-fold
169 lower (p values ≤ 0.03). Interestingly, there was no significant difference for the genes encoding resistance
170 to macrolides for both HWW over UWW and HWW over MWW comparisons (Table 1, Figure 3a and 3b).
171 The only resistance gene significantly lower in HWW compared to UWW or MWW ($p<0.0001$), was the
172 streptogramin resistance gene *vatB*. The *mecA* gene encoding resistance to methicillin was undetectable
173 in all UWW and in all but one MWW samples.
174 Altogether, these data indicate a significant dilution impact of UWW on the normalized abundance of the
175 targeted resistome of HWW when mixing at the experimental ratio of 1:2 (HWW:UWW).
176

177 **Table 1:** Average fold changes for gene classes cumulative abundance of untreated HWW over UWW (2012-2015;
178 significant differences indicated by asterisk *; *p* values ≤ 0.004) and HWW over MWW (2015, significant differences
179 indicated by asterisk *; *p* values ≤ 0.03) ± Standard Deviation. Significant differences were calculated by comparing
180 the normalized cumulative abundance values of individual gene classes for all samples belonging to each sample
181 group using the non-parametric Mann-Whitney test. Fold changes were calculated for individually paired samples
182 for each gene class / sample group. NA indicates that gene classes were undetectable in either one or both of the
183 sample groups.

Gene classes conferring resistance to:	**Fold change untreated Hospital WW/Urban WW**	**Fold change untreated Hospital WW/Mixed WW**
chloramphenicol	84 (±93)*	13 (±11)*
aminoglycosides	43 (±31)*	8 (±6)*
bacitracin	8 (±13)*	7 (±4)*
beta-lactams	26 (±22)*	9 (±6)*
macrolides	1 (±1)	0.6 (±0.3)
(multi-drug) Efflux	9 (± 11)*	4 (±4)*
quinolones (qnr)	161 (±326)*	10 (±8)*
heavy metals	7 (±9)*	4 (±3)*
quaternary ammonium compounds QACs)	18 (±14)*	7 (±4)*
vancomycin	12 (±35)*	18 (±20)*
tetracycline	4 (±3)*	3 (±2)*
polymixin	8 (±9)*	4 (±4)*
sulphonamides	19 (±12)*	7 (±4)*
methicillin	NA (undetectable in UWW)	NA (undetectable in all but one MWW sample)
streptogramin	0.2 (±0.2)*	0.2 ((±0.2)*
trimethoprim	8 (±6)*	5 (±3)*
Gene classes grouped according to function:		
transposase genes (MGEs)	3 (±1)*	2 (±1)
integron integrase genes	16 (±9)*	6 (±3)*

184

185 **Resistome reduction through WW treatment.** The bacterial biomass (copies of 16S rRNA / liter) for all
186 WW sources was decreased by 2-3 log after WW treatment (Supplementary Figure 5). To estimate the
187 impact of WW treatment on the resistome, fold changes of untreated over treated H, U and M WW were
188 calculated. The normalized cumulative abundance of all gene classes significantly decreased and was
189 between 78 times (for genes conferring resistance to quinolones) and 5 times (for genes conferring
190 resistance to QACs, sulphonamides and genes encoding transposase genes) lower in the treated HWW
191 compared to untreated HWW ($p<0.003$) (Table 2; Figure 3a and 3b).When comparing untreated UWW to
192 treated UWW, we showed a significant reduction ($p<0.05$) in the normalized cumulative abundance for 9
193 resistance gene classes with fold changes between 43 (for the streptogramin resistance gene *vatB*) and 3
194 (for genes encoding resistance to aminoglycosides) times (Table 2, Figure 3a and 3b). No significant
195 reduction was observed after treatment of UWW for gene classes conferring resistance to bacitracin,
196 beta-lactams, quinolones, heavy metals and quaternary ammonium compounds, and for genes encoding
197 integron integrases. Surprisingly, sulphonamide resistance encoding genes were found to be significantly
198 enriched after UWW treatment ($p<0.05$) (Table 2, Figure 3a and 3b). For MWW, a similar removal efficacy

199 as for UWW could be observed (Table 2, Fig. 3b), with significant decrease for the normalized cumulative
200 abundance of 10 genes classes and with fold changes between 41 (for the streptogramin resistance gene
201 *vatB*) and 3 (for the genes encoding integron integrase genes) times. No significant decrease for gene
202 classes conferring resistance to bacitracin, beta-lactams, tetracycline, heavy metals, QACs and
203 sulphonamides could be detected.
204

205 **Table 2:** Average fold changes for gene classes of untreated WW over treated WW for H, U and M WW (*p* values ≤
206 0.03) ± Standard Deviation. *= significantly lower; + = significantly higher. Significant differences were calculated by
207 comparing the normalized cumulative abundance values of individual gene classes for all samples belonging to each
208 sample group using the non-parametric Mann-Whitney test. Fold changes were calculated for individually paired
209 samples for each gene class / sample group. Average fold change ± Standard Deviation are depicted in the table for
210 comparison. NA indicates that gene classes were undetectable in either one or both of the sample groups.

Gene classes conferring resistance to:	**Fold change H untreated WW/H treated WW**	**Fold change U untreated WW/U treated WW**	**Fold change M untreated WW/ M treated WW**
chloramphenicol	34 (±47)*	5 (±12)*	16 (±27)*
aminoglycosides	15 (±14)*	3 (±5) *	7 (±10)*
bacitracin	12 (±30)*	1 (±2)	1 (±1)
beta-lactams	30 (±42) *	4 (±14)	10 (±21)
erythromycin (macrolides)	48 (±68)*	27 (±55)*	23 (+17)*
(multi-drug) efflux	29 (±43)*	9 (±27)*	13 (±17)*
quinolones	78 (±123)*	3 (±5)	19 (±29)*
heavy metals	15 (±32) *	3 (±3)	2 (±2)
quaternary ammonium compounds QACs)	5 (±5) *	1 (±2)	2 (±2)
vancomycin	NA (undetectable in all but one sample for treated HWW)	NA (undetectable in all but one sample for treated UWW)	NA (undetectable in all samples for treated MWW)
tetracycline	51 (±77)*	13 (±40)*	6 (±5)
polymixin	29 (±52)*	5 (±12) *	7 (±7)*
sulphonamides	5 (±6)*	1 (±1)$^{+}$	2 (±1)
Streptogramin	17 (±43)*	43 (±77)*	41 (±20)*
methicillin	NA (undetectable in all treated HWW samples)	NA (undetectable in all treated UWW samples)	NA (undetectable in all but one untreated MWW sample)
trimethoprim	66 (±101)*	9 (±28)*	6 (±7)*
Gene classes grouped according to function:			
transposase genes (MGEs)	5 (±5)*	6 (±7)*	5 (±3)*
integron integrase genes	7 (±6)*	1 (±1)	3 (±4)*

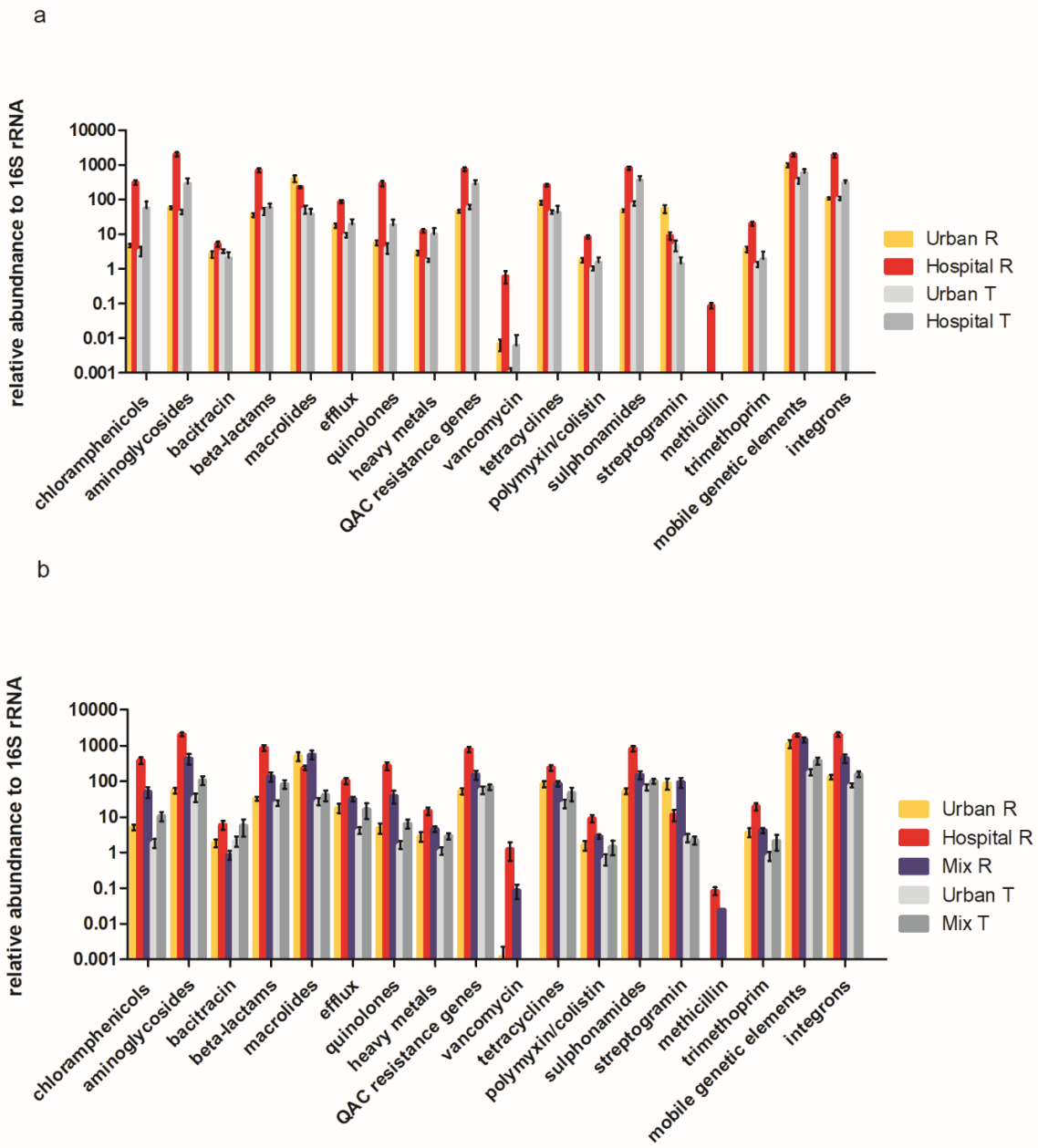


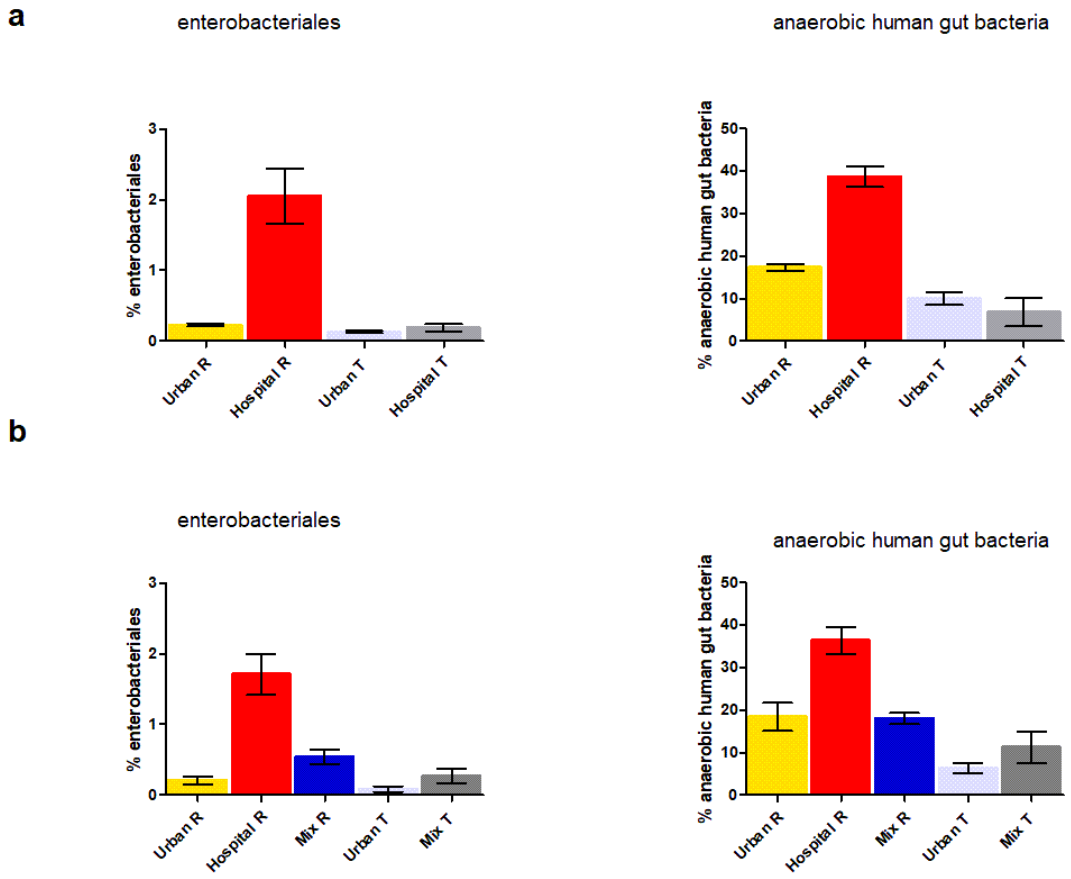
Figure 3: Averaged normalized abundance of ARG classes, heavy metals, MGEs and integrons over all collected samples per sample type +- standard deviation. a: relative abundance of ARG classes, heavy metals, QACs, MGEs and integrase genes in untreated (n21) and treated HWW (n=19), untreated (n=21) and treated (n=20) UWW averaged over the numbers of samples collected for each water type in the given time interval (March 2012 – December 2014) +- standard deviation. **b**: averaged normalized abundance of ARG classes, heavy metals, QACs, MGEs and integrase genes in untreated (n=10) HWW, untreated (n=9) and treated (n=8) UWW and untreated (n=10) and treated (n=8) MWW (at the experimental ratio of 1/3 HWW to 2/2 UWW) WW samples averaged over the numbers of samples collected for each sample group in the given time interval (January 2015- November 2015) +- standard deviation.

221 **No evident impact of treated hospital WW on the receiving river.** We also analyzed 12 river samples up
222 and downstream the WWTP to evaluate putatively associated risks with the release of the treated WW
223 into the effluent receiving river and downstream environment (Supplementary Figure 6). The resistome
224 for the river samples collected for the sites up (Ri) and downstream (Ri1 and Ri2) of the effluent release
225 pipes during winter season 2013 was not significantly different for either of the three sampling sites
226 (Supplementary Figure 6). There was no significant difference of the relative abundance for any of the
227 detected gene classes in the river samples compared to treated UWW. On the contrary, the relative
228 abundance of nine resistance gene classes including genes encoding MGEs and integron integrase genes
229 was significantly lower ($p<0.04$) in river samples when compared to treated HWW (Supplementary Figure
230 6).

231 **Human gut bacteria are enriched in HWW.** The human gut microbiota is an important reservoir of
232 ARGs[20,28,29] and bacteria of the human gut are likely to be shed into the environment via wastewaters that
233 contain at least partially human feces. We calculated the relative abundance of anaerobic human gut
234 bacteria, as well as Enterobacteriales in the respective WW. Enterobacteriales were specifically detected
235 as many of these Gram-negative bacteria are also pathogens. The orders Clostridiales, Bifidobacteriales
236 and Bacteroidales which represent the most important and abundant anaerobic human gut bacteria[30,31],
237 were grouped together and are referred to as anaerobic human gut bacteria. Untreated HWW contained
238 significantly higher levels of anaerobic human gut bacteria (38% ± 11 standard deviation) and
239 Enterobacteriales (2% ± 1.5) compared to all other samples (Figures 4a and 4b). Interestingly, these orders
240 are comparable in their relative abundance for untreated UWW and MWW, indicating a significant
241 dilution effect of UWW in HWW (Figure 4b), as observed for the resistome. The treatment allowed a
242 significant ($p<0.05$) decrease of the relative abundance of these orders for HWW and UWW (Figure 4a
243 and 4b) whereas the reduction for MWW was only marginally significant ($p=0.05$) (Figure 4b).

244

245



246

247 **Figure 4: a: Relative abundance of anaerobic human gut bacteria (Clostridiales, Bifidobacteriales and**
248 **Bacteroidales) and Enterobactertiales** in untreated (n=21) and treated HWW (n=19), UWW (n=21) and (n=20)
249 averaged over the numbers of samples collected for each sample group between March 2012 and December 2014
250 +- standard deviation. **b**: relative abundance of anaerobic human gut bacteria and Enterobactertiales in untreated
251 (n=10) HWW, untreated (n=9) and treated (n=8) UWW and untreated (n=10) and treated (n=8) MWW samples (at
252 the experimental ratio of 1:2 HWW:UWW) averaged over the numbers of samples collected for each sample group
253 between January 2015 and November 2015 +- standard deviation.

254

255 **The eco-exposome plays an important role in shaping the resistome and microbiota in hospital and**
256 **urban WW**. In order to estimate the eco-exposome in the WW, selected pharmaceutical and chemical
257 compounds including antibiotics, surfactants and heavy metals were quantified in untreated WW
258 (Supplementary Table 2).

259 The relationship between the resistome/microbiota and the eco-exposome was visualized by
260 means of PCA biplots. Here, HWW and UWW form two distinct clusters, while the MWW clusters closely
261 to the UWW for both resistome and microbiota (Figure 5a and 5b). Antibiotics, non-ionic and cationic

262 surfactants are the most important contributors to the respective HWW resistome and microbiota (length
263 of arrows), while anionic surfactants and some metals are more associated with the resistome and
264 microbiota of the UWW (Fig. 5a and 5b). Moreover, the putative impact of the measured chemicals (eco-
265 exposome) on the resistome and microbiota variation was statistically assessed by RDA. The results of the
266 Monte-Carlo permutations indicated that the eco-exposome significantly influences the resistome and
267 microbiota (p-values 0.002).

268 Finally, a variation partitioning analysis was performed to study which group of the measured
269 compounds (heavy-metals, pharmaceuticals, surfactants) might have a larger contribution on the
270 resistome and microbiota variation, and to explore whether the interaction between these compounds
271 has a stronger influence than the individually grouped compounds (Figure 5c and 5d). Pharmaceuticals
272 (the antibiotics ciprofloxacin, sulfamethoxazole and vancomycin, and the neurological drug
273 carbamazepine) explain the largest proportion of the variance for the resistome, while the surfactants
274 have the largest impact on the variation for the microbiota. Finally, we show that the interaction between
275 pharmaceuticals and surfactants contributes more to the variability in the resistome than the individual
276 compounds alone, while such interactions are less clear in the microbiota dataset (Fig. 5c and 5d).

277 We also collected data on the consumption of three antibiotics, ciprofloxacin, sulfamethoxazole
278 and vancomycin, by the hospital pharmacy over the period of 2012 to 2014, that were summarized here
279 as gram per season (Supplementary Table 3 and Supplementary Figure 3c). The studied hospital site has
280 just been opened in the winter month of February 2012 which is probably why antibiotic consumption by
281 the hospital pharmacy was low during winter 2012. No obvious correlation between summer peaks for
282 the measured antibiotics in WW and their respective consumption by the hospital pharmacy was shown.
283 (Supplementary Tables 3 and 4, and Supplementary Figure 3b and 3c). The observed peaks and variation
284 during summer season for antibiotics, individual resistance genes and gene classes in HWW
285 (Supplementary Figures 3a, b and 4) may be due to dry season during summer that could result in
286 decreased flow rate of HWW.

287

288

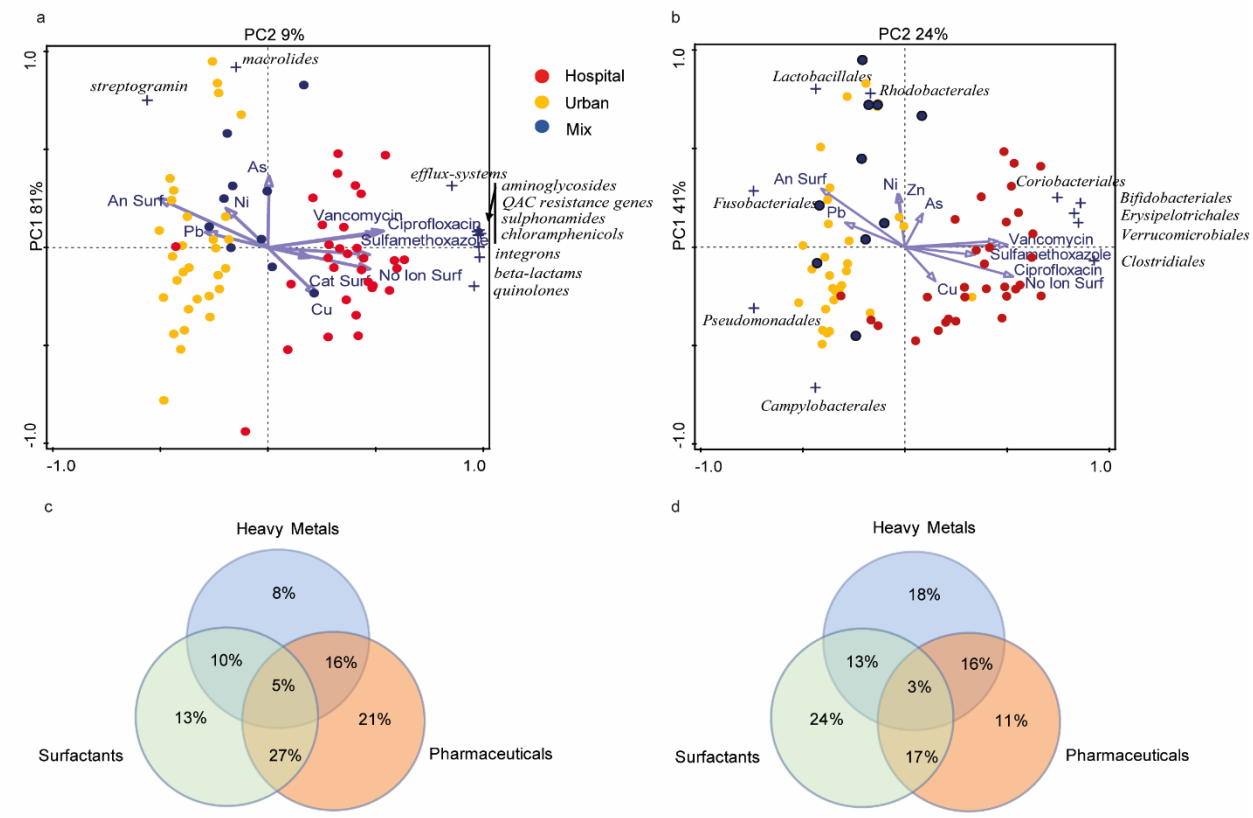


Figure 5: Principal component analysis showing the relationship between the eco-exposome (heavy metals, pharmaceuticals and surfactants) and the resistome (a) and microbiota (b) of untreated HWW, UWW and MWW samples; and Venn diagrams showing the results of the variation partitioning analysis with the different measured chemical classes and the resistome (c) and microbiota (d). In the PCA analysis, dots refer to urban (yellow), hospital (red) and mixed (blue) untreated WW samples.

296 **Discussion**

297 In the context of globally increasing AMR, wastewaters (WW) have been identified as sources for the
298 spread of AMR determinants (ARB, ARGs, MGEs) and chemical pollutants (often pharmaceutical residues)
299 that may favor AMR selection during wastewater treatment and in the receiving environment. In this
300 study we thoroughly monitored the resistome and microbiota dynamics of untreated and treated
301 (applying conventional secondary WWT) hospital and urban WW over four years throughout the seasons
302 in France. We identified distinct and robust resistome and microbiota signatures, in particular for
303 untreated HWW and UWW, indicating that HWW and UWW form distinct and stable ecological niches
304 over time. Performing machine learning (ML) classified each of the sources with high accuracy and
305 revealed top predictive genes, gene classes and taxa for the respective WW sources before and after WW
306 treatment (Supplementary Figure 8). Interestingly, when collapsing data-sets obtained from untreated
307 and treated samples of the respective WW sources, ML was able to predict HWW and UWW in general
308 with more than 93% certainty on all predictor levels (individual genes, gene classes, taxa) (Supplementary
309 Figure 9). These top 10 predictors for HWW and UWW (Supplementary Figure 8 and 9) provide
310 considerable marker gene classes, individual genes and taxa for the respective WW sources that indicate
311 underlying important differences in risk to the environment and that present targets for the monitoring
312 and management of HWW and UWW. A recent study assessed the removal efficacy of 62 Dutch
313 wastewater treatment plants (WWTPs), applying conventional secondary WW treatment, on a selected
314 panel of six ARGs and the gene encoding the class 1 integron integrase gene[15]. These genes (*emrB, sul1*
315 and *sul2* (in our study synonym with *sulA*), *qnrS*, *tetM*, *blaCTX-M* and *intI*), proposed as general WW
316 markers for risk assessment[5], were also monitored by our resistome approach and were similarly classified
317 as high predictors for the respective WW sources (Supplementary Figure 7, 8 and 9). All seven genes were
318 detected in both WW sources, however in different normalized abundances (e.g. *ermB* was more
319 abundant in UWW compared to HWW, whereas the genes *sul1, sul2, qnrS, blaCTX-M* and *intI1* were more
320 abundant in HWW; interestingly the normalized abundance of *tetM* was comparable in both sources). We
321 were also able to identify additional genes that could be implemented for the classification of HWW vs
322 UWW based on their normalized abundance. For example, the streptogramin resistance gene *vatB*, and
323 the transposase gene *ISS1N* were significantly more abundant in UWW compared to HWW, and seem to
324 be specifically indicative for UWW.

325 The HWW resistome was found to be significantly enriched with resistance gene classes and genes
326 encoding integron integrase genes compared to UWW except for the streptogramin resistance gene *vatB*.
327 For the gene class conferring resistance to macrolides, we did not observe a significant difference between

328 HWW and UWW; however, there is a trend for the genes encoding resistance to macrolides to have higher
329 normalized abundance in UWW than in HWW (Figure 3a and 3b; Table 1). Based on the machine learning
330 approach, macrolide resistance genes also contribute to the specific resistome signature for UWW
331 (Supplementary Figure 8 and 9). Considering the fact that macrolide and streptogramin antibiotics are
332 more frequently prescribed in the community compared to the hospital environment in France, could
333 explain the high abundance of these gene classes in UWW[32]. In the hospital environment, antibiotics such
334 as quinolones, beta-lactams, aminoglycosides and vancomycin are frequently used, which also could
335 explain the relatively high abundance of gene classes conferring resistance to those antibiotics in HWW
336 compared to UWW[32]. Furthermore, the antibiotics ciprofloxacin, sulfamethoxazole and vancomycin were
337 detected in higher concentrations in HWW compared to UWW, which may point towards a relationship
338 of the measured antibiotics and the detected resistome in HWW (Figure 5a and c). Interestingly, the *qnr*
339 genes encoding quinolone resistance in HWW were the ones with the highest fold increase (161-fold)
340 between HWW and UWW. As these genes are located on plasmids, their higher abundance in HWW
341 indirectly reflects the likely abundance of bacteria harboring genetic elements involved in resistance
342 dissemination as plasmids in HWW. Indeed, *qnr* genes have been described mainly in
343 Enterobacteriales[33,34] and we found that the relative abundance of Enterobacteriales is higher in HWW
344 than in UWW (Figure 4).
345 The human gut microbiota is an important reservoir for ARGs[28,29] and recently evidence-based data
346 showed that the occurrence and abundance of (human) fecal pollution is a likely explanation for the
347 detection of high amounts of ARGs in anthropogenically impacted environments[35,36]. The significant
348 dilution of human gut bacteria in MWW observed here, is hence also likely to explain the significant
349 reduction of the abundance of gene classes after mixing HWW with UWW.
350 A significant dilution of HWW-associated genes in UWW was described previously, but under
351 circumstances that reflected a much lower rate of contribution of HWW to UWW (between 0.8 and 2.2%)
352 at the respective study sites[9,15,19] compared to our study site (here HWW ~33.4%). Here we show that
353 UWW can dilute the normalized abundance of resistance genes, MGEs and integrons in untreated HWW
354 significantly when mixed, even with an increased proportional contribution of HWW to UWW. This implies
355 that on the scale of a small sized urbanized area as studied here, WW mixing of U and HWW bares no
356 greater risk than separate treatment. The increase of the effluent flow rate entering in the WWTP treating
357 the MWW due to the injected UWW may contribute largely to the observed significant dilution impact of
358 the UWW on the HWW. Previous studies conducted on the same experimental site for shorter time
359 periods (up to 2 years) coherently provided a detailed catalogue of pharmacological parameters, such as

360 specific surfactants, that could further aid to discriminate UWW from HWW[27,37–39]. They also concluded
361 that there is no greater advantage associated with separate treatment of HWW from UWW with respect
362 to their pharmaceutical discharges and ecotoxicological impacts[27].
363 With respect to the removal efficacy for ARGs and MGEs through secondary UWW treatment, we
364 observed that the normalized abundance of six gene classes did not significantly decrease while one gene
365 class (sulphonamides) did significantly increase (Table 2, Fig 3a and Fig 3b). The removal efficacy of
366 secondary WW treatment in terms of absolute abundance of ARGs and ARBs is significant by reducing the
367 overall release of ARGs and ARB into the downstream environment in general by more than 95%[15].
368 However, unchanged, or even increased relative or normalized abundance of ARGs and gene classes is
369 indicative for selective processes during secondary WW treatment favoring the spread of AMR into the
370 environment[15,40–42]. The resistome monitored here, exhibits a high mobilization potential due to the high
371 proportion of MGEs, integrons and plasmid borne ARGs detected in all WW samples, as well as in the
372 receiving river waters (Figure 1, MGEs and integrons account for up to 60% of the resistome of treated
373 effluents and river waters). The MGEs, more specifically transposase genes targeted here, represent genes
374 that are associated with the dissemination of ARGs by means of conjugative integrative elements or
375 plasmids, across environmental bacteria and human pathogens[25,40]. There is evidence that advanced WWT
376 such as disinfection by UV radiation or ozone treatment, or physical treatment by ultrafiltration of WW
377 are more efficient in reducing ARB and ARGs compared to conventionally applied secondary WW
378 treatment. However, despite reducing ARB and ARG loads significantly compared to secondary treatment
379 often ARBs and ARGs are not completely eliminated by advanced WWT either[43].
380 The exposome, a term originally coined in the context of human health epidemiology and referring to "the
381 totality of human environmental exposures"[44], here specifically refers to the chemical compounds
382 quantified in our longitudinal study that represent partially the environmental or "eco-exposome"[45] of
383 the WW microbiota. Anthropogenic pollution through micro pollutants present in WW (particularly
384 surfactants, antibiotics and heavy metals) has shown to have a negative impact on the environment and
385 significantly shape the terrestrial and aquatic microbial ecosystems[8,46–49]. For example, high
386 concentrations of antibiotics and cationic surfactants found in HWW[17,50,51], and pharmaceutical
387 production sites[52,53] have been correlated with a higher abundance of ARBS, ARGs, as well as higher
388 abundance of MGEs and integrons[17,19,54], whereas anionic surfactants that are generally abundant in urban
389 WW effluents (untreated and treated UWWs; "grey waters") are associated with toxicity to aquatic and
390 terrestrial environments[55–57]. Here, multivariate analysis (Figure 5) revealed a significant impact of the
391 eco-exposome on the resistome and microbiota signatures of all investigated WW sources. We observed

392 that cationic surfactants and antibiotics are specifically linked to HWW (Figure 5), which reflects the
393 frequent use of antibiotics and active-surface agents as quaternary ammonium compounds in hospitals.
394 Interestingly, we also found that *qac* genes that encode resistance to quaternary ammonium compounds
395 are significantly higher in HWW than in UWW. The urban WW eco-exposome on the other hand was found
396 to be enriched with anionic surfactants (Supplementary Table 2) which in turn were found to be
397 specifically linked to the UWW and MWW resistome and microbiota (Figure 5). This points towards an
398 important correlation of anionic surfactants detected in UWWs and the resistome and microbiota of
399 UWWs, in addition to other pharmaceuticals and antibiotics. Specific measures to reduce the emission of
400 surfactants into the environment by selective removal and improved WW treatment are currently
401 discussed[47] and warrant further attention. In addition, further research needs to be done to illuminate
402 the detailed mechanism of the synergetic effects of all compounds that make up the WW eco-exposome
403 on shaping the resistome and microbiota of WW effluents and the downstream environment.
404 Altogether, our findings give further emphasis to the requirement of
405 implementing and optimizing sanitation systems and operational WWTPs on a global level, particular in
406 countries/continents with poor water sanitation infrastructure and correlated high occurrence of multi-
407 resistant bacteria[58]. The elimination of pollution by means of human feces, and bacterial taxa associated,
408 through advanced or selective WW treatment, may further aid in limiting the release of ARGs associated
409 with human gut bacteria and pathogens[35,59,60]. Finally, the data generated by this study are of important
410 interest to policy makers concerning the risks associated with H and U WW, their putative implication into
411 the dissemination of AMR, and provide further evidence towards the necessity of environmental pollution
412 management in the battle of AMR and other important global health factors such as the preservation of
413 biodiversity and the prevention of climate change[49] (France national report:
414 hyperlink:https://www.tresor.economie.gouv.fr/Articles/a9782706-87a4-4cdc-9a4c-
415 86a72736315d/files/525144cc-24fb-4c93-aa5b-c67aefd3b40a).

416

417

418 # Methods

419 **Sampling and study design**

420 126 Urban and hospital wastewater (UWW and HWW) samples were collected in Scientrier (Bellecombe
421 WWTP), Haute-Savoie, France[27] as part of the multi-disciplinary project SIPIBEL. The study site was
422 implemented as an observatory for untreated and treated hospital and UWW and to evaluate their impact
423 (during separate and subsequently mixed treatment) on the environment (e.g. the effluent receiving
424 river). The CHAL (Centre Hospitalier Alpes Léman) hospital opened in February 2012 and includes 450 beds
425 (140 m3/d), whereas the Bellecombe WWTP was collecting UWW of approximately 21.000 inhabitants
426 (5200 m3/d). For more details of the SIPIBEL project, study set up, WWT and sample collection refer to
427 Chonova et *al*. 2018 and Wiest et *al.*[27,37], and to Figure 1. The samples included in our study were collected
428 in monthly intervals (untreated and treated samples) by flow proportional sampling, from March 2012
429 through November 2015[27]. From March 2012 to December 2014, UWW and HWW were treated
430 separately applying the same conventional (activated sludge) WWT[27,37]. Then, in the period from January
431 2015 through November 2015, UWW was mixed into the HWW (1:2 ratio HWW:UWW, the ratio was fixed
432 by a local operating constraint) and added to the separate HWW treatment line resulting in a controlled
433 mixed WW (MWW)[27]. In addition, 12 water samples of the effluent receiving river up (river upstream) and
434 downstream (river downstream sampling point 1 and 2) of the effluent release pipes have been collected
435 during the winter months of 2013 (January, February, November, and December). Activated sludge
436 samples have also been collected from both sludge basins throughout the sampling campaign. Due to
437 different resident times and flow sizes of each wastewater treatment basin, sludge dynamics for resistome
438 and microbiota were not directly comparable hence results will not be further discussed in this study. 16S
439 rRNA sequence data from all samples, including sludge samples, are publicly available.

440 **DNA isolation/ sample preparation**

441 Water samples were filtered for microorganisms, using a filtration ramp (Sartorius,
442 Göttingen, Germany), on sterile 47 mm diameter filter with pore size of 0.45 µm (Sartorius,
443 Göttingen, Germany). Microorganisms were recovered from filters and subject to DNA isolation for
444 downstream analysis, using the Power water DNA extraction kit (MoBio Laboratories Inc., Carlsbad, CA,
445 USA). For sludge samples, 2 ml of sludge were pelleted, and DNA was extracted by following the protocol
446 of the Fast DNA Spin kit for feces (MP Biomedicals, Illkirch, France). DNA concentration was determined
447 by Qubit Fluoremetric Quantitation (Thermo fisher scientific, Waltham, MA USA) assays according the

448 manufacturer's instructions. All DNA samples were diluted or concentrated to a final concentration of 10
449 ng/µl for downstream qPCR and 16S rRNA analysis.

450 **High-throughput qPCR**

451 Nanolitre-scale quantitative PCRs to quantify levels of genes that confer resistance to antimicrobials and
452 heavy metals were performed as described previously[9,20], with some modifications in the collection of
453 primers. The primer sequences and their targets are provided in the supplementary data (Supplementary
454 Table 5). The primer set used in the qPCR assays covered 78 genes conferring resistance to antibiotics,
455 quaternary ammonium compounds or heavy metals. This set includes genes encoding efflux pumps
456 (referred to as 'efflux', Supplementary Table 6) leading to multi-resistance at once to different antibiotic
457 families. We also added primers for genes encoding mobile genetic elements, namely nine transposase
458 genes[25,40], and the class 1, 2 and 3 integron integrase genes[61]. We also included primers targeting 16SrRNA
459 encoding DNA. Primer design and validation prior and after Biomark analysis has been done as described
460 earlier[20]. Real-Time PCR analysis was performed using the 96.96 BioMark™ Dynamic Array for Real-Time
461 PCR (Fluidigm Corporation, San Francisco, CA, U.S.A), according to the manufacturer's instructions, with
462 the exception that the annealing temperature in the PCR was lowered to 56°C. DNA was first subjected to
463 14 cycles of Specific Target Amplification using a 50 nM mixture of all primer sets, excluding the 16S rRNA
464 primer sets, in combination with the PreAmp Master Mix (100-5581, Fluidigm), followed by a 5-fold
465 dilution prior to loading samples onto the Biomark array for qPCR. Thermal cycling and real-time imaging
466 was performed at the Plateforme Génomique GeT – INRA Transfert (https://get.genotoul.fr/en/), and Ct
467 values were extracted using the BioMark Real-Time PCR analysis software.

468 **Calculation of normalized abundance and cumulative abundance**

469 Normalized abundance of individual resistance genes was calculated relative to the abundance of the 16S
470 rRNA gene (CTARG – CT16S rRNA) resulting in a log2-transformed estimate of gene abundance.
471 Cumulative abundance was calculated for resistance gene classes based on the sum of the normalized
472 abundance of individual ARGs. The differences in cumulative abundance over the indicated time periods
473 (2012-2014; 2015 for mixed WW) are shown as an averaged fold-change ± standard deviation. The non-
474 parametric Mann-Whitney test was used to test for significance; p values were corrected for multiple
475 testing by the Benjamin-Hochberg procedure (Benjamini & Hochberg, 1995) with a false discovery rate of
476 0.05. Averaged normalized abundance data for allocated gene classes is provided in Supplementary Table
477 7.

478 **qPCR to determine absolute copy numbers of 16S rRNA genes**

479 The qPCRs for the determination of 16S rRNA gene copy number as a proxy for the bacterial biomass was
480 performed as described previously by Stalder et. *al*.[19].

481 **16S rRNA gene sequencing and sequence data pre-processing**

482 Extracted DNA samples for 16S rRNA sequencing were prepared following a dual barcoded two-step PCR
483 procedure for amplicon sequencing for Illumina. Primers of the first PCR step included universal CS1 and
484 CS2 tags targeting the V4 region of the hypervariable region of the 16S rRNA gene using the 16SrRNA
485 primer sequences of the earth microbiota project (http://press.igsb.anl.gov/earthmicrobiota/protocols-
486 and-standards/16s/). During the second step of the PCR barcoded adapters suitable for multiplex illumina
487 sequencing were added. Following pooling of the barcoded samples, the amplicon pool was cleaned to
488 remove short undesirable fragments using Magbio HighPrep PCR beads (MagBio AC-60050), QC'ed on a
489 High Sensitivity NGS Fragment Analyzer, and then qPCR quantified using the Kapa kit for ABI optical-
490 cyclers. The pool was then normalized to 10nM, denatured using 0.1N NaOH followed by a 2min
491 incubation @96C followed by 5min in an ice-water bath just prior to sequencing as per the Illumina
492 protocol for a 2x301 MiSeq run (Illumina, Inc., San Diego, CA). DNA sequence reads from the Illumina
493 MiSeq were demultiplexed and classified in the following manner: The Python application dbcAmplicons
494 (https://github.com/msettles/dbcAmplicons) was used to identify and assign reads to the appropriate
495 sample by both expected barcode and primer sequences. Barcodes were allowed to have at most 1
496 mismatch (hamming distance) and primers were allowed to have at most 4 mismatches (Levenshtein
497 distance) as long as the final 4 bases of the primer matched the target sequence perfectly. Reads were
498 then trimmed of their primer sequence and merged into a single amplicon sequence using the application
499 FLASH[62]. Finally, the RDP Bayesian classifier was used to assign sequences to phylotypes[63]. Reads were
500 assigned to the first RDP taxonomic level with a bootstrap score >=50.

501 **16S rRNA data analysis**

502 Illumina MiSeq forward and reverse were processed using the MASQUE pipeline
503 (https://github.com/aghozlane/masque). Briefly, raw reads are filtered and combined followed by
504 dereplication. Chimera removal and clustering are followed by taxonomic annotation of the resulting
505 OTUs by comparison to the SILVA database. A BIOM file is generated that combines both OTU taxonomic
506 assignment and the number of matching reads for each sample. Relative abundance levels form bacterial
507 taxa (Order level) were obtained and analyzed. The obtained relative abundance OTU tables (Order level)

508 were analyzed with Microsoft excel (Supplementary Table 8), multi-variate analysis package (see below)
509 and by means of a machine learning approach employing a random forest algorithm (see below).

510 **Chemical analysis**

511 All chemical data measured here and used for analysis where extracted from the SIPIBEL database. Solid-
512 phase extraction (SPE) and liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS)
513 were used to measure the antibiotics ciprofloxacin, sulfamethoxazole and vancomycin and the
514 pharmaceutical carbamazepine as detailed elsewhere[26]. Heavy metals (Zn, Cu, Ni, Pb, Cr, Gd, Hg, As and Cd)
515 were measured with inductively coupled plasma combined with atomic emission spectroscopy (ICP-AES).
516 Concentration of surfactants (anionic, cationic and non-ionic surfactants) were measured following
517 standard methods approved by the French organization of standardization AFNOR as described by Wiest
518 et. *al*.[37].

519 **Multivariate analyses**

520 Multivariate statistical techniques were used to test the influence of waste water treatment or sampling
521 time (independent variables) on the microbiota and the resistome (dependent variables) of the different
522 sample groups (urban, hospital, mixed), including all individual genes and genes allocated into gene
523 classes in two independent datasets. Statistically significant influence of the treatment or the sampling
524 time on the microbiota and the resistome were assessed by Redundancy Analysis (RDA) with 499 Monte
525 Carlo permutations. For testing the influence of sampling time, we used sampling year or season as
526 independent variables, and all different sample groups together (i.e., to potentially identify general
527 patterns influencing all groups at the same time), and individually, as dependent variables. Such analysis
528 revealed the percentage of variance that is explained by sampling time or WW treatment in each case,
529 and whether the influence of the independent variables is statistically significant or not. The relationship
530 between the resistome and the microbiota and the measured chemicals in the different raw water
531 samples (eco-exposome) was visualized by means of Principal Component Analysis (PCA) biplots.
532 Moreover, the influence of the measured chemicals (eco-exposome) on the microbiota and resistome was
533 statistically assessed by RDA. A variation partitioning analysis was performed to assess which group of
534 chemicals (Metals, Pharmaceuticals, Surfactants) explains the largest share of the variation of the
535 microbiota and resistome datasets, and to explore whether the interactive effects of the groups of
536 chemicals would have a larger influence on those datasets than the individual groups themselves. All
537 multivariate analyses were performed with the Canoco v5.0 software[64], using a significance level of 0.05.

538 **Machine learning**

539 A Random Forest Algorithm (RFA) was used in order to predict a response variable (water sources) of each
540 sample independently, using measurements on individual gene, gene class and microbiota level (predictor
541 variables). To run the RFA the R-package randomForest was used: Breiman and Cutler's Random Forests
542 for Classification and Regression, a software package for the R-statistical environment[65]. In summary, the
543 RFA follows the pseudo-steps: (I) the response variable and predictor variables are chosen by the user; (II)
544 a predefined number of independent bootstrap samples are drawn from the dataset with replacement,
545 and a classification tree is fit to each sample containing roughly 2/3 of the data, for which predictor
546 variable selection on each node split in the tree is conducted using only a small random subset of predictor
547 variables; (III) the complete set of trees, one for each bootstrap sample, composes the random forest (RF),
548 from which the status (classification) of the response variable is predicted as an average (majority vote)
549 of the predictions of all trees. Compared to single classification trees, RFA increases prediction accuracy,
550 since the ensemble of slight different classification results adjusts for the instability of the individual trees
551 and avoids data overfitting[66]. The Mean Decrease Accuracy (MDA), or Breiman-Cutler importance, was
552 employed as a measure of predictor variable importance, for which classification accuracy after data
553 permutation of a predictor variable is subtracted from the accuracy without permutation, and averaged
554 over all trees in the RF to give an importance value [2]. It should be noted that since all predictor variables
555 were of numeric nature, using RFA is equivalent to regression over classification trees. For the results
556 presented here and in supplementary text, only the 2.5% of top RFA scores were considered (as presented
557 by the resulting MDA distribution of all predictor variables) , thus selecting the subset of predictor
558 variables which appear statistically more informative than expected in the background of all predictor
559 variables (i.e. we assume that 95% of the RFA scores fall between the 2.5th and 97.5th percentiles, as
560 done elsewhere[67]).

561

Data availability: 16S rRNA sequence data are available at the European Nucleotide Archive (ENA) under the accession number PRJEB29948. All other important raw data needed to reconstruct the findings of our study are made available in the supplementary material.

Funding:

E. Buelow has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement RESOLVE 707999-standard EF.

A. Rico is supported by a postdoctoral grant provided by the Spanish Ministry of Science, Innovation and University (IJCI-2017-33465).

J. Lourenço is supported by a Lectureship from the Department of Zoology, University of Oxford.

16S rRNA sequence Data collection and analyses performed by the **IBEST Genomics Resources Core at the University of Idaho** were supported in part by NIH COBRE grant P30GM103324.

Acknowledgements: The authors thank the SIPIBEL Consortium and the SIPIBEL field observatory on the hospital's effluents and urban wastewater treatment plants for displaying data and measurements.

Author contributions:
C.D., M-C.P., and E.B. designed the study, **C.D**. provided access to the SIPIBEL data collection, **M.G**. and **E.B.** performed experiments, **E.B.**, **A.R.**, **J.L.** and **S.P.K**. performed data analysis, **L.W.** performed chemical analysis (SIPIBEL project), **E.B.**, **M-C.P.** and **C.D.** wrote the manuscript with contribution of all other co-authors.

Competing interests: The authors declare no competing interests.

585 **References**

586 1. Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with
587 antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level
588 modelling analysis. *LancetInfectious Dis.* (2018).

589 2. Nadimpalli, M. *et al.* Combating Global Antibiotic Resistance: Emerging One Health Concerns in
590 Lower- and Middle-Income Countries. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **66**, 963–969
591 (2018).

592 3. Stemming the Superbug Tide | READ online. *OECD iLibrary* Available at: https://read.oecd-
593 ilibrary.org/social-issues-migration-health/stemming-the-superbug-tide_9789264307599-en.
594 (Accessed: 9th April 2019)

595 4. Larsson, D. G. J. *et al.* Critical knowledge gaps and research needs related to the environmental
596 dimensions of antibiotic resistance. *Environ. Int.* **117**, 132–138 (2018).

597 5. Berendonk, T. U. *et al.* Tackling antibiotic resistance: the environmental framework. *Nat. Rev.* **13**,
598 310–317 (2015).

599 6. Canton, R. Antibiotic resistance genes from the environment: a perspective through newly identified
600 antibiotic resistance mechanisms in the clinical setting. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc.*
601 *Clin. Microbiol. Infect. Dis.* **15 Suppl 1**, 20–25 (2009).

602 7. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev.* **8**,
603 251–259 (2010).

604 8. Flandroy, L. *et al.* The impact of human activities and lifestyles on the interlinked microbiota and
605 health of humans and of ecosystems. *Sci. Total Environ.* **627**, 1018–1038 (2018).

606 9. Buelow, E. *et al.* Limited influence of hospital wastewater on the microbiome and resistome of
607 wastewater in a community sewerage system. *FEMS Microbiol. Ecol.* (2018).
608 doi:10.1093/femsec/fiy087

609 10. Jin, L. *et al.* Emergence of mcr-1 and carbapenemase genes in hospital sewage water in Beijing,
610 China. *J. Antimicrob. Chemother.* **73**, 84–87 (2018).
611 11. Manaia, C. M. *et al.* Antibiotic resistance in wastewater treatment plants: Tackling the black box.
612 *Environ. Int.* **115**, 312–324 (2018).
613 12. Wang, Q., Wang, P. & Yang, Q. Occurrence and diversity of antibiotic resistance in untreated
614 hospital wastewater. *Sci. Total Environ.* **621**, 990–999 (2018).
615 13. Chonova, T. *et al.* Separate treatment of hospital and urban wastewaters: A real scale comparison of
616 effluents and their effect on microbial communities. *Sci. Total Environ.* **542**, 965–975 (2016).
617 14. Raheem, A. *et al.* Opportunities and challenges in sustainable treatment and resource reuse of
618 sewage sludge: A review. *Chem. Eng. J.* **337**, 616–641 (2018).
619 15. Pallares-Vega, R. *et al.* Determinants of presence and removal of antibiotic resistance genes during
620 WWTP treatment: A cross-sectional study. *Water Res.* (2019). doi:10.1016/j.watres.2019.05.100
621 16. Zhu, Y. G. *et al.* Continental-scale pollution of estuaries with antibiotic resistance genes. *Nat.*
622 *Microbiol.* **2**, 16270 (2017).
623 17. Rodriguez-Mozaz, S. *et al.* Occurrence of antibiotics and antibiotic resistance genes in hospital and
624 urban wastewaters and their impact on the receiving river. *Water Res.* **Volume 69**, 234--242 (2015).
625 18. Lienert, J., Koller, M., Konrad, J., McArdell, C. S. & Schuwirth, N. Multiple-criteria decision analysis
626 reveals high stakeholder preference to remove pharmaceuticals from hospital wastewater. *Environ.*
627 *Sci. Technol.* **45**, 3848–3857 (2011).
628 19. Stalder, T. *et al.* Quantitative and qualitative impact of hospital effluent on dissemination of the
629 integron pool. *ISME J.* **8**, 768–777 (2014).
630 20. Buelow, E. *et al.* Comparative gut microbiota and resistome profiling of intensive care patients
631 receiving selective digestive tract decontamination and healthy subjects. *Microbiome* **5**, 88-017-
632 0309-z (2017).

633 21. Forslund, K. *et al.* Country-specific antibiotic use practices impact the human gut resistome. *Genome*
634 *Res.* **23**, 1163–1169 (2013).
635 22. Hu, Y. *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut
636 microbiota. *Nat. Commun.* **4**, 2151 (2013).
637 23. Pal, C., Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. Co-occurrence of resistance genes to
638 antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC*
639 *Genomics* **16**, 964-015-2153–5 (2015).
640 24. Gnanadhas, D. P., Marathe, S. A. & Chakravortty, D. Biocides--resistance, cross-resistance
641 mechanisms and assessment. *Expert Opin. Investig. Drugs* **22**, 191–206 (2013).
642 25. Zhu, Y. G. *et al.* Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proc. Natl.*
643 *Acad. Sci. U. S. A.* **110**, 3435–3440 (2013).
644 26. Gillings, M. R. *et al.* Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution.
645 *ISME J.* **9**, 1269–1279 (2015).
646 27. Chonova, T. *et al.* The SIPIBEL project: treatment of hospital and urban wastewater in a
647 conventional urban wastewater treatment plant. *Environ. Sci. Pollut. Res. Int.* **25**, 9197–9206 (2018).
648 28. Sommer, M. O., Church, G. M. & Dantas, G. The human microbiome harbors a diverse reservoir of
649 antibiotic resistance genes. *Virulence* **1**, 299–303 (2010).
650 29. Ruppe, E. *et al.* Prediction of the intestinal resistome by a three-dimensional structure-based
651 method. *Nat. Microbiol.* (2018). doi:10.1038/s41564-018-0292-6
652 30. Rajilic-Stojanovic, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal
653 microbiota. *FEMS Microbiol. Rev.* **38**, 996–1047 (2014).
654 31. Costea, P. I. *et al.* Enterotypes in the landscape of gut microbial community composition. *Nat.*
655 *Microbiol.* **3**, 8–16 (2018).

656 32. Robert, J. *et al.* Point prevalence survey of antibiotic use in French hospitals in 2009. *J. Antimicrob.*
657 *Chemother.* **67**, 1020–1026 (2012).
658 33. Rozwandowicz, M. *et al.* Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J.*
659 *Antimicrob. Chemother.* **73**, 1121–1137 (2018).
660 34. Carattoli, A. Resistance Plasmid Families in Enterobacteriaceae. *Antimicrob. Agents Chemother.* **53**,
661 2227–2238 (2009).
662 35. Karkman, A., Parnanen, K. & Larsson, D. G. J. Fecal pollution can explain antibiotic resistance gene
663 abundances in anthropogenically impacted environments. *Nat. Commun.* **10**, 80-018-07992–3
664 (2019).
665 36. Su, J. Q. *et al.* Metagenomics of urban sewage identifies an extensively shared antibiotic resistome
666 in China. *Microbiome* **5**, 84-017-0298-y (2017).
667 37. Wiest, L. *et al.* Two-year survey of specific hospital wastewater treatment and its impact on
668 pharmaceutical discharges. *Environ. Sci. Pollut. Res. Int.* **25**, 9207–9218 (2018).
669 38. Chonova, T. *et al.* River biofilm community changes related to pharmaceutical loads emitted by a
670 wastewater treatment plant. *Environ. Sci. Pollut. Res. Int.* **25**, 9254–9264 (2018).
671 39. Laquaz, M. *et al.* Ecotoxicity and antibiotic resistance of a mixture of hospital and urban sewage in a
672 wastewater treatment plant. *Environ. Sci. Pollut. Res. Int.* **25**, 9243–9253 (2018).
673 40. Karkman, A. *et al.* High-throughput quantification of antibiotic resistance genes from an urban
674 wastewater treatment plant. *FEMS Microbiol. Ecol.* **92**, 10.1093/femsec/fiw014. Epub 2016 Jan 31
675 (2016).
676 41. Di Cesare, A. *et al.* Co-occurrence of integrase 1, antibiotic and heavy metal resistance genes in
677 municipal wastewater treatment plants. *Water Res.* **94**, 208–214 (2016).

678 42. Lee, J. *et al.* Quantitative and qualitative changes in antibiotic resistance genes after passing through
679 treatment processes in municipal wastewater treatment plants. *Sci. Total Environ.* **605–606**, 906–
680 914 (2017).
681 43. Jäger, T. *et al.* Reduction of Antibiotic Resistant Bacteria During Conventional and Advanced
682 Wastewater Treatment, and the Disseminated Loads Released to the Environment. *Front. Microbiol.*
683 **9**, (2018).
684 44. Wild, C. P. Complementing the Genome with an "Exposome": The Outstanding Challenge of
685 Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol. Prev.*
686 *Biomark.* **14**, 1847–1850 (2005).
687 45. Escher, B. I. *et al.* From the exposome to mechanistic understanding of chemical-induced adverse
688 effects. *Environ. Int.* **99**, 97–106 (2017).
689 46. Palmer, M. & Hatley, H. The role of surfactants in wastewater treatment: Impact, removal and
690 future techniques: A critical review. *Water Res.* **147**, 60–72 (2018).
691 47. Pereira, B. M. P. & Tagkopoulos, I. Benzalkonium chlorides: Uses, regulatory status, and microbial
692 resistance. *Appl. Environ. Microbiol.* AEM.00377-19 (2019). doi:10.1128/AEM.00377-19
693 48. Cairns, J., Becks, L., Jalasvuori, M. & Hiltunen, T. Sublethal streptomycin concentrations and lytic
694 bacteriophage together promote resistance evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**,
695 (2017).
696 49. Hendry, A. P., Gotanda, K. M. & Svensson, E. I. Human influences on evolution, and the ecological
697 and societal consequences. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).
698 50. Varela, A. R., Andre, S., Nunes, O. C. & Manaia, C. M. Insights into the relationship between
699 antimicrobial residues and bacterial populations in a hospital-urban wastewater treatment plant
700 system. *Water Res.* **54**, 327–336 (2014).

701 51. Szekeres, E. *et al.* Abundance of antibiotics, antibiotic resistance genes and bacterial community
702 composition in wastewater effluents from different Romanian hospitals. *Environ. Pollut. Barking*
703 *Essex 1987* **225**, 304–315 (2017).
704 52. Bengtsson-Palme, J. & Larsson, D. G. Concentrations of antibiotics predicted to select for resistant
705 bacteria: Proposed limits for environmental regulation. *Environ. Int.* **86**, 140–149 (2016).
706 53. Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Environmental factors influencing the
707 development and spread of antibiotic resistance. *FEMS Microbiol. Rev.* **42**, 10.1093/femsre/fux053
708 (2018).
709 54. Rowe, W. P. M. *et al.* Overexpression of antibiotic resistance genes in hospital effluents over time. *J.*
710 *Antimicrob. Chemother.* **72**, 1617–1623 (2017).
711 55. Shafran, A. W., Gross, A., Ronen, Z., Weisbrod, N. & Adar, E. Effects of surfactants originating from
712 reuse of greywater on capillary rise in the soil. *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.*
713 **52**, 157–166 (2005).
714 56. Siggins, A., Burton, V., Ross, C., Lowe, H. & Horswell, J. Effects of long-term greywater disposal on
715 soil: A case study. *Sci. Total Environ.* **557–558**, 627–635 (2016).
716 57. Jardak, K., Drogui, P. & Daghrir, R. Surfactants in aquatic and terrestrial environment: occurrence,
717 behavior, and treatment processes. *Environ. Sci. Pollut. Res. Int.* **23**, 3195–3216 (2016).
718 58. Burgmann, H. *et al.* Water and sanitation: an essential battlefront in the war on antimicrobial
719 resistance. *FEMS Microbiol. Ecol.* **94**, 10.1093/femsec/fiy101 (2018).
720 59. Pehrsson, E. C. *et al.* Interconnected microbiomes and resistomes in low-income human habitats.
721 *Nature* **533**, 212–216 (2016).
722 60. Gouliouris, T. *et al.* Detection of vancomycin-resistant *Enterococcus faecium* hospital-adapted
723 lineages in municipal wastewater treatment plants indicates widespread distribution and release
724 into the environment. *Genome Res.* (2019). doi:10.1101/gr.232629.117

725 61. Barraud, O., Baclet, M. C., Denis, F. & Ploy, M. C. Quantitative multiplex real-time PCR for detecting
726 class 1, 2 and 3 integrons. *J. Antimicrob. Chemother.* **65**, 1642–1645 (2010).
727 62. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome
728 assemblies. *Bioinforma. Oxf. Engl.* **27**, 2957–2963 (2011).
729 63. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of
730 rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
731 64. Braak, C. J. F. ter & Smilauer, P. *Canoco reference manual and user's guide: software for ordination,*
732 *version 5.0*. (Microcomputer Power, 2012).
733 65. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
734 66. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference,*
735 *and Prediction, Second Edition*. (Springer-Verlag, 2009).
736 67. Lourenco, J. *et al.* Lineage structure of *Streptococcus pneumoniae* may be driven by immune
737 selection on the groEL heat-shock protein. *Sci. Rep.* **7**, 9023-017-08990-z (2017).
738