

Interconnected microbiomes and resistomes in low-income human habitats

Erica C. Pehrsson^{1*}, Pablo Tsukayama^{1*}, Sanket Patel^{1,2}, Melissa Mejía-Bautista^{1,3}, Giordano Sosa-Soto^{1,3}, Karla M. Navarrete³, Maritza Calderon⁴, Lilia Cabrera⁵, William Hoyos-Arango³, M. Teresita Bertoli³, Douglas E. Berg^{6,7}, Robert H. Gilman^{4,5,8} & Gautam Dantas^{1,2,6,9}

Antibiotic-resistant infections annually claim hundreds of thousands of lives worldwide. This problem is exacerbated by exchange of resistance genes between pathogens and benign microbes from diverse habitats. Mapping resistance gene dissemination between humans and their environment is a public health priority. Here we characterized the bacterial community structure and resistance exchange networks of hundreds of interconnected human faecal and environmental samples from two low-income Latin American communities. We found that resistomes across habitats are generally structured by bacterial phylogeny along ecological gradients, but identified key resistance genes that cross habitat boundaries and determined their association with mobile genetic elements. We also assessed the effectiveness of widely used excreta management strategies in reducing faecal bacteria and resistance genes in these settings representative of low- and middle-income countries. Our results lay the foundation for quantitative risk assessment and surveillance of resistance gene dissemination across interconnected habitats in settings representing over two-thirds of the world's population.

Antibiotic resistance in bacterial pathogens causes hundreds of thousands of annual fatalities globally^{1–3}. The spread of resistant organisms and their antibiotic resistance genes occurs by direct contact between humans and via interactions with environmental microbiota^{4–6}. Horizontal gene transfer (HGT) facilitates this dissemination, and the reservoir of antibiotic resistance genes (the ‘resistome’⁷) in the environment, from which pathogens could theoretically draw, is ancient, diverse, and widespread^{8–10}. Characterizing resistome distributions and their potential for dissemination across diverse habitats can identify the microbiota and antibiotic resistance genes that pose the highest risks to human health.

Most resistome studies have focused on either industrialized^{11–13} or remote, ‘pristine’ settings^{8,9,14}. However, most of the world's people reside outside these extremes: ~5.8 billion live in low- and middle-income countries¹⁵, with 863 million people living in slums¹⁶. Here, we characterized the microbiomes and resistomes of human faecal and co-localized, ecologically diverse environmental microbiota from two low-income, resource-limited Latin American settings: (1) a rural village of subsistence farmers in El Salvador (RES) and (2) a peri-urban shanty-town (slum) in Lima, Peru (PST), which represent critically understudied microbial ecosystems (Extended Data Fig. 1a). Aspects of life in such settings that are distinct from both hunter-gatherer and industrialized populations include crowding (in PST), limited access to clean drinking water and sanitation, supplementation of personally grown produce and livestock with processed foods, and ready access to antibiotics without prescriptions¹⁷. Furthermore, such industrializing countries are responsible for the majority of the worldwide 36% increase in antibiotic use between 2000 and 2010 (ref. 18), making investigation of antibiotic resistance transfer in these settings a global public health priority.

We analysed 263 faecal samples from 115 individuals in 27 houses over two years from RES and PST, as well as 209 environmental samples from donor households and surrounding areas in these communities.

The environmental samples included faeces from domesticated animals, soil, water, and samples from the sanitation facilities of each community: composting latrines in RES and a district-wide sewage system with a modern wastewater treatment plant (WWTP) in PST. We used a combination of 16S sequencing^{10,19}, high-throughput functional metagenomic selections^{10,20,21}, and whole-metagenome shotgun sequencing^{11,22} to compare the phylogenetic architectures of these microbial populations and their associated resistomes.

Resistome correlates with phylogeny across habitats

Small-insert metagenomic expression libraries in *Escherichia coli* constructed from 51 human faecal and 45 environmental samples from RES and PST (representing 258 Gb) were selected for functional resistance against 17 antibiotics (Supplementary Table 1). Sequencing and annotation^{4,23} of these selections identified 1,100 unique (100% amino acid identical) encoded antibiotic resistance proteins collectively conferring resistance against all antibiotics except meropenem (see Methods; Supplementary Table 2). A total of 121 of these proteins were novel (<60% amino acid identity to any protein in NCBI nr), the majority of which (72%) were predicted antibiotic modifiers, including 57 class A β -lactamases. RES latrine libraries yielded the most novel proteins (46%), proportionally more than expected (Pearson's chi-squared test, $P < 2 \times 10^{-5}$; Extended Data Fig. 1c).

To further characterize RES and PST resistome diversity and abundance, we performed whole-metagenome shotgun sequencing on 191 human faecal and 94 environmental samples (representing 344 Gb; Extended Data Fig. 1b; see Methods). We used ShortBRED²⁴ to quantify translated antibiotic resistance gene abundance in all sequenced metagenomes using a custom antibiotic resistance database that included antibiotic resistance genes identified here (see Methods; Supplementary Tables 3–5). RES and PST human-associated and environmental resistomes were related along an ecological gradient

¹Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Facultad de Ciencias de la Salud “Dr. Luis Edmundo Vásquez”, Universidad Dr. José Matías Delgado, El Salvador. ⁴Laboratorios de Investigación y Desarrollo, Universidad Peruana Cayetano Heredia, San Martín de Porres, Lima 31, Peru. ⁵Asociación Benéfica PRISMA, San Miguel, Lima 32, Peru. ⁶Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ⁷Department of Medicine, University of California San Diego, La Jolla, California 92093, USA. ⁸Department of International Health, Johns Hopkins School of Public Health, Baltimore, Maryland 21205, USA. ⁹Department of Biomedical Engineering, Washington University, St Louis, Missouri 63105, USA. *These authors contributed equally to this work.

in terms of input from human faeces (Bray–Curtis; Extended Data Fig. 2a), with habitat explaining 22.4% of resistome variation (adonis, $P < 0.001$, Bray–Curtis).

A similar ecological gradient by habitat was observed when considering phylogenetic composition, based on 16S sequencing of 228 human faecal and 203 environmental samples from RES and PST (Extended Data Fig. 1b, see Methods), with habitat explaining even more of the variation between samples (41.9%; adonis, $P < 0.001$, weighted UniFrac; Extended Data Fig. 2b). Procrustes analysis confirmed that antibiotic resistance is significantly correlated with community composition (Bray–Curtis, $M^2 = 0.360$, $P < 0.001$; Extended Data Fig. 2c), not randomly distributed across habitats^{10,23}.

RES/PST versus global human faecal microbiota

The human gut microbiota is highly diverse globally, with phylogenetic and functional variation potentially driven by age, diet, cultural traditions, pathogen carriage, and periodic perturbation (for example, by antibiotic exposure)^{19,22}. We compared the RES and PST human faecal microbiota to published microbiota data sets from the USA, Malawi, Venezuela, and Peru (Amazonian hunter–gatherers and Andean highlands agriculturalists) (see Methods)^{14,19,25}, classifying PST as ‘peri-urban industrializing’ and RES as ‘rural agriculturalist’. Faecal microbiota composition clustered by host lifestyle^{19,25}, despite differences in geographic origin and study: RES microbiota clustered with other rural agriculturalists and hunter–gatherers, while peri-urban PST microbiota were intermediate between these and the urban industrialized cohorts (adonis $R^2 = 23.8\%$, $P < 0.001$, weighted UniFrac; Fig. 1a and Extended Data Fig. 3).

To evaluate whether antibiotic resistance burden correlated with industrialization^{14,26,27}, we compared resistomes from RES and PST to industrialized USA, traditional hunter–gatherer, and rural agriculturalist communities from²⁵, two of which are also Peruvian (Extended Data Fig. 4a–f; see Methods). PST had the greatest number of antibiotic resistance proteins per person, despite not being the most industrialized cohort analysed (Fig. 1b and Extended Data Fig. 4g), consistent with Chinese and Hadza hunter–gatherer faecal microbiota harbouring at least as much or more antibiotic resistance as Western industrialized microbiota^{11,22}.

RES human and environmental microbiota

Frequent contact with environmental reservoirs during subsistence farming and inadequate excreta management²⁸ likely promote antibiotic resistance exchange in rural settings. The RES community had only one source of filtered water, and contents of composting latrines were used to fertilize household agricultural plots after attempted sterilization, potentially recycling antibiotic resistance. Accordingly, we compared the microbiomes and resistomes of RES human faecal samples and their surroundings, including latrines, animal faeces, soils, and drinking water sources (see Methods).

The RES human faecal microbiota separated from soil and water along PC1 in a principal coordinate analysis (PCoA) visualization of their phylogenetic composition, and habitats differed at the phylum level (Fig. 2a, Extended Data Fig. 5a and Supplementary Tables 6–10). Latrines were equidistant to human faecal microbiota and soil (weighted UniFrac, non-parametric Student's *t*-tests with Bonferroni correction) and were enriched for Halomonadaceae, a family of halophilic organisms potentially selected by the alkaline latrine environment (Extended Data Fig. 5b, Supplementary Discussion). Cow and dog faecal microbiota, which are ecologically similar to human faecal microbiota but with greater environmental exposure, were also intermediate along PC1. Soil and water were closest to each other ($P < 0.05$, non-parametric Student's *t*-tests with Bonferroni correction) and varied along PC3 (6.3% of variation). Human faecal microbiota were more homogenous in phylogenetic composition than latrine, soil, and water microbiota, potentially because environmental samples encountered more diverse and variable conditions.

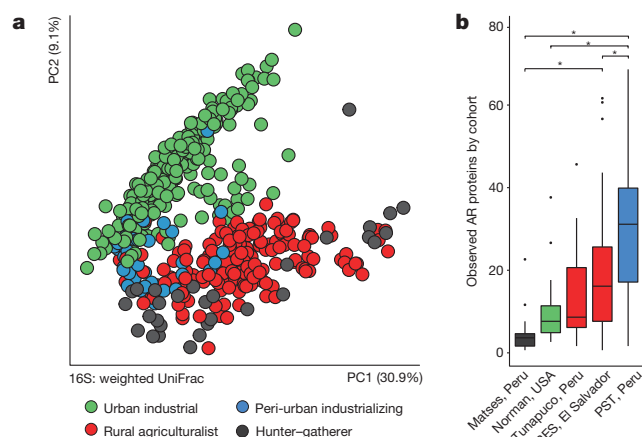


Figure 1 | RES and PST human faecal microbiota and resistomes versus global populations. **a**, PCoA of weighted UniFrac distances between RES ($n = 60$) and PST ($n = 46$) microbiota and published human faecal microbiota^{14,19,25} ($n = 446$; see Supplementary Table 14), coloured by host lifestyle. Adonis $R^2 = 15.4\%$, $P < 0.001$. **b**, Number of antibiotic resistance proteins per person in RES ($n = 42$) and PST ($n = 44$) and published human faecal microbiota (ref. 25) ($n = 53$; see Supplementary Table 15), coloured by host lifestyle. Error bars, s.d.; centre bars, median. * $P < 0.05$, non-parametric Student's *t*-tests, Bonferroni correction. AR, antibiotic resistance.

The RES resistomes exhibited a similar ecological gradient (Bray–Curtis, Fig. 2b). Although RES soil had the highest phylogenetic diversity (Fig. 2c), it contained fewer antibiotic resistance proteins per sample than all habitats but water (Fig. 2d). In contrast, human faecal microbiota had the lowest phylogenetic diversity, but more antibiotic resistance proteins per sample than both soil and water. However, soil and latrine resistomes were more heterogeneous than human and animal faecal resistomes. In particular, non-human RES resistomes were enriched in aminoglycoside nucleotidyltransferases, class D β -lactamases, SMR transporters, and resistant dihydropteroate synthetases and dihydrofolate reductases (Fig. 2b, Extended Data Fig. 5c, d; see Methods).

By analysing RES habitats prone to exchange with human faeces using SourceTracker²⁹, we found that direct input from chicken faeces represents a potent avenue for the introduction of antibiotic resistance genes compatible with the human faecal microbiota into soil (see Methods). The contribution of RES human faecal resistomes to environmental resistomes (Fig. 2f) roughly recapitulated phylogenetic trends, with significantly higher input to latrines than soil or water (Fig. 2e; see Supplementary Information). Only the two soils collected adjacent to chicken coops had $>1\%$ human faecal resistome input. These two samples clustered away from the other samples along PC3 (5.3% of variation) and were closer to human samples along PC1 in the PCoA visualization of all RES samples (Fig. 2a). Overall, RES human faecal microbiota were closer to soil collected near chicken coops than to any other soil location in both phylogenetic and antibiotic resistance composition (Bray–Curtis, non-parametric Student's *t*-tests with Bonferroni correction). They shared 80 antibiotic resistance proteins at $>99\%$ identity, including three class C β -lactamases, which are common in Enterobacteriaceae, in contrast to two antibiotic resistance proteins shared between RES humans and soil from the pond edge. On average, human faecal and chicken coop soil resistomes shared 10 antibiotic resistance proteins, significantly more than with any other soil type (1–2 antibiotic resistance proteins, non-parametric Student's *t*-tests with Bonferroni correction).

PST human and sewage microbiota

To monitor the impact of sewage treatment on resistomes, we sampled wastewater at street-level access points nearest to participating PST households and influent and effluent from the local WWTP and compared their phylogenetic composition and resistome to the

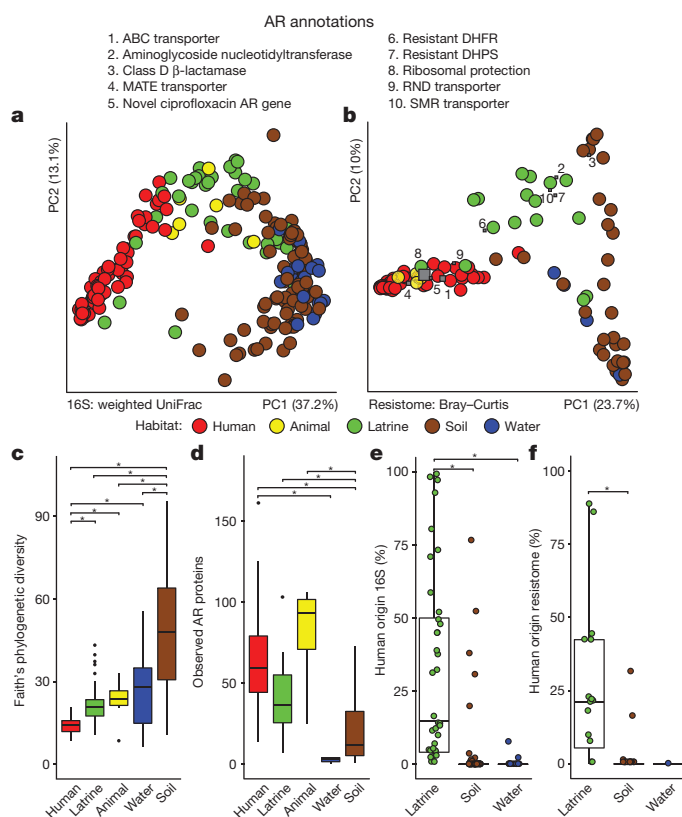


Figure 2 | Salvadoran rural agriculturalist (RES) human faecal and environmental microbiota and resistomes. **a, c, e,** Phylogenetic composition ($n = 60, 6, 36, 84$ and 22 for human, animal, latrine, soil and water, respectively). **b, d, f,** Resistome ($n = 42, 4, 16, 30$ and 4 for human, animal, latrine, soil and water, respectively). **a,** PCoA of weighted UniFrac distances between microbiota. Adonis $R^2 = 43.2\%$, $P < 0.001$. **b,** PCoA of Bray–Curtis distances between resistomes with abundance-weighted coordinates of the top five most discriminative antibiotic resistance categories enriched in human or non-human habitats (squares, size proportional to overall abundance). Adonis $R^2 = 26.6\%$, $P < 0.001$. **c, d,** $*P < 0.05$, non-parametric Student's t -tests, Bonferroni correction. **c,** Faith's phylogenetic diversity. **d,** Observed antibiotic resistance proteins. **e, f,** Percentage of latrine, soil, and water microbiota (**e**) and resistomes (**f**) attributable to human faeces, as determined by SourceTracker²⁹. $*P < 0.05$, pairwise Wilcoxon tests, Bonferroni correction. Error bars, s.d.; centre bars, median.

faecal microbiota of PST residents (see Methods and Supplementary Information). Portions of treated wastewater effluent are discharged into the Pacific Ocean and also used to irrigate public parks and agricultural fields, potentially enabling re-introduction of antibiotic residues and antibiotic resistance genes into human communities^{30–32}.

Although geographically closest to human donors, street-access sewage was more similar in microbial composition to WWTP influent (non-parametric Student's t -tests with Bonferroni correction, $P < 0.001$), implying that even relatively brief periods in this non-human, aerobic environment caused a greater shift in bacterial composition than all changes downstream during transit through the sewage system. Overall, PST human waste underwent drastic changes in microbial composition as it progressed through sewage treatment, decreasing in similarity to human faecal microbiota at each subsequent stage (Student's non-parametric t -tests with Bonferroni correction, $P < 0.001$, weighted UniFrac; Fig. 3a, Extended Data Fig. 6a, b and Supplementary Table 11).

Sewage resistomes also decreased in similarity to PST human faecal resistomes at each treatment stage, although street-access sewage and WWTP influent were equally similar to human faeces in antibiotic resistance content (non-parametric Student's t -tests with Bonferroni

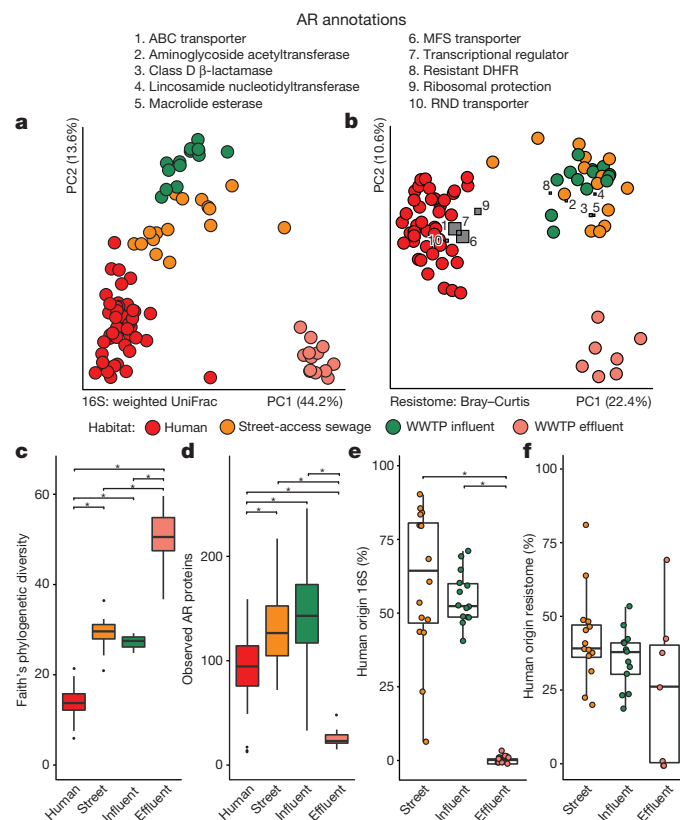


Figure 3 | Peruvian peri-urban slum (PST) human faecal and sewage microbiota and resistomes. **a, c, e,** Phylogenetic composition ($n = 45, 16, 14$ and 13 for human, street-access, influent and effluent, respectively). **b, d, f,** Resistome ($n = 44, 14, 13$ and 7 for human, street-access, influent and effluent, respectively). **a,** PCoA of weighted UniFrac distances between microbiota. Adonis $R^2 = 58.0\%$, $P < 0.001$. **b,** PCoA of Bray–Curtis distances between resistomes with abundance-weighted coordinates of the top five most discriminative antibiotic resistance categories enriched in human or sewage habitats (squares, size proportional to overall abundance). Adonis $R^2 = 32.3\%$, $P < 0.001$. **c, d,** $*P < 0.05$, non-parametric Student's t -tests, Bonferroni correction. **c,** Faith's phylogenetic diversity. **d,** Observed antibiotic resistance proteins. **e, f,** Percentage of sewage microbiota (**e**) and resistomes (**f**) attributable to human faeces at each sewage treatment stage, as determined by SourceTracker. $*P < 0.05$, pairwise Wilcoxon tests, Bonferroni correction. Error bars, s.d.; centre bars, median.

correction, Bray–Curtis; Fig. 3b). Although WWTP effluent had the highest phylogenetic diversity (Fig. 3c), it had the fewest antibiotic resistance proteins per sample (Fig. 3d), as observed for soil in RES. In contrast, street-access sewage and WWTP influent had both higher phylogenetic diversity and more antibiotic resistance proteins per sample than PST human faeces. Drug efflux antibiotic resistance mechanisms were higher overall in PST human faecal vs sewage resistomes ($P < 0.05$, pairwise Wilcoxon tests with Bonferroni correction; Fig. 3b and Extended Data Fig. 6c, d). In contrast, sewage was enriched for aminoglycoside acetyltransferases, class D β -lactamases, and resistant dihydrofolate reductases.

As expected, the human faecal contribution to WWTP effluent phylogenetic composition was lower than for street-access sewage or WWTP influent ($P < 0.05$, pairwise Wilcoxon tests with Bonferroni correction; Fig. 3e). Interestingly, this difference was not significant for resistomes, where human faecal input was high at all sewage treatment stages ($P > 0.05$; Fig. 3f). Thus, although sewage treatment was successful in reducing the overall carriage of antibiotic resistance genes as well as depleting human faecal microbes, antibiotic resistance genes of faecal origin survived and could be reintroduced into the natural environments where effluent is discharged.

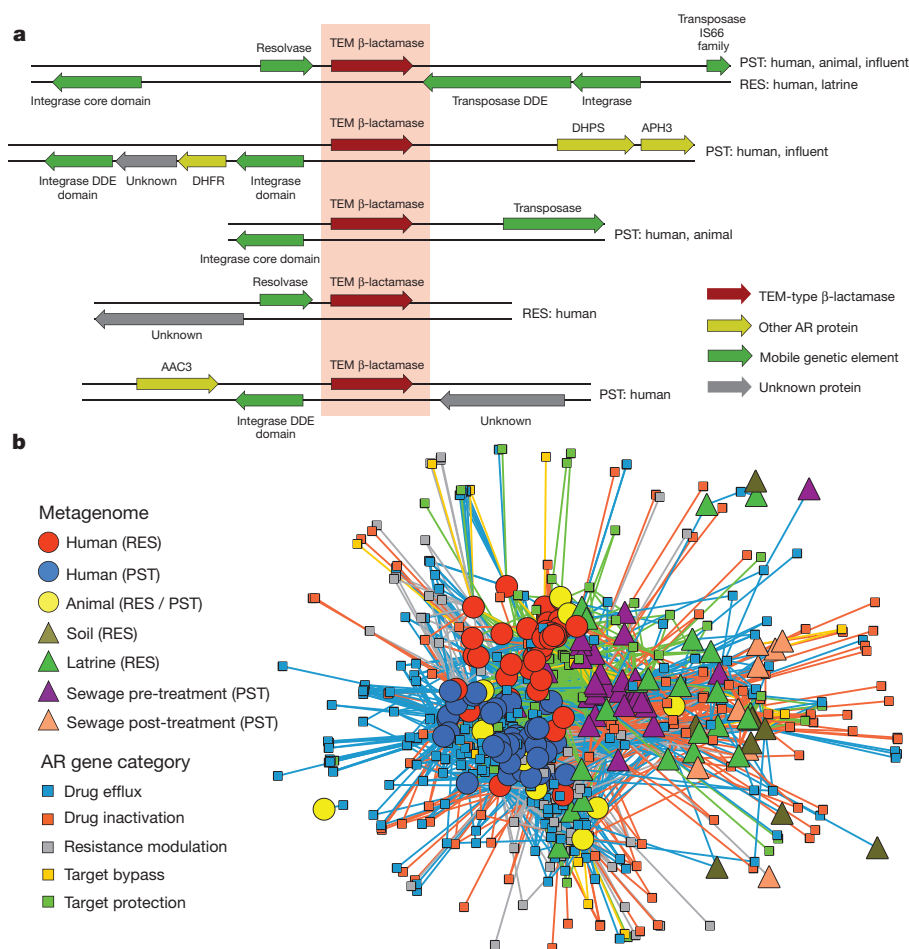


Figure 4 | Antibiotic resistance proteins found in multiple habitats and genetic contexts in RES and PST. a, Representative alignment of 5 of 25 contigs encoding a TEM-type β -lactamase at 99.9% nucleotide identity (full list of contigs in Methods). Contigs were annotated with Resfams v1.2. Source metagenomic libraries are indicated on the left. **b**, Antibiotic resistance networks between human and environmental metagenomes in RES and PST. Small nodes (squares) represent unique antibiotic resistance proteins found in at least one sampled metagenome, coloured by predicted resistance mechanism. Large nodes represent individual human/animal (circle) or environmental (triangle) metagenomes, coloured by habitat/cohort. Lines connecting samples and antibiotic resistance proteins represent a ShortBRED hit with an RPKM (reads per kilobase per million reads) of ≥ 10 , coloured by mechanism.

As the presence of antibiotics in sewage systems may influence resistome diversity and select for HGT, we used a modified solid phase extraction and mass spectrometry protocol³³ to detect 16 antibiotics from seven classes in 22 WWTP influent and effluent samples. Chloramphenicol, ciprofloxacin, tetracycline, trimethoprim, and sulfamethoxazole were consistently detected in influent throughout the sampling period, while erythromycin was detected in 36% of influent samples. Only sulfamethoxazole was detected in effluent samples (concentrations of 18–26,000 $\mu\text{g l}^{-1}$) (Supplementary Table 12). Antibiotics detected in WWTP samples were among the highest selling antibiotics in Peru¹⁸ and may enrich for antibiotic resistance in these bacterial communities and in those exposed to effluent. However, no β -lactams were detected in any WWTP sample, despite the high abundance of β -lactam resistance genes found in our metagenomic surveys and amoxicillin being the highest-selling antibiotic in Peru¹⁸. This suggests that β -lactam antibiotics are degraded to undetectable levels in humans or the sanitation system before reaching the WWTP.

Highly cosmopolitan AR and HGT across microbiota

Although resistome and phylogenetic composition appear tightly linked in most microbial communities^{10,23}, some clinically relevant antibiotic resistance proteins (for example, TEM, CTX-M, KPC, AAC-6') have been extremely successful in global dissemination via clonal expansion and HGT between multiple pathogen hosts³⁴. We identified highly cosmopolitan antibiotic resistance proteins by comparing the prevalence of genes encoding them across all RES and PST habitats (Extended Data Fig. 7a). Two proteins, including the sulfonamide-resistant dihydropteroate synthetase (DHPS) Sul2, were found in 50% of samples in six of seven habitats. On our functional metagenomic contigs, ten of the twelve DHPSs were $>98\%$ amino acid identical to Sul1, Sul2, or Sul3 and were co-localized with integrases and numerous

other antibiotic resistance genes, suggesting multidrug-resistance integrons may facilitate their broad distribution in these settings³⁵.

To further investigate antibiotic resistance exchange potential in RES and PST, we examined flanking genetic sequences in the contig assemblies from our functional selections for evidence of past HGT. A total of 120 (11%) of our unique antibiotic resistance proteins were encoded in more than one genetic context (contigs with $<90\%$ local nucleotide identity; Extended Data Fig. 7b), and the number of contexts was positively correlated with the number of metagenomic libraries (Spearman's $\rho = 0.59$, $P < 2.2 \times 10^{-16}$) and habitats (Spearman's $\rho = 0.47$, $P < 2.2 \times 10^{-16}$) in which an antibiotic resistance protein was encoded. One TEM β -lactamase (TEM-1) was encoded in 25 contexts (Fig. 4a). In contrast, 41% of antibiotic resistance proteins found in multiple habitats were always encoded in the same genetic context. For instance, a TetX (a tetracycline-inactivating enzyme) variant was encoded in the same context in human and animal faeces, latrines, and sewage influent. Three of the six antibiotic resistance proteins encoded by both human faecal and soil microbiota were encoded in the same genetic context (CblA and TEM class A β -lactamases and a class D β -lactamase).

We revealed a large network of antibiotic resistance gene sharing between microbial communities of human, animal, and environmental origin (Fig. 4b), facilitated by HGT between bacterial genomes and spread of bacterial hosts across communities. To further assess the potential mobility of the antibiotic resistance genes found in our contigs, we identified putative mobile genetic elements (MGEs) and multidrug resistance clusters (MDRCs) by annotation (Supplementary Table 13; see Methods). There was a small but significant positive correlation between the proportion of antibiotic resistance contigs with an MGE or MDRC and the number of libraries and habitats in which the antibiotic resistance protein was encoded (Spearman's $\rho = 0.11$ – 0.17 ,

$P < 4.3 \times 10^{-4}$), which supports a role for MGEs and MDRCs in antibiotic resistance transfer across environments and increased accessibility to pathogens³⁶. When ecological analyses were restricted to only antibiotic resistance genes adjacent to an MGE on one of our functional contigs, the results largely recapitulated the trends observed with the full antibiotic resistance set, with resistome correlating with phylogenetic composition across ecological habitats. Additionally, the RES human faecal contribution was reduced in soil compared to latrine resistomes, while the contribution of PST human faeces to sewage was not significantly different before or after treatment (Supplementary Discussion and Extended Data Fig. 8).

Conclusions

Our characterization of human faecal and environmental microbiota and their resistomes from two low-income settings in Latin America is particularly relevant to global public health. Billions of people currently live in rural or transitional areas around large urban centres, where unregulated access to antibiotics and limited access to clean water and improved sanitation increase the risk of pathogen transmission. Future studies on the factors that promote or restrict antibiotic resistance exchange between environmental microbiota, human commensals, and pathogens, particularly during waste treatment, are merited. These would involve real-time molecular surveillance of 'high-risk' environments (for example, hospitals, large-scale animal feeding operations) to identify specific routes for the spread of resistant bacteria and antibiotic resistance genes and inform the design of public health interventions to decrease their global enrichment and dissemination.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 January; accepted 16 March 2016.

1. *Antimicrobial Resistance: Global Report on Surveillance* 1st edn (World Health Organization, 2014).
2. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2013* (2013).
3. The Review on Antimicrobial Resistance. *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations* (2014).
4. Forsberg, K. J. *et al.* The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107–1111 (2012).
5. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nature Rev. Microbiol.* **8**, 251–259 (2010).
6. Martinez, J. L. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proc. R. Soc. Lond. B* **276**, 2521–2530 (2009).
7. Wright, G. D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Rev. Microbiol.* **5**, 175–186 (2007).
8. D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457–461 (2011).
9. Allen, H. K., Moe, L. A., Roddumr, J., Gaarder, A. & Handelsman, J. Functional metagenomics reveals diverse β -lactamases in a remote Alaskan soil. *ISME J.* **3**, 243–251 (2009).
10. Forsberg, K. J. *et al.* Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**, 612–616 (2014).
11. Hu, Y. *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Commun.* **4**, 2151 (2013).
12. Li, B. *et al.* Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J.* **9**, 2490–2502 (2015).
13. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241–244 (2011).
14. Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Science Advances* **1**, e1500183 (2015).
15. The World Bank Group. *Data: Countries: High Income* (<http://data.worldbank.org/income-level/HIC>) (2015).
16. World Health Organization. *Global Health Observatory (GHO) Data: Urban Health* (http://www.who.int/gho/urban_health/en/) (2015).
17. Okeke, I. N. *et al.* Antimicrobial resistance in developing countries. Part I: recent trends and current status. *Lancet Infect. Dis.* **5**, 481–493 (2005).
18. Van Boeckel, T. P. *et al.* Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *Lancet Infect. Dis.* **14**, 742–750 (2014).
19. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
20. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998).
21. Sommer, M. O., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128–1131 (2009).
22. Rampelli, S. *et al.* Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr. Biol.* **25**, 1682–1693 (2015).
23. Gibson, M. K., Forsberg, K. J. & Dantas, G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* **9**, 207–216 (2014).
24. Kaminski, J. *et al.* Fast and accurate metagenomic search with ShortBRED. *PLoS Comp. Biol.* **11**, e1004557 (2015).
25. Obregon-Tito, A. J. *et al.* Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Commun.* **6**, 6505 (2015).
26. Walson, J. L., Marshall, B., Pokhrel, B. M., Kafle, K. K. & Levy, S. B. Carriage of antibiotic-resistant fecal bacteria in Nepal reflects proximity to Kathmandu. *J. Infect. Dis.* **184**, 1163–1169 (2001).
27. Pallecchi, L. *et al.* Quinolone resistance in absence of selective pressure: the experience of a very remote community in the Amazon forest. *PLoS Negl. Trop. Dis.* **6**, e1790 (2012).
28. *Millennium Development Goals Report 2015* (United Nations, 2015).
29. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **8**, 761–763 (2011).
30. Stalder, T. *et al.* Quantitative and qualitative impact of hospital effluent on dissemination of the integron pool. *ISME J.* **8**, 768–777 (2013).
31. Baquero, F., Martinez, J. L. & Canton, R. Antibiotics and antibiotic resistance in water environments. *Curr. Opin. Biotechnol.* **19**, 260–265 (2008).
32. Munck, C. *et al.* Limited dissemination of the wastewater treatment plant core resistome. *Nature Commun.* **6**, 8452 (2015).
33. Li, B., Zhang, T., Xu, Z. & Fang, H. H. Rapid analysis of 21 antibiotics of multiple classes in municipal wastewater using ultra performance liquid chromatography-tandem mass spectrometry. *Anal. Chim. Acta* **645**, 64–72 (2009).
34. Hawkey, P. M. & Jones, A. M. The changing epidemiology of resistance. *J. Antimicrob. Chemother.* **64**, i3–i10 (2009).
35. Huovinen, P., Sundstrom, L., Swedberg, G. & Skold, O. Trimethoprim and sulfonamide resistance. *Antimicrob. Agents Chemother.* **39**, 279–289 (1995).
36. Martinez, J. L., Coque, T. M. & Baquero, F. What is a resistance gene? Ranking risk in resistomes. *Nature Rev. Microbiol.* **13**, 116–123 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the residents of our study communities in El Salvador and Peru for their generosity and trust, without which this study would not have been possible; Epilogos Charities Inc. for on-site logistical support and community networking; the Fundación Luis Edmundo Vásquez (FUNDALV), Universidad Dr. José Matías Delgado, Asociación Benéfica Prisma, and Universidad Peruana Cayetano Heredia for logistical support in the collection and shipment of samples; S. del Pilar Basilio at SEDAPAL in Lima for facilitating access and sample collection at the 'PTAR San Juan' WWTP; J. Hoisington-Lopez at the Center for Genome Sciences and Systems Biology and staff at the Genome Technology Access Center at Washington University School of Medicine for generating Illumina sequencing data; S. Alvarez and staff at the Proteomics & Mass Spectrometry Facility at the Donald Danforth Plant Science Center for mass-spectrometry analyses of water samples; and members of the Dantas laboratory for discussions of the results and analyses. This work is supported in part by awards to G.D. through the Edward Mallinckrodt, Jr. Foundation (Scholar Award), the Children's Discovery Institute (MD-II-2011-117), and the National Institute of General Medical Sciences of the National Institutes of Health (R01-GM099538). Work at the DDPSC was supported by the National Science Foundation (DBI-0521250) for acquisition of the QTRAP LC-MS/MS instrument. E.C.P. is funded by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate (NDSEG) Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author Contributions D.E.B., G.D., M.T.B., and E.C.P. planned the RES study; D.E.B., G.D., R.H.G., and P.T. planned the PST study; M.T.B. and W.H.A. implemented the RES study approval in El Salvador; E.C.P. implemented the RES study approval in the USA; R.H.G. and L.C. implemented the PST study approval in Peru; P.T. implemented the PST study approval in the USA; M.T.B., W.H.A., K.M.N., M.M.B., G.S.S., and E.C.P. collected surveys and samples in RES; P.T., M.C., and L.C. collected samples in PST; E.C.P., M.M.B., G.S.S., and S.P. extracted DNA and generated 16S, functional metagenomic, and shotgun data for RES samples; P.T. and S.P. extracted DNA and generated 16S, functional metagenomic, and shotgun data for PST samples; E.C.P. and P.T. performed analyses and interpreted results; and E.C.P., P.T., and G.D. wrote the paper with input from other co-authors.

Author Information Assembled functional metagenomic contigs and 16S and shotgun metagenomic reads have been deposited to NCBI GenBank and SRA (PRJNA300541). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.D. (dantas@wustl.edu).

METHODS

Site and population overview. The community sampled in Lima (PST) is a 'pueblo joven' (or shantytown) built on the slopes of the desert hills ~15 miles southwest of Lima, Peru (Extended Data Fig. 1a). The area was largely settled in the 1980s; the 2007 census recorded 56,915 inhabitants in an area of ~5 km², a population density approximately four times higher than the rest of the Lima Metropolitan Area. Although accurate estimates of socioeconomic status were not available, the community is considered low income (although not uniformly) with an average family income of ~\$130 per month. Although the majority of the households have access to electricity, water and sewage, many households in the poorer hilltop settlements still lack access to these services. Most households are now linked to a district-wide sewage system that funnels waste from PST and neighbouring communities to the nearby 'San Juan' WWTP, one of 19 treatment plants serving the city of Lima. Sewage influent is collected in large aeration lagoons and subjected to stages of oxidation, settling, and chlorination in subsequent lagoons. Diarrheal diseases are common in this community, in particular among children, with an average of three episodes of diarrhoea per year due to high prevalence of various bacterial, viral and protozoan infectious agents^{37–39}. Commonly used antibiotics are commonly purchased without prescription at the local pharmacies.

The rural Salvadoran village (RES) is home to approximately 100 people and is located in the mountains outside of San Salvador. Most individuals are subsistence farmers, although some are employed outside the village. They also participate in communal small-scale commercial enterprises, such as fish cultivation. Although the community has benefitted from infrastructure improvements, including pre-fabricated houses, drinking water is available only from a sand-filtered communal tap. Drinking water is stored in containers, and washing of dishes and clothing is primarily performed with unfiltered water or rainwater stored in outdoor barrels or reservoirs, which can be visibly contaminated with plant material. Each household has a double-vault composting latrine, a recommended method of waste disposal in low-income areas⁴⁰, but which was available to only ~14% of rural Salvadorans in 2013 (ref. 41). Of the almost half (47%) of all people who live in rural areas worldwide⁴², 16% do not have drinking water sources protected from contamination with human excreta, and 50% lack sanitation facilities that separate excrement from human contact²⁸. Urine is diverted away from the latrine, and wood ash is added to the latrine compartment after each use to increase the internal pH. After the compartment is full, it is sealed off to allow heat, desiccation, and alkalization to destroy faecal microbes. Sterilized waste is then spread onto household agricultural plots as fertilizer, but sterilization may be incomplete⁴³. The village is a two-hour walk from the nearest primary health care centre, but many antibiotics are available for purchase over-the-counter. Depending on the season (rainy or dry), villagers consume a combination of food grown in individual household plots and food purchased from the town. They primarily consume beans, as well as starches such as tortillas and rice, and rarely consume meat. Chickens and cows are the most common domestic animals.

Study design. In the Salvadoran village (RES), the study was explained to the community in a public forum at a preliminary visit, and all members of the community were invited to participate. In the Peruvian community (PST), ten households with a minimum of four members and one child <10 years old were randomly selected and invited to participate in the study. All individuals living in the same household were invited, but were not required, to participate. Written informed consent from each participant was obtained before asking survey questions or faecal sample collection. Both studies conform to the guidelines set forth in the Helsinki Declaration. Prior to initiation, the studies were reviewed and received approval from the National Ethics Committee of El Salvador (Comité Nacional de Ética para la Investigación, Acta number 039-2012), the Institutional Review Board of Asociación Benéfica Prisma in Lima (Act CE0809.12), and The Washington University in St Louis Institutional Review Board (IRB ID numbers 201301049/201206094).

After the initial collection in January 2013, human faecal samples from RES were collected one week, three weeks, and one year later. Human faecal samples from PST were collected between January and May 2012. We sampled 3.6 ± 2.0 individuals per house in RES and 5.7 ± 3.0 in PST. Household environmental samples (soil, water, latrine compartments, and animal faeces from dogs, cows, chickens, guinea pigs) were collected with the permission of the residents. In RES, we sampled soils from washing areas adjacent to each house, adjacent to the latrine compartment where treated waste is removed, from urine-diverting tube exits, and where possible, from agricultural plots and chicken coops, as well as mud from the community pond's edge. We also collected water from the piped source of sand-filtered drinking water, stored drinking water, wash water from outdoor storage barrels, reservoirs, and other containers, and the community pond and its inflow.

Sewage influent (pre-treatment) and effluent (post-treatment) samples from the wastewater treatment plant 'PTAR San Juan', located in the vicinity of PST, were collected in collaboration with Water and Sewage Authority of Lima (SEDAPAL).

Twelve influent and effluent samples were collected between May 2012 and January 2013 to assess the stability of sewage communities over time. We collected influent and effluent from the districts of San Juan de Miraflores (SJM; which PST is part of) and Villa El Salvador (VES; a neighbouring pueblo joven with similar demographic characteristics), which converge in this WWTP and serve a population of ~700,000.

RES demographic survey. In RES, a survey was administered to each study participant at the first three sample collections inquiring about household structure, occupation, diet, and medications, among other topics. Self-reported answers were used to determine participant age, household association, and frequency of travel outside the community, as well as latrine and animal ownership for each household.

DNA extraction. Samples were collected in sterile containers, immediately frozen, and stored at -20 to -80 °C until shipment to Washington University in St Louis, USA where samples were stored at -80 °C until DNA extraction. Metagenomic DNA was extracted from approximately 400 to 600 mg of each faecal and latrine sample with the phenol-chloroform bead-beating protocol described previously⁴⁴. Metagenomic DNA was extracted from soil and latrine samples with high ash content using MO BIO PowerSoil and PowerMax Soil DNA isolation kits. Water samples were filtered with sterile 0.22 µm filters, and metagenomic DNA was extracted from the filter membranes using the MO BIO PowerWater DNA isolation kit. For PST sewage influent and effluent, 50 ml of sample was centrifuged at 10,000 r.p.m. for 10 min; pellets were resuspended in 0.5 ml PBS, transferred to microcentrifuge tubes, and metagenomic DNA was isolated using the phenol-chloroform extraction protocol used for faecal samples. Although extraction method does have an effect in metagenome studies, large differences in community composition (such as those found between different microbial habitats) have a greater influence on variation between samples⁴⁵.

Functional metagenomic selection. Small-insert shotgun expression libraries were created from metagenomic DNA in the vector pZE21 in *E. coli* DH10B as previously described^{4,44}. 68 libraries were created from 51 human faecal and 45 soil, latrine, sewage, and animal faecal metagenomes. Libraries were created for all human faecal samples ($n = 20$) and for soil, latrine, and cow faecal samples from two houses in RES ($n = 14$). House 4 was a family of four adults, and House 6 was a family of two adults and four children where one parent routinely worked outside of the community. Libraries were also created from latrine samples from four additional houses ($n = 4$). Six of the human faecal libraries, one soil library, and one animal faecal library from RES were each created from two pooled metagenomes from the same individual/location at different time points. Libraries were created from human ($n = 31$) and animal ($n = 5$) faecal samples from four houses in PST, as well as pooled sewage treatment plant influent and effluent. Metagenomic DNA from thirteen sewage influent and nine effluent samples were combined into a separate pool for each stage before library creation and selection because of low per-sample DNA yield.

Libraries were screened as previously described on Mueller-Hinton agar containing $50 \mu\text{g ml}^{-1}$ kanamycin and another antibiotic at concentrations inhibitory to *E. coli* DH10B harbouring pZE21 without an insert (Supplementary Table 1). The surviving colonies for each selection were pooled. Metagenomic inserts from each pool were isolated via PCR with vector-specific primers, barcoded, and sequenced in parallel with the Illumina HiSeq 2000 (2×101 or 2×150 bp reads). Reads were demultiplexed by barcode, assembled into contigs with PARFuMS⁴, and annotated with Resfams v1.2 (ref. 23).

A selection was excluded from analysis of antibiotic resistance if: (a) more than 100 contigs were assembled; (b) the number of contigs assembled was more than ten times the number of colonies on the selection plate.

With these criteria, 16 out of 568 selections (2.8%) were excluded. Antibiotic resistance genes were identified by Resfam annotation (Supplementary Table 1). If a core, hand-curated, Resfam annotation specific to the antibiotic class was present on a contig, it was preferentially considered the causative resistance gene. Other Resfam annotations plausible for that antibiotic class were then identified from any contig that did not already contain a resistance gene. With this method, less-specific annotations such as efflux pumps were only identified as resistance genes if they were not co-localized with an antibiotic class-specific, canonical resistance gene.

Lipopolysaccharide modification is a conserved mechanism of antimicrobial peptide resistance in the phylum Bacteroidetes⁴⁶. In the colistin selections, 25 ORFs were annotated as 'PAP2 superfamily' (PF01569.16) and shared 30.4–38.5% global amino acid identity with the *Bacteroides thetaiotaomicron* resistance gene *lpxF* (AAO76961.1)⁴⁶. These open reading frames (ORFs) were also considered antibiotic resistance.

To confirm the function of the ciprofloxacin resistance gene from library 01C_014, the plasmid was isolated from the resistant colony and reintroduced into a susceptible strain of *E. coli*, and the transformed cells were confirmed to grow in liquid and solid media containing $0.5 \mu\text{g ml}^{-1}$ ciprofloxacin at 48 and 72 h

after inoculation, which was not observed for the negative control. The insert was amplified from the plasmid via PCR and Sanger-sequenced from both ends of the pZE21 vector. The combined Sanger sequence was manually trimmed to remove vector sequence, and the resulting contig (1,043 bp) was annotated with Resfams v1.2. The contig and its single ORF were included with the resistant contigs and ORFs identified by annotation above.

A total of 2,075 antibiotic resistance ORFs were identified on 1,955 contigs. The ORFs and the MetaGeneMark-generated protein sequences for each resistance ORF were each clustered at 100% identity over the entire length of the shorter sequence to identify identical sequences, collapsing to 1,245 unique (100% nucleotide identical) ORFs and 1,100 unique (100% amino acid identical) proteins. The contigs were clustered at 90% local identity (cd-hit-est parameters: -c 0.9 -d 0 -r 1 -G 0 -n 8 -uS 0.05 -aS 0.5) to identify different genetic surroundings.

Identification of top hits in NCBI nr. The MetaGeneMark-generated protein sequence for each ORF was compared to NCBI nr (accessed on 15 September 2014) with blastp to identify the top local hit(s). A Needleman-Wunsch alignment was generated with EMBOSS needle for each protein and top hit (default parameters), and the global percent identity was calculated as the number of identities over the length of the shorter sequence.

Whole metagenome shotgun sequencing. Metagenomic DNA was sheared to 300–400 bp, barcoded by sample, and sequenced on an Illumina HiSeq or NextSeq with 2 × 150 bp paired reads. Reads were demultiplexed by barcode with no mismatches, retaining reads whose pair did not contain the same barcode as unpaired reads. Demultiplexed reads were trimmed with Trimmomatic-0.30 to remove Illumina adaptor and low-quality bases (<Q13) from the ends, with default ILLUMINACLIP parameters and a minimum trimmed read length of 36 bases. Paired reads were trimmed in palindrome mode, while single reads were trimmed in simple mode. Human sequences were removed with DeconSeq by mapping to the human reference genome (GRCh38)⁴⁷. Any paired read whose pair was a human sequence was also removed. Samples with fewer than 1 million total reads (paired and unpaired) were excluded from further analysis. 85 of the 98 samples interrogated with functional metagenomics were shotgun sequenced.

Assembly of metagenomes from low-diversity metagenomes. Nine shotgun metagenomes from children <3 years old with 150 OTUs/sample or fewer based on 16S data (see below) were assembled using Velvet (Supplementary Table 3). VelvetOptimiser was run on each sample with hash values from 19 to 141 in steps of 2, with both paired and unpaired reads, using n50 as the optimization function. Assembled contigs were annotated with ResFams²³, and ORFs with core ResFams annotations were included in the ShortBRED markers.

Quantification of antibiotic resistance genes in metagenomes with ShortBRED. ShortBRED²⁴ was used to quantify the abundance of antibiotic resistance genes in the metagenomes. ShortBRED identifies unique marker sequences for clustered proteins that distinguish them from close homologues and maps reads to only those markers. This technique has greater accuracy than mapping to the entire protein, especially for antibiotic resistance genes, many of which evolved from genes performing non-resistance functions in the host.

ShortBRED markers were identified from the antibiotic resistance proteins (1) isolated from the functional selections performed in this study (2,075) (2) identified from the human faecal metagenome assemblies in this study (132), (3) the Comprehensive Antibiotic Resistance Database (CARD) (downloaded 20 October 2014; 2,972 proteins)⁴⁸, and (4) the Lahey β-lactamase database (<http://www.lahey.org/studies/>; 1,145 proteins; one short protein, VEB-6, removed)⁴⁹ (Supplementary Table 4), clustered at 100% identity. The reference database was the modified version of the Integrated Microbial Genomes database, version 3.5., described in ref. 24. ShortBRED produced 2,275 markers when clustered at 100% identity (Supplementary Table 5) and 1,266 markers when clustered at 90% identity. Unless noted, the 100% identity markers were used for all analyses.

We quantified translated antibiotic resistance gene abundance in all metagenomes by mapping paired and unpaired fastq reads to the ShortBRED markers with 99% sequence identity. This extended our resistome investigation to individuals and sites not interrogated using functional metagenomics and to antibiotics that target Gram positive bacteria (for example, vancomycin, macrolides) and are not detectable in functional selections in our Gram negative *E. coli* host⁵⁰. All analyses were performed on marker abundances normalized to reads per kilobase per million reads (RPKM).

For antibiotic resistance proteins from the CARD and Lahey databases, metadata (resistance category, mechanism of action, antibiotic target(s)) was hand-curated from information available on the CARD website. For antibiotic resistance proteins identified through functional metagenomic selections and shotgun assemblies, resistance category and mechanism of action were assigned based on Resfams annotation. Antibiotic target(s) for the former were the antibiotics to which they conferred resistance in our functional selections, while the latter were

not assigned an antibiotic target. Annotations for ShortBRED markers were drawn from all constituent proteins.

Comparison of human faecal resistomes to published cohorts. For comparison to the resistomes in ref. 25, whole metagenome shotgun reads were downloaded from SRA (accession PRJNA268964). Fastq reads were trimmed with Trimmomatic in simple mode using the same parameters as for the RES and PST reads. Samples with fewer than 1 million total reads and individuals <3 years old were excluded. ShortBRED markers were quantified as above. For ref. 25, the average read length of the paired reads for each sample was specified as the average read length during ShortBRED quantification. Only ShortBRED markers that included proteins from the CARD and Lahey databases were considered for this analysis to avoid bias towards our cohorts.

16S gene V4 amplification, sequencing, and preprocessing. The 16S gene V4 region (515–806) was amplified using the original Earth Microbiome Project (EMP) protocol (<http://www.earthmicrobiome.org/emp-standard-protocols/16s/>) or with Takara Taq DNA polymerase premix, with barcoded primers designed in ref. 51, from 228 human faecal and 203 environmental samples. 245 samples interrogated by whole metagenome shotgun sequencing were also 16S sequenced. Barcoded amplicons were pooled and sequenced on an Illumina MiSeq with 2 × 250 bp paired-end reads.

Barcoded Illumina reads were demultiplexed with QIIME version 1.8.0, *split_libraries_fastq.py*⁵². Paired reads were truncated at the first base with quality score ≤ Q3 and merged using usearch⁵³, requiring 100% identity in the overlap region and a merged length of 253 bp ± 5 bp. Merged reads were filtered with QIIME to remove reads with three or more contiguous bases with quality score ≤ Q20.

Open OTU picking with UPARSE for RES and PST comparisons. OTUs were picked from all 12,797,788 merged and filtered RES and PST reads with the UPARSE pipeline⁵⁴. Singletons were excluded from OTU picking, and a reference-based chimera check against the GOLD database (downloaded 1 September 2013) was performed on OTUs as recommended. Reads were assigned to OTUs at 97% identity using usearch. Representative sequences from each OTU were assigned taxonomy with uclust against the Greengenes database (release 13_8, 97% clusters), aligned, and used to create a phylogenetic tree using QIIME. 19,301 OTUs were picked across all samples. Biom tables were rarefied to 7,000 sequences per sample, which excluded eight samples.

Comparison of phylogenetic composition of human faecal samples to published cohorts. Raw 16S amplicons from ref. 19 were downloaded from MG-RAST (accession number qiime:850). 16S amplicons from faecal samples from ref. 14 were downloaded from the European Nucleotide Archive (ENA) (ERP008799). 16S amplicons from ref. 25 were provided by the authors. For all studies, reads were generated with primers F515/R806 on an Illumina platform. We classified each cohort's lifestyle as described in ref. 25. Only samples from individuals 3 years or older were included in analysis. Individuals labelled with an adult-specific keyword (Family Member: "Mother", "Father", or Sample Identifier: "adlt") in ref. 19 supplementary table 2 were also included. Samples from ref. 19 were rarefied to 50,000 reads each to reduce computational load. The merged reads from RES and PST human samples and the reads from ref. 14 were truncated at 101 bp.

OTUs were picked from all reads with the reference-based protocol described in ref. 19 against the Greengenes database (13_8, 97% clusters). Reverse strand alignment was permitted, and new clusters were suppressed. The OTU table was rarefied to 5,000 sequences per sample, which excluded two samples from ref. 14.

Sample filters. To better characterize within- and between-sample diversity, we limited both the shotgun and 16S analysis to a single sample per human individual except where noted. Infant faecal microbiota undergo large and frequent shifts in microbial composition before stabilizing into an adult-like configuration around three years of age¹⁹. Therefore, except where noted, we limited all analyses with human samples to those from individuals 3 years old or older.

Alpha diversity. For 16S, biom tables were rarefied ten times to 7,000 sequences per sample, and equitability, observed species, and Faith's phylogenetic distance were calculated on all rarefactions and averaged using the QIIME script *alpha_diversity.py*. For resistome, tables were not rarefied, and only the number of ARGs present in each sample was compared. Alpha diversity was compared between categories using non-parametric Student's *t*-tests with 999 permutations at a depth of 7,000 sequences per sample with Bonferroni correction for multiple hypothesis testing using the QIIME script *compare_alpha_diversity.py*. For Extended Data Fig. 4g, the total antibiotic resistance RPKM for each sample was summed, multiplied by one million, and divided by the total number of reads for that sample to confirm that the results were robust to sequencing depth.

Beta diversity. Unweighted UniFrac, weighted UniFrac, Sørensen-Dice, and Bray-Curtis dissimilarity matrices were calculated from biom tables with the QIIME script *beta_diversity.py*, using the Greengenes 13_8 97% phylogenetic tree (for closed reference OTU picking) and the phylogenetic tree generated during

de novo OTU picking for UniFrac distances. Principal coordinates analysis was performed with the QIIME script *principal_coordinates.py*.

ANOSIM and adonis analyses were performed on dissimilarity matrices with the QIIME script *compare_categories.py* or in R (vegan package functions *anosim()* and *adonis()*) with 999 permutations.

Distance to centroid was calculated in R with the vegan package function *betadist()*, and significance was tested with a permutation test (*permutest()*).

Average pairwise dissimilarities between categories were compared with non-parametric Student's *t*-tests with 999 Monte Carlo permutations with Bonferroni correction using code adapted from the QIIME script *make_distance_boxplots.py*. For shared antibiotic resistance proteins, the number of antibiotic resistance proteins shared between each pair of samples was calculated with *shared_phylotypes.py*, and significance was calculated as above.

Procrustes. The 16S OTU table rarefied to 7,000 sequences/sample and the antibiotic resistance abundance table were both filtered to include only samples interrogated with both methods ($n = 172$). Bray–Curtis dissimilarity matrices were calculated for both, principal coordinates analysis was performed, and procrustes analysis was performed on the PCoA results (QIIME script *transform_coordinate_matrices.py*) across 172 dimensions with 999 permutations to determine significance.

Identification of discriminatory taxa using LEfSe. Taxa summaries were created from the rarefied OTU table and filtered to levels L1 to L5 (kingdom – family). LEfSe⁵⁵ was used to identify taxa that were discriminative between categories. Alpha was 0.05 for both the Kruskal–Wallis and pairwise Wilcoxon rank-sum tests. The Linear Discriminant Analysis (LDA) effect size threshold was set at 3.0 or 4.0, and all-against-all comparison was performed.

Identification of discriminatory antibiotic resistance functions. The antibiotic resistance table was filtered to only antibiotic resistance proteins detected in the sample set under consideration and summarized by each metadata category with the QIIME script *summarize_taxa.py*. Absolute abundances rather than relative abundances were calculated, because antibiotic resistance genes represent a small fraction of the total metagenome. Supervised learning with the Random Forests classifier was performed to identify antibiotic resistance categories that most discriminated between sample categories (*supervised_learning.py*, 500 trees, and tenfold cross-validation). The top discriminatory antibiotic resistance categories (as determined by feature importance scores) enriched in subsets of the samples were plotted onto the principal coordinates of all samples, using the abundance-weighted average for the coordinates. The size of the point is proportional to the overall abundance across all samples under consideration (1/100 RPKM).

SourceTracker. We estimated the proportion of each latrine, soil, and water microbiota attributable to RES human faeces and the proportion of each sewage microbiota attributable to PST human faeces using SourceTracker²⁹. Antibiotic resistance tables were filtered to remove any samples and markers without any observations. SourceTracker was run through QIIME with default settings using human faecal microbiota as source. Jigger was introduced in boxplots to allow visualization, but some samples with very low abundance are not distinguishable.

Mass spectrometry-based detection of antibiotics in WWTP samples. Solid phase extraction and ultra-performance liquid chromatography–tandem mass spectrometry (UPLC–MS/MS) were used to detect the presence of 16 antibiotics from 7 classes in 11 influent and 11 effluent sample supernatants using a modified protocol³³. Fifty ml of influent or effluent sample were centrifuged at 10,000 r.p.m. for 10 min; supernatants were concentrated using solid phase extraction cartridges (6cc, 200 mg, Waters, Milford, MA). Samples were analysed on a 4000 QTRAP LC/MS/MS instrument (AB Sciex). The data was normalized based on the internal standard (isotopically labelled caffeine, 13C3, Cambridge Isotope Laboratories) to account for experimental variation and antibiotic extraction/ionization efficiency. A mixture of the antibiotic standards was also processed along with the samples as positive controls to test for recovery. Standards for amoxicillin, cefotaxime, cefoxitin, chloramphenicol, ciprofloxacin, clindamycin, erythromycin, penicillin G, sulfamethoxazole, tetracycline, and trimethoprim were purchased from Sigma-Aldrich (St Louis, MO). Standards for azithromycin, aztreonam, cefepime, doxycycline, and tigecycline were purchased from AK Scientific (Union City, CA). All analyses were performed at the Proteomics and Mass Spectrometry Facility at the Donald Danforth Plant Science Center (St Louis, MO).

Network creation. Antibiotic resistance gene networks were generated based on ShortBRED results of human and environmental metagenomes, filtering hits with

RPKM value of <0.1 to determine sharing across samples. Graphics were generated in Cytoscape 3.2.1 using an edge-weighted spring-embedded layout.

Mobile genetic elements (MGEs). Putative MGEs were identified from the functional selection contigs by Pfam and TIGRfam annotation (Supplementary Table 13). Antibiotic-resistant ORFs were considered co-localized with an MGE if they shared a contig with a MGE ORF. 365 (19%) antibiotic resistance contigs contained at least one of 236 unique MGEs (100% amino acid identical).

Multidrug resistance clusters. Multidrug resistance clusters were identified as contigs containing multiple antibiotic resistance proteins identified from our selections and/or annotated with a core Resfams profile HMM (ref. 23), regardless of the antibiotic used in the selection. 200 (10%) contigs contained multidrug resistance clusters.

Mobilome analysis. For ‘mobilome’ analyses, the ShortBRED markers were restricted to those that included an antibiotic resistance protein encoded in our functional metagenomic selection contigs adjacent to a putative MGE.

Phylogenetic classification of contigs with PhyloPythia. Antibiotic-resistant contigs and metagenomic assembly contigs were classified by PhyloPythia^{56,57} using the recommended model (800 genera, 2013).

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

37. Checkley, W. *et al.* Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community. *Lancet* **363**, 112–118 (2004).
38. Cooper, M. A. *et al.* Molecular analysis of household transmission of *Giardia lamblia* in a region of high endemicity in Peru. *J. Infect. Dis.* **202**, 1713–1721 (2010).
39. Oswald, W. E. *et al.* Fecal contamination of drinking water within peri-urban households, Lima, Peru. *Am. J. Trop. Med. Hyg.* **77**, 699–704 (2007).
40. Water Sanitation and Health Unit Organization (http://www.who.int/water_sanitation_health/sanitproblems/en/index4.html) (World Health Organization, 2002).
41. WHO / UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation. *El Salvador: estimates on the use of water sources and sanitation facilities (1980–2015)*. (2015).
42. The World Bank Group. *Data: Topics: Urban Development* (<http://data.worldbank.org/topic/urban-development>) (2015).
43. Corrales, L. F., Izurieta, R. & Moe, C. L. Association between intestinal parasitic infections and type of sanitation system in rural El Salvador. *Trop. Med. & Int. Health* **11**, 1821–1831 (2006).
44. Moore, A. M. *et al.* Pediatric fecal microbiota harbor diverse and novel antibiotic resistance genes. *PLoS ONE* **8**, e78822 (2013).
45. Lozupone, C. A. *et al.* Meta-analyses of studies of the human microbiota. *Genome Res.* **23**, 1704–1714 (2013).
46. Cullen, T. W. *et al.* Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. *Science* **347**, 170–175 (2015).
47. Schmieder, R. & Edwards, R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* **6**, e17288 (2011).
48. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
49. Bush, K., Palzkill, T. & Jacoby, G. <http://www.lahey.org/studies/> (Lahey Clinic, 2015).
50. Pehrsson, E. C., Forsberg, K. J., Gibson, M. K., Ahmadi, S. & Dantas, G. Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front. Microbiol.* **4**, 145 (2013).
51. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
52. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
53. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
54. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods* **10**, 996–998 (2013).
55. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
56. McHardy, A. C., Martin, H. G., Tsirigos, A., Hugenholtz, P. & Rigoutsos, I. Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* **4**, 63–72 (2007).
57. Patil, K. R., Roune, L. & McHardy, A. C. The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS ONE* **7**, e38581 (2012).

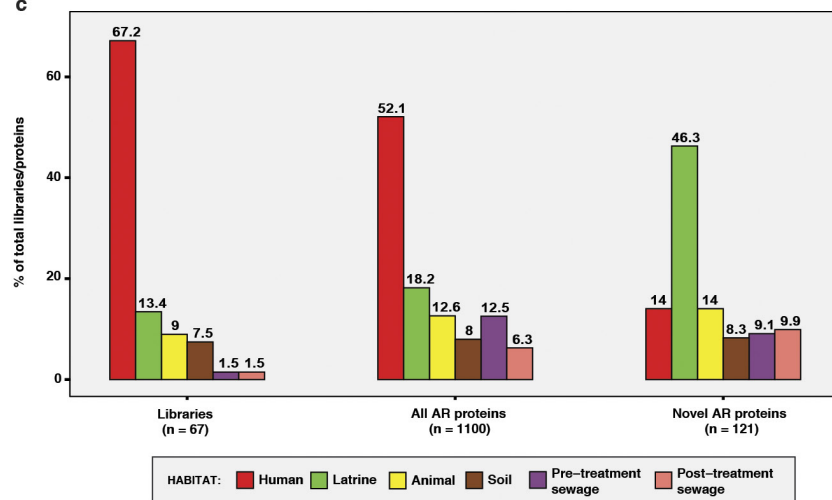
a



b

	Samples interrogated with 16S seq.	OTUs	Samples interrogated with shotgun seq.	AR genes
Human	226	1641	191	584
Post-treatment sewage	13	2220	7	82
Pre-treatment sewage	30	2324	27	514
Animal	14	2409	10	283
Water	22	3025	4	8
Latrine	36	3245	16	242
Soil	84	12503	30	206
Total	425	19301	285	797

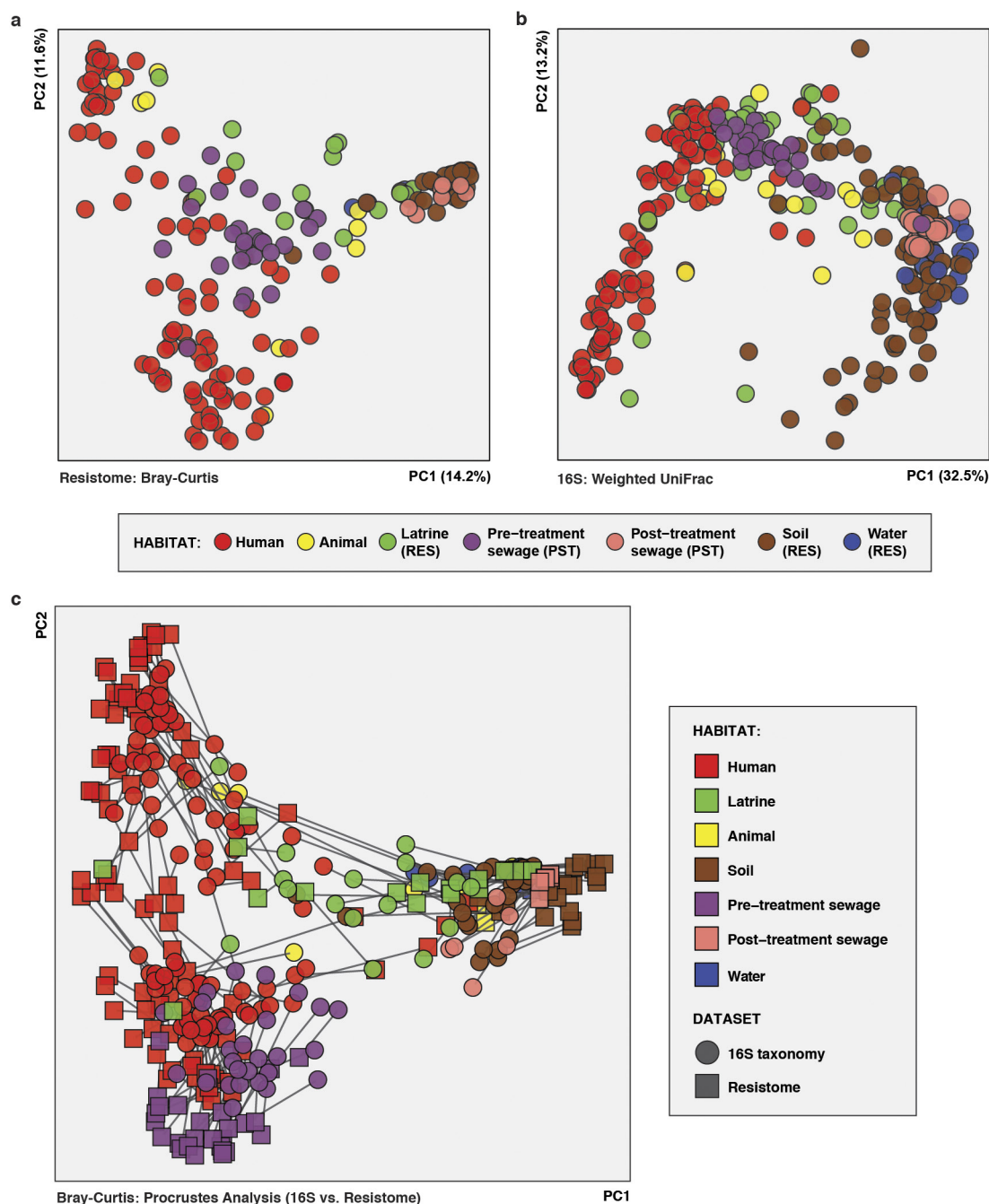
c



Extended Data Figure 1 | Overview of study and methods. a, Location and overview of study sites in El Salvador and Peru. RES photographs by the authors G.S.S. and M.M.B., PST photographs by the author P.T.

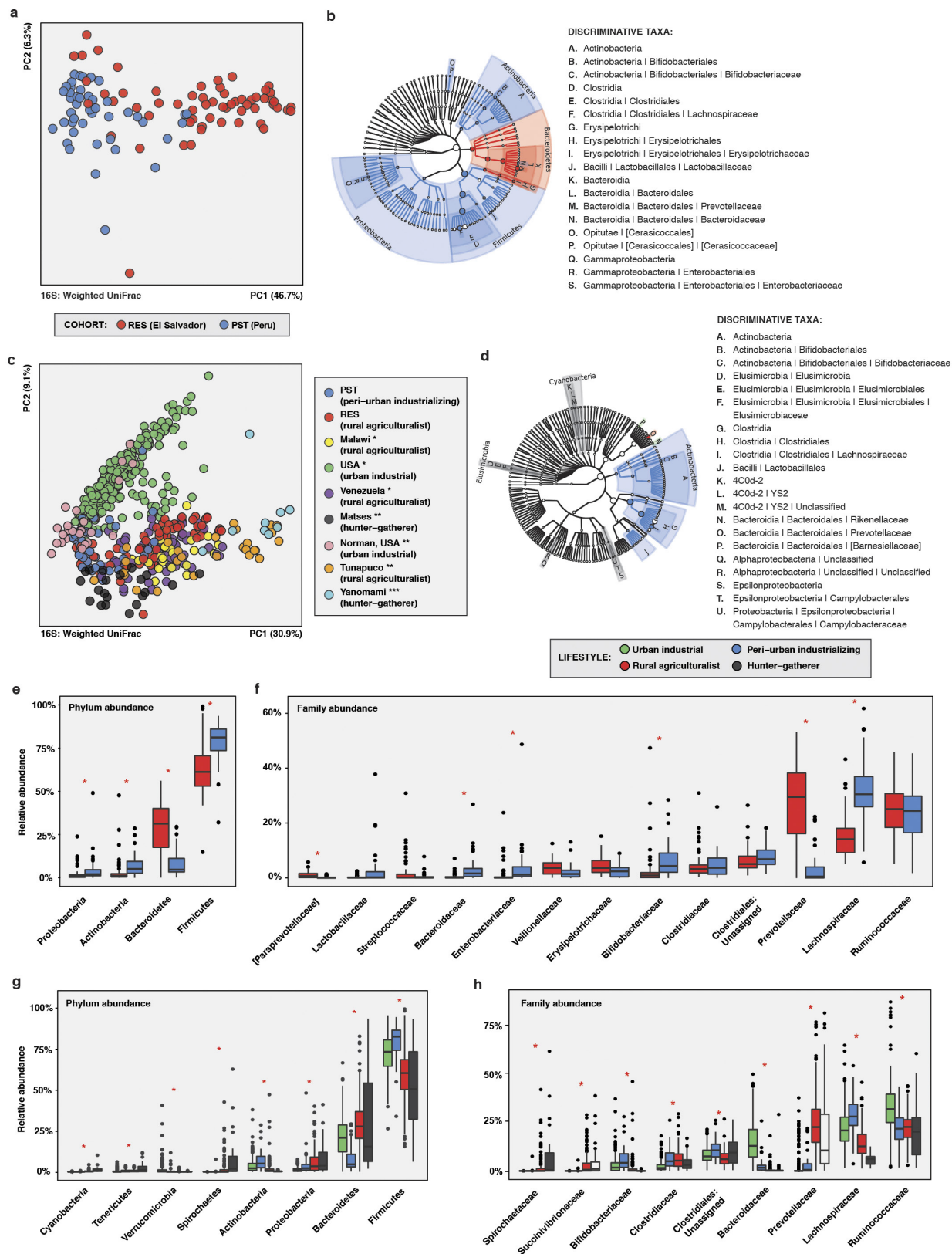
b, Antibiotic resistance markers and OTUs detected vs number of samples interrogated by whole metagenome and 16S sequencing by habitat in RES and PST. c, Proportion of metagenomic libraries ($n = 67$), all antibiotic resistance proteins identified from functional metagenomic selections ($n = 1,100$), and novel antibiotic resistance proteins identified from

functional metagenomic selections ($n = 121$) originating from each microbial habitat. The percent of total libraries/proteins in that category originating from each microbial habitat is listed above the bar. For all antibiotic resistance markers and novel antibiotic resistance proteins, the total sums to $>100\%$ due to proteins identified in more than one habitat. The number of novel antibiotic resistance proteins vs libraries screened was significantly different than expected compared to the total for human and latrines (chi-squared test, $P < 0.005$).



Extended Data Figure 2 | Human faecal and environmental microbiota from RES and PST. Microbiota are coloured by habitat. **a**, PCoA of Bray-Curtis distances between resistomes. ($n = 86$, $n = 10$, $n = 16$, $n = 30$, $n = 4$, $n = 27$ and $n = 7$ for human, animal, latrine, soil, water, pre-treatment sewage and post-treatment sewage, respectively) Adonis $R^2 = 22.4\%$, $P < 0.001$. **b**, PCoA of weighted UniFrac distances between

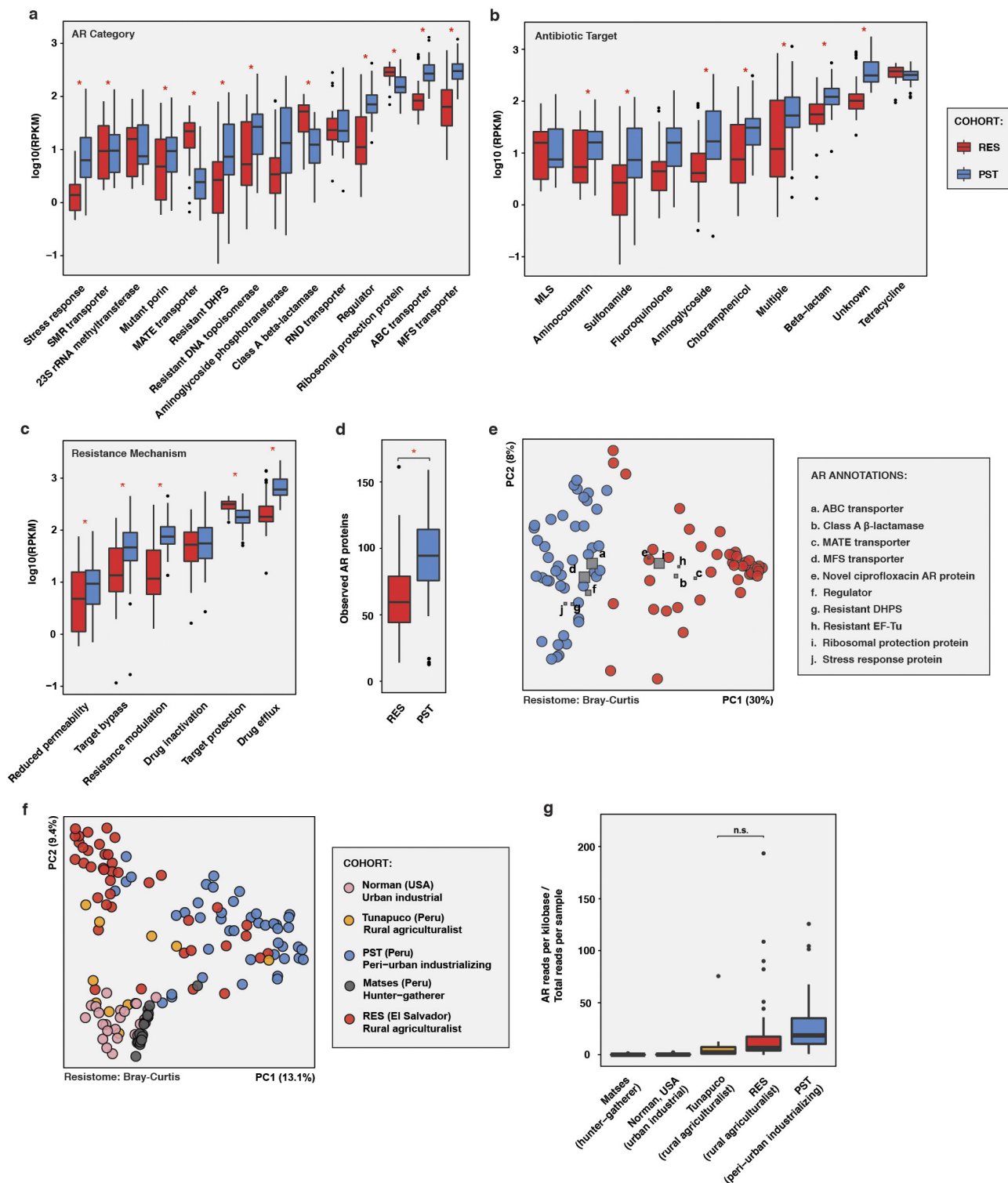
microbiota. ($n = 105$, $n = 14$, $n = 36$, $n = 84$, $n = 22$, $n = 30$ and $n = 13$ for human, animal, latrine, soil, water, pre-treatment sewage and post-treatment sewage, respectively.) Adonis $R^2 = 41.9\%$, $P < 0.001$. **c**, Procrustes transformation of taxonomic composition vs resistome. Only samples interrogated with both methods were included ($n = 172$). $M^2 = 0.360$, $P < 0.001$ (172 dimensions, 999 permutations).



Extended Data Figure 3 | See next page for caption.

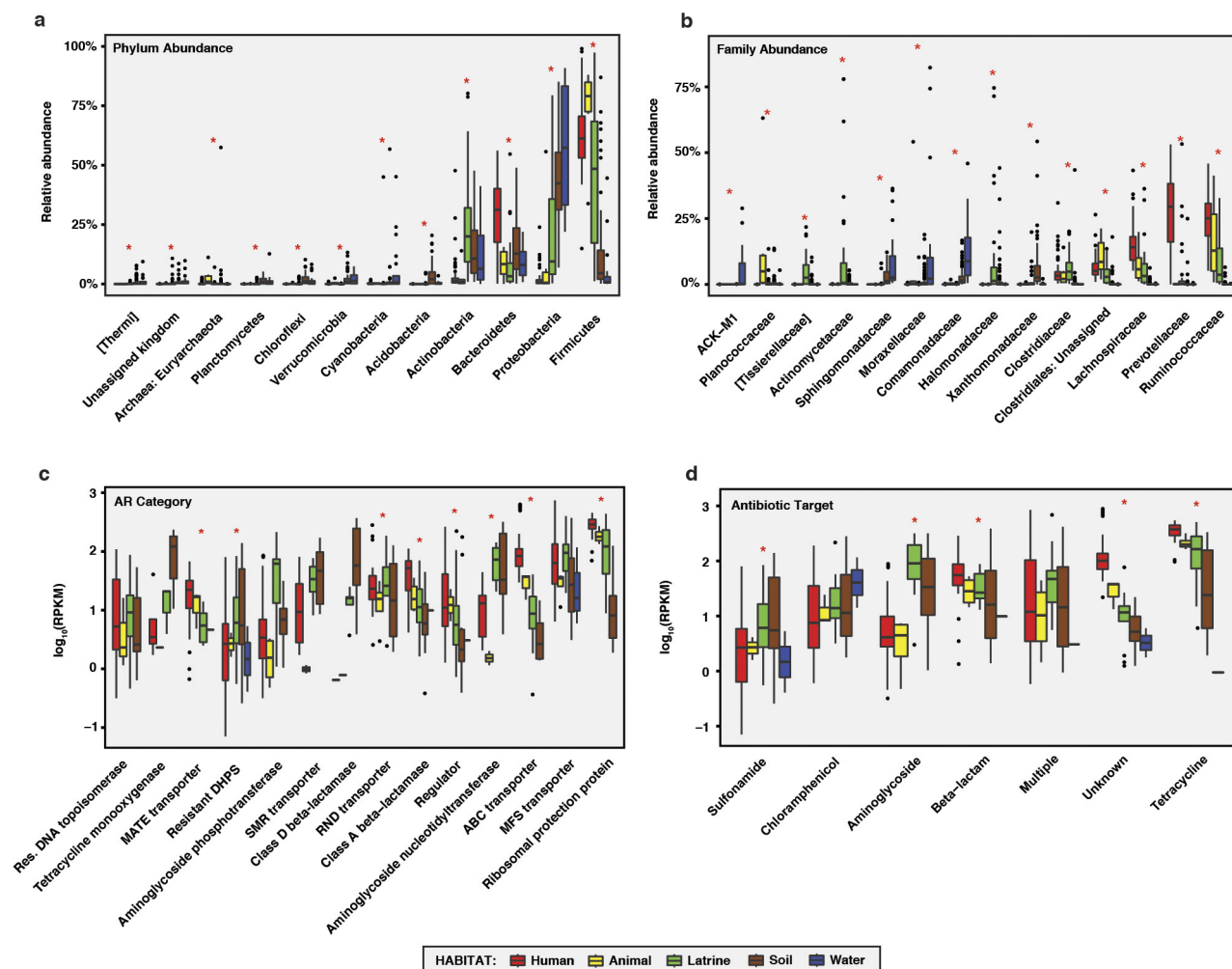
Extended Data Figure 3 | Phylogenetic composition of RES and PST human faecal microbiota and published microbiota from previous studies^{14,19,25}. **a, b, e, f**, RES vs PST. (RES $n = 60$, PST $n = 45$) **c, d, g, h**, RES and PST vs published human microbiota. (RES $n = 60$, PST $n = 46$, other $n = 446$; see Supplementary Table 14) **a**, PCoA of weighted UniFrac distances between RES and PST human faecal microbiota, coloured by cohort. Adonis $R^2 = 29.7\%$, $P < 0.001$. **b**, Taxa discriminating between RES and PST human faecal microbiota as determined by LEfSe. The phylogenetic tree includes all kingdom- to family-level taxa present in any sample. Coloured taxa are discriminative between cohorts and have an LDA effect size of ≥ 4.0 ; they are coloured by the cohort in which they have the highest abundance. Circle size is relative to the highest abundance in either cohort. **c**, PCoA of weighted UniFrac distances between RES and PST human faecal microbiota and published human faecal microbiota,

coloured by cohort. Cohorts are labelled by lifestyle and study (* (ref. 19), ** (ref. 35), *** (ref. 14)). Adonis $R^2 = 37.6\%$, $P < 0.001$. **d**, Taxa discriminating between host lifestyles for RES and PST and published human faecal microbiota as determined by LEfSe, effect size threshold 3.0. Discriminative taxa are coloured by the host lifestyle in which they are most abundant. **e, f**, Relative abundances of microbial phyla (**e**) and families (**f**) in human faecal microbiota from RES and PST. * $P < 0.05$, Wilcoxon test with Bonferroni correction. **g, h**, Relative abundances of microbial phyla (**g**) and families (**h**) in human faecal microbiota from RES and PST and published human faecal microbiota, by lifestyle. * $P < 0.05$, Kruskal–Wallis test with Bonferroni correction. **e–h**, Only taxa with a mean relative abundance of $\geq 1\%$ in one cohort/lifestyle are shown. Taxa are in order of increasing overall mean relative abundance. Error bars, s.d.; centre bars, median.



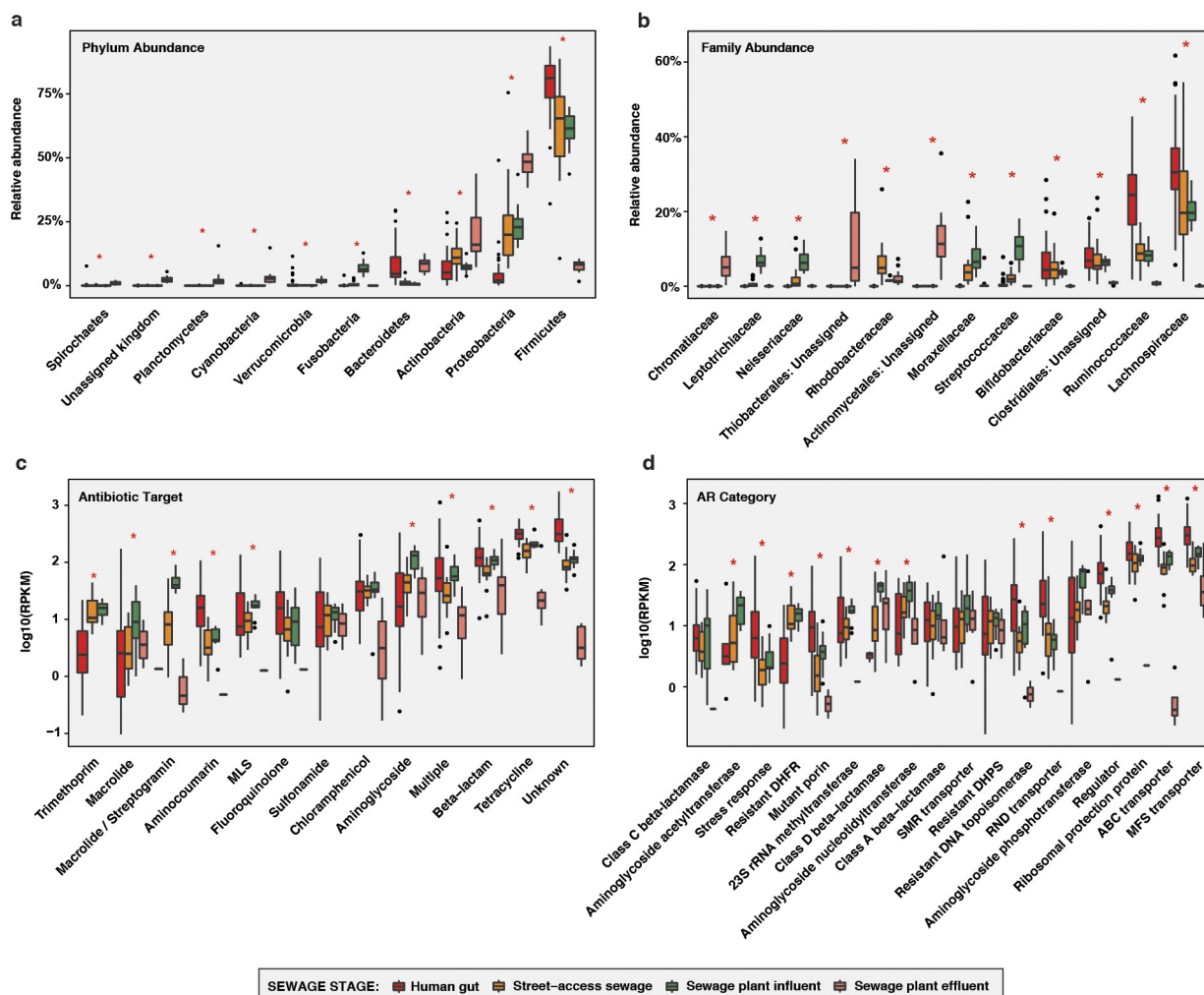
Extended Data Figure 4 | RES and PST human faecal resistomes and comparison to the published data sets from ref. 25. **a–e**, RES and PST resistomes, coloured by cohort. (RES $n = 42$, PST $n = 44$) **f**, **g**, RES and PST vs published human data sets, coloured by cohort. (RES $n = 42$, PST $n = 44$, other $n = 53$; see Supplementary Table 15) **a–c**, Absolute abundances of antibiotic resistance categories (**a**), antibiotic targets (**b**), and mechanisms of action (**c**) in human faecal resistomes from RES and PST. Only categories with a mean RPKM of >10 in one cohort are shown. Categories are in increasing order of overall mean absolute abundance. Abundances are plotted in \log_{10} scale. $*P < 0.05$, Wilcoxon test with Bonferroni correction. **d**, Number of antibiotic resistance proteins per RES and PST human faecal resistome. $*P < 0.05$, non-parametric Student's t -tests. **e**, PCoA of Bray–Curtis distances between RES and PST resistomes,

with abundance-weighted coordinates of the top five most discriminative antibiotic resistance categories enriched in each cohort (squares, size proportional to overall abundance). Adonis $R^2 = 25.0\%$, $P < 0.001$. **f**, PCoA of Bray–Curtis distances between human faecal resistomes from RES and PST and ref. 25. Adonis $R^2 = 19.7\%$, $P < 0.001$. **g**, Total reads mapping to antibiotic resistance markers per person (normalized by marker length) normalized by the total reads in that sample in RES and PST and published human faecal microbiota, by cohort. Includes both paired and unpaired reads. The overall distribution of normalized antibiotic resistance read depth was significantly different than expected (Kruskal–Wallis, $P < 1 \times 10^{-15}$). n.s., not significant. All other comparisons are $P < 0.05$, Wilcoxon test with Bonferroni correction. **a–d**, **f**, Error bars, s.d.; centre bars, median.



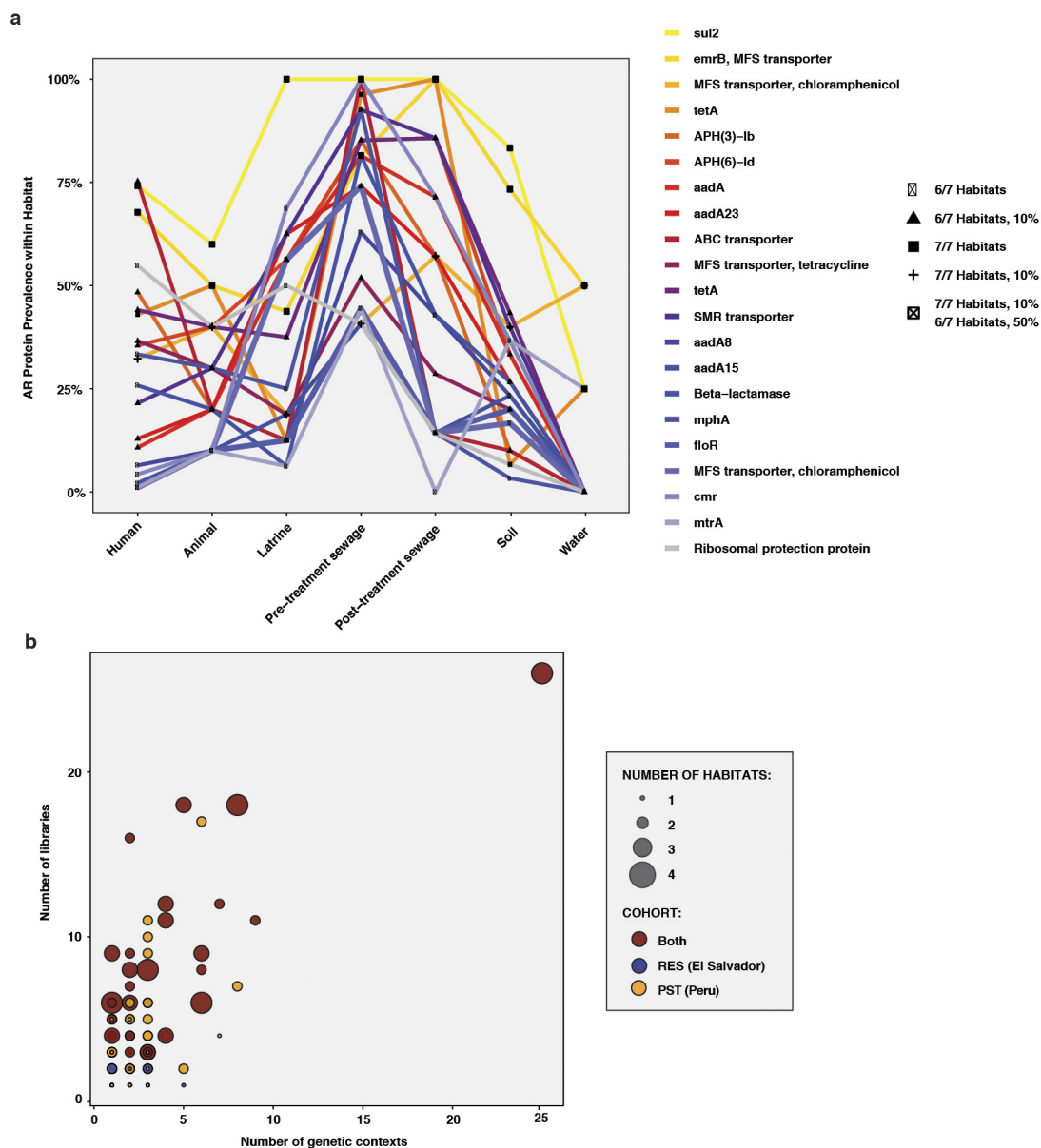
Extended Data Figure 5 | RES human faecal and environmental microbiota and resistomes. **a, b**, Relative abundances of microbial phyla (**a**) and families (**b**) in RES microbiota, by habitat. ($n = 60$, $n = 6$, $n = 36$, $n = 84$ and $n = 22$ for human, animal, latrine, soil, water, respectively) Only taxa with a mean relative abundance of $\geq 1\%$ in one habitat are shown. Taxa are in increasing order of overall mean relative abundance. $*P < 0.05$, Kruskal–Wallis test with Bonferroni correction. **c, d**, Absolute

abundances of antibiotic resistance categories (**c**) and antibiotic targets (**d**) in RES resistomes, by habitat. ($n = 42$, $n = 4$, $n = 16$, $n = 30$ and $n = 4$ for human, animal, latrine, soil, water, respectively). Only categories with a mean RPKM of > 10 in one habitat are shown. Categories are in increasing order of overall mean absolute abundance. Abundances are plotted in log₁₀ scale. $*P < 0.05$, Kruskal–Wallis test with Bonferroni correction. **a–d**, Error bars, s.d.; centre bars, median.



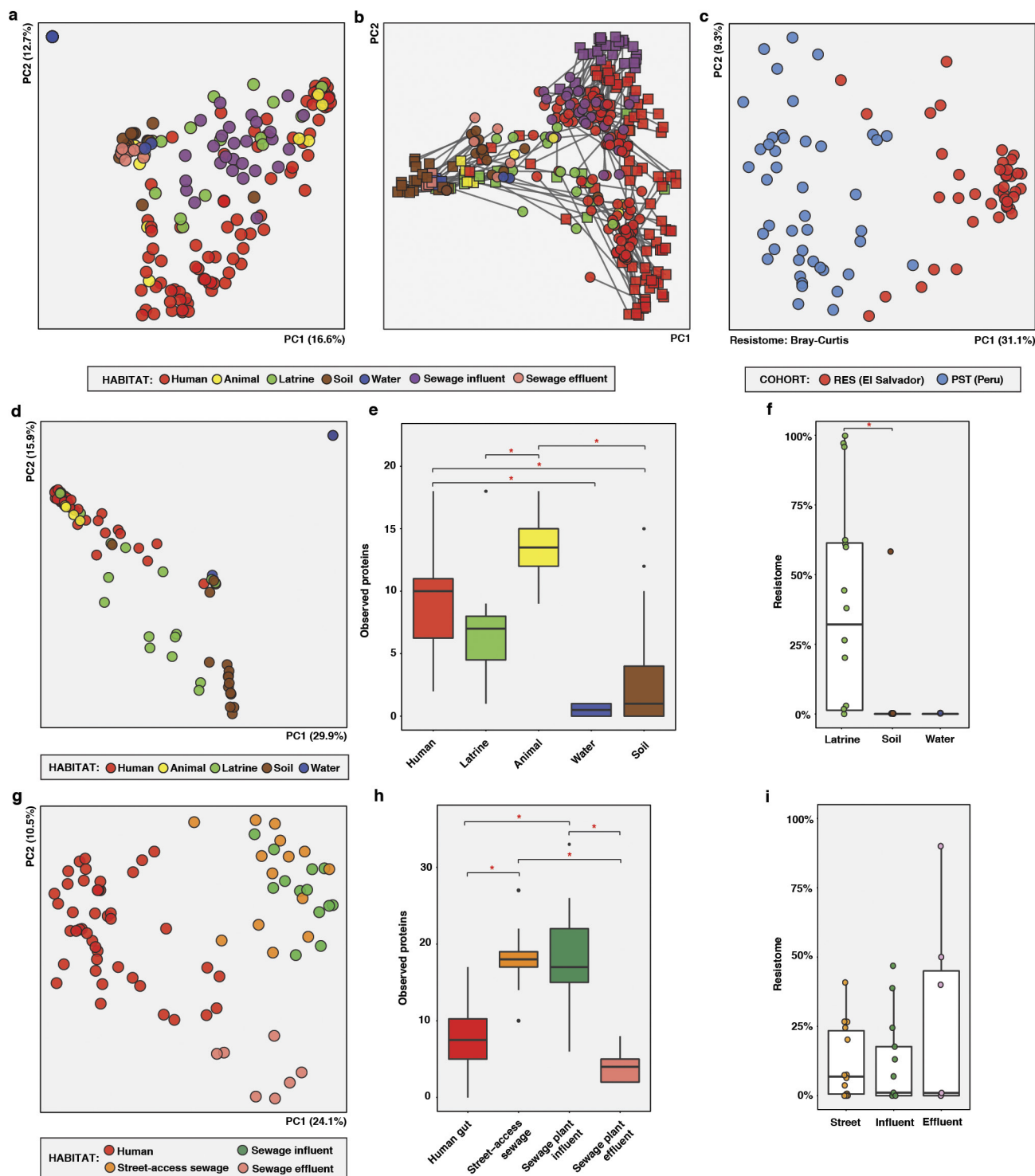
Extended Data Figure 6 | PST human faecal and environmental microbiota and resistomes. **a**, **b**, Relative abundances of microbial phyla (**a**) and families (**b**) in human faecal and sewage microbiota from PST, by stage. ($n = 45$, $n = 16$, $n = 14$ and $n = 13$ for human, street-access, influent and effluent, respectively) Only taxa with a mean relative abundance of $\geq 1\%$ in one stage are shown. Taxa are in increasing order of overall mean relative abundance. $*P < 0.05$, Kruskal–Wallis test with Bonferroni correction. **c**, **d**, Absolute abundances of antibiotic resistance categories

(**c**) and antibiotic targets (**d**) in PST resistomes, by stage. ($n = 44$, $n = 14$, $n = 13$ and $n = 7$ for human, street-access, influent and effluent, respectively). Only categories with a mean RPKM of > 10 in one stage are shown. Categories are in increasing order of overall mean absolute abundance. Abundances are plotted in \log_{10} scale. $*P < 0.05$, Kruskal–Wallis test with Bonferroni correction. **a–d**, Error bars, s.d.; centre bars, median.



Extended Data Figure 7 | Antibiotic resistance gene sharing across habitats. **a**, Highly cosmopolitan antibiotic resistance proteins. The prevalence of each antibiotic resistance protein in metagenomes from each microbial habitat is depicted for all proteins detected in six of the seven habitats ($n = 21$). Detection was based on ShortBRED quantification of the protein in each metagenome. Prevalences for an antibiotic resistance protein are linked by lines of the same colour. The shape of each point reflects the number of habitats in which it was found, as well as the minimum prevalence within each habitat. The legend lists the annotation for each protein. **b**, Protein sequences of antibiotic resistance genes isolated from functional metagenomic selections were clustered

at 100% amino acid identity, and the number of metagenomic libraries, microbial habitats (for example, human faecal, soil), and cohorts in which each unique protein ($n = 1,100$) was encoded were calculated across all members of the cluster. Antibiotic resistance contigs ($n = 1,955$) were clustered at 90% local identity to identify different genetic contexts, and the number of genetic contexts in which each unique protein was encoded was calculated across all contigs encoding a protein in that cluster. Spearman's $\rho = 0.59$, $P < 2.2 \times 10^{-16}$, number of genetic contexts vs libraries; $\rho = 0.47$, $P < 2.2 \times 10^{-16}$, number of genetic contexts vs. habitats; Wilcoxon test, $P < 2.2 \times 10^{-16}$, number of genetic contexts vs cohorts (one or both).



Extended Data Figure 8 | Mobilome analyses. **a**, PCoA of Bray-Curtis distances between RES and PST human and environmental resistomes, coloured by habitat. (n = 86, n = 10, n = 16, n = 30, n = 4, n = 27 and n = 7 for human, animal, latrine, soil, water, pre-treatment sewage and post-treatment sewage, respectively) Adonis $R^2 = 24.1\%$, $P < 0.001$. **b**, Procrustes transformation of taxonomic composition vs resistome. Only samples interrogated with both methods were included (n = 172). $M^2 = 0.493$, $P < 0.001$ (172 dimensions, 999 permutations). **c**, PCoA of Bray-Curtis distances between RES (n = 42) and PST (n = 44) resistomes, coloured by cohort. Adonis $R^2 = 31.0\%$, $P < 0.001$. **d**–**f**, RES human faecal and environmental microbiota and resistomes, coloured by habitat. (n = 42, n = 4, n = 16, n = 30 and n = 4 for human, animal, latrine, soil, water, respectively). **d**, PCoA of Bray-Curtis distances between resistomes.

Adonis $R^2 = 32.0\%$, $P < 0.001$. **e**, Observed antibiotic resistance proteins.

* $P < 0.05$, non-parametric Student's *t*-tests, Bonferroni correction.

f, Percentage of latrine, soil, and water resistomes attributable to human faeces, as determined by SourceTracker²⁹. * $P < 0.05$, pairwise Wilcoxon tests, Bonferroni correction. **g**–**i**, PST human faecal and sewage microbiota and resistomes, coloured by stage. (n = 44, n = 14, n = 13 and n = 7 for human, street-access, influent, effluent, respectively). **g**, PCoA of Bray-Curtis distances between resistomes. Adonis $R^2 = 34.8\%$, $P < 0.001$. **h**, Observed antibiotic resistance proteins. * $P < 0.05$, non-parametric Student's *t*-tests, Bonferroni correction. **i**, Percentage of sewage resistomes attributable to human faeces at each sewage treatment stage, as determined by SourceTracker. * $P < 0.05$, pairwise Wilcoxon tests, Bonferroni correction. Error bars, s.d.; centre bars, median.