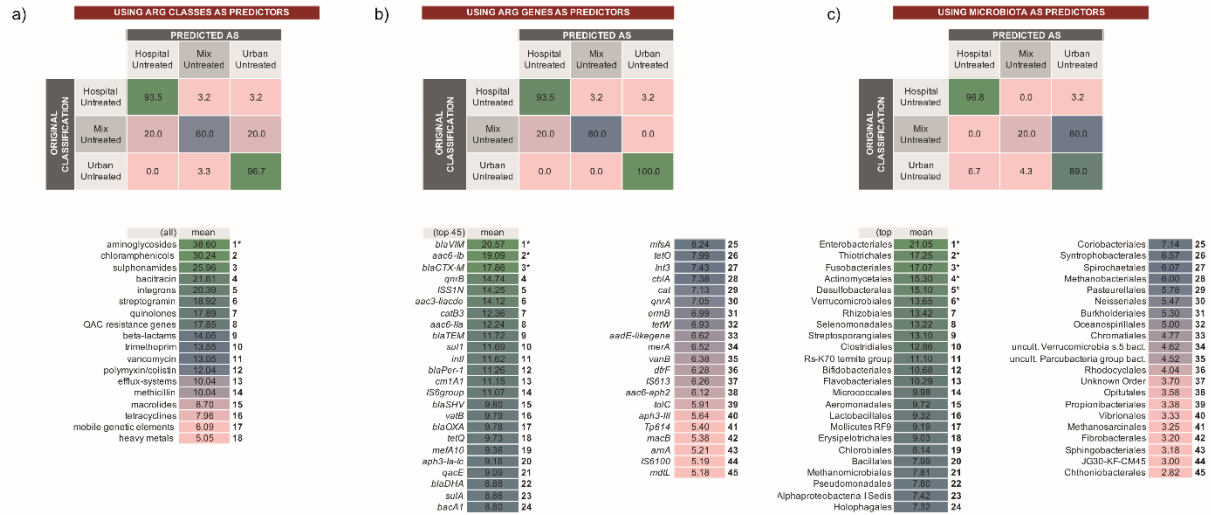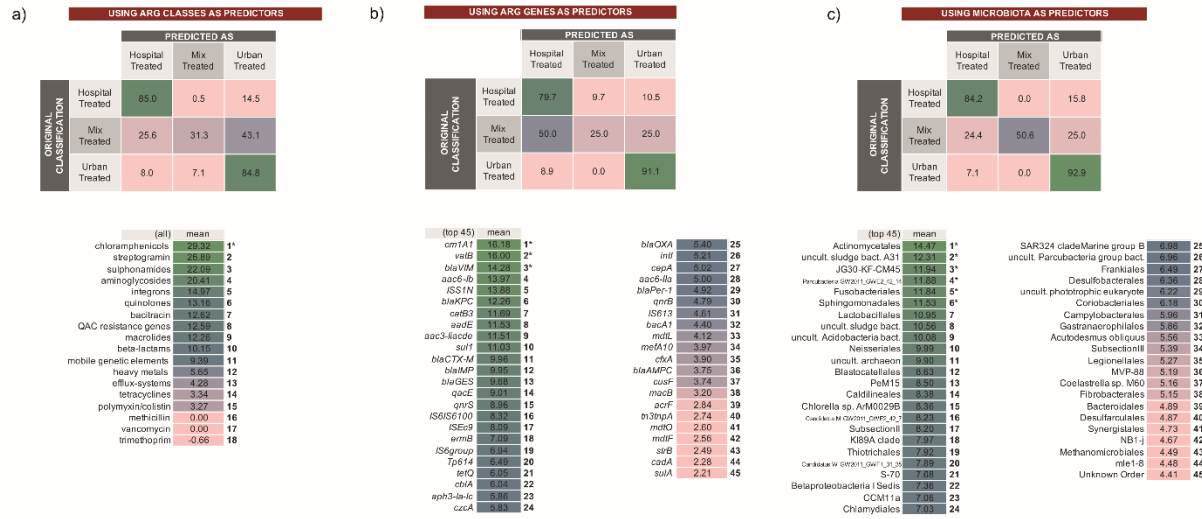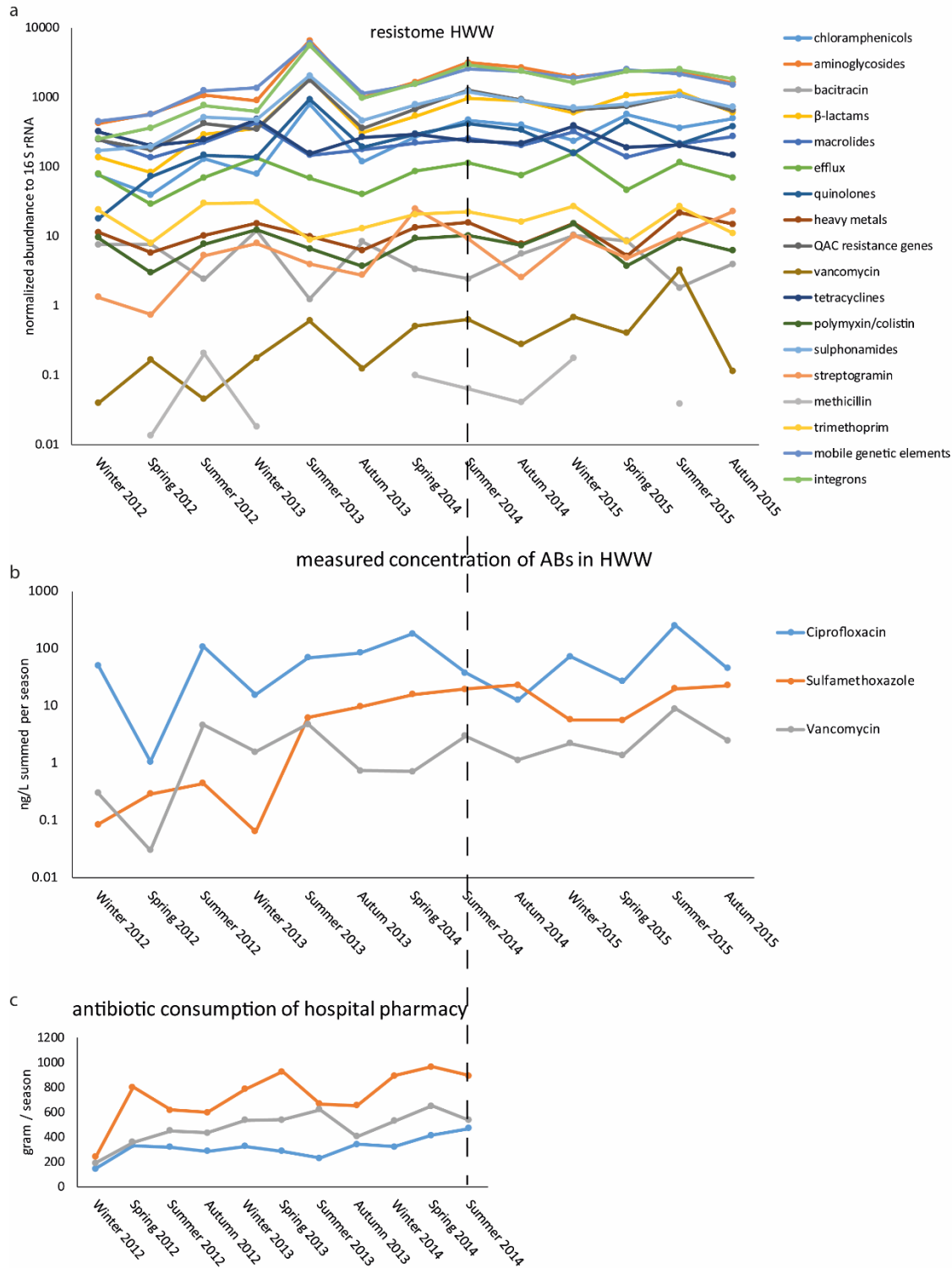# Supplementary Figures:

Supplementary Figure 1. Classification success of untreated wastewater from hospital, mix and urban sources. Top panels present the RFA classification accuracy of raw (untreated) wastewater sources when setting the predictor variables as **(a)** ARG classes, **(b)** ARG genes, and **(c)** the microbiota. The values within cells are the mean % prediction accuracy of 20 independent RFA runs with 1000 trees and default mtry. Actual wastewater source of samples in rows, predictions in columns. Colour scale is used for visualisation purposes only, from light red at 0%, blue at 50% and green at 100% of accuracy. Lower panels present the top predictor variables (capped at N=45 for visualization purposes) when setting the predictor variables as **(a)** ARG classes, **(b)** ARG genes, and **(c)** the microbiota. The values within cells are the mean MDA score of 20 independent RFA runs with 1000 trees and default mtry. Color scale is used for visualization purposes only, from light red at 0, blue at percentile 70% and green at percentile 97.25% of all MDA scores. Variables marked with * are the ones that appear more informative than expected given all the MDA scores of all predictor variables (above the cut off of percentile 97.25% of all scores, see method description).

a) USING ARG CLASSES AS PREDICTORS

	PREDICTED AS		
ORIGINAL CLASSIFICATION	Hospital Treated	Mix Treated	Urban Treated
Hospital Treated	85.0	0.5	14.5
Mix Treated	25.6	31.3	43.1
Urban Treated	8.0	7.1	84.8

b) USING ARG GENES AS PREDICTORS

	PREDICTED AS		
ORIGINAL CLASSIFICATION	Hospital Treated	Mix Treated	Urban Treated
Hospital Treated	79.7	9.7	10.5
Mix Treated	50.0	25.0	25.0
Urban Treated	8.9	0.0	91.1

c) USING MICROBIOTA AS PREDICTORS

	PREDICTED AS		
ORIGINAL CLASSIFICATION	Hospital Treated	Mix Treated	Urban Treated
Hospital Treated	84.2	0.0	15.8
Mix Treated	24.4	50.6	25.0
Urban Treated	7.1	0.0	92.9

a) (all)

	mean	
chloramphenicols	29.32	1*
streptogramin	26.89	2
sulphonamides	22.09	3
aminoglycosides	20.41	4
integrons	14.97	5
quinolones	13.16	6
bacitracin	12.62	7
QAC resistance genes	12.59	8
macrolides	12.26	9
beta-lactams	10.15	10
mobile genetic elements	9.30	11
heavy metals	5.65	12
efflux-systems	4.28	13
tetracyclines	3.34	14
polymyxin/colistin	3.27	15
methicillin	0.00	16
vancomycin	0.00	17
trimethoprim	-0.66	18

b) (top 45)

	mean			mean	
crntA1	16.18	1*	blaOXA	5.40	25
vatB	16.00	2*	intI	5.21	26
blaVIM	14.28	3*	cepA	5.02	27
aac6-Ib	13.97	4	aac6-IIa	5.00	28
ISS1N	13.88	5	blaPer-1	4.92	29
blaKPC	12.26	6	qnrB	4.79	30
catB3	11.69	7	IS613	4.61	31
aadE	11.53	8	bacA1	4.40	32
aac3-IIacde	11.51	9	mdtL	4.12	33
sul1	11.03	10	metfA10	3.97	34
blaCTX-M	9.96	11	cfxA	3.90	35
blaIMP	9.95	12	blaAMPC	3.75	36
blaGES	9.88	13	cusF	3.74	37
qacE	9.01	14	macB	3.20	38
qnrS	8.96	15	acrF	2.84	39
IS6IS6100	8.32	16	tr13trpA	2.74	40
ISEc9	8.09	17	mdtO	2.60	41
ermB	7.09	18	mdtF	2.56	42
IS6group	6.94	19	strB	2.49	43
Tp614	6.49	20	cadA	2.28	44
tetQ	6.05	21	sulA	2.21	45
cblA	6.04	22			
aph3-Ia-Ic	5.86	23			
czcA	5.83	24			

c) (top 45)

	mean			mean	
Actinomycetales	14.47	1*	SAR324 cladeMarine group B	6.96	25
uncult. sludge bact. A31	12.31	2*	uncult. Parcubacteria group bact.	6.96	26
JG30-KF-CM45	11.94	3*	Frankiales	6.49	27
Parcubacteria GW2011_GWC2_t2_11	11.88	4*	Desulfobacterales	6.36	28
Fusobacteriales	11.84	5*	uncult. phototrophic eukaryote	6.22	29
Sphingomonadales	11.53	6*	Coriobacteriales	6.18	30
Lactobacillales	10.95	7	Campylobacterales	5.96	31
uncult. sludge bact.	10.56	8	Gastranaerophilales	5.86	32
uncult. Acidobacteria bact.	10.08	9	Acutodesmus obliquus	5.56	33
Neisseriales	9.99	10	Subsection III	5.39	34
uncult. archaeon	9.90	11	Legionellales	5.27	35
Blastocatellales	8.81	12	MVP-88	5.19	36
PeM15	8.50	13	Coelastrella sp. M60	5.16	37
Caldilineales	8.38	14	Fibrobacterales	5.15	38
Chlorella sp. ArM0029B	8.36	15	Bacteroidales	4.89	39
Candidatus M GW2011_GWF2_42_7	8.23	16	Desulfarculales	4.87	40
Subsection II	8.20	17	Synergistales	4.73	41
KI89A clade	7.97	18	NB1-j	4.67	42
Thiotrichales	7.92	19	Methanomicrobiales	4.49	43
Candidatus W GW2011_GWF1_31_35	7.89	20	mle1-8	4.48	44
S-70	7.88	21	Unknown Order	4.41	45
Betaproteobacteria I Sedis	7.36	22			
CCM11a	7.06	23			
Chlamydiales	7.03	24			

Supplementary Figure 2. Classification success of treated wastewater from hospital, mix and urban sources.
Upper panels present the RFA classification accuracy of treated wastewater sources when setting the predictor variables as **(a)** ARG classes, **(b)** ARG genes, and **(c)** the microbiota. The values within cells are the mean % prediction accuracy of 20 independent RFA runs with 1000 trees and default mtry. Actual water source of samples in rows, predictions in columns. Colour scale is used for visualisation purposes only, from light red at 0%, blue at 50% and green at 100% of accuracy. Lower panels present the top predictor variables (capped at N=45 for visualisation purposes) when setting the predictor variables as (a) ARG classes, (b) ARG genes, and (c) the microbiota. The values within cells are the mean MDA score of 20 independent RFA runs with 1000 trees and default mtry. Colour scale is used for visualisation purposes only, from light red at 0, blue at percentile 70% and green at percentile 97.25% of all MDA scores. Variables marked with * are the ones that appear more informative than expected given all the MDA scores of all predictor variables (above the cut off of percentile 97.25% of all scores, see method description).

Supplementary Figure 3: **Hospital wastewater (HWW) resistome, eco-exposome and distribution of measured antibiotics by the hospital pharmacy displayed per season**. **(a)** normalized cumulative abundance of gene classes; **(b)** measured concentration of antibiotics in HWW (eco-exposome: ciprofloxacin, sulfamethoxazole and vancomycin) and **(c)** the distribution of those antibiotics in the hospital pharmacy per season. Vertical black line indicates summer season 2014 which represents the last data point available for hospital pharmacy distribution.

a

b
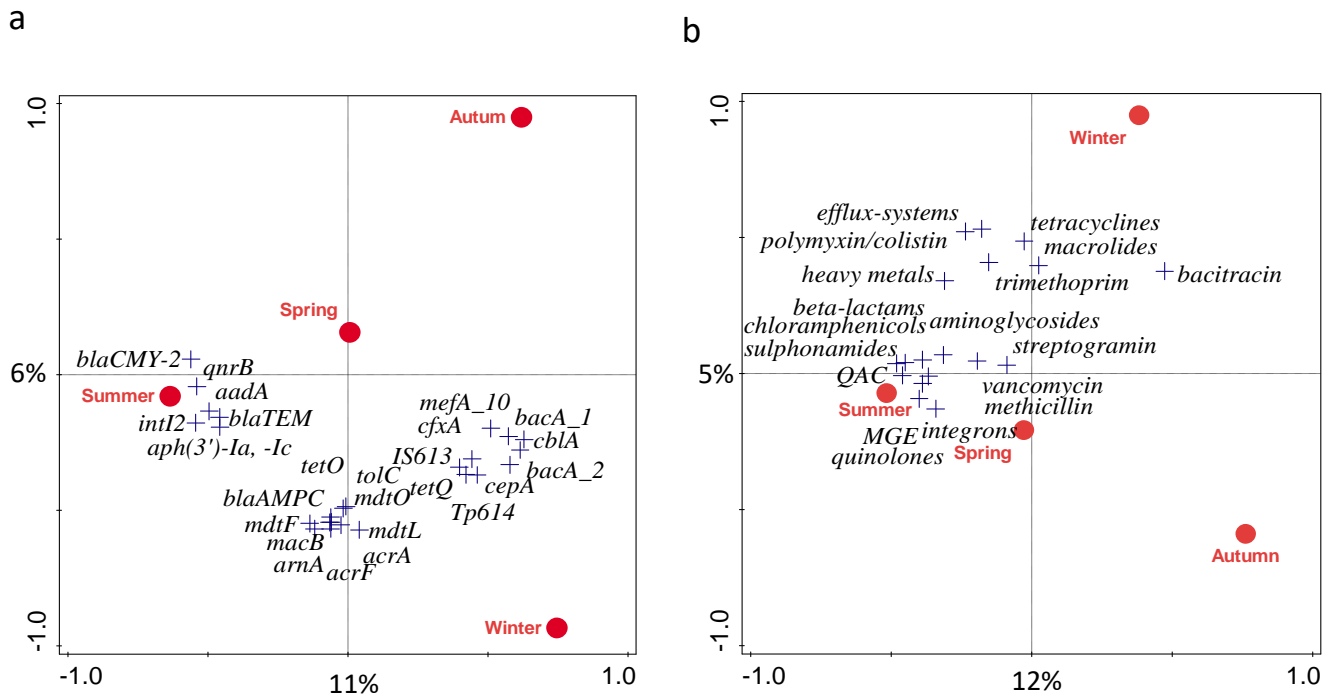
Supplementary Figure 4: Redundancy analysis (RDA) exploring the relationship of the different seasons to individual genes and allocated gene classes in untreated HWW. a: RDA for individual ARGs. Season explains 18% of the variability (MC p-value: 0.038), of which 11% is on the x-axis and 6% on the y-axis. Only the 30 genes most representative are displayed. **b**: RDA for allocated gene classes. Season explains 19% of the variability (MC p-value: 0.06, marginally significant), of which 12% is on the x-axis and 5% on the y-axis.
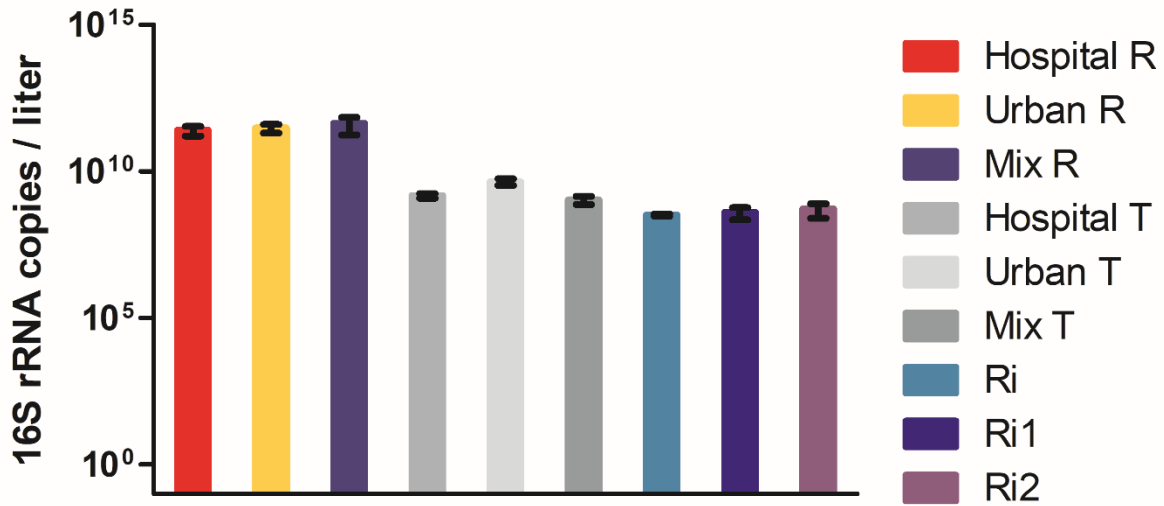
Supplementary Figure 5: Absolute copy number of 16S rRNA genes (Biomass; copy numbers of 16S rRNA genes/ liter) for the different WW and river samples averaged over all samples) +- standard deviation. R= raw (untreated WW); T= treated WW, Ri= river upstream samples; Ri1= river downstream near effluent release pipe; Ri2= river downstream of effluent release pipe.
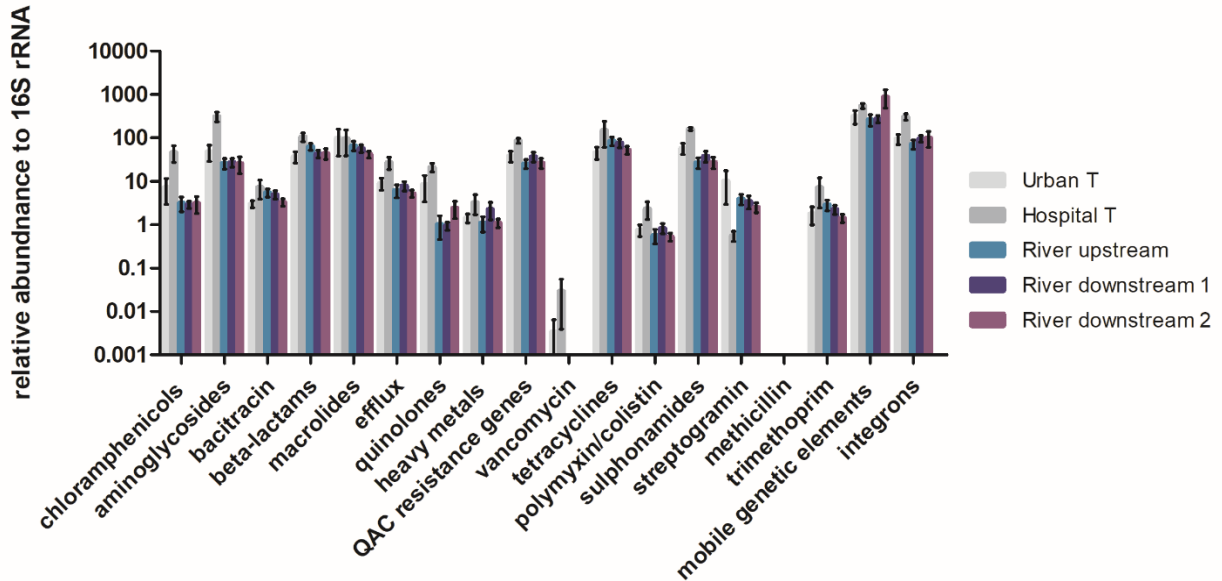
Supplementary Figure 6: Averaged normalized abundance of ARG classes, heavy metals, MGEs and integrons over all collected river samples and their paired treated hospital and urban WW samples +- standard deviation. Normalized abundance levels of gene classes averaged over n=4 samples collected during winter season of 2013 for River, treated HWW and treated UWW samples +- standard deviation. T= treated WW.

a) USING ARG CLASSES AS PREDICTORS

ORIGINAL CLASSIFICATION (rows) vs PREDICTED AS (columns)

	Hospital Untreated	Hospital Treated	Mix Untreated	Mix Treated	Urban Untreated	Urban Treated
Hospital Untreated	87.7	9.0	0.0	0.0	3.2	0.0
Hospital Treated	10.5	73.7	0.0	0.0	5.3	10.5
Mix Untreated	20.0	0.0	60.0	0.0	20.0	0.0
Mix Treated	0.0	25.0	12.5	15.6	0.0	46.9
Urban Untreated	0.0	0.0	3.3	0.0	90.0	6.7
Urban Treated	0.0	7.1	0.0	4.6	7.1	81.1

b) USING ARG GENES AS PREDICTORS

	Hospital Untreated	Hospital Treated	Mix Untreated	Mix Treated	Urban Untreated	Urban Treated
Hospital Untreated	93.5	3.2	0.0	0.0	3.2	0.0
Hospital Treated	10.5	73.7	0.0	5.3	5.3	5.3
Mix Untreated	20.0	0.0	73.5	0.0	6.5	0.0
Mix Treated	2.5	38.1	2.5	31.3	0.0	25.6
Urban Untreated	0.0	0.0	0.0	0.0	93.3	6.7
Urban Treated	0.0	10.7	0.0	0.0	3.6	85.7

c) USING MICROBIOTA AS PREDICTORS

	Hospital Untreated	Hospital Treated	Mix Untreated	Mix Treated	Urban Untreated	Urban Treated
Hospital Untreated	96.8	3.2	0.0	0.0	0.0	0.0
Hospital Treated	5.3	84.2	0.0	0.0	0.0	10.5
Mix Untreated	0.0	0.0	20.0	0.0	80.0	0.0
Mix Treated	0.0	25.0	0.0	50.0	0.0	25.0
Urban Untreated	3.3	0.0	6.0	3.3	84.0	3.3
Urban Treated	0.0	3.6	0.0	0.0	3.6	92.9

Top predictor variables (MDA scores)

Panel a) (all) mean:

| variable	mean	rank
macrolides	46.13	1*
aminoglycosides	43.79	2
streptogramin	42.79	3
chloramphenicols	39.01	4
sulphonamides	31.81	5
quinolones	28.80	6
trimethoprim	24.28	7
tetracyclines	24.17	8
bacitracin	24.08	9
integrons	23.16	10
beta-lactams	19.81	11
QAC resistance genes	19.25	12
vancomycin	17.56	13
mobile genetic elements	16.67	14
polymyxin/colistin	14.41	15
efflux-systems	12.75	16
heavy metals	10.80	17
methicillin	7.16	18

Panel b) (top 45) mean:

variable	mean	rank	variable	mean	rank
blaVIM	23.21	1*	aadE-likegene	12.06	25
ISS1N	21.30	2*	cat	11.77	26
aac6-Ib	21.77	3*	tetO	11.71	27
vadB	20.70	4	ISSWt	11.59	28
ermB	19.00	5	aadE	11.54	29
blaCTX-M	18.85	6	qacE	11.11	30
aac3-IIacde	18.52	7	qnrS	11.04	31
cmlA1	17.97	8	blaTEM	10.83	32
aph3-Ia-Ic	17.80	9	blaCMY-2	10.73	33
ISEc9	17.57	10	aph3-III	10.53	34
catB3	17.07	11	aac6-IIa	10.47	35
blaIMP	16.83	12	tetQ	10.45	36
mefA10	16.55	13	blaSHV	10.35	37
IS6group	16.46	14	blaACC	10.32	38
sul1	14.31	15	blaPer-1	10.28	39
aph2-Ib	14.27	16	aac6-aph2	10.05	40
tetM	14.10	17	blaGES	9.84	41
tetW	14.10	18	intI2	9.73	42
qnrB	13.79	19	bacA1	9.20	43
mfsA	13.31	20	vanB	8.95	44
blaKPC	13.11	21	dfrF	8.72	45
intI	13.05	22			
IS6100	12.94	23			
blaOXA	12.75	24			

Panel c) (top 45) mean:

variable	mean	rank	variable	mean
Enterobacteriales	21.64	1*	Planctomycetales	10.40
Thiotrichales	19.82	2*	uncult. sludge bact. A31	10.18
Fusobacteriales	19.01	3*	Coriobacteriales	10.02
Actinomycetales	18.43	4*	Rs-K70 termite group	10.01
Clostridiales	16.12	5*	Pseudomonadales	9.87
Rhizobiales	14.44	6*	Candidatus W OW2011_GWF1_31_35	9.80
Lactobacillales	14.27	7	SAR324 cladeMarine group B	9.67
PaM15	13.78	8	Mollicutes RF9	9.51
Pedosphaerales group CW2011_GWF2_42_18	13.62	9	Phycisphaerales	9.47
Desulfobacterales	13.08	10	SubsectionII	9.47
SubsectionIII	12.69	11	Acidimicrobiales	9.34
Caldilineales	12.64	12	Candidatus M bact. GW2011_GWE2_42_7	9.32
Sphingomonadales	12.45	13	Oligoflexales	9.31
Selenomonadales	12.22	14	Myxococcales	9.23
Streptosporangiales	12.19	15	Obscuribacterales	9.23
Solirubrobacterales	12.18	16	Legionellales	9.11
Verrucomicrobiales	12.04	17	Nitrospirales	9.10
JG30-KF-CM45	11.87	18	uncult. archaeon	9.10
Chlorobiales	11.86	19	Micrococcales	9.07
Bifidobacteriales	11.86	20	Chlorella sp. ArM0029B	8.88
Frankiales	11.67	21		
Neisseriales	11.07	22		
Flavobacteriales	11.02	23		
uncult. sludge bact.	10.90	24		

Supplementary Figure 7. Classification success of treated and untreated wastewater from hospital, mix and urban sources. Upper panels present the RFA classification accuracy of six water sources (Hospital Untreated / Treated, Mix Untreated / Treated, Urban Untreated / Treated) when setting the predictor variables as **(a)** ARG classes, **(b)** ARG genes, and **(c)** the microbiota. The values within cells are the mean % prediction accuracy of 20 independent RFA runs with 1000 trees and default mtry. Actual water source of samples in rows, predictions in columns. Colour scale is used for visualisation purposes only, from light red at 0%, blue at 50% and green at 100% of accuracy. Top Panels present the top predictor variables (capped at N=45 for visualisation purposes) when setting the predictor variables as **(a)** ARG classes, **(b)** ARG genes, and **(c)** the Microbiota. The values within cells are the mean MDA score of 20 independent RFA runs with 1000 trees and default mtry. Colour scale is used for visualisation purposes only, from light red at 0, blue at percentile 70% and green at percentile 97.25% of all MDA scores. Variables marked with * are the ones that appear more informative than expected given all the MDA scores of all predictor variables (above the cut off of percentile 97.25% of all scores, see method description).

a) USING ARG GENES AS PREDICTORS

all	Hospital Untreated	Hospital Treated	Mix Untreated	Mix Treated	Urban Untreated	Urban Treated	
blaVIM	aph3-la-lc	vatB	ISS1N	blaIMP	blaVIM	blaVIM	1
ISS1N	IS6group	ISS1N	blaVIM	ermB	aac6- lb	aac6-lb	2
aac6.lb	catB3	ISEc9	vatB	ISEc9	blaCTX-M	cm1A1	3
vatB	intI	ermB	ermB	ISS1N	ISS1N	ISS1N	4
ermB	mfsA	mefA10	ISEc9	blaVIM	ermB	ermB	5
blaCTX.M	cm1A1	IS6100	blaCTX-M	aadE	aac3-liacde	blaIMP	6
aac3.liacde	tetW	aph2-lb	mefA10	aac6-lb	vatB	catB3	7
cm1A1	tetO	tetW	blaIMP	cepA	mefA10	blaKPC	8
aph3.la.lc	cat	ISSW1	aac6-lb	qacE	ISEc9	ISEc9	9
ISEc9	aac6- lb	sul1	ISSW1	ermC	blaIMP	tetM	10

b) USING ARG CLASSES AS PREDICTORS

all	Hospital Untreated	Hospital Treated	Mix Untreated	Mix Treated	Urban Untreated	Urban Treated	
macrolides	aminoglycosides	streptogramin	streptogramin	streptogramin	aminoglycosides	macrolides	1
aminoglycosides	quinolones	macrolides	macrolides	macrolides	macrolides	chloramphenicols	2
streptogramin	chloramphenicols	sulphonamides	aminoglycosides	vancomycin	streptogramin	aminoglycosides	3
chloramphenicols	trimethoprim	tetracyclines	bacitracin	sulphonamides	chloramphenicols	quinolones	4
sulphonamides	tetracyclines	beta-lactams	mobile genetic elements	chloramphenicols	sulphonamides	streptogramin	5
quinolones	bacitracin	trimethoprim	chloramphenicols	QAC resistance genes	quinolones	integrons	6
trimethoprim	sulphonamides	bacitracin	vancomycin	bacitracin	integrons	sulphonamides	7
tetracyclines	QAC resistance genes	mobile genetic elements	tetracyclines	aminoglycosides	beta-lactams	vancomycin	8
bacitracin	macrolides	quinolones	sulphonamides	quinolones	QAC resistance genes	beta-lactams	9
integrons	polymyxin/colistin	vancomycin	trimethoprim	polymyxin/colistin	tetracyclines	QAC resistance genes	10

c) USING MICROBIOTA AS PREDICTORS

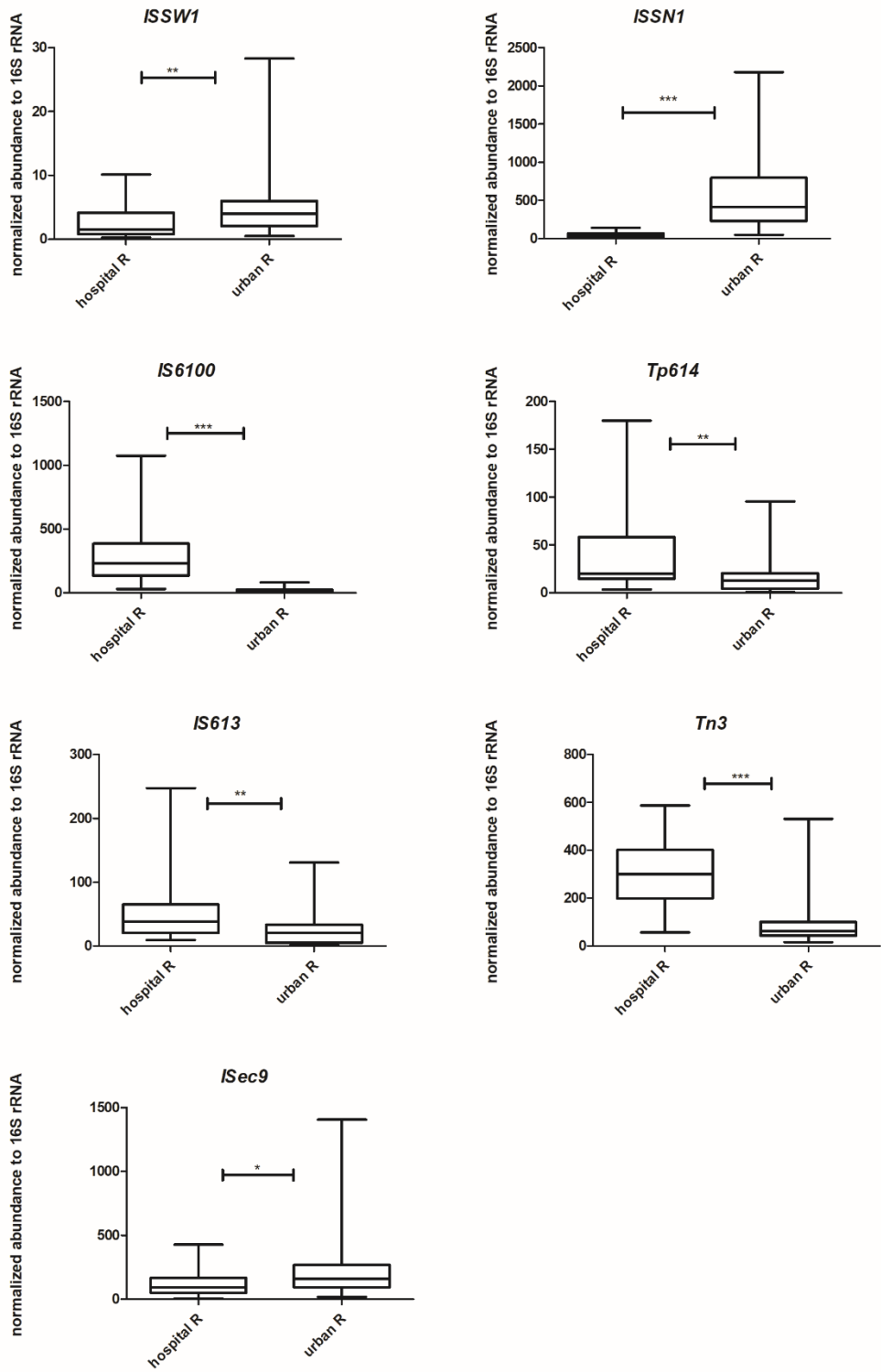| all	Hospital Untreated	Hospital Treated	Mix Untreated	Mix Treated	Urban Untreated	Urban Treated	
Enterobacteriales	Enterobacteriales	Lactobacillales	Fusobacteriales	uncult. sludge bact. A31	Enterobacteriales	PeM15	1
Thiotrichales	Thiotrichales	Actinomycetales	Actinomycetales	Blastocatellales	Fusobacteriales	Parcubacteria group GW2011_GWE2_42_14	2
Fusobacteriales	Fusobacteriales	Sphingomonadales	Lactobacillales	SAR324 clade Marine group B	Thiotrichales	Caldilineales	3
Actinomycetales	Clostridiales	Fusobacteriales	Alphaproteobacteria Incertae Sedis	KI89A clade	Clostridiales	SubsectionIII	4
Clostridiales	Rhizobiales	JG30-KF-CM45	Pasteurellales	Solirubrobacterales	Verrucomicrobiales	Solirubrobacterales	5
Rhizobiales	Actinomycetales	Frankiales	Desulfobacterales	uncult. archaeon	Actinomycetales	Thiotrichales	6
Lactobacillales	Desulfobacterales	SubsectionII	Streptosporangiales	uncult. Acidobacteria bact.	Selenomonadales	uncult. sludge bact.	7
PeM15	Bifidobacteriales	Oligoflexales	Myxococcales	Phycisphaerales	Solirubrobacterales	Candidatus W bact. GW2011_GWF1_31_35	8
Parcubacteria group GW2011_GWE2_42_14	Streptosporangiales	Coriobacteriales	Enterobacteriales	JG30-KF-CM45	Rhizobiales	Planctomycetales	9
Desulfobacterales	Chlorobiales	Neisseriales	Solirubrobacterales	Obscuribacterales	Pseudomonadales	Sphingomonadales	10

Figure 8. Top 10 predictors for Classification success of treated and untreated wastewater from hospital, mix and urban sources. Panels present the top 10 predictor variables for RFA classification accuracy of six water sources Hospital Untreated / Treated, Mix Raw Untreated / Treated, Urban Untreated / Treated) when setting the predictor variables as **(a)** ARG genes, **(b)** ARG classes, and **(c)** the microbiota. The column 'all' presents the top 10 predictors from classification exercise and accuracy presented in Supplementary Figure 7. The other columns present the top 10 predictors (from partial dependence) for particular states of the response variable (water source); Colours are solely to facilitate visualisation.
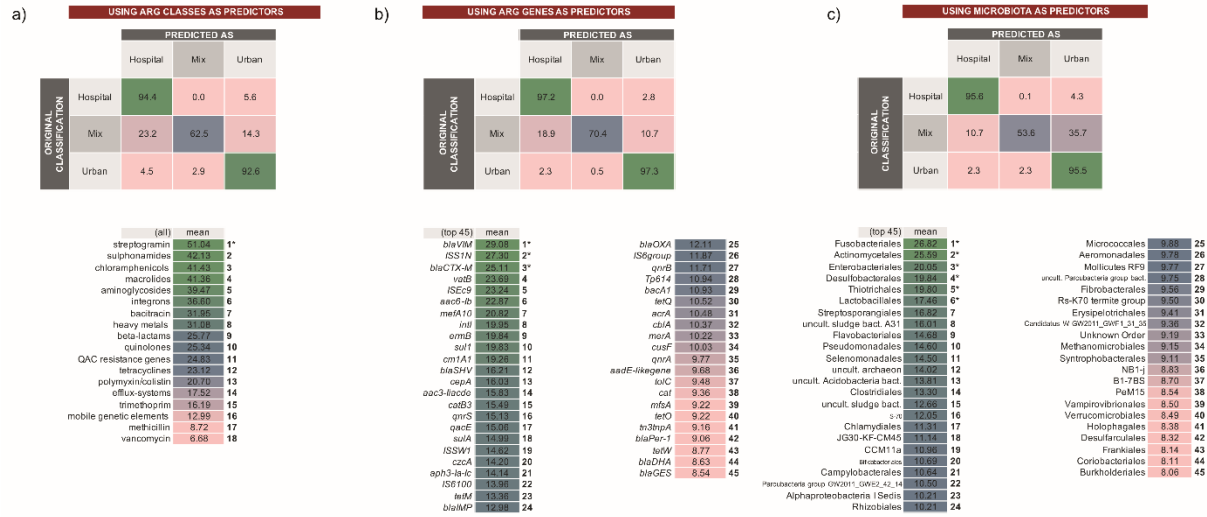
Supplementary Figure 9. Classification success of hospital, mix, and urban sources when ignoring wastewater type (treated, untreated). Upper panels present the RFA classification accuracy of three water sources (Hospital, Mix, Urban) when setting the predictor variables as **(a)** ARG classes, **(b)** ARG genes, and **(c)** the Microbiota. For this classification exercise, the original Hospital Untreated / Treated, Mix Untreated / Treated, Urban Untreated / Treated classifications of the response variable (as in Supplementary Figure 7) are aggregated to Hospital, Mix and Urban only (Untreated / Treated are collapsed). The values within cells are the mean % prediction accuracy of 20 independent RFA runs with 1000 trees and default mtry. Actual water source of samples in rows, predictions in columns. Colour scale is used for visualisation purposes only, from light red at 0%, blue at 50% and green at 100% of accuracy. Top Panels present the top predictor variables (capped at N=45 for visualisation purposes) when setting the predictor variables as (a) ARG classes, (b) ARG genes, and (c) the Microbiota. The values within cells are the mean MDA score of 20 independent RFA runs with 1000 trees and default mtry. Colour scale is used for visualisation purposes only, from light red at 0, blue at percentile 70% and green at percentile 97.25% of all MDA scores. Variables marked with * are the ones that appear more informative than expected given all the MDA scores of all predictor variables (above the cut-off of percentile 97.25% of all scores, see method description).

Supplementary Figure 10: Normalized abundance of individual transposase genes (from IS or transposons (MGEs)) in untreated hospital and urban WW. R= untreated. The non-parametric Mann-Whitney test was used to test for significance: *=P < 0.05; **= P> 0.005; ***= P<0.0001.