Supplementary Materials and Methods:

Overview of dataset

Stool samples (266 total, 252 after quality control, see below) from 207 volunteers as part of the MetaHIT (85 Danish and 39 Spanish samples) (HMP Consortium 2012), HMP (139 US American samples) (Qin et al. 2010), and twin gut microbiome studies (3 US samples) (Schloissnig et al. 2012) have previously been sequenced using Illumina technology. See the referenced works for database accessions to this sequence data. We included only samples with read lengths of 45 nt or higher, and 14 Solexa sequenced MetaHIT samples were therefore not used. We additionally included stool samples (14 total) from US American and Malawian individuals that were sequenced using 454 pyrosequencing technology (Yatsunenko et al. 2012) and 28 samples from Danish, Spanish, Italian, French, and Japanese individuals (Arumugam et al. 2011; Kurokawa et al. 2007) (see Supplementary Table S4) that have previously been Sanger sequenced, respectively. Any samples from non-adult donors (younger than 18 years) or samples with less than 50Mbp of sequence after processing were excluded, as was one Italian sample where the species composition suggested pathogen infection, making it non-representative for healthy gut microbiota. While one of the source datasets also included Venezuelan samples (Yatsunenko et al. 2012), fewer than 4 of these were large enough to be included and as such, they were not used in this comparison.

Reference genomes and construction of a gut metagenomic reference gene catalog

Prokaryotic genome sequences were downloaded from the NCBI genomes database on 27 March 2012 and annotated with eggNOG (version 3) (Powell et al. 2012) to identify orthologous groups including a set of 40 universal single copy marker genes (Ciccarelli et al 2006, Sorek et al. 2007). Complete genomes with fewer than 30 of these marker genes and any genome with more than 500 contigs were removed to yield a set of 3496 reference genomes (Supplementary Table S5). Raw sequencing reads from the 252 metagenomic samples that have previously been sequenced using the Illumina platform were processed using the MOCAT software package (Kultima et al. 2012) to generate a gut metagenomic reference gene catalog. Briefly, raw sequencing reads were quality trimmed and filtered using the FastX software (http://hannonlab.cshl.edu/fastx_toolkit) with a base quality and read length cut-off of 20 and 30, respectively. The resulting high-quality reads were assembled into scaftigs using SOAPdenovo 1.05 (Li et al. 2010) and genes predicted on scaftigs longer than 500 nt using MetaGeneMark v1.0 (Zhu et al. 2010). All predicted genes were clustered using CD-HIT v4.5.447 at 95% sequence identity with 90% of the shorter sequence required to be covered in pair-wise alignments, yielding a catalog of 3,661,047 non-redundant gut metagenomic reference genes.

Metagenomic abundance estimation

A dataset of 139 HMP samples, 110 MetaHIT samples, and 3 additional samples from US individuals were mapped against the metagenomic reference gene catalog using the MOCAT package (Kultima et al. 2012), and gene abundances estimated by calculating base coverages per site (Supplementary Table S6). Similarly, taxonomic abundances for the species in the reference

genome collection were determined by mapping metagenomic reads to a set of 40 universal, single-copy marker genes (Ciccarelli et al. 2006; Sorek et al. 2009) using MOCAT, requiring 95% sequence identity for species-level abundance estimates and 90% sequence identity for genus-level abundance estimates (Supplementary Table S8). Effectively, since the marker genes are assumed to exist in single copy, their abundances are directly proportional to species abundances, and proportional to the amount of sequence from each species there is in each sample by a factor that is the inverse ratio of the concatenated marker gene lengths to the genome length as a whole. This allowed estimation of how much sequence material in each sample came from each species. Downsampling was carried out for the concatenated marker gene base counts in the same way as for other gene base counts (see below), in order to make taxonomic composition estimates comparable between samples with different sequencing depth.

Processing of Sanger and 454 reads

We used control datasets that were generated using the Sanger/454 pyrosequencing platforms (Supplementary Table S4), which generate much longer reads than Illumina does. To allow for consistent application of the same analysis pipeline, as well as to make the comparison fair, these longer reads were split into non-overlapping fragments, generally 75 nt (the average read length within the 252 Illumina samples), but always at least 40 nt in length, discarding any remaining fragments. Following this, analysis proceeded identically to the Illumina-based analysis.

Gene abundance thresholds/noise filter

There is a risk of spurious matches when mapping reads to the gene catalog, and also of spurious matches in a biological sense, such as matches to gene material from a bacterium the donor has ingested, but which is not actually resident in the gut. Therefore, all gene base counts that were smaller than 10% (cut-off chosen based on previous in-house benchmarking) of the length of the reference gene sequence were discarded, and all other gene base counts were reduced by this amount to avoid any systematic bias from this procedure. However, this step had only a minor impact on the analysis of the Illumina data (data not shown).

Sequence stochastic downsampling

The Illumina samples used vary much in sequencing depth, ranging from ~10Gbp for the largest HMP samples to ~726Mbp for the smallest MetaHIT sample. Due both to these differences between the source studies and in order to make comparisons between samples fair, gene and taxa abundances (base counts) were stochastically downsampled. This was done by comparing the amount of sequence in a sample to a downsampling target size, yielding the fraction of reads to be randomly retained or discarded in order to achieve the same approximate sequencing depth. Gene and taxa abundances (i.e. concatenated marker gene base counts, which were subsequently used to determine taxa base counts) were thus stochastically downsampled according to this sample-specific probability, by discarding blocks of bases of the same size as the average read length for that sample.

This procedure can reduce the base count of a gene to zero, which corresponds to a case where the signal would have been lost due to the small size of the dataset. We performed 100 repeated downsamplings for the three Illumina country datasets from 726Mbp to 50Mbp to test the impact of this procedure on the observed country differences (Supplementary Figure S17). For several antibiotic classes or subclasses, the median resistance potential over samples falls to zero for all country datasets under these conditions, reducing the number of antibiotics for which we have the resolution to observe country differences. However, the country subsets remain the same in relative degree of resistance potential as in the larger dataset.

Similarly, downsampling this dataset from 726Mbp to 50Mbp, then repeatedly subsampling only 4 samples for each country (1000 iterations), show that the country differences we observe are robust also to reductions in sample size (Supplementary Figure S18). Downsampling further to 10Mbp or to only 3 samples per country does not eliminate the country differences either, but does reduce resolution even further (data not shown).

Gene family-wise coverage measurement, definition of a gene being fully covered

In some cases, a resistance gene (each distinct gene symbol listed in the ARDB was considered a separate resistance gene/resistance gene family) is represented by only a single reference gene catalog entry, whereas in others, it is represented by several sequences. A gene was here considered fully covered when its per-site base coverage (total number of bases divided by gene length) was >= 1. To integrate this for a family of homologous genes, we computed its effective gene length as the base count-weighted average of the length of each of its member genes. The total number of bases mapped to any of the family members was counted, and the sum was divided by the effective gene length to yield the composite per-site base coverage for that resistance gene family.

Identifying antibiotic resistance genes

Protein sequences from the reference set provided for the ARDB antibiotic resistance gene database were used as queries to search for homologs within the proteomes of the reference genome collection (requiring BLASTP E < $1e^{-5}$, and additionally either sequence identity above the family-specific threshold suggested in ARDB for at least 80% of the length of either sequence, or annotation with the same gene symbol). The matches (Supplementary Table S9), together with the ARDB reference protein collection sequences, were used as queries to search the gut metagenomic reference gene catalog using the same inclusion criteria. Additionally, any potential resistance gene homolog in the gut metagenomic reference gene catalog was searched using BLAST both against the resistance genes previously compiled and the NCBI non-redundant protein collection (nr), and rejected if the top annotated hit contradicted the family assignment. This resulted in a final curated set of resistance gene family matches in the metagenomic reference gene catalog (Supplementary Table S10, Supplementary Table S11).

Determining the taxonomic distribution of each resistance gene family

To minimize the risk of bias due to uncommon species composition in a sample, we

constructed a map of the distribution of resistance gene homologs across bacteria, including plasmid genomes. For each resistance gene family, the taxonomic distribution (with respect to the 3496 species represented in the reference genome collection) was determined from the results of the initial search for resistance gene homologs, and supplemented by curated information that was text-mined from the ARDB website, as well as by searching a representative sequence for each resistance homolog from among the gut metagenomic reference proteins against the NCBI non-redundant protein database (requiring BLASTP $E < 1e^{-10}$ and alignment along at least 95% of the length of either sequence). Supplementary Table S12 lists the distribution of each resistance gene family across the species represented in the reference genome collection, determined using the above procedure.

Curation of resistance gene annotations from the ARDB database

The resistance genes that we analyzed were mostly imported directly from the ARDB database (Liu and Pop 2009), but were given additional curation as follows. While the ARDB distinguishes some specific subclasses of cephalosporins (e.g. first generation, second generation…), annotation of gene products differed widely in whether specific or general such terms were used and all "cephalosporin" references were therefore collapsed to just a single class (including third-generation cephalosporins, but not cephamycins). The ARDB contains resistance genes for both specific compounds and for broader classes. In the country difference analysis, if both a specific compound (which could be considered a very narrow antibiotic class, encompassing all drugs which use that compound) and its more general class was analyzed, only the more general class was tested for differences unless the dataset also contained at least one resistance gene that was explicitly annotated as providing resistance to the more specific subclass or individual compound. The more general class was then represented as a 'generic' member, listed as e.g. "miscellaneous beta-lactam". This was done so that parallel tests of country difference hypotheses would not be hierarchically dependent, which is achieved since the 'generic' antibiotics then are distinct from the specific class member compounds included. Even so, the large overlap between 'generic' and specific class members, as well as between antibiotics where there is cross-resistance, means that parallel tests are not fully independent. To account for this, multiple testing correction was performed using a method shown to control the false discovery rate even under arbitrary dependence structures between hypotheses (Benjamini and Yekutieli 2001).

Some antibiotic classes (anion and cation antibiotics) were considered too vaguely defined and were discarded from analysis of effects of animal use approval or time in use; they were however included when computing overall resistance potentials. Some genes were listed as being multidrug efflux pumps but had no associated antibiotic range, these were included as providing resistance to a proxy "multidrug" antibiotic class, but were excluded from the usage/age-based analysis as, again, the class was too vaguely defined to allow definite statements on age or usage approval.

For a small number of ARDB gene families, resistance was not associated with presence of the gene, but rather with its mutation or loss. These genes, which included penicillin-binding proteins and the *KsgA* kasugamycin resistance determinant, were also excluded.

The ARDB also includes some resistance annotation for compounds that are not widely

considered antibiotics (the dye acriflavin, bile salt, tetrarsonium chloride, fusaric acid and the anticancer drug doxorubicin). These annotations were excluded, but doing so did not change the set of resistance genes found in the gut reference gene catalog.

Incorporating cross-resistances

Wherever a gene was annotated as providing resistance to a broad class of antibiotics, it was considered a resistance gene for all such class members as well. Those antibiotics that are considered members of more general classes for this purpose are listed in Supplementary Table S1. For the special classes mentioned above (the broad categories of anion or cation antibiotic resistance, or multidrug efflux pump with no range specified), these descriptions from the ARDB were considered too vague to allow genes annotated with them to contribute to cross-resistance to other antibiotic classes or subclasses. See Supplementary Figure S1 for an illustration of how resistance genes, broad and narrow antibiotic classes or subclasses, and single antibiotics are linked, using the beta-lactams as an example.

Defining antibiotic age and usage

For each antibiotic, the first reliable publication describing it was noted (provided as either direct or indirect references in Supplementary Table S1). The age of a more general class was taken as the age of its oldest member. Cross-resistance could be seen as increasing the effective age of an antibiotic; therefore an adjusted age was assigned which is the highest age of any member of the class. For animal use approval, the FDA Green Book (Shields 2009) was downloaded (in May 2012) and searched for products including each antibiotic as an active ingredient, or including compounds that were considered analogs of that antibiotic. A more general class was considered to be approved for animal use if any of its member antibiotics were, whereas specific antibiotics were considered to have analogs in animal use if any member of their more general class, or the more general class as a whole, were approved for animal use.

Definition of the taxonomy-adjusted antibiotic resistance potential

We defined the antibiotic resistance potential of a metagenomic sample as the amount of resistance gene sequence it contained relative to the amount of sequence it contained from any species that may (but need not necessarily) carry such genes (the set of potential carrier species). While considering the ratio between these two quantities would provide an intuitive measure of this type, biases could result from differences between samples in sequencing depth or taxonomic composition, particularly in the case of very small abundances of applicable species. To avoid this, the pooled resistance gene base counts were subjected to a secondary stochastic downsampling, separately for each antibiotic. This was done by randomly discarding short read-sized blocks of sequence in proportion to how much sequence from carrier species the sample contained, relative to a common “target number” for each antibiotic. This “target number” was set as the smallest number of bases mapping to carrier species from any sample that i) was no smaller than 1% of the highest corresponding value across the samples and ii) was no smaller than 0.1% of the total sample size. Samples with carrier species sequence totals falling below this target for a given antibiotic were excluded for that particular comparison. For

most antibiotics where significant country differences were found, virtually all samples met these inclusion criteria. The ratio of the downsampled resistance gene base count to this "target number" of bases from potentially resistance-carrying species was taken as the final resistance potential. Results were similar if raw ratios were used instead of downsampled ones (Supplementary Figure S21). By measuring resistance gene abundances relative to the pooled abundances of potential carrier species, this taxonomy-adjusted resistance potential aims to broadly disentangle, to the extent that current metagenomic data allows, the effects of selection towards higher relative prevalence of resistance determinants from the effects of broad changes in gut microbial species composition. Simulations show that a method not considering taxonomic composition in this manner will falsely conclude country differences, whereas our resistance potential analysis will not (Supplementary Table S13). The term "resistance potential", rather than simply "resistance", was chosen to reflect how differences in resistance gene expression and regulation also affect phenotypic resistance in vivo, in addition to the role the resistome plays.

Integrating gene abundances by antibiotic

Most of the present study was carried out on the level of antibiotics (either specific compounds or 'generic' representatives of broader classes, see above), by summing the abundances of all resistance genes active against each antibiotic, rather than on the level of individual genes. This was motivated by the fact that we searched for the effects of selection pressures caused by the presence of antibiotics, which could act through increasing the presence of any of several potential resistance genes for a particular antibiotic. Therefore, our power to detect such an effect should be higher for the integrated gene abundances than for the abundances of individual genes. Supplementary Figure S2 shows medians and 25%-75% quartiles for taxonomy-adjusted resistance potential calculated for each resistance gene found in the gut dataset. The significant country differences found when integrating all genes providing resistance to each antibiotic are reflected in country differences at a single gene level as well, though the trend is less clear. This may be interpreted that selection for resistance to an antibiotic can result in different adaptations in different cases.

This diversity is also seen when decomposing the resistome into genes that provide resistance by different mechanisms, i.e. by degrading antibiotics, by protecting their microbial targets, or by removing them from the cell through export (Figure 2B).

Controlling for number of resistance gene families per antibiotic

Antibiotics approved for animal use tend to have more known resistance gene families (see Supplementary Table S1). To test whether the overrepresentation of resistance to animal-use antibiotics is seen independently of the number of known resistance gene families for each antibiotic, we measured the average sequencing coverage/position (by summing together base counts and dividing by the total length of the set of genes) across all genes that contribute to resistance to antibiotics falling into each animal use approval or age category. The distribution of this measure across the 252 samples (Figure 1B) is significantly different (Kruskal-Wallis $P<2.2e^{-16}$) for antibiotics used in animals and their analogs than those that are not. Similarly, it is significantly different across age categories (Figure 1C) (Kruskal-Wallis $P<2.2e^{-16}$). To verify that these findings are not artifacts of differences in numbers of genes in each category, we

repeated the above analysis ten times using random selections of only ten genes for each age or animal use approval category each time. Even with this degree of subsampling, the differences between categories remain stable (Supplementary Figures S4-5).

Controlling for confounding factors and testing for country differences

Metadata for potential confounding factors (gender, inflammatory disease status, enterotype (Arumugam et al. 2011), and overweight status) were available for the MetaHIT and HMP US samples. To determine enterotypes for samples, the procedure described at http://enterotype.embl.de/ was followed, using genus-level taxonomic abundance data and setting the number of clusters to 3. The cluster assignment for each sample was taken as its enterotype.

For each antibiotic, resistance potential was represented as proportion of bases encoding resistance genes out of bases mapping to potential carrier species (i.e. bacterial species potentially carrying any of these resistance genes), following a secondary downsampling to a common taxonomic target step, as described above. These proportions were modeled using mixed-effects linear models with a binomial error model, using the R (R Development Core Team 2012) *lme4* (Bates and Maechler 2009) and *languageR* (Baayen 2009) packages. Sample identity was included as a random intercept component in all models. As a first step, a full model, containing country of origin and all potential confounders (gender, inflammatory disease status, enterotype and overweight status (BMI > 25)) was compared against a null model with one confounder removed, for every combination of antibiotic and confounding factor. Significant influence of the confounder was concluded if a likelihood ratio test of the fit of data to the model showed significantly better fit to the model containing the confounder, at Benjamini-Hochberg-Yekutieli (Benjamini and Yekutieli 2001) FDR <= 0.05. Enterotype was the only confounder that significantly improved model fit, therefore the final test for country differences compared a model with enterotype and country as independent variables to a model with country omitted. As in the first step, significant country differences were concluded if the model containing country yielded significantly better fit in a likelihood ratio test of the two models, at Benjamini-Hochberg-Yekutieli FDR <= 0.05.

To illustrate the influence of the confounding factors, we also determined resistance potentials separately for the samples stratified into subsets based on levels of the confounding factors. We plot these sub-analyses, with results shown for every antibiotic class where country differences were concluded in the main analysis, in Supplementary Figures S8-11, along with the results of a similar analysis where US HMP samples were stratified based on which city they were sampled from (Supplementary Figure S12). This was done to ensure that regional variability, where we could determine it, was not large enough to compromise country-level conclusions.

Resistance isolates from slaughterhouses

For the fraction of positive resistance tests data shown in Figure 2C, bottom left panel, the data in de Jong et al. (2012) was used. For each bacterial species considered in that study, the isolates were considered separately for each tested antibiotic and added together, yielding a total fraction of antibiotic resistance tests, regardless of antibiotic or species, in which resistance was

concluded.

Time series analysis

To investigate the persistence of resistance determinants over time, taxonomy-adjusted resistance gene profiles were computed for the US HMP samples. For each of 99 resistance gene families detected in the gut samples, we considered the quantity $\mathbf{M}_{species}$ of bases sequenced from any of the potential carrier species. The lowest $\mathbf{M}_{species}$ across samples was taken for each gene, excluding any samples with $\mathbf{M}_{species}$ smaller than 1% of the largest $\mathbf{M}_{species}$ for that gene, and samples with $\mathbf{M}_{species}$ smaller than 0.1% of the total sample size. For any sample so excluded for a resistance gene family, the relative abundance of that resistance gene was considered missing, as it could not be reliably determined. For the remaining samples, the relative resistance gene abundance was computed by secondary downsampling, randomly reducing it by read-sized blocks by a proportion relative to how much larger $\mathbf{M}_{species}$ for that sample was than that of the smallest included sample. For each gene in a sample not considered missing, then, the profile contains its relative abundance under equivalent taxonomic composition as measured through $\mathbf{M}_{species}$. These resistance gene profiles were then compared across samples using Kendall τ rank correlation, with genes excluded in any pairwise comparison where their taxonomically adjusted abundance was missing in either profile.

There were 88 samples from 43 individuals taken at two or three different time points, resulting in 47 pairwise comparisons, for each of which a time interval between samplings was defined. We compared the resistance gene profile similarity for each such pair with the average similarity of the two samples to the other 137 HMP samples (Figure 3, green and grey markers), finding that same-donor pair resistance gene profiles are on average more similar to each other in this respect than they are to those from other donors, and that this higher similarity holds during at least a year.

Comparison of different sequencing platforms

Three datasets, Spanish (MetaHIT, N=39), Danish (MetaHIT, N=71) and American (US) (HMP+, N=142), were previously Illumina sequenced. Of the others for which resistance potentials are shown in Figure 2C, the Japanese (kurokawa07, N=7), French (MicroObes, N=8) and Italian (MicroAge, N=5) were Sanger sequenced, whereas the Malawi (Gordon, N=4) were sequenced using 454 technology. Median resistance potentials for three additional datasets are also shown as dotted lines in Figure 2C: American (Gordon, N=10) sequenced using 454 technology, Danish (MetaHIT, N=4) and Spanish (MetaHIT, N=4) sequenced using Sanger technology. Sanger/454 samples, with longer read lengths before processing, appear to detect higher resistance potentials than Illumina samples for the same countries, implying somewhat higher sensitivity. Where Illumina and Sanger/454 data are both available for a country, however, they are broadly in agreement.

Testing for country differences using non-downsampled data

To verify that downsampling of the data does not cause artificial country differences in

resistance potential to appear, we tested each antibiotic class for which such differences were concluded from the 726Mbp downsampled data, but using the original, non-downsampled gene and taxa abundances. Resistance potentials for non-downsampled data were calculated by dividing the number of bases mapped to resistance determinants for each antibiotic by the total number of bases mapped to potential carrier species. To test non-downsampled resistance potentials for country differences while taking into account differences in taxonomic composition and sequencing depth, we modeled the proportions of resistance-encoding bases as a mixed-effect linear model using the R (R Development Core Team 2012) *lme4* (Bates and Maechler 2009) package. The proportion was taken as the dependent variable, and modeled with country as an independent variable and sample as a random intercept component. A binomial error model was chosen to allow for scale-aware proportion modeling. Corresponding null models omitting the country independent variable were likewise constructed, and significant country differences were concluded where comparison of null and alternative models under χ^2 likelihood ratio tests yielded false discovery rates <= 5% (Benjamini-Hochberg-Yekutieli procedure). Comparisons with the results from downsampled data are shown in Supplementary Figure S19.

Fermented foods and antibiotic resistance gene transfer

Earlier studies have reported that food products common in the Mediterranean region, such as aged cheeses, cured sausages or olive fermentations, may contain high quantities of bacteria, such as enterococci (Franz et al. 2011), due to the fermentation processes involved in their production. Resistance was found in enterococci isolated from French cheeses to tetracycline, chloramphenicol, erythromycin, kanamycin and minocycline (Jamet et al. 2012). Transfer of resistance elements for erythromycin and tetracycline from enterococci to other genera was experimentally shown to occur during sausage production (Gazzola et al. 2012), and a study of Portuguese fermented meat products found strains resistant to chloramphenicol, ciprofloxacin, erythromycin, tetracycline, rifampicin and nitrofurantoin (Barbosa et al. 2009). For all of these antibiotics, where we detected resistance genes, the median resistance potential was highest in the Spanish samples (Figure 2A). This may imply that transfer of resistance determinants from farm animals to the human gut is more likely in regions where fermented meat and dairy products are common, which is consistent with the country differences we observe.

Modeling resistance potential as a function of veterinary antibiotic consumption

Antibiotic sales figures for veterinary use in 2009 for Denmark were taken from the ESVAC project report (European Medicines Agency 2011), with corresponding Spanish figures taken from the Spanish department of agriculture (http://www.aemps.gob.es/en/informa/notasInformativas/medicamentosVeterinarios/2011/docs/ventas-antimicrobianos_Espana-2009.pdf; Sep 2012, date last accessed), for eleven antibiotic classes. The Spanish data does not separate sulfonamide and trimethoprim, as these are frequently sold in mixture. For these two antibiotics, then, the combined total exposure was used. Following the procedure of Grave et al. (2010), we calculated effective treated livestock biomass by totaling weight at slaughter and adding estimated total weight of dairy cows, using data

acquired from the EuroStat database (http://epp.eurostat.ec.europa.eu/portal/page/portal/agriculture/data/main_tables; 17 Sep 2012, date last accessed). These data are shown in Supplementary Table S2, and compared with median resistance potentials across the 39 Spanish and 71 Danish samples, averaged over antibiotics in each broad class, in Supplementary Figure S20.

Mixed-effect modeling was done using the R (R Development Core Team 2012) *lme4* (Bates and Maechler 2009) and *languageR* (Baayen 2009) packages, with median resistance potential for each included antibiotic across country samples as the dependent variable. Values were centered at the mean (subtraction of the average Spanish and Danish value) for each antibiotic, separately for antibiotic exposure and resistance potential, due to the large differences in scale between different antibiotics. However, results remained highly similar also when this step was omitted (data not shown). Country and antibiotic class were included in the model as random intercept components, whereas biomass-normalized antibiotic sales (tons sold per ton of treated livestock weight) were included as fixed effect components. Inspection of residuals against fitted values revealed no significant trends, indicating that assumptions of normality and homogeneity hold sufficiently well that use of a linear mixed effects model is justified. A null model using only country and antibiotic class as random intercepts was likewise fitted. The nested model taking biomass-normalized antibiotic sales data into account as an independent variable has significantly better fit (Likelihood ratio test, $\chi^2 = 6.0514$, df = 1, P = 0.0139). This model was evaluated using MCMC sampling, showing a significant positive effect of biomass-normalized veterinary antibiotic sales on antibiotic-specific resistance potential (P (MCMC) = 0.0254, P (t-test) = 0.0223).

Expected effect of country differences in gut microbiome taxonomic composition

There are large differences between the country datasets in average species composition, particularly such that the HMP US samples stand out. For 45 of 54 tested antibiotics, there is a significant (Kruskal-Wallis test, Benjamini-Hochberg-Yekutieli FDR <= 5%) difference between the country subsets with respect to fraction of the samples made up of potential resistance determinant carrier species (Supplementary Table S13, second column. Whether this reflects differences between European and American gut microbiota or merely technical variation is not presently clear. However, it can cause apparent differences in antibiotic resistance gene carriage unless accounted for.

To test the extent to which this factor impacts our tests for country differences in resistance potential, we simulated resistance gene abundance data in the presence of taxonomic composition differences between samples and subsets of samples, but using the same model for all three countries for the relationship between gene and taxa abundances. For each resistance gene family found in the gut reference gene catalog, its coverage was computed for the set of 252 Illumina samples. Likewise the total coverage of potential carrier species was computed for each gene and sample.

A linear regression model $\mathbf{M}_{gene}$ was built for every resistance gene family, modeling resistance gene (per-site) coverage as a function of species (per-site) coverage, and requiring the intercept to be zero. Only samples with pooled carrier species coverage above 1X was used for training the model. If no such samples existed, the resistance gene family was treated as absent in

the simulated samples, with a coverage of zero. The model $\mathbf{M}_{gene}$ was used to generate simulated gene coverages based on the original species coverages for each of the 252 samples, corresponding to a situation where the original taxonomic composition is kept but where relative density of resistance determinants is forced to be the same in all three countries. Following this step, a noise term, sampled from a Gaussian distribution with standard deviation equal to the mean residual when fitting the original gene coverages to predictions from the model $\mathbf{M}_{gene}$, was added to the simulated coverage values, and any resulting negative coverage values were set to zero. The resulting gene coverages were divided evenly in case of multiple reference gene catalog entries for each family, then multiplied by the corresponding gene lengths to produce simulated gene abundances. The resulting data approximately corresponds to a situation where there are no differences in relative resistance gene carriage between the countries, but where differences in taxonomic composition, biological or technical, remain. Since this is a null hypothesis for a test of country differences in relative resistance potential, any apparent significant difference then becomes a false positive inference for the approach used.

The same model comparison likelihood ratio test for significant resistance potential country differences as for the main 726Mbp dataset was performed on this simulated dataset. This test, calculating the resistance potential using secondary downsampling to a common taxonomic target, reports no significant country differences (Supplementary Table S13, fourth column), indicating that this test is relatively robust to taxonomic composition shifts, at least on the scale tested for here. In contrast, the same type of test applied using the “raw” proportion of resistance gene sequence out of total amount sequenced, and without including enterotype as a model parameter, reports falsely significant (Benjamini-Hochberg-Yekutieli FDR < 5%) country differences for a majority of antibiotics in the simulated dataset (Supplementary Table 13, third column), revealing this as a relevant error source unless it is controlled for.

To put this similarity into perspective, taxonomic composition profiles, measured as number of bases from each sample mapped to marker genes for each bacterial genus in the reference genome collection, were likewise compared using Kendall τ rank correlations and shown in Figure 3 (red markers) as well for the 47 pairwise comparisons of same-donor samples. These correlations were generally lower than the resistance profile correlations, consistent with expectations that resistance adaptations persist over time.

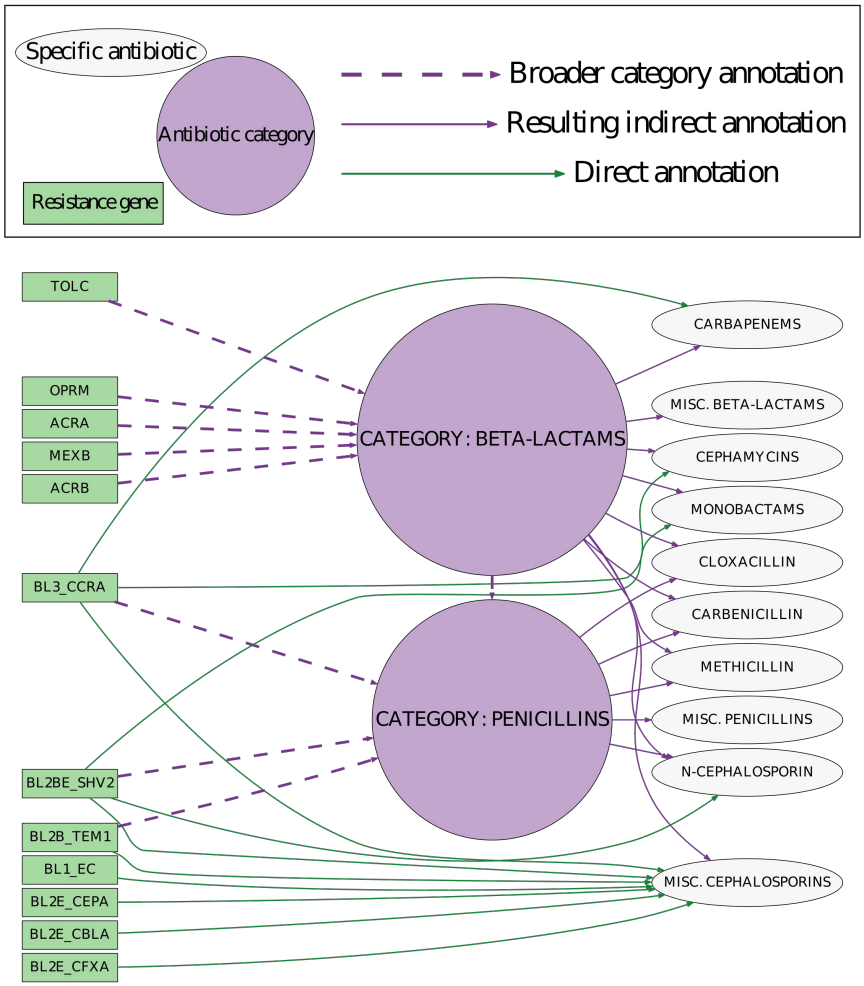
Simulation of short reads to estimate risk of computational cross-hybridization

To determine the extent to which the analysis pipeline risks erroneously mapping reads originating from a gene in one of the resistance gene families to genes from another such family, a simulation of the short-read sequencing process was carried out and the metagenomic analysis pipeline applied to its output. We simulated 100bp reads without sequencing errors from the set of known antibiotic resistance genes found in the set of 3496 reference genomes at 10X coverage using the *cMessi* application (Mende et al. 2012). These reads were mapped to the metagenomic reference gene catalog using MOCAT. Using these mappings, we determined for each read the number of resistance gene families to which the read was mapped. Reads that were "multiple mappers", i.e. which mapped to other resistance gene families beside the one they were drawn from by the simulator, were counted, yielding an estimate of the expected degree of computational cross-hybridization between these families under conditions matching those of our study. Of these cases, we further counted the fraction of cases where the read mapped not only to

another resistance gene family, but to a resistance gene family with a different antibiotic resistance effect. These results are shown in Supplementary Table S7, and reveal a very modest effect of multiple mappers on predicted resistance potentials.

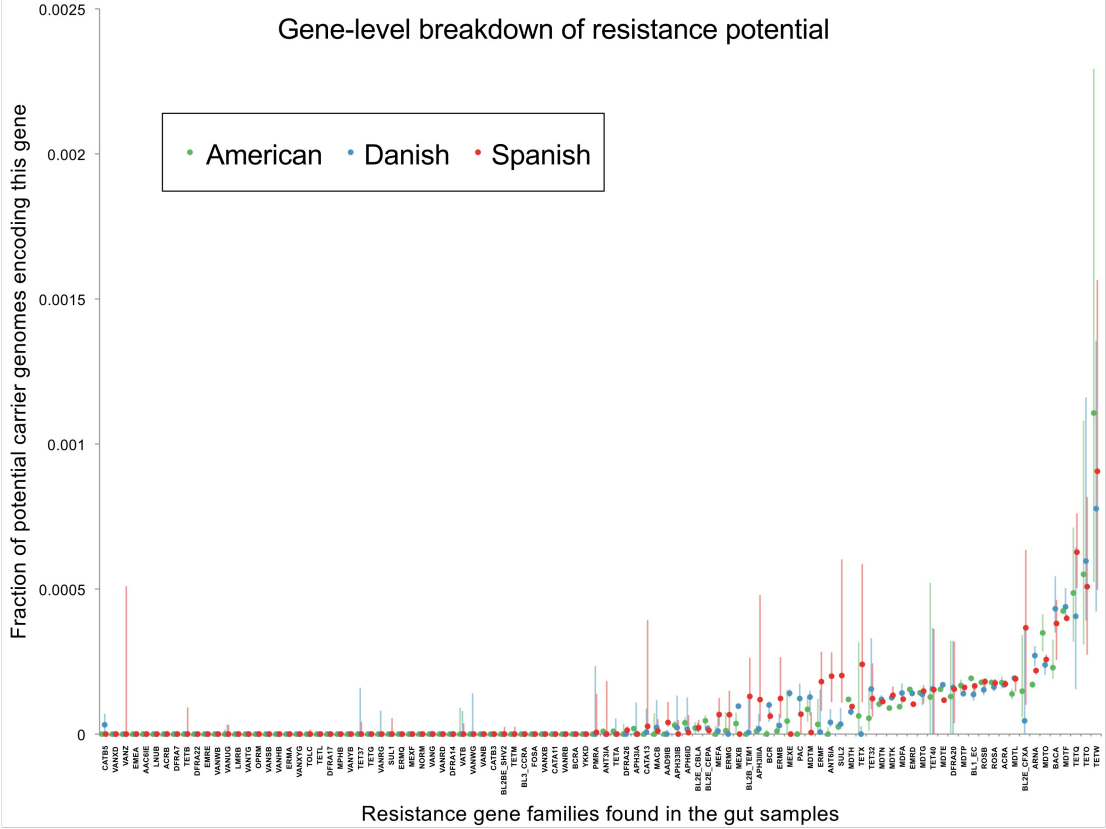
Supplementary Figures and Legends:

Supplementary Figure S1



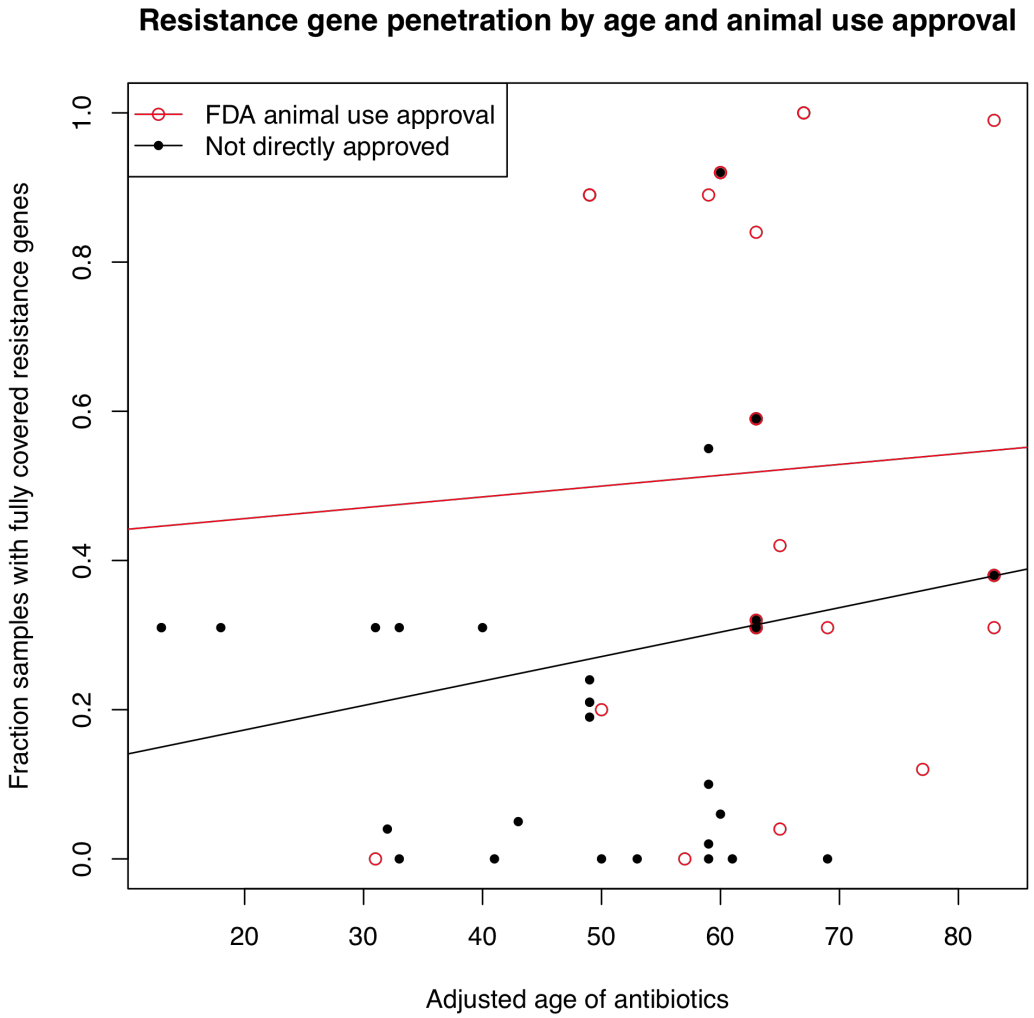
This graph shows a subset of the resistance gene annotation underlying the study, limited to beta-lactam antibiotics and to genes directly or indirectly providing resistance to beta-lactams. Some genes cause resistance to only specific drugs, some to all members of a class of antibiotics. These classes may in turn contain subclasses, such as the penicillins as a subset of beta-lactam antibiotics. The result is a network of many-to-many relationships between resistance genes and antibiotics, with potentially several genes contributing directly or indirectly for each antibiotic, as well as with resistance to several antibiotics directly or indirectly provided by each gene. The contribution of each gene to each antibiotic is counted only once, even where there are several paths through the network connecting them. Member antibiotics of broader classes are included separately in the resulting resistance profile if there are resistance gene annotations calling them out specifically, or else represented through composite, 'generic' classes (e.g. "miscellaneous penicillins" representing the resistance potential for any penicillin not uniquely addressed in any resistance annotation).

Supplementary Figure S2



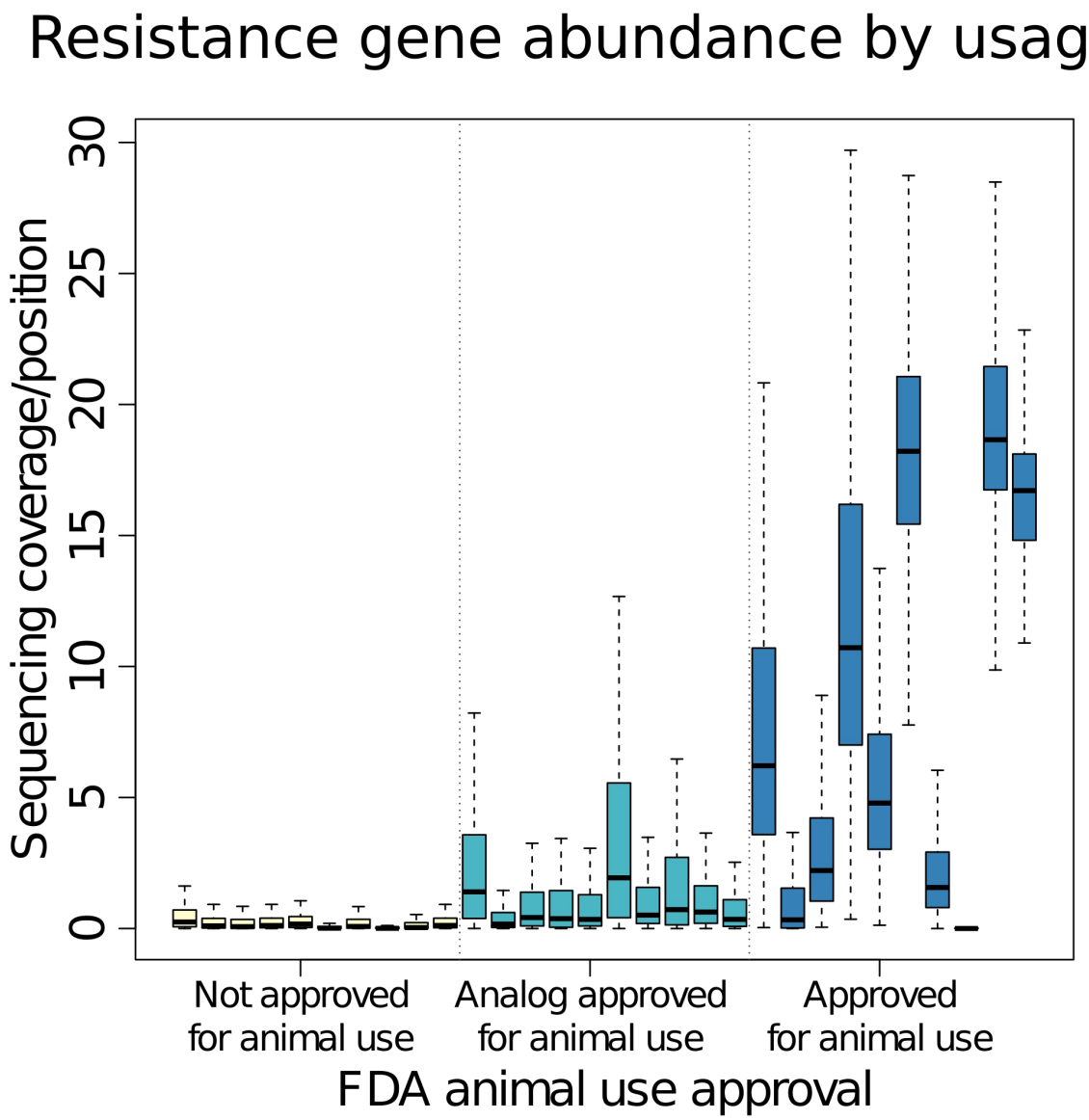
Country differences in resistance potential are visible at the level of genes as well as when integrated by antibiotic class. This figure shows a gene-level view of taxonomy-adjusted antibiotic resistance potential across three countries in the dataset of 252 previously sequenced Illumina samples. 98 gene families were found in at least one sample following compensation for differences in taxonomic composition. Notably, the gene families *ermB*, *ermF*, *ermG* and *tetQ*, where a dramatic increase in carriage has been seen over the past 50 years in retrospective analysis of fecal samples (Shoemaker et al. 2001), are prevalent in these samples as well.

Supplementary Figure S3



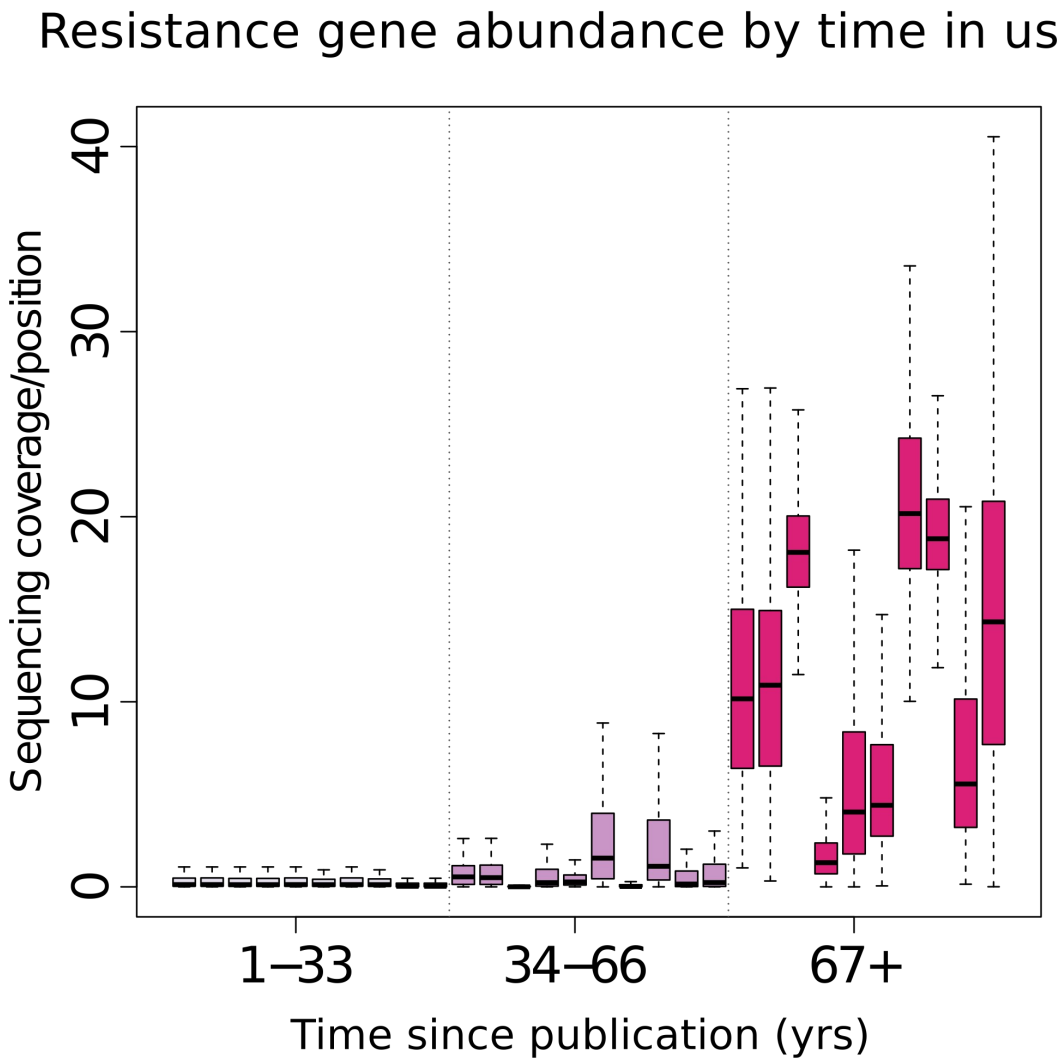
Antibiotic resistance potentials in the gut samples are higher both for antibiotics used in animals and for antibiotics that have been longer available, and furthermore, older antibiotics are more likely to have been approved for animal use. However, these factors appear to impact resistance potentials separately. This figure shows a regression of population resistance (measured as the fraction of samples with at least one fully covered resistance gene) as a function of the age of an antibiotic (taken as the time since first publication for each family of analogous compounds). Antibiotics are separated into those that are approved for animal use (N=26) and those that are not (N=40). ANCOVA analysis of resistance as a function of age with usage category as a covariate shows a highly significant ($P<0.006$, partial $\chi^2 = 0.11$) effect of animal use approval and a weakly significant ($P<0.025$, partial $\chi^2 = 0.007$) effect of age. The lines in the figure are separate regressions for antibiotics with and without animal use approval.

Supplementary Figure S4



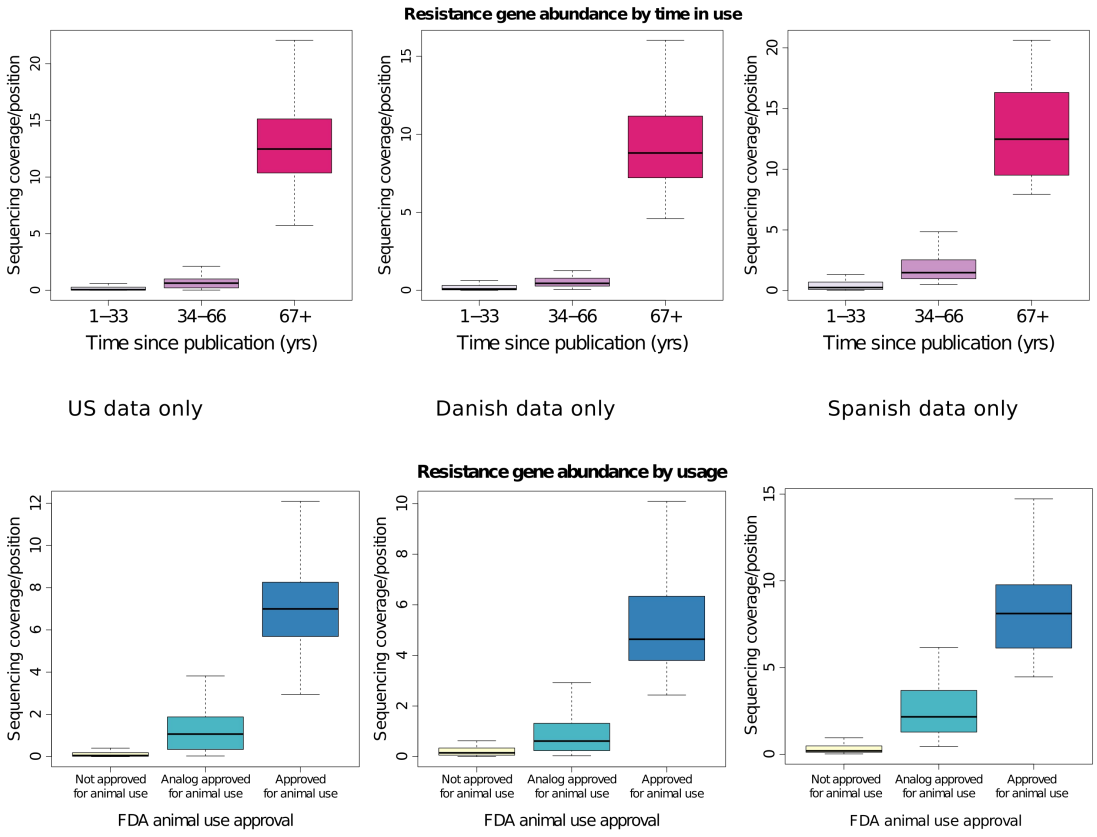
Our dataset contains 66 gene families providing resistance to antibiotics approved for animals, 33 for antibiotics with analogs approved for animal use, and 36 for antibiotics not approved for animal use. To verify differences between usage categories is not merely an effect of the higher number of genes active against animal use-approved antibiotics, the analysis from Figure 1B was repeated ten times, each time randomly sampling 10 genes from each category instead of using all genes. Each box plot within a set corresponds to a random selection of 10 genes active against the group of antibiotics in question, showing average sequencing coverage (total gene base counts divided by concatenated gene length) across these genes across the 252 Illumina samples. While variable within each category depending on the genes chosen, the observation of higher abundance of resistance genes for antibiotics directly or indirectly approved for use in animals cannot be reduced to this difference in numbers of genes active against them.

Supplementary Figure S5



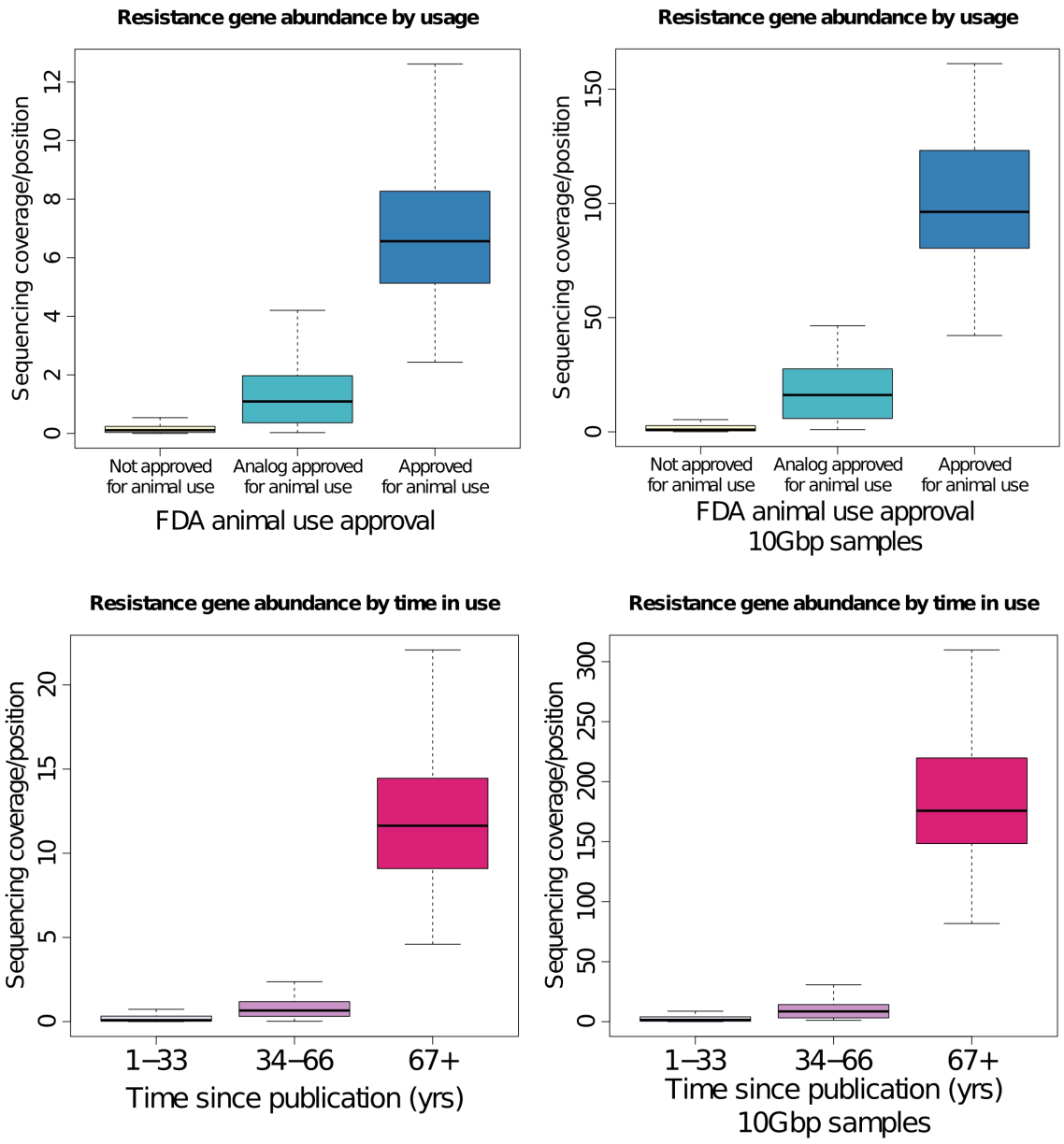
Our dataset contains 11 gene families providing resistance to antibiotics that have been in use less than 33 years, 58 for antibiotics in use between 34 and 66 years, and 46 for antibiotics in use for 67 years or longer. To verify differences between antibiotic age groups is not merely an effect of the higher number of genes active against older antibiotics, the analysis from Figure 1C was repeated ten times, each time randomly sampling 10 genes from each category instead of using all genes. Each box plot within a set corresponds to a random selection of 10 genes active against the group of antibiotics in question, showing average sequencing coverage across these genes across the 252 Illumina samples. As with animal use approval (Supplementary Figure S4), the differences between the categories with regards to numbers of genes are not sufficient to explain the observed differences in sequencing coverage between the categories.

Supplementary Figure S6



Significant differences in average sequence coverage/site, corresponding to relative abundance, are seen between antibiotic categories of different age (in terms of years since first publication of the compound) and animal use approval status. These results can be replicated using only one country subset at a time (US American, N=142; Spanish, N=39; Danish, N=71).

Supplementary Figure S7

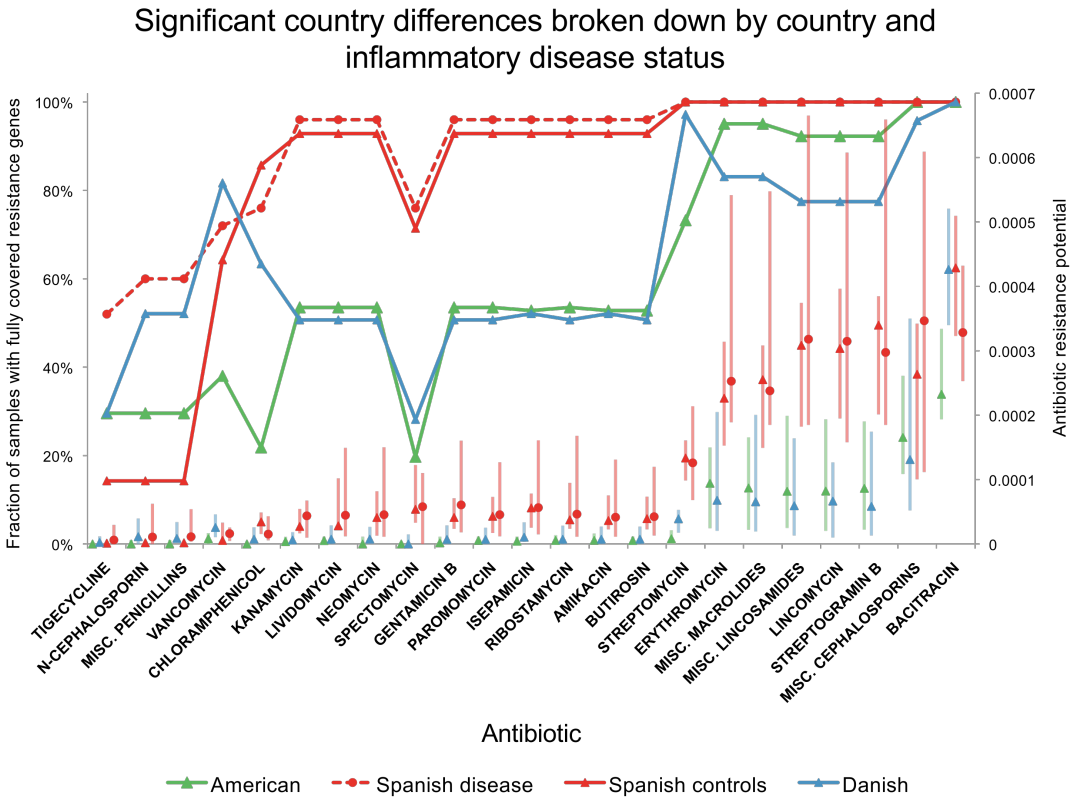


Significant differences in average sequence coverage/site, corresponding to relative abundance, are seen between antibiotic categories of different age (in the sense of years since first publication of the compound) and animal use approval status. These results can be replicated using only the subset of samples that are larger than 10Gbp, downsampled to this size, which is more than tenfold the 726Mbp dataset size used elsewhere in this work.

Supplementary Figures S8-15

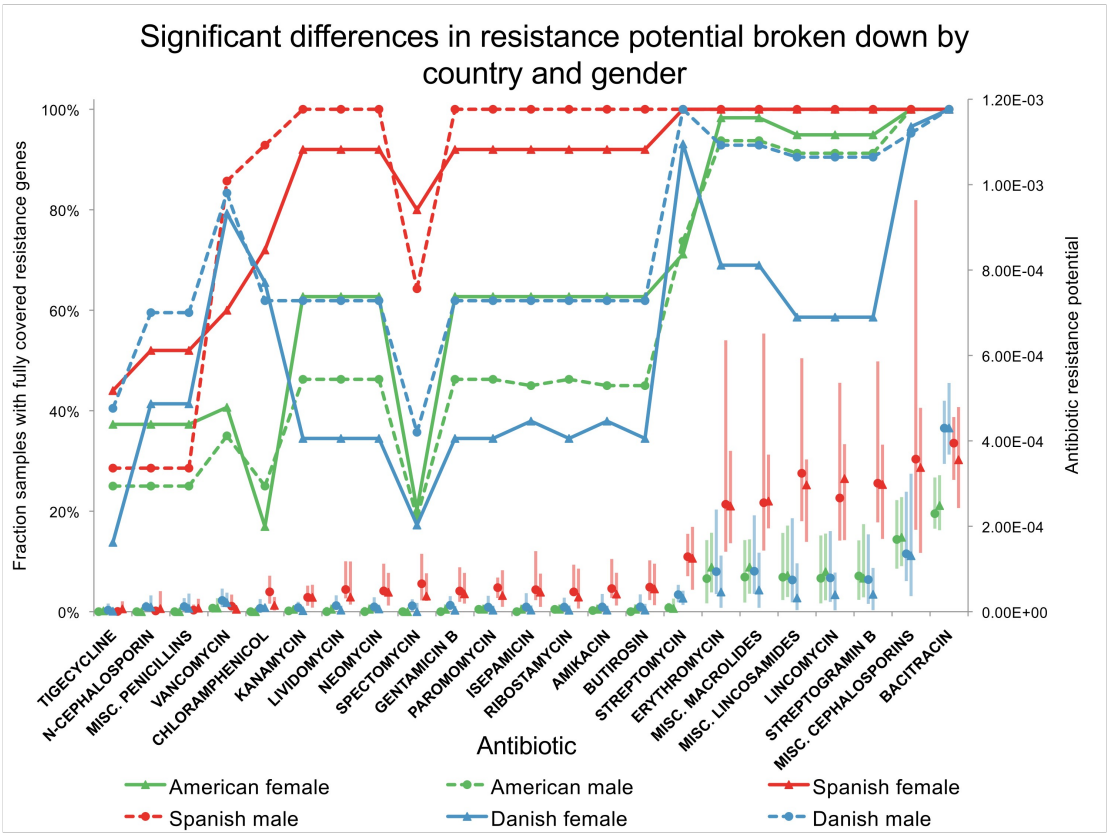
The country differences we observe in antibiotic resistance remains stable across a range of possible confounding factors, both with regards to population penetration (left axes, line plots, measured as fraction of samples with fully covered resistance genes) and taxonomy-adjusted resistance potential (right axes, point markers, population median shown). For Supplementary Figures S8-12, these results are shown for those antibiotics where significant country differences were concluded in the main analysis (Figure 2A in main manuscript). For Supplementary Figures S13-15, results are shown where independent analysis using only resistance genes from one of each of the three broad mechanistic categories was used at a time, with those antibiotics displayed where significant country differences (at FDR <= 5%) was concluded using those genes only.

Supplementary Figure S8



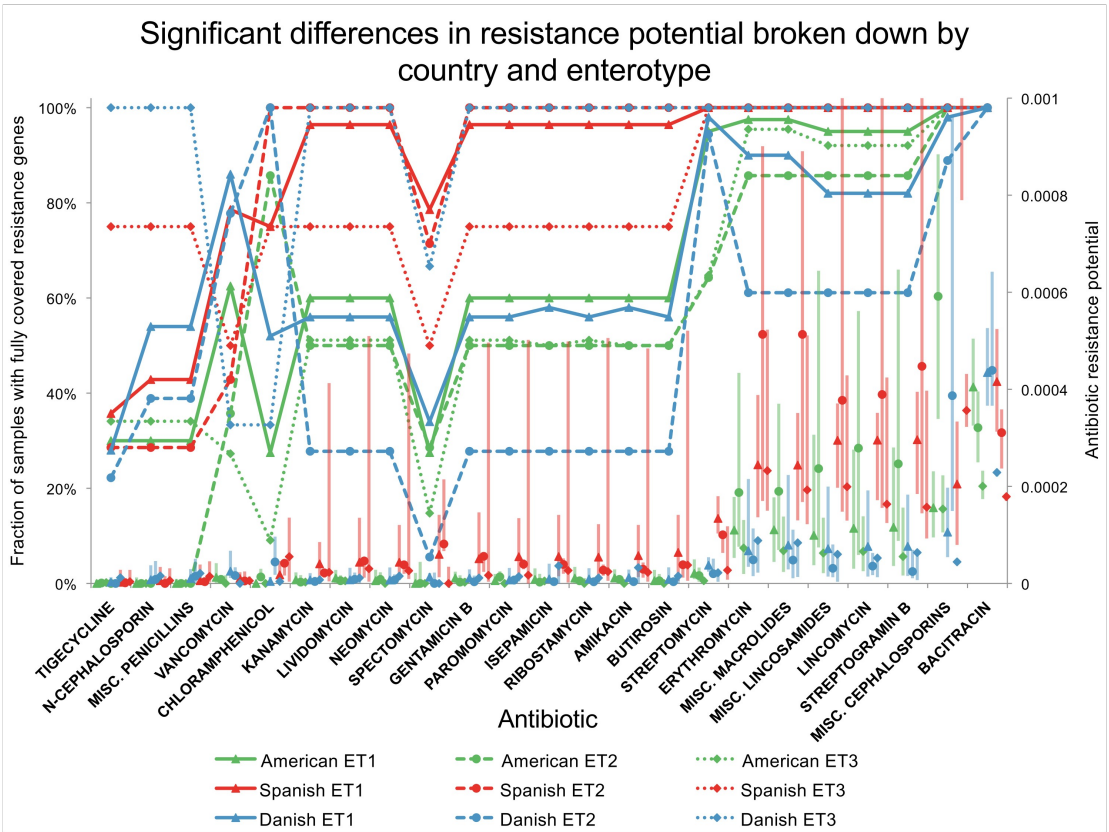
Separating the Spanish MetaHIT samples based on inflammatory bowel disorder (25 sufferers, 14 controls). The resistance potential difference between the countries is much larger in scope than between IBD and control samples, possibly reflecting the fact that antibiotics play only a limited role in treatment of the disease.

Supplementary Figure S9



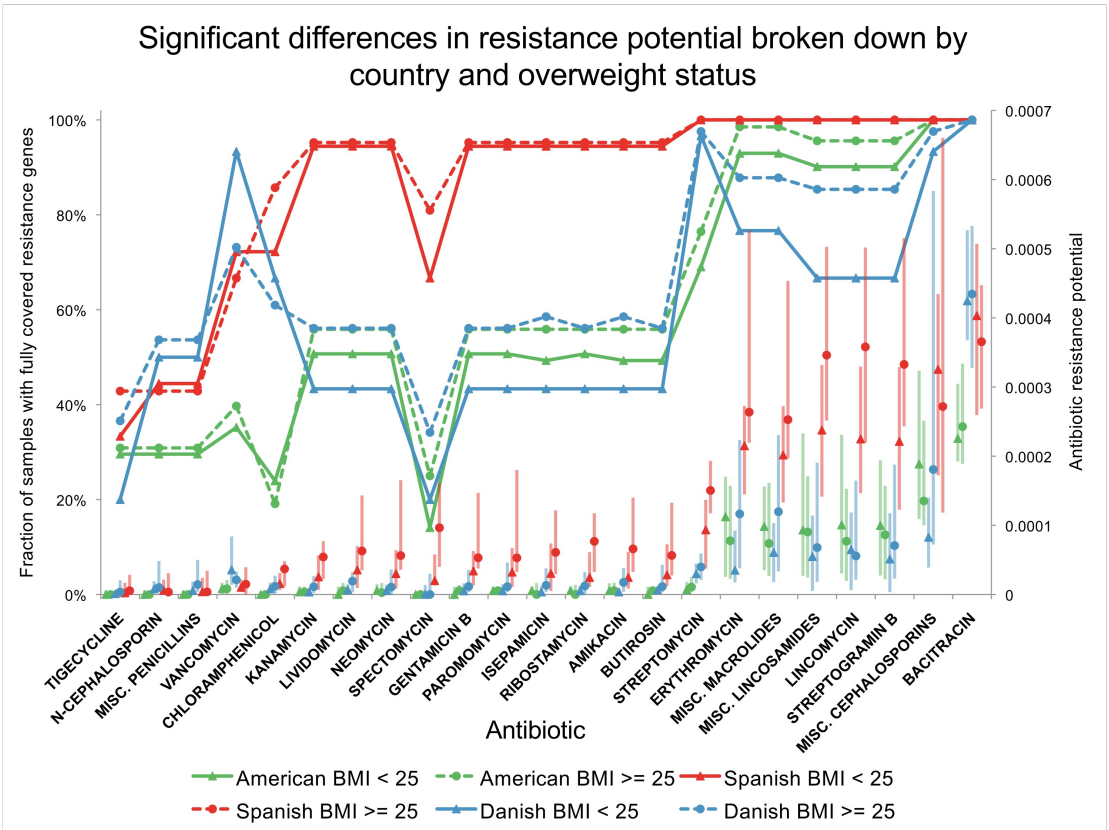
Separating the samples based on donor gender (Spanish; female N=25, male N=14, Danish; female N=29, male N=42, American; female N = 59, male N=80). Gender differences in resistance potential are much smaller in scope than the difference between Spanish and Danish/American samples.

Supplementary Figure S10



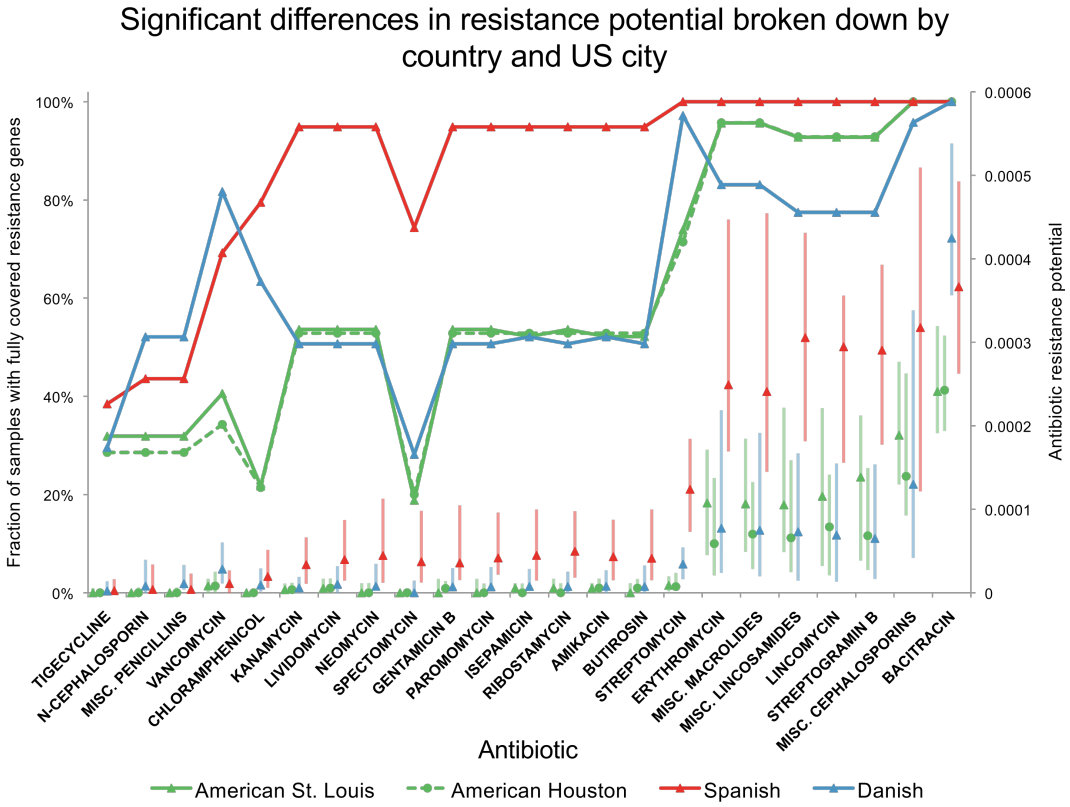
Separating the HMP and MetaHIT samples based on enterotype (Spanish; ET1 N=28, ET2 N=7, ET3 N=4, Danish; ET1 N=50, ET2 N=18, ET3 N=3, American; ET1 N=40, ET2=14, ET3=88). Resistance potential differences between enterotypes are visible for several antibiotics, but the pattern of observed country differences remains.

Supplementary Figure S11



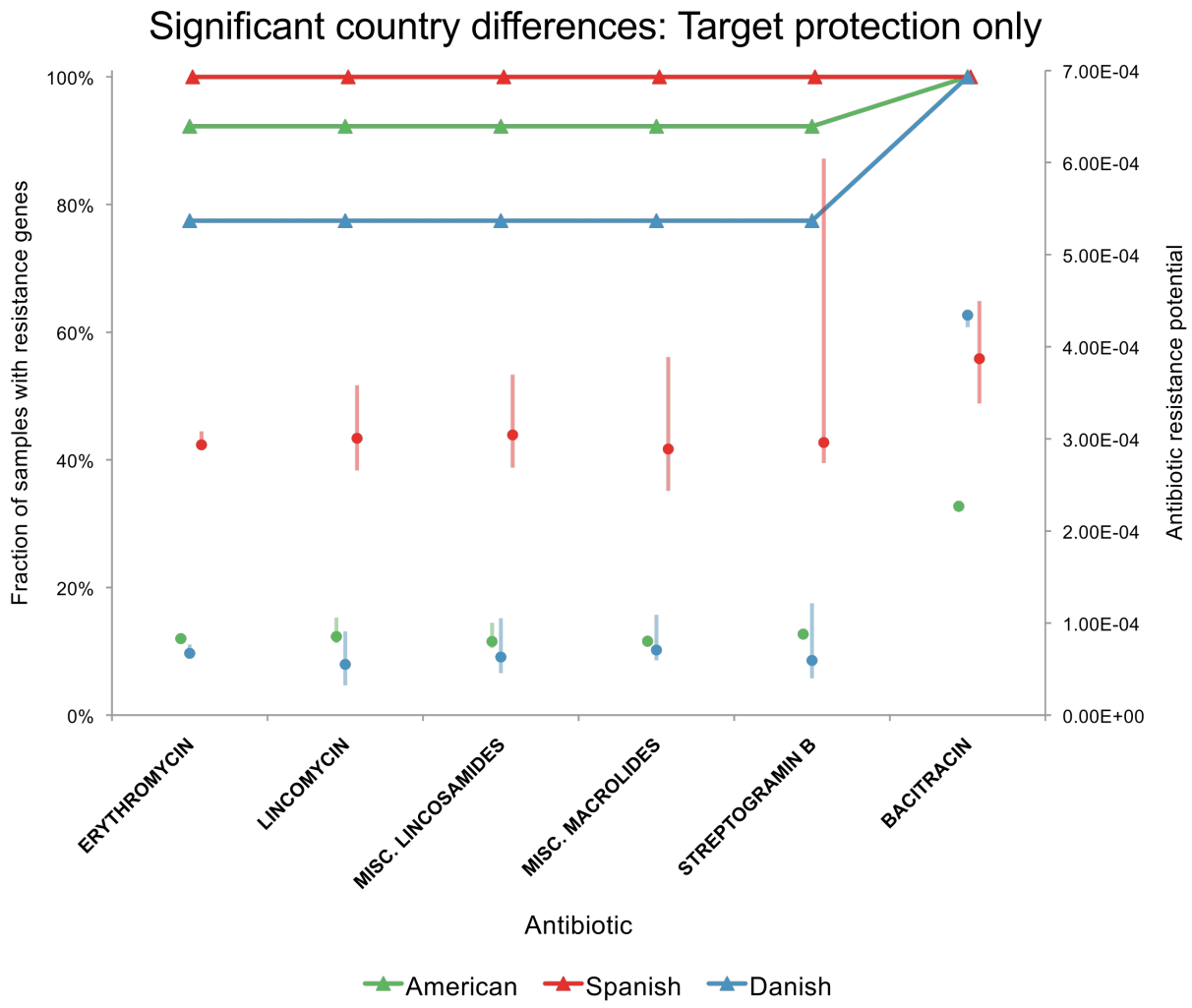
Separating the Spanish and Danish MetaHIT samples based on BMI (Spanish; BMI < 25 N=18, BMI >= 25 N=21, Danish; BMI < 25 N=30, BMI >= 25 N=41, American; BMI < 25 N=71, BMI >= 25 N=68). The resistance potential difference between overweight status and controls is smaller than country separation.

Supplementary Figure S12



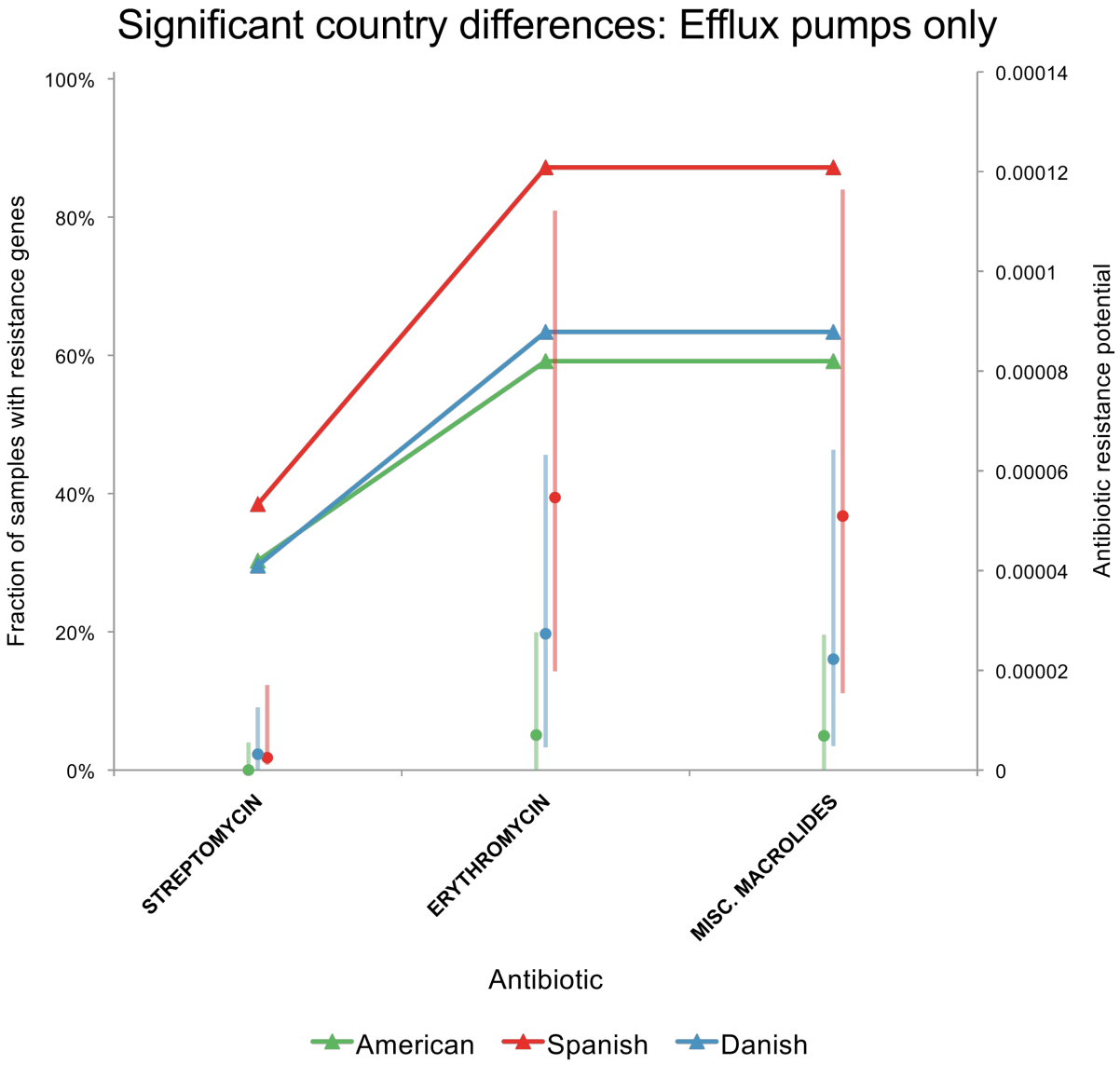
Separating the US samples based on sampling location. US samples not from Houston (N=70) or St Louis (N=69) were excluded for this comparison. Resistance potential differences between US cities are generally smaller than country differences.

Supplementary Figure S13



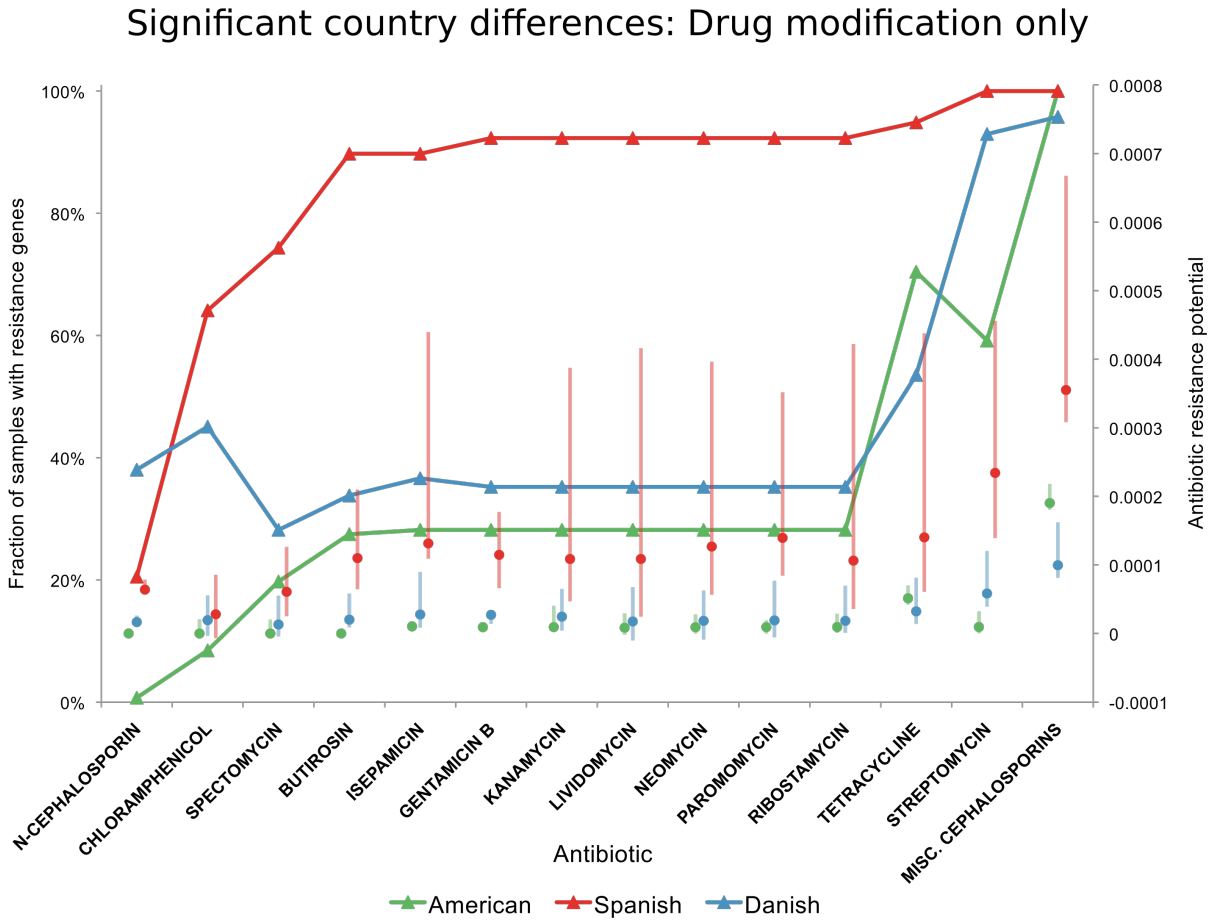
Significant country differences in resistance potential when using only target protection-type resistance genes. While only some of the antibiotics for which country differences were found in the main analysis also display such differences here, the pattern of country differences remains.

Supplementary Figure S14



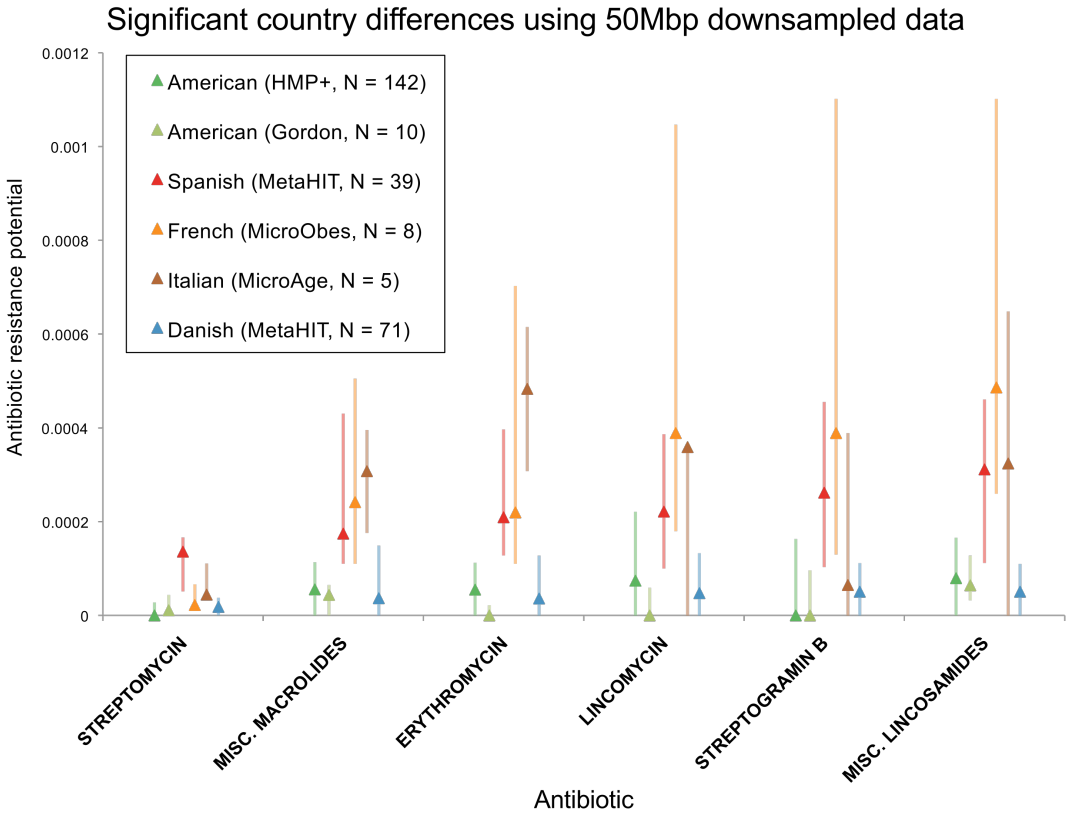
Significant country differences in resistance potential when using only efflux pump-type resistance genes. While only some of the antibiotics for which country differences were found in the main analysis also display such differences here, the pattern of country differences remains.

Supplementary Figure S15



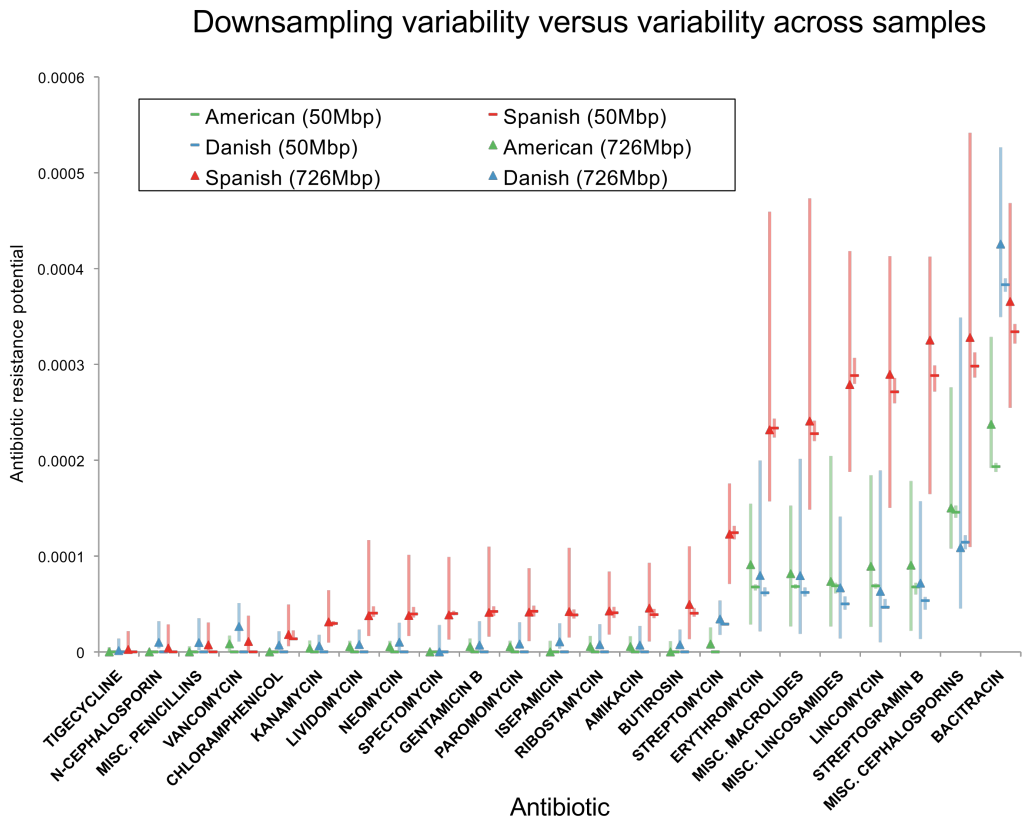
Significant country differences in resistance potential when using only antibiotic modification-type resistance genes. While only some of the antibiotics for which country differences were found in the main analysis also display such differences here, the pattern of country differences remains.

Supplementary Figure S16



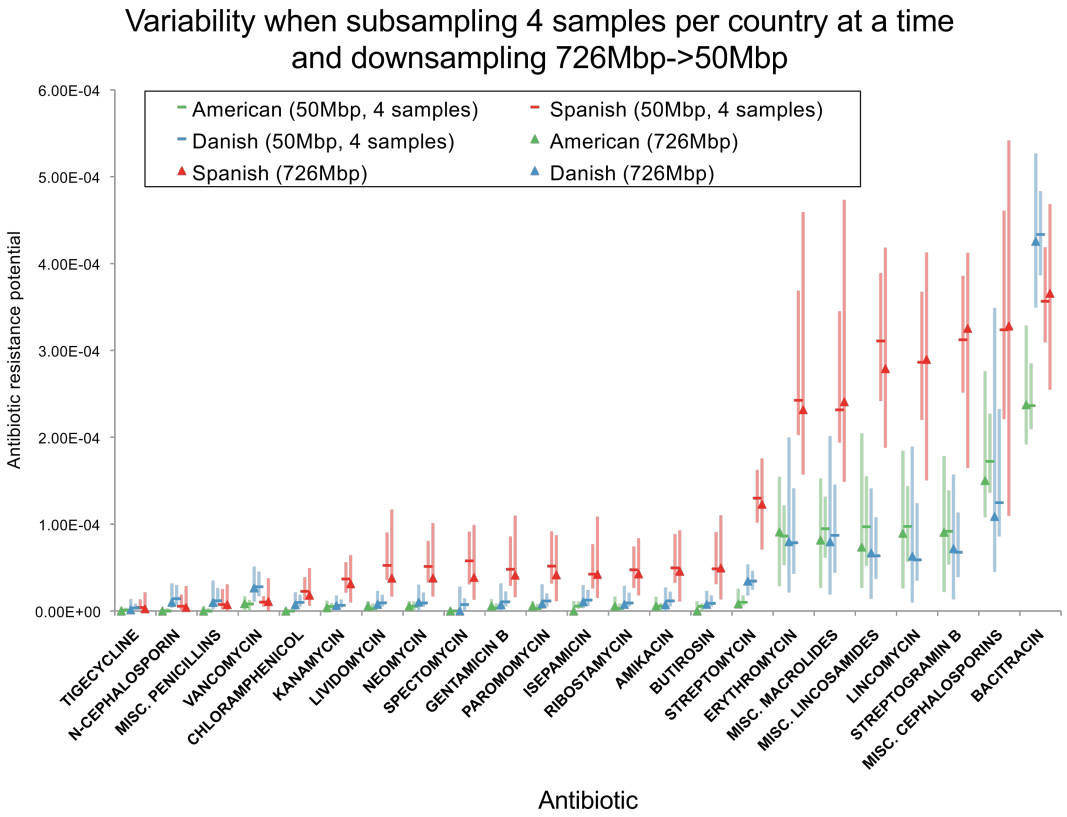
Significant country differences in taxonomy-adjusted resistance potentials (vertical bars and triangle markers showing 25%-50%-75% percentiles across samples) are visible for several antibiotics, even using a dataset downsampled to only 50Mbp sequencing depth. Spanish samples consistently show higher resistance potentials than Danish or US samples. Additionally, three further datasets were added: French (N=8) and Italian (N=5) samples, and a second set of recently published US samples (N=10). The French and Italian samples are broadly similar to the Spanish samples, and the two US datasets are likewise not radically different. Significance testing was carried out for differences between the Spanish, American (N = 142) and Danish samples using the same criteria as for the larger 726Mb dataset (see Methods Summary), with the antibiotics for which significance was achieved shown here.

Supplementary Figure S17



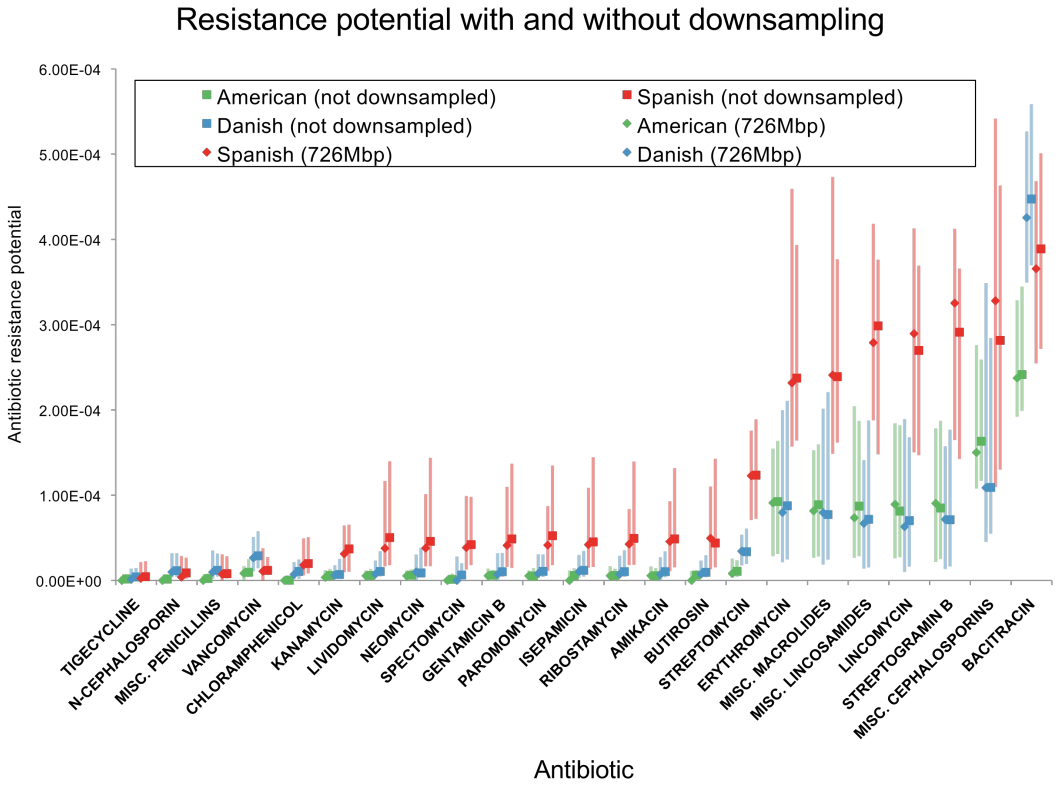
Stochastic downsampling of the data from 726Mbp to 50Mbp yields different outcomes for different randomizations, but typically preserves the observed country differences. As such, results from the 50Mbp Sanger/454 datasets should broadly reflect what would have been observed with greater sequencing depth for these samples. For this figure, downsampling was carried out 100 times in repetition with different random seeds, and the distribution of resistance potentials that follow for each antibiotic are shown in the figure for each country subset (American N=142, Spanish N=39, Danish N=71). The circles show the median (across samples) resistance potential for the original 726Mbp dataset, whereas the bar markers show the median (across randomizations) of median (across samples) resistance potentials. The vertical bars show 25% and 75% percentiles (across randomizations) for the median (across samples) resistance potential. The scale of this variation is mostly smaller than the difference between Spanish and American/Danish samples, though across randomizations, the country signal may be lost for lower-abundance resistance genes, in randomizations where the resistance load becomes zero for every country subset. Antibiotics shown are those for which country differences were concluded in the 726Mbp dataset.

Supplementary Figure S18



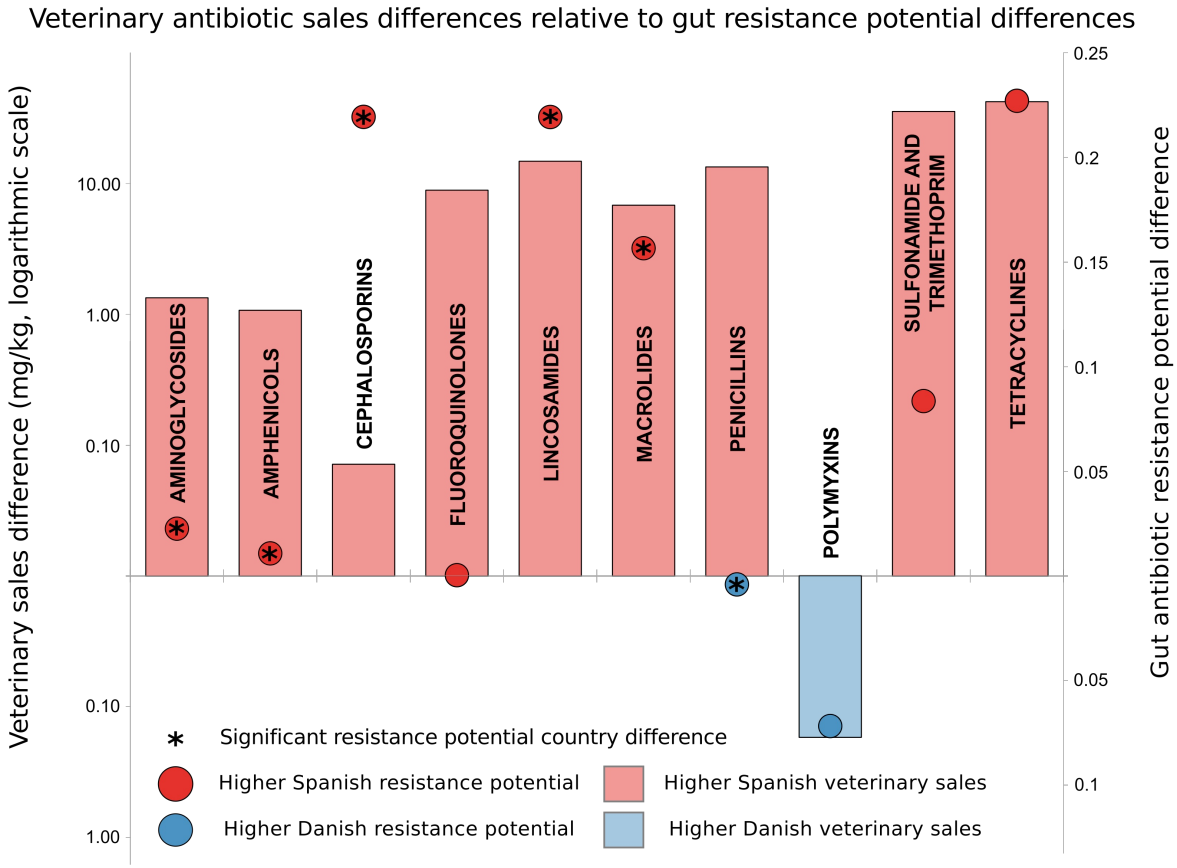
Subsampling only a small number of samples at a time reduces resolution but preserves the observed country differences. This supports the reliability of the Sanger/454 control datasets despite their relatively small size, since it implies country differences in resistance load are not driven by a small number of anomalous samples. This figure shows the resistance potential for the three country subsets (American N=142, Spanish N=39, Danish N=71) after having downsampled the data from 726Mbp to 50Mbp, but with median resistance potentials measured only from 4 randomly chosen samples from each country at a time. This randomization is repeated 1000 times for each antibiotic category, and the median (across randomizations) of median (across the 4 chosen samples) resistance potential is shown as horizontal bar markers. The vertical bar shows the corresponding 25% and 75% percentiles (across randomizations) for the same measure. The circle markers show the resistance potentials for each antibiotic in each country subset computed from the original 726Mbp dataset with all 252 samples included.

Supplementary Figure S19



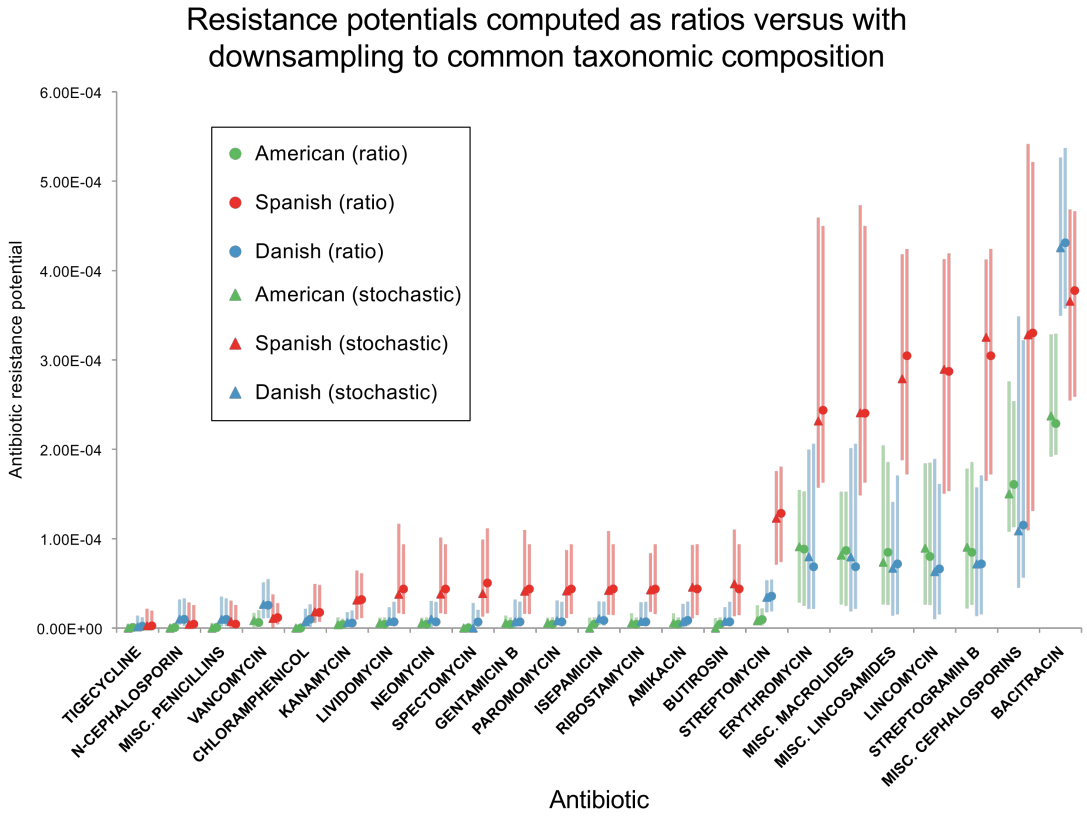
For the set of antibiotics where country differences were detected, results are only very slightly affected by downsampling. Median antibiotic resistance potentials are shown for the country subsets downsampled to 726Mbp as well as for non-downsampled data. In 22 of 23 cases (excepting tigecyclines) a proportion-aware significance test on the non-downsampled data supports the same country differences as determined from the 726Mbp datasets at Benjamini-Hochberg-Yekutieli FDR <= 5%.

Supplementary Figure S20



When comparing Denmark and Spain with regards to livestock antibiotic consumption and resistance potential for each antibiotic category in human fecal samples, higher levels of the former corresponds to higher levels of the latter. Livestock antibiotic consumption was calculated from veterinary antibiotic sales as measured in the ESVAC project, and normalized by approximate total mass of treated animals. Median resistance potentials were computed as described elsewhere in this work, and averaged across class members for comparison with the ESVAC classes. Modeling median resistance potential as a function of livestock antibiotic consumption (generalized linear model with antibiotic class and country included as random intercepts) confirms higher antibiotic consumption as significantly increasing resistance potential in the gut (P (MCMC) < 0.03).

Supplementary Figure S21



Resistance potential distributions are shown as 25%-50%-75% quantiles for the antibiotics where significant country differences were concluded. The triangle markers show series where resistance potential is taken as a simple ratio of resistance-encoding sequence amount to amount of sequence from potentially resistant taxa, whereas the circle markers use the more complex procedure where samples are stochastically downsampled to the same effective taxonomic composition for each antibiotic class. For these antibiotics, there is very little difference between the results of the different methods.

Supplementary Tables:

Supplementary Table S1

(File: Forslund_TableS1.xls)

This table shows 66 antibiotic classes or subclasses analyzed in this study for which time in use and animal use approval status is available. The remaining classes with resistance genes in the ARDB (cation and anion antibiotics and multidrug efflux pumps with no specified antibiotic range) were excluded from analyses that required age or usage information, but were included in overall resistance potential determination. For 50 of the 66 antibiotic classes or subclasses, as well as for anion antibiotics and unspecified efflux pumps, resistance determinants were detected in the gut, with each sample on average having 21 categories with fully covered resistance genes. For each antibiotic, the reference provided in this table either is the earliest published reference, or contains a reference to it. “Adjusted age” refers to the time since first publication of the oldest member of each group of antibiotics considered analogous for this purpose, which is listed in the “Analogs” column. While some newer beta-lactam antibiotics belong to the same class as older antibiotics like the penicillins, they were explicitly designed to circumvent existing beta-lactam resistance mechanisms, and as such are not considered their analogs for this purpose. Exceptions of this type are noted in the “Notes” column. The “Cross-resistance” column shows which broader class resistance annotations of genes in the ARDB are considered to apply to each more specific antibiotic class or compound.

Supplementary Table S2

(File: Forslund_TableS2.xls)

This table lists, for antibiotic classes where sales are measured as part of the ESVAC surveillance project, the Danish and Spanish sales for 2009. In addition, livestock head count and weight at slaughter data from the EuroStat (http://epp.eurostat.ec.europa.eu/portal/page/portal/agriculture/data/main_tables; 17 Sep 2012, date last accessed) database are shown, and the resulting effective treated biomass estimates according to the procedure used by Grave et al. (2010). Median resistance potentials are shown for comparison.

Supplementary Table S3

(File: Forslund_TableS3.xls)

This table lists every case where one of the 252 Illumina samples exhibited more than 10 X the population median resistance potential for an antibiotic. The last column lists which resistance gene families, out of those contributing to the resistance in question, that were fully covered by sequencing in that sample, using the dataset downsampled to 726 Mbp.

Supplementary Table S4

No of samples in final set	Nationality	Study	Downsampled sizes	Platform	Reference
139	US American	HMP	726Mbp/50Mbp	Illumina	HMP Consortium 2012
3	US American	Twin gut microbiome	726Mbp/50Mbp	Illumina	Schloissnig et al. 2012
39	Spanish	MetaHIT	726Mbp/50Mbp	Illumina	Qin et al. 2010
71	Danish	MetaHIT	726Mbp/50Mbp	Illumina	Qin et al. 2010
10	US American	Geography	50Mbp	454	Yatsunenko et al. 2012
4	Malawi	Geography	50Mbp	454	Yatsunenko et al. 2012
4	Danish	MetaHIT	50Mbp	Sanger	Arumugam et al. 2011
4	Spanish	MetaHIT	50Mbp	Sanger	Arumugam et al. 2011
5	Italian	MicroAge	50Mbp	Sanger	Arumugam et al. 2011
8	French	MicroObes	50Mbp	Sanger	Arumugam et al. 2011
7	Japanese	kurokawa07	50Mbp	Sanger	Kurokawa et al. 2007

This table shows the various gut metagenome sample sets that were included in the study. The number of samples following quality control are shown in the first column, followed by the nationality of the donors, the study the data originated from, the size that the data was downsampled to, the sequencing platform used, and a reference for the origin of the samples.

Supplementary Table S5

(File: Forslund_TableS5.xls)

This table shows the species represented in our microbial reference genome collection. The first column shows the NCBI taxonomy ID, the second the project ID, and the third the species name.

Supplementary Table S6

(File: Forslund_TableS6.xls)

This file lists the number of bases mapped to each reference gene that was classified as a resistance gene family member, in the set of 252 US, Danish and Spanish samples.

Supplementary Table S7

(File: Forslund_TableS7.xls)

This table shows the result of simulating a large number of short reads drawn from resistance gene family matches in the collection of reference genomes used in this study, then supplying these reads to the metagenomic gene mapping pipeline. Only a small fraction of reads (of a length equal to average read length in the 252 samples Illumina dataset used elsewhere in this manuscript) mapped to another gene family in addition to the one they were sampled from; these are referred to as multiple mappers. Of the multiple mappers, only an even smaller fraction of reads change the resulting resistance profile.

Supplementary Table S8

(File: Forslund_TableS8.xls)

This file lists the number of bases mapped to each concatenated set of species-specific universal marker genes, at a species level, for the set of 252 US, Danish and Spanish samples. The -1 fraction corresponds to such sequences that cannot be mapped to any of the species in the reference genome collection. Normalizing these data will yield relative species proportions.

Supplementary Table S9

(File: Forslund_TableS9.xls)

This file lists all proteins (first and third column) in the set of reference organisms that were mapped to resistance gene families (second column).

Supplementary Table S10

(File: Forslund_TableS10.xls)

This table lists each entry in the gut metagenomic reference gene catalog (first column) that is assigned to a resistance gene family (second column), the ARDB resistance annotation for that family (third column, comma-separated antibiotic classes or subclasses), and the ARDB class that each gene family is assigned to (fourth column).

Supplementary Table S11

(File: Forslund_TableS11.xls)

This table lists the reference gene catalog entries that were mapped to resistance gene families, along with the eggNOG 3 entries they were mapped to. A small fraction of reference genes, listed at the bottom, were novel relative to eggNOG 3.

Supplementary Table S12

(File: Forslund_TableS12.xls)

This table lists, for each resistance gene family included in the study (first column), the bacterial species, out of the 3496 represented in the reference genome collection, where members of this family were found (second column, separated by slash characters).

Supplementary Table S13

	Significant country differences in species composition?	Taxonomy-blind test: falsely significant country differences in resistance gene carriage?	Taxonomy-aware test: falsely significant country differences in resistance gene carriage?
Antibiotic	**BHY-adjusted P-value**	**BHY-adjusted P-value**	**BHY-adjusted P-value**
AMIKACIN	**3.7E-19***	>=0.05	>=0.05
ANIONIC ANTIMICROBIALS	**5.3E-4***	>=0.05	>=0.05
BACITRACIN	**1.6E-17***	**2.1E-6***	>=0.05
BUTIROSIN	**3.7E-19***	>=0.05	>=0.05
CARBAPENEM	**3.7E-19***	>=0.05	>=0.05
CEPHALOSPORIN	**3.7E-19***	**2.4E-9***	>=0.05
CEPHAMYCIN	**3.7E-19***	>=0.05	>=0.05
CHLORAMPHENICOL	**3.4E-18***	>=0.05	>=0.05
CIPROFLOXACIN	**1.5E-17***	>=0.05	>=0.05
DIBEKACIN	**3.7E-19***	>=0.05	>=0.05
ENOXACIN	**6.9E-19***	>=0.05	>=0.05
ERYTHROMYCIN	**3.7E-19***	**3.3E-3***	>=0.05
FOSFOMYCIN	>=0.05	>=0.05	>=0.05
FOSMIDOMYCIN	>=0.05	>=0.05	>=0.05
GENTAMICIN	**3.7E-19***	>=0.05	>=0.05
GENTAMICIN B	**3.7E-19***	>=0.05	>=0.05
ISEPAMICIN	**3.7E-19***	>=0.05	>=0.05
KANAMYCIN	**2.6E-17***	>=0.05	>=0.05
LINCOMYCIN	**7.0E-14***	**3.9E-3***	>=0.05
LIVIDOMYCIN	**3.7E-19***	>=0.05	>=0.05
MISC. AMINOGLYCOSIDES	**3.7E-19***	>=0.05	>=0.05
MISC. BETA-LACTAMS	**3.7E-19***	>=0.05	>=0.05
MISC. FLUOROQUINOLONES	**6.6E-19***	>=0.05	>=0.05
MISC. GLYCYLCYCLINES	**3.7E-19***	>=0.05	>=0.05
MISC. LINCOSAMIDES	**3.7E-14***	**5.9E-3***	>=0.05
MISC. MACROLIDES	**3.7E-19***	**1.4E-3***	>=0.05
MISC. PENICILLINS	**3.7E-19***	>=0.05	>=0.05
MONOBACTAM	**3.7E-19***	>=0.05	>=0.05
N-CEPHALOSPORIN	**3.7E-19***	>=0.05	>=0.05
NEOMYCIN	**3.7E-19***	>=0.05	>=0.05
NETILMICIN	**3.7E-19***	>=0.05	>=0.05
NORFLOXACIN	**1.5E-17***	>=0.05	>=0.05
PAROMOMYCIN	**3.7E-19***	>=0.05	>=0.05
POLYMYXIN	>=0.05	>=0.05	>=0.05
PUROMYCIN	>=0.05	>=0.05	>=0.05
RIBOSTAMYCIN	**3.7E-19***	>=0.05	>=0.05
SISOMICIN	**3.7E-19***	>=0.05	>=0.05
SPECTINOMYCIN	**3.7E-19***	>=0.05	>=0.05
SPECTOMYCIN	**5.4E-15***	>=0.05	>=0.05
STREPTOGRAMIN A	**9.6E-18***	>=0.05	>=0.05
STREPTOGRAMIN B	**3.7E-14***	**5.9E-3***	>=0.05
STREPTOMYCIN	**2.7E-17***	**7.5E-3***	>=0.05
SULFONAMIDE	>=0.05	>=0.05	>=0.05

TEICOPLANIN	**3.1E-17***	>=0.05	>=0.05
TETRACYCLINE	**3.6E-16***	**3.9E-3***	>=0.05
TIGECYCLINE	**2.2E-17***	>=0.05	>=0.05
TOBRAMYCIN	**3.7E-19***	>=0.05	>=0.05
TRIMETHOPRIM	**1.5E-16***	>=0.05	>=0.05
UNSPECIFIC MULTIDRUG EFFLUX	>=0.05	>=0.05	>=0.05
VANCOMYCIN	**2.4E-12***	**1.1E-2***	>=0.05

Differences in taxonomic composition can cause country differences in resistance gene carriage, even for data simulated using the same model for all countries. Resistance gene abundances were generated based on carrier species abundances, assuming the same relative resistance gene densities in all three countries. Since this is a null hypothesis for country differences in resistance potential, any apparently significant difference will constitute a false positive for the method used to conclude it. Carrier species abundances are significantly different (Kruskal-Wallis test, Benjamini-Hochberg-Yekutieli FDR <= 5%) between the country datasets (142 US HMP+ samples, 71 Danish and 39 Spanish MetaHIT samples), possibly reflecting differences between datasets in sample preparation protocols as well as genuine biological factors. Testing for country differences in raw gene abundances using this simulated data, without taking taxonomy into account (Likelihood ratio test comparing mixed-effects models with and without country as predictor variable), yields falsely significant (Benjamini-Hochberg-Yekutieli FDR <= 5%) country differences for a majority of antibiotic classes or subclasses. Taking taxonomy into account by downsampling pooled gene abundances to the same pooled carrier species abundance across samples, as well as controlling for enterotype in the model comparison test, none of these differences are significant, demonstrating that the test used in the main analysis is more robust to these error sources.

Supplementary Bibliography

Barbosa J, Ferreira V, Teixeira P. 2009. Antibiotic susceptibility of enterococci isolated from traditional fermented meat products. *Food Microbiol* **26**: 527–532.

Bates DM, Maechler M. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

Baayen RH. 2009. languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 0.955.

Ciccarelli FD, Doerks T, Von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**: 1283–1287.

European Medicines Agency (2011). Trends in the sales of veterinary antimicrobial agents in nine European countries (2005-2009). EMA/238630/2011.

Gazzola S, Fontana C, Bassi D, Cocconcelli PS. 2012. Assessment of tetracycline and erythromycin resistance transfer during sausage fermentation by culture-dependent and -independent methods. *Food Microbiol* **30**: 348–354.

Grave K, Torren-Edo J, Mackay D. 2010. Comparison of the sales of veterinary antibacterial agents between 10 European countries. *J Antimicrob Chemother* **65**: 2037–2040.

Jamet E, Akary E, Poisson M-A, Chamba J-F, Bertrand X, Serror P. 2012. Prevalence and characterization of antibiotic resistant Enterococcus faecalis in French cheeses. *Food Microbiol* **31**: 191–198.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20**: 265–272.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Mende DR, Waller AS, Sunagawa S, Järvelin AI, Chan MM, Arumugam M, Raes J, Bork P. 2012. Assessment of metagenomic assembly using simulated next generation sequencing data. *PLoS One* **7**: e31386.

Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, et al. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* **40**: D284–D289.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**: 1449–1452.

Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* **38**: e132.