Supporting Online Materials

Materials and Methods

Isolation of Highly-Resistant, Soil-Derived Organisms for Construction of the AB95 Metagenomic Library

Highly antibiotic-resistant, soil-derived microorganisms were isolated by first inoculating one of eleven soils (table S1) into defined liquid media containing 1g/L of various antibiotics as a source of carbon (table S2) as well as 5g $(NH_4)_2SO_4$, 3g KH_2PO_4, 0.5g $MgSO_4 \cdot 7H_2O$, 15mg EDTA, 4.5mg $ZnSO_4 \cdot 7H_2O$, 4.5mg $CaCl_2 \cdot 2H_2O$, 3mg $FeSO_4 \cdot 7H_2O$, 1mg $MnCl_2 \cdot 4H_2O$, 1mg H_3BO_3, 0.4mg $Na_2MoO_4 \cdot 2H_2O$, 0.3mg $CuSO_4 \cdot 5H_2O$, 0.3mg $CoCl_2 \cdot 6H_20$ and 0.1mg KI per liter water. The pH was adjusted to 5.5 using HCl, and the media was sterilized through a 0.22 μm filter. Initial inputs of approximately 125mg of each of the eleven soils were inoculated into the aforementioned media containing one of 18 selected antibiotics (table S2), as previously described (*21*). After seven days of growth at room temperature, 2.5μL of all cultures were transferred into 5mL of fresh media with 1g/L of antibiotic, resulting in a $5x10^4$ dilution at each passage. After seven days of further growth at room temperature, this transfer was repeated again, and the passaged culture allowed to grow for another seven days (room temperature). In previous published work *(21)*, a subset of these cultures were plated and further passaged to select for individual clones capable of subsisting on various antibiotics as sole carbon sources. Due to the high antibiotic concentrations used, we reasoned that our initial culturing steps would also select for highly-resistant organisms. Accordingly, 2.5μL of the defined-media cultures were inoculated into 1.5mL Luria-Bertani broth (5g Yeast Extract, 5g NaCl and 10g of Tryptone in 1L H2O), and allowed to grow for another four days at room temperature. At this stage, the 95 cultures with the highest optical density were selected for construction of the AB95 metagenomic library. Freezer stocks in 15% glycerol were created from these cultures and stored at -80°C.

Construction of AB95 metagenomic library

For construction of the AB95 metagenomic library, 1mL of each of the 95 highly antibiotic-resistant, soil-derived cultures was pooled into a total volume of 95mL. Cells were then immediately pelleted from this mixture, and re-suspended in 10mL of Luria-Bertani broth. Half of this volume was used as input for metagenomic DNA isolation using the PowerMax Soil DNA Isolation Kit (MoBio Laboratories Inc) per suggested protocols (http://www.mobio.com/images/custom/file/protocol/12900-10.pdf). Subsequently, DNA was sheared using a custom made semi automated DNA shearer to a size range of 500 – 3500bp. Sheared DNA was end-repaired using the End-It DNA End Repair kit (Epicentre) with the following protocol:

1. Combine:
    a. 210μL of sheared DNA (30ng/μL)
    b. 30μL 10X End Repair Buffer
    c. 30μL 2.5mM dNTP mix
    d. 30μL 10mM ATP
    e. 3μL of End-It enzyme mix
2. Incubate at room temp for 45 minutes
3. Heat inactivate at 70°C for 20 minutes

End-Repaired DNA was size selected by electrophoresis through a 1% low melting point agarose gel in 0.5X Tris-Borate-EDTA (TBE) buffer. A gel slice corresponding to 1000-3000bp was excised from the gel and DNA was extracted using a QIAquick Gel Extraction Kit (Qiagen). DNA was ligated into pZE21 MCS 1 vector (*30*) at the HincII site using the Fast Link ligation kit (Epicentre) with the following ligation protocol:

1. Combine:
    a. 0.5µL 10X Ligation buffer
    b. 0.25µL 10 mM ATP
    c. 0.25µL dH20
    d. 0.35µL HincII cut pZE21 MCS 1 vector (240 ng/µL)
    e. 2.90µL Sheared, End-repaired, Gel purified DNA insert (153 ng/µL)
    f. 1.0µL Fast-Link ligase (2 units/µL)
2. Incubate at room temperature for 16 hours
3. Heat inactivate at 70°C for 20 minutes

Four microliters of the fresh ligation mixture was used for transformation by electroporation into 50µL of electrocompetent *E. coli* TOP10 cells (Invitrogen). After transformation using standard protocols for a 1mm electroporation cuvette, cells were recovered in 1mL SOC medium for one hour at 37°C. Libraries were titered by plating out 1µL and 0.01µL of recovered cells onto LB agar (5g Yeast Extract, 5g NaCl, 10g of Tryptone, 12g Agar in 1L H_2O) plates containing 50µg/mL kanamycin. For each library, insert size distribution was estimated by gel electrophoresis of PCR products obtained by amplifying the insert from 16 randomly picked clones using primers flanking the HincII site of the multiple cloning site of the pZE21 MCS1 vector (which contains a selectable marker for kanamycin resistance). The average insert size was found to be 1.8Kb. The total size of the AB95 metagenomic library was determined to be 2.57Gb by multiplying the average PCR based insert size by the number of titered colony forming units after transformation recovery. The rest of the recovered cells were inoculated into 10mL of LB containing 50µg/mL kanamycin and grown overnight. The overnight culture was frozen down with 15% glycerol and stored at -80°C for subsequent screening.

Selection of Antibiotic Resistant Clones from AB95 and HG1

Together, this manuscript and its supporting online material discuss the selections of two libraries on various antibiotics. A 1.1Gb human fecal metagenomic library (HG1) *(20)* (described previously) was used to validate the PARFuMS assembly pipeline while AB95 (described above) represents the highly antibiotic-resistant, soil-derived, library detailed throughout the manuscript. For each metagenomic library tested, 100 µL of library freezer stock corresponding to $0.6\text{-}1.4\text{x}10^8$ cfu (representing approximately 10x the number of unique clones titered in each library) was plated out on Luria-Bertani agar plates containing binary combinations of kanamycin (50 µg/mL) and one of various other antibiotics (see tables S4, S14 for the exact antibiotics used during AB95 and HG1 functional selections). The plates were subsequently incubated at 37°C for 16 hours to allow for growth of clones containing an antibiotic resistant metagenomic insert. After overnight growth, all colonies from a single antibiotic plate were collected by adding 1 mL of sterile water to the plate, and scraping with an L-shaped cell scraper to gently remove colonies and collecting the liquid 'plate scrape culture'. This process was repeated a second time to ensure that all colonies were removed from the plate. The bacterial cells were then pelleted by centrifugation at 13,000rpm, the supernatant was discarded and the pellet re-suspended in 100 µL of nuclease free water.

Selection of Six Resistant Clones for PARFuMS Development and to Serve as Internal Illumina Controls

Internal Illumina sequencing control samples were created using six previously identified antibiotic resistant clones *(20)*. This collection of clones was also used to develop and optimize PARFuMS. All clones contained β-lactamase resistance genes encoding resistance to penicillin. The metagenomic inserts of these clones ranged in size from 1.3 Kbp to 3.3 Kbp. Pooled cultures of these clones were plated at 5 different relative ratios on Luria-Bertani agar plates containing 50 µg/mL

Penicillin-G (table S15) and grown for 16 hours at 37ºC. Colonies were collected and processed in the same manner as the AB95 and HG1 library resistant colonies.

Amplification of Antibiotic Resistant Metagenomic DNA Fragments

Each metagenomic fragment integrated in the pZE21-MCS1 vector conferring antibiotic resistance was amplified from plasmids by colony PCR. In brief, for plates with 0-50 colonies, 5μL of plate scrape culture was used; for plates with 50-500 colonies, 1μL was used; and for plates with 500-1500 colonies, 0.1μL was used as template. Colony PCR was performed using 1.5μL of 10μM primers (5'CCCCCCCTCGAGGTC-3' and 5'ATCAAGCTTATCGATACCGTC-3'), 0.2mM dNTPs, 2.5 units NEB *taq* polymerase, and 2.5μL thermopol buffer in a final volume of 25μL. DNA was amplified in a thermocycler at 94ºC for 10 min followed by 25 cycles of 94ºC for 45 seconds, 55ºC for 30 seconds, 72ºC for 5.5 minutes followed by 10 minutes at 72ºC for final extension. PCR products were visualized on a 1% agarose gel to ensure amplification of DNA. These colony PCR samples were cleaned using the Qiagen QIAquick PCR purification kit following the manufacturer's instructions.

Illumina Sample Preparation and Sequencing

Two micrograms of DNA amplified from the colony PCR were diluted to a final volume of 120μL in Qiagen EB buffer and then sheared to 100 bp fragments using the Covaris S2 sonicator for 20 minutes, intensity 5, duty cycle 10% and 100 cycles per burst. Subsequently, DNA was concentrated using QIAGEN MinElute PCR Purification Kit and all samples were diluted to 40ng/μL. Libraries were prepared as described previously (*31*) A subset of 400ng of sheared DNA was blunted using 2μL T4 DNA ligase buffer with 10mM ATP (10X, NEB), 5 units T4 PNK (NEB), 2.5 units Klenow DNA polymerase (NEB) and 0.1μL 10mM dNTPs (NEB) in a final volume of 20uL. The reaction was incubated at 25ºC for 30 minutes followed by 20 minutes at 75ºC.

Excess nucleotides were de-phosphorylated by adding 1U SAP (Promega) to the reaction and incubating at 37ºC for 30 minutes followed by 30 minutes at 75ºC. An adenosine was added to the 3' end of the fragments by adding 6μL of 1X T4 DNA ligase buffer with 10mM ATP (NEB), 0.6μL 5mM dATP and 10 units of Klenow Fragment (3´→5´ exo–) (NEB) to 21μL of the heat-killed SAP reaction for a final volume of 30μL. Barcoded adapters were annealed in a 19μL reaction containing 2.5μM of each adapter strand and 5 μL of 10X T4 DNA ligase buffer. The mixture was heated to 94ºC and slowly cooled to room temperature (1ºC/s). The annealed adapters were mixed with the end-repaired DNA and 1μL of 10X T4 DNA ligase (NEB). The reaction was then incubated at 16ºC for 1 hour, 65ºC for 10 minutes, and purified using the Qiagen MinElute PCR purification kit. A single barcode was used for each antibiotic plate.

The DNA fragments were then size selected on a 2% agarose gel in 1X TBE buffer stained with Biotium GelGreen dye (Biotium). DNA fragments were combined with 2.5uL 6X Orange loading dye before loading on to the gel. Adaptor-ligated DNA was extracted from gel slices corresponding to DNA of 150-220bp using a QIAGEN MinElute Gel Extraction kit. The purified DNA was enriched by PCR using 12.5μL 2X Phusion HF Master Mix and 1μL of 10μM Illumina PCR Primer Mix in a 25μL reaction using 1μL of purified DNA as template. DNA was amplified at 98ºC for 30 seconds followed by 18 cycles of 98ºC for 10 seconds, 65ºC for 30 seconds, 72ºC for 30 seconds with a final extension of 5 minutes at 72ºC. Afterwards, the DNA concentration was measured using the Qubit fluorometer and 10 nanomoles of each sample were pooled into a final volume of 190.9μL. Subsequently, samples were submitted for Illumina Hi-Seq Pair-End (PE) 76bp sequencing using the GAIIx Genome Analyzer at GTAC (Genome Technology Access Center, Washington University in St. Louis) at 5pmol per lane.

Experimental Development of PARFuMS

In overview, our strategy selects a small-insert (1-3Kb) library of *Escherichia coli* metagenomic transformants on solid media for a desired phenotype (e.g. antibiotic resistance) before all surviving

colonies are pooled into a liquid slurry and metagenomic inserts amplified via PCR. Inserts from a given selection are prepared into Illumina sequencing libraries and barcoded with a unique 5-8bp tag, allowing up to hundreds of selections to be pooled on a single lane for sequencing. To simultaneously assemble multiple variable-length inserts, we implemented an iterative approach in which intermediate-length contigs generated from multiple rounds of assembly with the short-read assembler Velvet (*32*) are assembled into full-length contigs using the long-read assembler Phrap (*33*).

In initial attempts to develop a high-throughput protocol for the assembly of functional metagenomic selections, we used the short-read assembler Velvet (*32*) to assemble functional selections of a defined metagenomic library comprised of six previously Sanger-sequenced, β-lactamase containing fecal metagenomic fragments *(20)*. The six fragments were chosen to evenly represent the range of sizes expected in a standard small-insert library (1350 - 3300bp); each fragment was cloned into the pZE21 expression vector (*30*), transformed into *E. coli* TOP10 cells (Invitrogen), and plated individually or in an equimolar pool on solid media containing 50 µg/mL Penicillin-G. Surviving colonies from all seven selection conditions were collected as liquid slurries, resistant fragments amplified and barcoded separately, and the samples prepared for Illumina sequencing. The sequencing run yielded approximately 1 million paired-end reads per barcode, which were subsequently assembled independently using Velvet. While all six individual samples successfully assembled into full-length contigs using this simple protocol, 36 short contigs were generated for the pooled sample, with an N50 of 190bp, and the largest contig assembled at 1037 bp. Therefore, we determined Velvet alone was insufficient to *de-novo* assemble diverse metagenomic fragments from functional selections.

To simultaneously assemble multiple variable-length fragments more completely, we implemented an iterative assembly approach in which intermediate-length contigs generated from multiple rounds of assembly with Velvet are assembled into full-length contigs using the long-read assembler Phrap (*33*). Our scheme begins with three iterations of assembly with Velvet: the first iteration assembles all reads, while the second and third use only reads that were not present in any previously assembled contig. Each round of assembly is split into jobs of a defined number of reads, assembly of each job occurs in parallel, and assemblies are combined at the end of each round. From one round of assembly to the next, the number of reads per job varies, decreasing as assembly progresses and dataset complexity is reduced. This approach regulates the coverage over any given contig, ensuring the abundance of reads with sequencing error in a particular assembly remains low, thus preventing inappropriate splitting of contigs. After each round of assembly, redundant contigs are collapsed to one sequence and chimeras removed using a window-based coverage approach. The Velvet-assembled contigs are further assembled using two iterations of Phrap. The first iteration assembles the Velvet output into contigs that are subsequently linked together by overlap with raw reads. Phrap is run a final time using the linked contigs as input; the output from this assembly is then annotated using a BlastX-similarity method against the COG functional database (*34*). Subsequently, individual contigs are joined based on sequence similarity and common annotation to generate the final assemblies, which are re-annotated through similarity to the COG functional database.

To establish our assembly and annotation pipeline (PARFuMS), we tested it on five defined metagenomic libraries of the six previously Sanger-sequenced, β-lactamase containing fecal metagenomic fragments. We pooled the six fragments together at varying relative abundance: in one pool all fragments were mixed in equal proportions while, in four others, one fragment was inoculated at 10x abundance relative to the other five (table S15). Each pool was plated on solid media with 50 µg/mL Penicillin-G and resistance fragments were amplified from surviving colonies, barcoded by pool, sequenced, and assembled. Despite sharing a lane with five unrelated samples, no contig was assembled that did not map to one of the original six metagenomic fragments, indicating the barcoding strategy

successfully binned reads by sample. Assembly with PARFuMS revealed that for each of the 30 fragments (five libraries, six fragments per library), the annotated β-lactamase was entirely contained in a single contig and at least 95.7% of the original fragment was covered by at most two contigs, demonstrating successful *de-novo* assembly for the recovery of full-length resistance elements in all samples. Additionally, in 23 of the 30 cases, fragments were assembled into a single contig covering at least 96.2% of the original sequence, yielding near-complete assembly of the original metagenomic fragment (fig. S6, table S16). Therefore, PARFuMS was able to simultaneously and completely assemble an array of functionally selected metagenomic inserts varying not only in length, but also by up to an order of magnitude in relative abundance.

We subsequently selected and re-sequenced a previously Sanger-sequenced metagenomic library *(20)* before assembling and annotating all resistance-conferring metagenomic DNA fragments using PARFuMS. This fecal metagenomic library ('HG1') was selected on solid media containing one of eight antibiotics (table S14), and resistance genotypes determined using PARFuMS. From these eight selections, we assembled 38 contigs with an aggregate length of 45.7Kb and an N50 length of 1.7Kb, within the expected size range for a library made from 1-3Kb fragments (table S17). Upon comparison with the previously-generated Sanger data *(20)*, we observed 95% of the original data present in the newly-assembled contigs (subjected to the same eight antibiotic selections *(20)*), indicating our methods sufficiently recapitulate traditional strategies. Furthermore, these data demonstrated an increased capacity to discover novel function: we observed that 19.4% of the assembled sequence data generated with our high-throughput protocol (which pools all selected clones) was not present in the original Sanger dataset (where a ‘representative’ number of clones were picked for sequencing). Importantly, 15% of the ORFs annotated as resistance genes from the new data were not present in the original Sanger dataset (below 55% identity over the length of the ORF), confirming the novel sequence data contains functional resistance determinants (fig. S7). For example, a 648bp chloramphenicol-O-acetyltransferase with 69.5% similarity to *Clostridium saccharolyticum* WM1 was fully encoded within a 1564bp fragment undiscovered in the original Sanger sequenced experiment *(20)* (table S18). These results indicate our approach was able to recapitulate the data generated using traditional methods, and uncover novel resistance elements, all at less than 1/1000th the per-base cost of Sanger-based methods of discovery.

Note on Comparative Analysis of Sanger-Sequenced vs. PARFuMS-Processed HG1 Functional Selections

For two antibiotics (amoxicillin and carbenicillin), the MICs in this experiment were at least two-fold higher than previously determined in the earlier, Sanger sequencing project *(20)*. In these instances, significant background growth with empty vector controls was observed when libraries were plated on solid media containing the previously-determined MICs *(20)*, necessitating an increased concentration of antibiotic in selections. We observed that these increased concentrations of antibiotic resulted in an concomitant decrease in the number of colonies following selection, indicating a potential decrease in the number of metagenomic fragments that withstood selection. As an increased capacity for antibiotic resistance is required for survival on the higher MICs, metagenomic fragments that previously conferred resistance to *E. coli* during the original Sanger-sequencing experiments may not appear on selections using increased antibiotic concentrations. Thus, metagenomic fragments from amoxicillin and carbenicillin selections that were not recovered using the elevated MICs (three total contigs) were not considered in comparative analyses.

Assembly of Illumina Short-Reads and Subsequent Annotation (Details of PARFuMS)

Raw Illumina PE reads were initially binned by barcode (exact match required) such that independent selections are assembled and annotated in parallel. Subsequently, reads with similarity to either the expression vector or Illumina adapters were removed from further analysis using cross_match (*33*) with the following options: -gap1_only -minmatch 6 -minscore 10 -gap_init -3. Then, reads were

assembled using Velvet (*32*) in sets of 10,000 reads using the following parameters: (a) velveth: hash_length: 31, -shortPaired flag enabled, (b) velvetg: -cov_cutoff: 10 -ins_length: 100 -min_contig_lgth: 100. The resulting contigs from all the sub-assemblies per selection were stored in a single file and redundant contigs were collapsed using CD-HIT (*35*) with the following options: -g 1 -r 1 –c 0.9 (contigs are collapsed if at least 90% identical over the length of the shortest sequence). Raw reads were then mapped to the assembled contigs using FR-HIT (*36*) (80% similarity over at least 30 nuc.) and chimeric contigs identified based on extreme drops in coverage within the contig. Chimeric contigs were then split at the chimeric position and all the contigs were dereplicated again using CD-HIT (90% identity). Subsequently, raw reads were mapped to the current contig set with FR-HIT (same parameters as earlier) to identify reads that do not map to any contig. These reads were then seperately assembled into contigs using Velvet in jobs of 1,500 reads with the following parameters: (a) velveth: hash_length: 31, -shortPaired flag enabled, (b) velvetg: -cov_cutoff: 7 -ins_length: 80 -min_contig_lgth: 100. These contigs were then added to the current contig set and all contigs were subsequently collapsed by 90% identity. FR-HIT was then used to identify reads that still did not map to any assembled contig, and the entirety of these reads were assembled into contigs a final time using Velvet (parameters: (a) velveth: hash_length: 31, -shortPaired flag enabled, (b) velvetg: -cov_cutoff: 10 -ins_length: 80 -min_contig_lgth: 100). These newly assembled contigs were then added to the total set, and any contigs that map to either adapter or vector sequence were removed. This final Velvet-assembled set was collapsed on 90% identity a final time to generate a non-redundant set for assembly using Phrap (*33*).

The non-redundant, Velvet-assembled contig set was then checked for chimeras by mapping raw reads to the contigs (with FR-HIT) and searching for drops in coverage, as performed earlier. After splitting any chimeric sequences, the contig set served as input for the long-read assembler Phrap, using the following options: -minmatch 25 -maxmatch 40 -bandwidth 1 -minscore 30 -penalty -5 -gap_init -4 -gap_ext -3. Assembled contigs from the *.contigs, *.singlets, and *.problems output were saved to a single file and redundancy checked by collapsing at 90% sequence identity. Raw reads were then mapped to the Phrap-assembled contigs with FR-HIT (same parameters), and if any two contigs were bridged by a sufficient number of reads (default >= 5), the two contigs were linked, fusing inappropriate contig splits. The bridging was performed using pairs of PE reads for which each read maps to the edge of different contigs pointing outwards. Subsequently, Phrap was run a second time (using the parameters: -minmatch 25 -maxmatch 40 -bandwidth 1 -minscore 30) to promote the more complete assembly of the linked contigs. The contig set was then collapesed for redundancy at 90% sequence identity before it was annotated using BlastX against the COG functional database. In cases where two ends of different contigs had significant similarity to non-overlaping fragments from a reference sequence in the appropriate orientation, they were joined together as a final measure to fuse incompletely assembled contigs. Redundant contigs were removed from the final set via clustering at 90% identity with CD-HIT. Lastly, the final contigs were re-annotated via BlastX comparison against COG. Annotations were outputted to a tab-delimited file and the corresponding contig sequences were stored in FASTA format.

Creation of Approximate Maximum-Likelihood β-lactamase Trees

Four datasets of β-lactamase amino acid sequences were used to create the approximate-maximum likelihood tree depicted in figure S3. The datasets used are as follows:

Dataset A: All genes in our set that were at least 350nt in length and annotated as a β-lactamase were selected for multiple alignment. Subsequently, these genes were compared against NCBI protein Non-Redundant (NR) database (retrieved June 6, 2011) with BlastX. The translation frame that produced the best blast hit was used to translate the nucleotide sequence into an amino acid sequence. These amino acid sequences (55 total) represent the AB95 β-lactamases used to create the approximate maximum-likelihood in figure S3A.

Dataset B: From the BlastX query performed between the AB95 β-lactamase gene sequences and NCBI protein NR database, every entry in GenBank that corresponded to (i) the best hit to an AB95 β-lactamase or (ii) an entry with 99% identity to an AB95 β-lactamase and a bit score greater than 200 was selected. The amino acid sequences from these entries were used to create the maximum-likelihood tree in figure S3A (72 total). If the sequence originated from a clinical isolate of a human pathogen, the branch on the tree was colored blue. Otherwise, the branch was colored black.

Dataset C: A *Streptomyces* β-lactamase protein set was obtained as follows: in NCBI-protein, the following search was used to obtain a 616-member database of β-lactamases originating from *Streptomyces*: txid1883[Organism] AND beta lactamase[Protein Name]. A BlastP query was then performed against this database using the 55 translated amino-acid sequences from dataset A. Any unique amino acid sequence from a *Streptomyces* β-lactamase mapping to an AB95 β-lactamase with an e-value below 1e-10 was used in the creation of the maximum-likelihood tree in figure S3A (10 total).

Dataset D: Fourteen β-lactamase amino acid sequences isolated from remote Alaskan soil were used as a reference set to represent the resistome of pristine soil (*11*). The GI numbers of these β-lactamases are as follows: 169647647, 197296820, 197296825, 197296827, 197296828, 197296829, 197296830, 197296831 197296834, 197296837, 197296838, 197296839, 197296842, 197296845.

All β-lactamase amino-acid sequences from datasets A-D (described above, 151 total sequences) were collapsed into clusters of sequences over 99% identical over 80% of the length of the shortest sequence with CD-HIT (*35*). The collapse was performed in a stepwise fashion: first, sequences were clustered on 100% identity (-c 1) over the 80% the length of the shortest sequence (-aS 0.8), and the longest representative member of each of these clusters served as input for clustering on 99% identity (-c 0.99). Each stepwise cluster was generated using the following additional options: -g 1 -d 0 -G 0.

Subsequently, a multiple alignment of the longest representative sequence from each 99%ID cluster (including all singletons) was performed using Muscle v3.7 (*37*) with the default parameters. This multiple alignment was used to create an approximate maximum-likelihood tree using FastTree v2.1 (*38*) with default parameters. In order to display all sequences from datasets A-D (rather than only the representative sequences from the 99% cluster), the following actions were taken: (i) the longest representative sequence of a 99% cluster (the sequence for which multiple alignment was performed) served as a node from which all members of that cluster branched, (ii) if sequences were also a member of the 100% cluster, their representative sequence in the 99% cluster acted as a node from which all 100%ID sequences branched. No sequence was represented more than once on the tree and each branch that was added due to membership in a CD-hit cluster was given a length of 0.01. All sequences added to the tree in this fashion were at least 99% identical to a sequence used in the multiple alignment. These sequences were then collapsed into triangles in order to (i) collapse all supplemental branches added to the tree into one node, thus preserving the tree's original topology, and (ii) emphasize the most closely related sequences from datasets A-D. Branches and triangles were subsequently colored based on the dataset from which they originated, and the number of sequences in any given triangle were denoted to its right.

The tree in figure S3B was created from a multiple alignment of the sequences represented by the boxed triangle in figure S3A. Partial amino acid sequences from Dataset B were not included in this alignment as the end-gaps that resulted from incomplete sequence data dramatically affected tree topology. Muscle v3.7 was used for multiple alignment and FastTree v2.1 was used to create the approximate maximum-likelihood tree. Default parameters were used for both programs.

Note on the Resuscitation of Frozen AB95 Cultures

The AB95 metagenomic library was created with DNA extracted from an equal abundance pool of 95 highly antibiotic resistant, soil-derived cultures. The antibiotic-selected cultures comprising this pool were frozen and stored at -80°C days prior to the creation of the library from the mixed culture. Of these

95 cultures, 80 supported freeze/thaw under standard procedures (i.e. were revived from frozen stock and became available for analyses beyond the initial selections). Of the nine unique antibiotic resistance genes with approximately 100% identity to human pathogens, eight were identified in this set of 80 cultures, consistent with the proportion of samples successfully recovered from frozen stock.

BlastP Comparisons of β-lactamases from Various Microbial Populations

β-lactamase proteins from both Proteobacteria and Acidobacteria were downloaded from NCBI-protein with the following searches: "Proteobacteria[Organism] AND beta lactamase[Protein Name]"; "Acidobacteria[Organism] and beta lactamase[Protein Name]". These databases (with 144 and 2757 sequences respectively), in addition to the *Streptomyces* β-lactamases (Dataset C, page 7), were compared against the 55 AB95 β-lactamase sequences used to generate figure S3 (Dataset A, page 7) using BlastP. The best Blast hit for each AB95 β-lactamase was used to generate the average and median bit scores reported in table S6.

Percent Identity Comparisons of AB95 Open Reading Frames against Various Databases

BlastX comparisons between 110 AB95 open reading frames and amino acid sequences from both the NCBI protein Non-Redundant (NR) database (retrieved June 6, 2011) and various other intestinal datasets were performed using default parameters. The best local alignment from this output was subsequently used to compute global identities over the full length of the shorter of the two sequences. The intestinal datasets queried against are as follows: (A) a set of antibiotic resistant genes from two metagenomic libraries prepared with a set of isolates cultured from the fecal samples of two healthy adults *(20)*, (B) a set of 128 representative organisms of the human gastrointestinal tract (table S12), (C) a set of antibiotic resistance genes from two metagenomic libraries prepared from bulk DNA isolated from fecal samples of two healthy adults *(20)*.

Determination of Source Organisms for Selected Resistance Genes from the AB95 Metagenomic Library

To determine the source organism(s) from the AB95 culture collection containing antibiotic resistance genes referenced in the main manuscript, PCR primers were designed specifically to the boundaries of the predicted ORF (table S19). PCRs (using these primers) were performed with the PFX polymerase using genomic DNA extracted from each of the original AB95 cultures as template, per manufacturer recommendations (http://tools.invitrogen.com/content/sfs/manuals/platinumpfx_pps.pdf). Genomic DNA was extracted using the UltraClean Microbial DNA Isolation Kit (MoBio Laboratories Inc), per suggested protocols (http://www.mobio.com/images/custom/file/protocol/12224.pdf). Cultures for which PCR yielded an amplicon of the expected size were streaked onto Luria-Bertani agar plates, and PCR used to identify single colonies containing a given resistance gene. The identity of the resistance gene was further confirmed via Sanger sequencing of the corresponding PCR amplicon, while the taxonomy of the source organism was determined via 16S PCR performed on genomic DNA extracted from axenic culture, grown from individual colonies (gDNA isolated as described above). The D-cycloserine resistance gene originated from AB95 culture H9 whereas both the *Ochrobactrum* and *Pseudomonas* species referenced herein (as well as their respective resistance genes) originated from AB95 culture H2 (table S3).

16S and rpoB PCR and Sequence Analysis

16S PCR was accomplished using the following universal primers: 27F (5' AGAGTTTGATCCTGGCTCAG-3') and 1391R (5' GACGGGCGGTGWGTRCA-3'), along with the 1.1X ReddyMix PCR Master Mix (Thermo Scientific), according to the manufacturer's recommendations (https://fscimage.fishersci.com/images/D11108_U~.pdf). PCR to amplify a portion of the *rpoB* gene for phylogenetic analysis was performed using the following primers: Forward – (5'

CGAACATCGGTCTGATCAACTC-3') and Reverse – (5' CGCTGCATGTTCGAACCCAT-3'). All Sanger sequencing was performed either at the Protein and Nucleic Chemistry Laboratory (PNACL) sequencing core at Washington University or with Beckman Coulter Genomics (Danvers, MA).

Determination of Minimum Inhibitory Concentrations

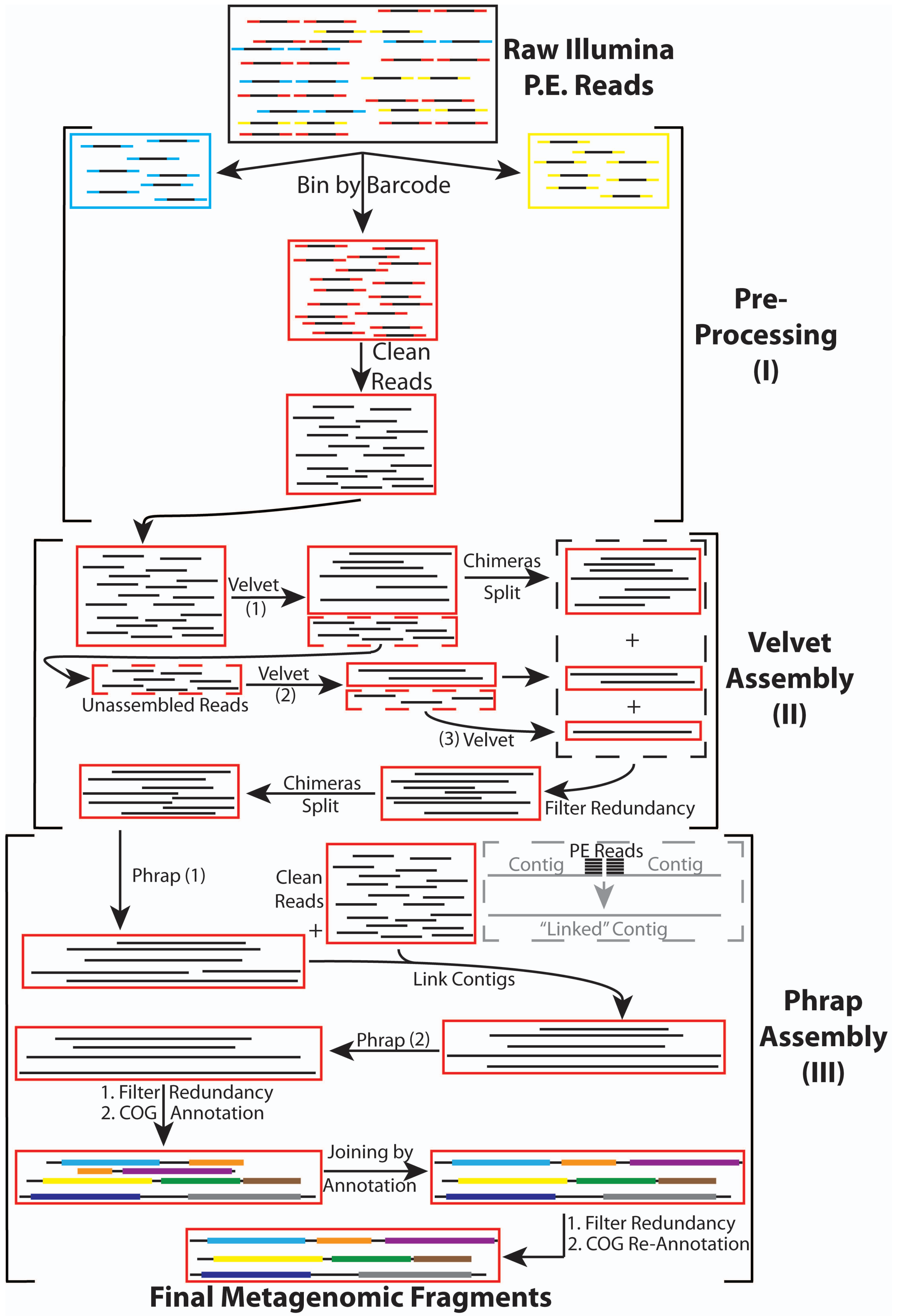
Minimum inhibitory concentrations were performed in liquid media, in the case of the original soil isolates, the indicated antibiotics were prepared in Luria-Bertani broth at the reported concentrations. In the case of the resistance genes cloned in to *E. coli*, ORFs were initially amplified via PCR from genomic DNA using a blunt-end polymerase (PFX, Invitrogen) and 5'-phosphorylated primers (table S19), according to suggested protocols (http://tools.invitrogen.com/content/sfs/manuals/platinumpfx_pps.pdf). Subsequently, the ORFs were cloned into the HincII site of the pZE21 expression vector (*30*) and transformed into *E. coli* MegaX cells (Invitrogen) before minimum inhibitory concentrations were determined. Determinations were performed using Mueller-Hinton broth (2g Beef Infusion Solids, 1.5g Starch, 17.5g Casein hydrolysate, pH 7.4, in a final volume of 1L) and the appropriate antibiotics at the reported concentrations. All minimum inhibitory concentrations were performed in a final volume of 200µL, with a 2.5% initial cellular inoculum. Growth was profiled using the Synergy H1 microplate reader (Biotek Instruments, Inc) for a minimum of 24 hours. In the case of the original soil isolates, growth assays were performed at room temperature, whereas *E. coli* experiments were performed at 37°C.

Identification of Resistance Genes in Plasmid Fraction of the AB95 *Pseudomonas* Isolate

Overnight cultures of the AB95 isolate most closely related to *Pseudomonas sp.* K94.23 were centrifuged at 4,000 RPM for 5 minutes. Plasmid DNA was recovered from the resultant cell pellets using the Qiagen Spin Miniprep Kit (Cat. No. 27106), according to manufacturer's recommendations. To permit the recovery of larger plasmids (>10Kb), elution was performed using buffer preheated to 70ºC. Subsequently, PCR was used to determine the presence or absence of seven antibiotic resistance genes shared between the *Pseudomonas* isolate and human pathogens, listed in table S11. PCR was performed using *taq* polymerase (NEB cat. No. M0267L) and the gene-specific primers listed in table S19, following manufacturer's recommendations (http://www.neb.com/nebecomm/products/protocol567.asp).

Determination of Fold-Reduction in Cost of Characterizing Resistomes with PARFuMS

A single lane of sequence data from the Illumina Hi-Seq was estimated to cost $3250 (current price at GTAC) and allows 200 selections to be processed using PARFuMS with an average of 700,000 reads per selection (assumes 140 million reads per lane). To calculate the number of functional metagenomic selections that could be sequenced using traditional Sanger-based methods for the same price, several assumptions were made. First, it was assumed six Sanger sequencing reactions (using four custom-ordered primers and two universal primers designed against the expression vector) would be required to entirely sequence any given metagenomic insert from a resistant clone. If costs average $3 per sequencing reaction (current price at Beckman Coulter) and $2 per custom-ordered primer (current price at Integrated DNA Technologies), the cost of sequencing the metagenomic insert from any given clone would be $26. Assuming the average functional metagenomic selection produces 100 resistant clones (an underestimate compared to the mean colony counts observed in selections with the AB95 library), a single selection would cost $2600 to sequence entirely with Sanger-based methods. Thus, 1.25 selections could be sequenced for the same $3250 it costs to sequence 200 functional metagenomic selections with PARFuMS. This results in a 160-fold decrease in cost, per selection, using PARFuMS.



Raw Illumina P.E. Reads
Bin by Barcode
Clean Reads
Pre-Processing (I)
Velvet (1)
Chimeras Split
Unassembled Reads
Velvet (2)
(3) Velvet
Velvet Assembly (II)
Chimeras Split
Filter Redundancy
Phrap (1)
Clean Reads
Contig
PE Reads
Contig
"Linked" Contig
Link Contigs
Phrap (2)
Phrap Assembly (III)
1. Filter Redundancy
2. COG Annotation
Joining by Annotation
1. Filter Redundancy
2. COG Re-Annotation
Final Metagenomic Fragments

Figure S1: Scheme by which PARFuMS assembles and annotates short-read sequencing data into full-length metagenomic inserts. Our method begins with a Pre-processing stage (I) in which reads are binned by their sequence barcodes and quality filters applied. The clean reads then go through with three iterations of assembly with Velvet (II): the first iteration assembles all reads, while the second and third use only reads that were not present in any previously assembled contig. After each round of assembly, redundant contigs are collapsed to one sequence and chimeras removed using a window-based coverage approach. The Velvet-assembled contigs are further assembled using two iterations of Phrap (III). The first iteration assembles the raw Velvet output into contigs that are subsequently linked together by overlap with raw reads. Phrap is run a final time using the linked contigs as input; the output from this assembly is then annotated using a BlastX-similarity method against the COG functional database (*34*). Subsequently, individual contigs are joined based on sequence similarity and common annotation to generate the final assemblies, which are re-annotated through similarity to the COG functional database. Note that some methodological details have been excluded from the above figure to promote clarity (e.g. number of sequence reads per job, several filters for redundancy). Please see methods within the supporting online material for a more detailed description.

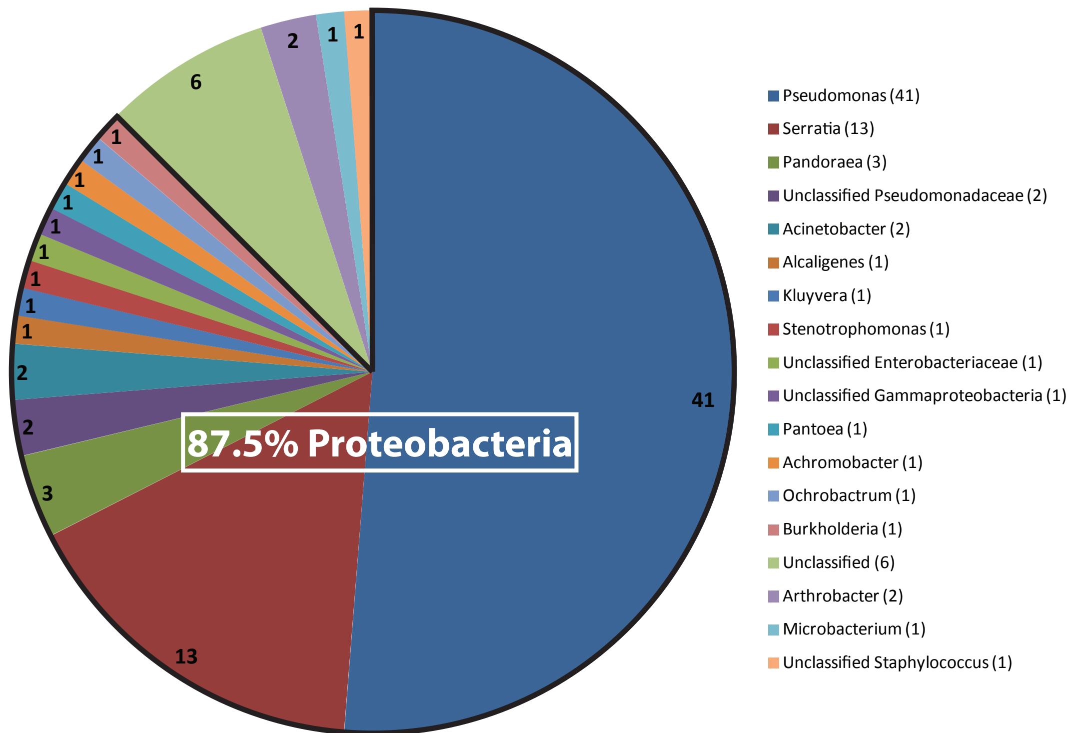


Figure S2: Eighty representative isolates were obtained from the AB95 culture collection by streaking for individual colonies on LB agar plates. Subsequently, 16S rDNA gene sequence was used to classify all 80 organisms by genus. The number of isolates in a given genus is indicated with numbers both on the pie chart and to the right of genera in the legend. Proteobacterial isolates are displayed within the black outline (87.5% of total isolates are proteobacteria).

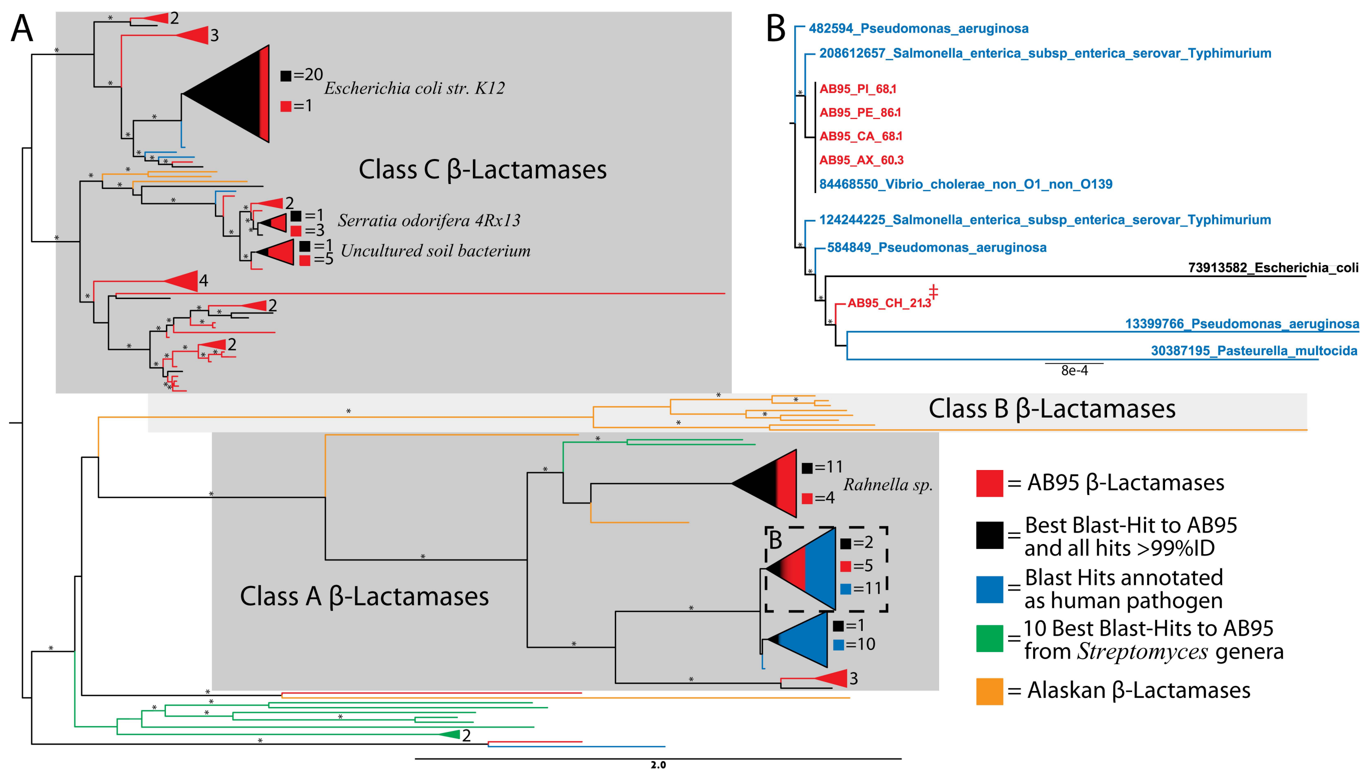


Figure S3: Approximate maximum-likelihood trees of aligned amino acid sequences from AB95 β-lactamases, closely related β-lactamases (determined by BlastX), top hits from *Streptomyces*, and recently published β-lactamases from Alaskan soil (*11*). The amino acid sequences from 55 AB95 β-lactamases were compared to three datasets: (i) the top BLAST hit to each enzyme, as well as GenBank entries with over 99% identity, (ii) β-lactamases from *Streptomyces* with the closest identity to AB95 genes, allowing comparison against enzymes from antibiotic producers, and (iii) β-lactamases from a recent Alaskan soil metagenomics project (*11*). An approximate maximum-likelihood tree was generated from a multiple alignment including all 151 sequences. Nodes with Shimodaira-Hasegawa like branch supports greater than 0.7 are marked with asterisks. (**A**) Tree formed from a multiple alignment of all β-lactamase sequences. Shading highlights clades of canonical β-lactamase families. Un-shaded nodes represent sequences not confidently categorized into a particular class. Sequences with over 99% identity over 80% of the length of the shorter sequence are collapsed into triangles. The number of sequences in each triangle belonging to a particular color group is indicated to the right of the triangle, as are the identities of GenBank entries with >99%ID to AB95 β-lactamases. For example, 20 sequences deposited in GenBank from *E. coli* str. K12 are over 99% identical to a single class C β-lactamase from the AB95 dataset. Similarly, five AB95 β-lactamases are greater than 99% identical to 11 sequences from human pathogens as well as 2 non-pathogenic resistance genes. (**B**) Tree formed from a multiple alignment of only the sequences in the collapsed triangle indicated by the dashed box. Where applicable, GI numbers for GenBank sequences are given. Partial amino acid sequences from the collection of top Blast hits were removed before alignment of this tree, and thus, the number of nodes does not match the numbers to the right of the corresponding collapsed triangle in (**A**). The AB95 sequence marked with "‡" represents a partial gene, all other AB95 sequences are full genes. Node-labels aligned with the right edge of the figure represent sequences with <100% identity to AB95 β-lactamases, all other nodes are 100% identical across the length of the shortest sequence, varying in length due to the reported start codon (tree topology in these cases is determined by the number of end-gaps in multiple alignment). The

two-letter codes in the names of AB95 nodes represent the antibiotic selection from which the sequence was isolated (PI=piperacillin, PE=penicillin, CA=carbenicillin, AX=amoxicillin, CH=chloramphenicol (β-lactamase captured as a partial gene in CH selection due to proximity to a chloramphenicol O-acetyltransferase)). The scale bars represent the number of substitutions per site.

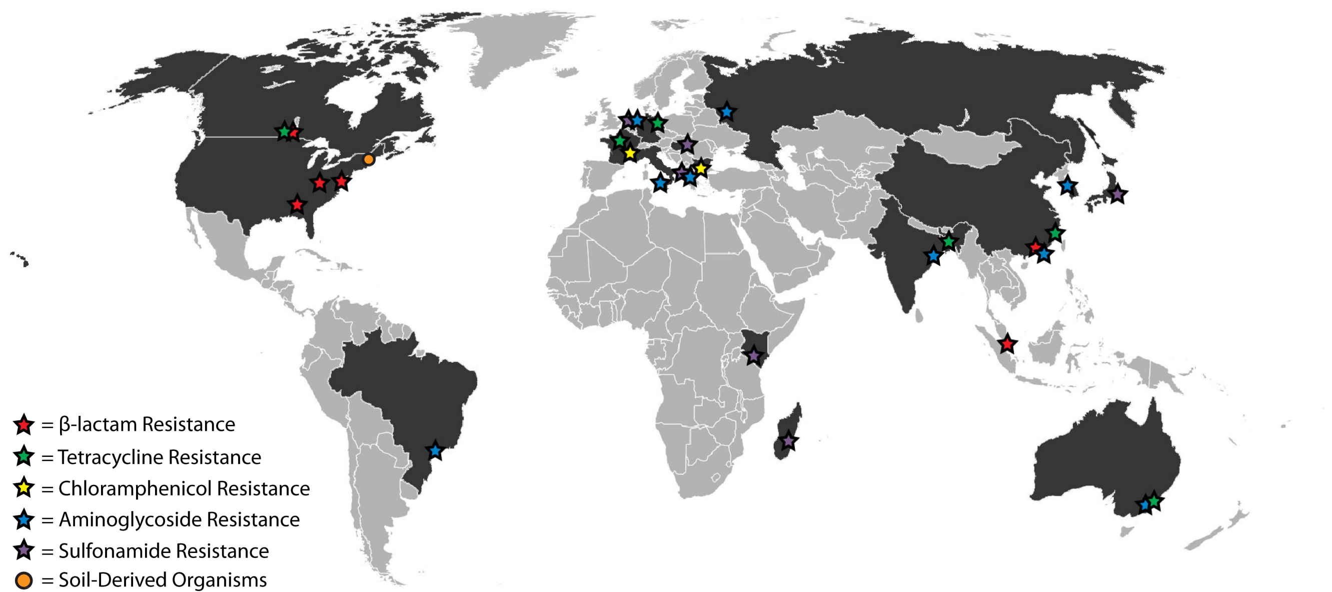
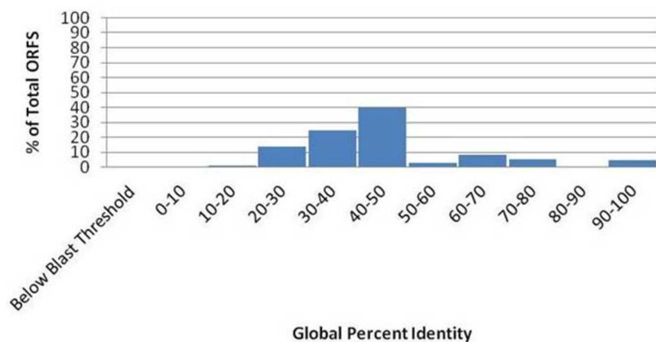
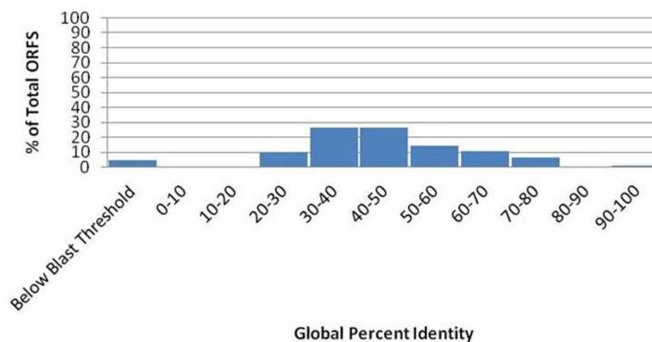


Figure S4: Geographical distribution of pathogenic resistance genes with perfect nucleotide identity to antibiotic resistance in soil. Stars depict locations of clinical isolates containing the indicated resistance and shaded countries are those from which the isolates originated. All geographical information pertaining to clinical pathogens was obtained from the NCBI nucleotide collection (nr/nt). In many cases, more pathogens exist with perfect identity to soil-derived resistance then are shown; in these instances only representative isolates are pictured. The circle depicts the origin of soil-derived organisms that share resistance with pathogens.





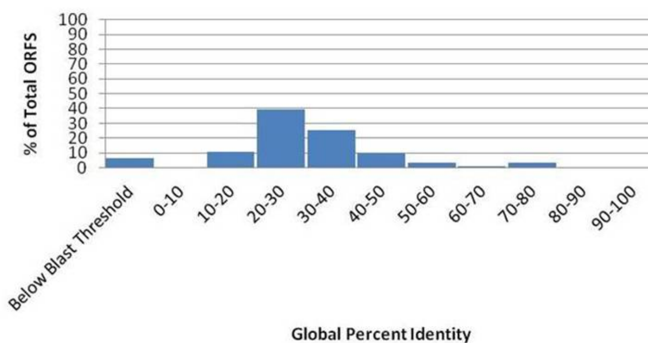


Figure S5: Percent identity between 110 AB95 antibiotic resistant ORFs and three complementary datasets profiling the human intestinal microbiome. Global amino acid identities are indicated on the x-axis, binned in 10% increments. The y-axis displays the percentage of total ORFs queried contained in a given identity bin. Comparisons were performed over entire length of the shorter sequence using BlastX.

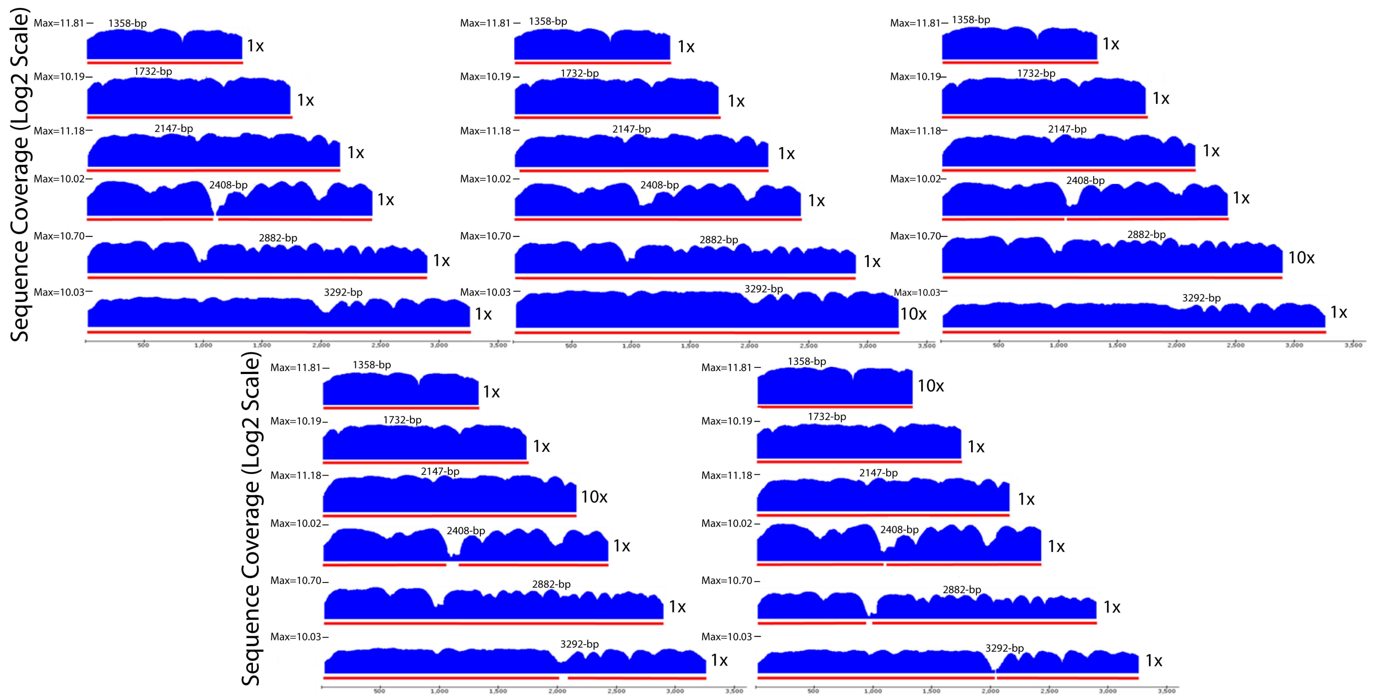


Figure S6: Sequence coverage, per base, of six pooled β-lactamase containing fragments is depicted in log2 scale (blue); red horizontal bars represent the contigs generated. Results from all five pools are shown above, for numeric values corresponding to these images, refer to table S16. The relative proportion of a particular fragment in a given pooling experiment is indicated to the right of the fragment while the maximum coverage (log2 scale) across all pools for a specific fragment is noted on the y-axis. The x-axis indicates position along the length of the original metagenomic fragments in base pairs. Images generated using the Lightweight genome viewer (*39*).

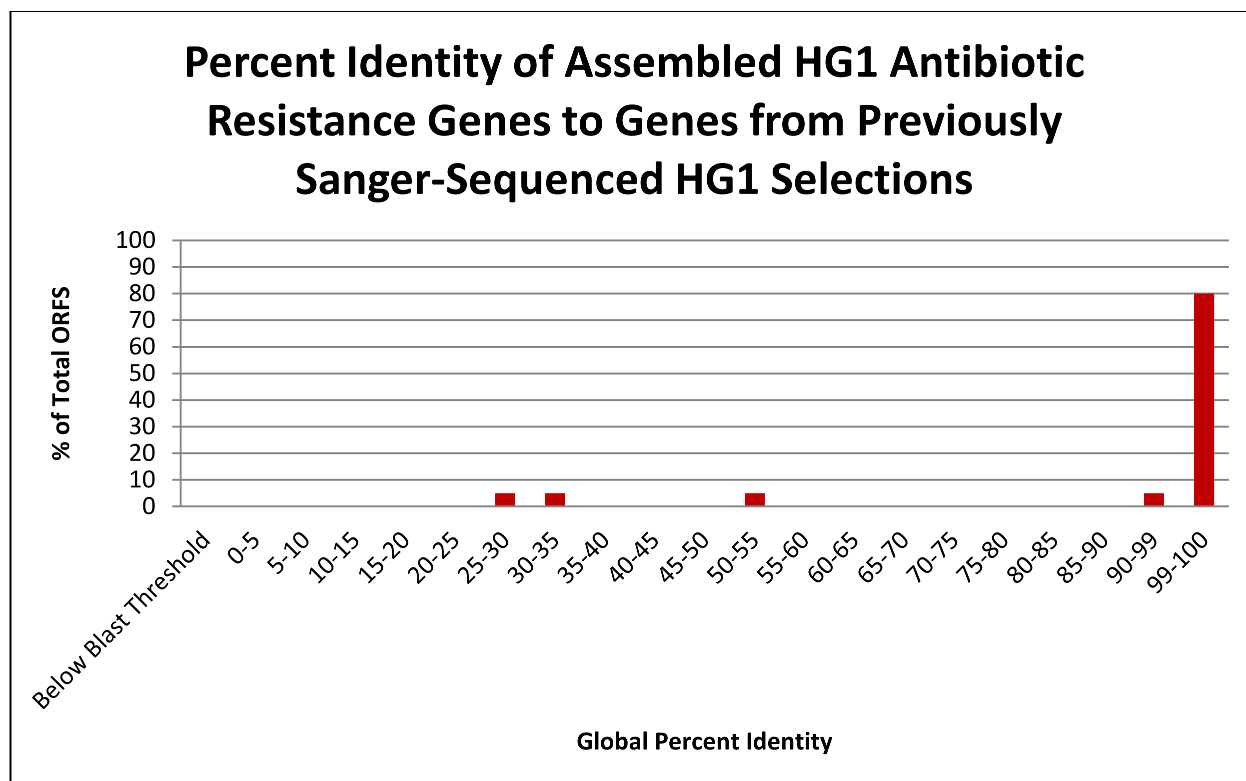


Figure S7: All antibiotic-resistant ORFs from functional selections of HG1 against eight antibiotics were compared against a collection of antibiotic resistant genes generated from previously-reported Sanger-sequenced selections using the same library, against the same antibiotics *(20)*. The percent identity between an Illumina-sequenced resistance ORF and its best hit from the database of Sanger-sequenced ORFs was recorded and is displayed as a fraction of total ORFs queried, binned in 5% increments (with the exception of the final two bins). Any gene less than 55% identical to its best hit from the previously-sequenced database was deemed novel and reported as such. In total, 15% of the antibiotic resistance genes discovered using the Illumina-based method were not present in the original Sanger dataset, indicating the capacity of the newer method to uncover novel function. Comparisons were performed using BlastX, where the nucleotide query was an Illumina-sequenced resistance ORF and the subject was the translation of a Sanger-sequenced open reading frame.

Table S1: Sample information on 11 soils from which antibiotic-resistant bacteria were cultured, and pooled together, to create the metagenomic library 'AB95'.

Soil Name	Soil Type	Soil Location
S1	Farmland	Corn Field with Antibiotic Treated Manure, Great Brook Farm, Carlisle, MA
S2	Farmland	Alfalfa Field with Manure Treatment, Northcroft Farm , Pelican Rapids, MN
S3	Farmland	Alfalfa Field without Manure Treatment, Northcroft Farm, Pelican Rapids, MN
S4	Forest	Raccoon Ledger, Rockport, MA
S5	Prairie	Pelican Rapids, MN
S6	Swamp	Brier's Swamp, Rockport, MA
S7	Forest	Alan Seeger Natural Area, PA
S8	Forest	Toftrees State Gameland Area, PA
S9	Urban	Waste Water Treated Area, Toftrees State Gameland Area, PA
S10	Urban	Boston Fens, MA
S11	Urban	Boston Public Garden, MA

Table S2: Catalog of the 18 antibiotics used to enrich culture for antibiotic resistant bacteria from 11 soils in creating the AB95 metagenomic library.

Antibiotics	Class
Ciprofloxacin	Fluoroquinolone
Levofloxacin	Fluoroquinolone
Sisomicin	Aminoglycoside
Gentamicin	Aminoglycoside
Kanamycin	Aminoglycoside
Amikacin	Aminoglycoside
PenicillinG	β-lactam
Carbenicillin	β-lactam
Dicloxacillin	β-lactam
Chloramphenicol	Amphenicol
Nalidixic acid	Quinolone
Thiamphenicol	Amphenicol
Sulfisoxazole	Sulfonamide
Trimethoprim	Dihydrofolate Reductase Inhibitor
Mafenide	Sulfonamide
Sulfamethizole	Sulfonamide
D-Cycloserine	Amino Acid Derivative
Vancomycin	Glycopeptide

Table S3: Soils and antibiotics used in the original isolation of the 95 AB95 cultures. Refer to table S1 for descriptions of soils.

AB95 Culture	Antibiotic Selection	Soil Origin	AB95 Culture	Antibiotic Selection	Soil Origin
A1	Levofloxacin	S8	E1	Vancomycin	S9
A2	Gentamicin	S8	E2	Levofloxacin	S10
A3	Kanamycin	S8	E3	Sulfisoxazole	S10
A4	Amikacin	S8	E4	Trimethoprim	S10
A5	PenicillinG	S8	E5	Mafenide	S10
A6	Carbenicillin	S8	E6	Sulfamethizole	S10
A7	Chloramphenicol	S8	E7	Vancomycin	S10
A8	Nalidixic acid	S8	E8	Ciprofloxacin	S6
A9	Thiamphenicol	S8	E9	Amikacin	S6
A10	Trimethoprim	S8	E10	PenicillinG	S6
A11	Mafenide	S8	E11	Dicloxacillin	S6
A12	D-Cycloserine	S8	E12	Mafenide	S6
B1	Vancomycin	S8	F1	Vancomycin	S6
B2	Sisomicin	S3	F2	PenicillinG	S2
B3	Gentamicin	S3	F3	Thiamphenicol	S2
B4	Kanamycin	S3	F4	Trimethoprim	S2
B5	PenicillinG	S3	F5	D-Cycloserine	S2
B6	Dicloxacillin	S3	F6	Vancomycin	S2
B7	Thiamphenicol	S3	F7	Kanamycin	S7
B8	Trimethoprim	S3	F8	PenicillinG	S7
B9	Mafenide	S3	F9	Carbenicillin	S7
B10	Sulfamethizole	S3	F10	Thiamphenicol	S7
B11	Vancomycin	S3	F11	Sulfisoxazole	S7
B12	Dicloxacillin	S5	F12	Mafenide	S7
C1	Thiamphenicol	S5	G1	Vancomycin	S7
C2	Trimethoprim	S5	G2	Kanamycin	S4
C3	Mafenide	S5	G3	Amikacin	S1
C4	Vancomycin	S5	G4	PenicillinG	S1
C5	Ciprofloxacin	S11	G5	Dicloxacillin	S4
C6	PenicillinG	S11	G6	Thiamphenicol	S4
C7	Carbenicillin	S11	G7	Trimethoprim	S4
C8	Chloramphenicol	S11	G8	Mafenide	S4
C10	Trimethoprim	S11	G9	Sulfamethizole	S4
C11	Sulfamethizole	S11	G10	D-Cycloserine	S4
C12	D-Cycloserine	S11	G11	Ciprofloxacin	S4
C9	Thiamphenicol	S11	G12	Levofloxacin	S4
D1	Vancomycin	S11	H1	Sisomicin	S1
D2	Ciprofloxacin	S9	H2	Gentamicin	S1
D3	Levofloxacin	S9	H3	Kanamycin	S1
D4	Sisomicin	S9	H4	Amikacin	S1
D5	PenicillinG	S9	H5	PenicillinG	S1
D6	Carbenicillin	S9	H6	Carbenicillin	S1
D7	Dicloxacillin	S9	H7	Chloramphenicol	S1
D8	Chloramphenicol	S9	H8	Nalidixic acid	S1
D9	Thiamphenicol	S9	H9	Thiamphenicol	S1
D10	Trimethoprim	S9	H10	Sulfisoxazole	S1
D11	Mafenide	S9	H11	Mafenide	S1
D12	D-Cycloserine	S9			

Table S4: Antibiotics used for functional selections of antibiotic resistance determinants from the AB95 metagenomic library. Concentrations indicated are those used during functional selections and also represent the minimum concentrations at which growth of *E. coli* Top10 (Invitrogen) was 100% inhibited.

Antibiotic	Code	Selection Concentration	Antibiotic Class	Antibiotic Type
Amoxicillin	AX	50ug/mL	Beta-lactam	Semi-synthetic
Carbenicillin	CA	100ug/mL	Beta-lactam	Semi-synthetic
Cefdinir	CF	20ug/mL	Beta-lactam	Semi-synthetic
Chloramphenicol	CH	20ug/mL	Amphenicol	Natural
Minocycline	MN	20ug/mL	Tetracyclines	Semi-synthetic
Penicillin G	PE	100ug/mL	Beta-lactam	Natural
Piperacillin	PI	20ug/mL	Beta-lactam	Semi-synthetic
Sisomicin	SI	20ug/mL	Aminoglycoside	Natural
D-Cycloserine	CY	64ug/mL	Amino Acid Derivative	Natural
Gentamicin	GE	20ug/mL	Aminoglycoside	Natural
Oxytetracycline	OX	20ug/mL	Tetracyclines	Natural
Tetracycline	TE	20ug/mL	Tetracyclines	Natural

Table S5: Assembly statistics and annotation categorizations from sequencing 12 selections of the metagenomic library AB95 on various antibiotics. The 12 selections are presented in three groups of four, listed in each third of the table from top to bottom. Antibiotic abbreviations: AX-amoxicillin, CA-carbenicillin, CF-cefdinir, CH-chloramphenicol, CY-cylcoserine, GE-gentamicin, MN-minocycline, OX-oxytetracycline, PE-penicillin-G, PI-piperacillin, SI-sisomicin, TE-tetracycline. ORF=Open Reading Frame.

Category	All AB95	AB95_AX	AB95_CA	AB95_CF	AB95_CH
Number of Contigs:	161	27	18	17	15
Average Assembled Contig Length (bp):	961.47826	1062	1139.16667	348.88235	763.6
n50 Length of Assembled Contigs (bp):	1728	1723	1705	773	1829
Average Annotated ORF Length (bp):	524.95294	565.41860	560.03226	295.15789	464.26316
Antibiotic Resistance	110	21	16	15	5
Resistance-Related	62	8	4	3	10
Mobility Genes	14	2	2	0	1
Transporters	17	4	2	1	0
Other Known Annotations	27	4	4	0	2
Unknown	22	1	3	0	1
TOTAL ORFs:	252	40	31	19	19
Category	**—**	**AB95_CY**	**AB95_GE**	**AB95_MN**	**AB95_OX**
Number of Contigs:	—	24	3	3	2
Average Assembled Contig Length (bp):	—	809.29167	1592	1557.33333	2531.5
n50 Length of Assembled Contigs (bp):	—	1389	3035	2246	2747
Average Annotated ORF Length (bp):	—	470.92683	525.77778	793.2	876.6
Antibiotic Resistance	—	9	3	2	2
Resistance-Related	—	11	1	1	2
Mobility Genes	—	0	3	0	0
Transporters	—	4	0	0	0
Other Known Annotations	—	8	1	1	0
Unknown	—	9	1	1	1
TOTAL ORFs:	—	41	9	5	5
Category	**—**	**AB95_PE**	**AB95_PI**	**AB95_SI**	**AB95_TE**
Number of Contigs:	—	28	19	3	2
Average Assembled Contig Length (bp):	—	897.64286	1019.31579	1570	2544.5
n50 Length of Assembled Contigs (bp):	—	1729	1768	3058	2772
Average Annotated ORF Length (bp):	—	535.46154	523.03226	486.375	876.6
Antibiotic Resistance	—	19	13	3	2
Resistance-Related	—	11	8	1	2
Mobility Genes	—	2	2	2	0
Transporters	—	2	4	0	0
Other Known Annotations	—	3	3	1	0
Unknown	—	2	1	1	1
TOTAL ORFs:	—	39	31	8	5

Table S6: Average bit scores from the top Blast hits generated from comparing AB95 β-lactamases against β-lactamases from various phylogenetic groupings. As these groupings increased in phylogenetic relatedness to the AB95 culture collection (fig. S2), so too did the identity of the β-lactamases. All comparisons were performed with BlastP, numbers in parentheses indicate standard deviation.

	Average Bit Score	Median Bit Score
Proteobacteria β-Lactamases	480 (201)	553
Acidobacteria β-Lactamases	105 (59)	92
Streptomyces β-Lactamases	96 (56)	84

Table S7: Non-redundant antibiotic resistance genes from the AB95 dataset with >99% identity, but under 100% identity to human pathogens.

Gene Name	GenBankID	Resistance Type	Annotation	Pathogens Hit (GI#)	Percent Identity
AB95_CH_21.2	JX009371	Amphenicol	Chloramphenicol acetyltransferase	*K. pneumoniae (114147191 & 114147177)*	99.7%
AB95_CH_33.1	JX009370	Amphenicol	Chloramphenicol Efflux	*S. typhimurium* (12719011), *S. newport* (198240902)	99.8%

Table S8: Source organisms from the soil-derived AB95 culture collection for eight resistance genes with over 99% identity to clinical pathogens. For explanation of the absence of AB95_TE_1.1, see the section entitled 'Note on the Resuscitation of Frozen AB95 Cultures', within the online methods.

Gene Name	GenBankID	Resistance Type	Annotation	Highest % Nucleotide Identity to Pathogen	Nearest Neighbor to Source Soil Isolate (determined via 16S)
AB95_PI_68.1	JX009363	β-lactamase	blaP1	100	*Pseudomonas* K94.23
AB95_CH_13.1	JX009364	Amphenicol	Chloramphenicol Efflux	100	*Pseudomonas* K94.23
AB95_TE_2.2	JX009366	Tetracycline	tetA(G)	100	*Pseudomonas* K94.23
AB95_GE_3.3	JX009367, JX009373	Aminoglycoside	aadB	100	Both Organisms[1]
AB95_GE_3.1	JX009368, JX009374	Sulfonamide	sul1	100	Both Organisms[1]
AB95_CH_21.1	JX009369	Aminoglycoside	aacA4	100	*Pseudomonas* K94.23
AB95_CH_21.2	JX009371	Amphenicol	Chloramphenicol acetyltransferase	99.7	*Pseudomonas* K94.23
AB95_CH_33.1	JX009370	Amphenicol	Chloramphenicol Efflux	99.8	*Ochrobactrum anthropi*

1) Gene confirmed to be present in two isolates: an organism most closely related to *O. anthropi* and another most similar to *Pseudomonas* K94.23

Table S9: Percent nucleotide identity of two housekeeping genes between soil isolates and representative pathogens. Identities are calculated over a minimum of 1100bp in the case of the 16S rRNA gene and 190bp in the case of the *rpoB* gene. Full gene lengths are used for antibiotic resistance genes.

Soil Isolates	**16s rRNA Gene**	***rpoB* gene**	**Multiple Antibiotic Resistance Genes**	**Representative Pathogens (GI Number)**
Ochrobactrum soil isolate	82%	63%	100%	*Salmonella typhimurium* (16445344)
	81%	69%	100%	*Pseudomonas aeruginosa* (110227054)
	83%	65%	100%	*Acinetobacter baumannii* (169794206)
Pseudomonas soil isolate	86%	66%	100%	*Salmonella typhimurium* (16445344)
	94%	83%	100%	*Pseudomonas aeruginosa* (110227054)
	88%	63%	100%	*Acinetobacter baumannii* (169794206)

Table S10: Longest stretch of continuous nucleotide sequence from soil-derived isolates with perfect identity to human pathogens. Only contigs containing antibiotic resistance genes with perfect identity to those of pathogens, and for which this identity comprises 80% of the full contig length, are shown. An asterisk (*) denotes a single mismatch over the length of identity. Two asterisks (**) denote two mismatches over the span of identity (a 2217bp stretch of AB95_GE_3 and a 1811bp stretch of AB95_TE_2 are entirely without mismatches to pathogens).

Contig Name	GenBankID	Resistance Gene(s) housed	Length (bp)	Longest Pathogen-Identical Region (bp)	Pathogens Hit (GI#)
AB95_PI_68	JX009329	AB95_PI_68.1	1300	1261	K. pneumoniae (114147191 & 114147177), P. aeruginosa (117321883)
AB95_CH_13	JX009249	AB95_CH_13.1	1829	1599	P. aeruginosa (260677483)
AB95_TE_2	JX009344	AB95_TE_2.2	2772	2569	S. typhimurium (12719011)*
AB95_TE_1	JX009345	AB95_TE_1.1	2317	2277	E. coli (312949035)
AB95_GE_3	JX009286	AB95_GE_3.1 & AB95_GE_3.3	3035	2415	E. cloacae (71361871)**

Table S11: Antibiotic resistance genes from the AB95 isolate most closely related to *Pseudomonas sp.* K94.23 that are also found in human pathogens. Four of the seven resistance genes were also identified in a plasmid preparation from the *Pseudomonas* isolate.

Gene Name	GenBankID	Resistance Type	Annotation	Presence in Genomic DNA Preparation	Presence in Plasmid DNA Preparation
AB95_PI_68.1	JX009363	β-lactamase	blaP1	+	-
AB95_CH_13.1	JX009364	Amphenicol	Chloramphenicol Efflux	+	+
AB95_TE_2.2	JX009366	Tetracycline	tetA(G)	+	+
AB95_GE_3.3	JX009367, JX009373	Aminoglycoside	aadB	+	+
AB95_GE_3.1	JX009368, JX009374	Sulfonamide	sul1	+	+
AB95_CH_21.1	JX009369	Aminoglycoside	aacA4	+	-
AB95_CH_21.2	JX009371	Amphenicol	Chloramphenicol acetyltransferase	+	-

Table S12: 128 representative organisms of the human intestinal microflora.

GenBank ID	Organism Name
NC_011740	Escherichia fergusonii ATCC 35469
NZ_ABXY00000000	Bifidobacterium catenulatum DSM 16992
NZ_ABXB00000000	Bifidobacterium gallicum DSM 20093
NZ_ABXA00000000	Anaerococcus hydrogenalis DSM 7454
NC_009513	Lactobacillus reuteri DSM 20016
NZ_ACDL00000000	Shigella sp. D9
NZ_ACGA00000000	Bacteroides sp. D2
NZ_ACAB00000000	Bacteroides sp. D1
NZ_ABYI00000000	Clostridium hylemonae DSM 15053
NZ_ABXX00000000	Bifidobacterium pseudocatenulatum DSM 20438
NZ_ACBW00000000	Bacteroides coprophilus DSM 18228
NZ_ACCF00000000	Holdemania filiformis DSM 12042
NZ_ACEC00000000	Clostridium methylpentosum DSM 5476
NZ_ACCH00000000	Bacteroides cellulosilyticus DSM 14838
NZ_ACBX00000000	Prevotella copri DSM 18205
NZ_ABYU00000000	Blautia hansenii DSM 20583
NZ_ABYH00000000	Parabacteroides johnsonii DSM 18315
NZ_ABYJ00000000	Roseburia intestinalis L1-82
NZ_ABXH00000000	Collinsella intestinalis DSM 13280
NZ_ABYV00000000	Methanobrevibacter smithii DSM 2374
NZ_ACCI00000000	Providencia rettgeri DSM 1131
NZ_ABXW00000000	Providencia alcalifaciens DSM 30120
NZ_ABYT00000000	Eubacterium biforme DSM 3989
NZ_ACCJ00000000	Clostridium asparagiforme DSM 15981
NZ_ABYS00000000	Bifidobacterium angulatum DSM 20098
NZ_ACCG00000000	Bifidobacterium breve DSM 20213
NC_012778	Eubacterium eligens ATCC 27750
NC_012781	Eubacterium rectale ATCC 33656
NZ_ABWN00000000	Butyrivibrio crossotus DSM 2876
NC_000913	Escherichia coli str. K-12 substr. MG1655
NZ_ABWL00000000	Citrobacter youngae ATCC 29220
NZ_ABXV00000000	Providencia rustigianii DSM 4541
NZ_ABWK00000000	Mitsuokella multacida DSM 20544
NZ_ABWP00000000	Clostridium hiranonis DSM 13275
NZ_ABWO00000000	Clostridium nexile DSM 1787
NZ_ABOT00000000	Bifidobacterium animalis subsp. lactis HN019
NZ_ABQC00000000	Bacteroides plebeius DSM 17135
NZ_ABWZ00000000	Bacteroides dorei DSM 17855
NZ_ABVO00000000	Bacteroides eggerthii DSM 20697
NZ_ABXI00000000	Bacteroides finegoldii DSM 17565
NZ_ABYW00000000	Methanobrevibacter smithii DSM 2375
NZ_ACCL00000000	Bryantella formatexigens DSM 14469

NZ_ACBZ00000000	Blautia hydrogenotrophica DSM 10507
NZ_ABVP00000000	Proteus penneri ATCC 35198
NZ_ABOU00000000	Ruminococcus lactaris ATCC 29176
NZ_ABJD00000000	Providencia stuartii ATCC 25827
NZ_ABJK00000000	Streptococcus infantarius subsp. infantarius ATCC BAA-102
NZ_ABKW00000000	Clostridium sporogenes ATCC 15579
NZ_ABJL00000000	Bacteroides intestinalis DSM 17393
NZ_ABVR00000000	Coprococcus comes ATCC 27758
NZ_ABIY00000000	Bacteroides coprocola DSM 17136
NZ_ACIE00000000	Fusobacterium varium ATCC 27725
NZ_ACDE00000000	Fusobacterium sp. 4_1_13
NZ_ACIC00000000	Bacteroides sp. 1_1_6
NZ_ABZX00000000	Bacteroides fragilis 3_1_12
NZ_ACCK00000000	Catenibacterium mitsuokai DSM 15897
NZ_ABFZ00000000	Bacteroides stercoris ATCC 43183
NZ_ABXJ00000000	Collinsella stercoris DSM 13279
NZ_ABFX00000000	Clostridium ramosum DSM 1402
NZ_ABEZ00000000	Clostridium bartlettii DSM 16795
NZ_ABGD00000000	Anaerotruncus colihominis DSM 17241
NZ_ABIL00000000	Anaerofustis stercorihominis DSM 17244
NZ_ABFK00000000	Alistipes putredinis DSM 17216
NC_011835	Bifidobacterium animalis subsp. lactis AD011
NC_009615	Parabacteroides distasonis ATCC 8503
NC_009614	Bacteroides vulgatus ATCC 8482
NZ_ABAW00000000	Eubacterium dolichum DSM 3991
NZ_ABIK00000000	Clostridium spiroforme DSM 1552
NZ_ABCB00000000	Clostridium leptum DSM 753
NC_009515	Methanobrevibacter smithii ATCC 35061
NC_009004	Lactococcus lactis subsp. cremoris MG1363
NZ_AAVN00000000	Collinsella aerofaciens ATCC 25986
NZ_ABCC00000000	Clostridium bolteae ATCC BAA-613
NZ_AAVM00000000	Bacteroides caccae ATCC 43185
NZ_ABAX00000000	Anaerostipes caccae DSM 14662
NZ_AAYW00000000	Clostridium sp. L2-50
NZ_ACFX00000000	Clostridium sp. M62/1
NZ_ABED00000000	Faecalibacterium prausnitzii M21/2
NZ_ABGC00000000	Clostridium sp. SS2/1
NZ_ACOP00000000	Faecalibacterium prausnitzii A2-165
NZ_AAXD00000000	Bifidobacterium adolescentis L2-32
NZ_AAYH00000000	Bacteroides uniformis ATCC 8492
NZ_AAXE00000000	Parabacteroides merdae ATCC 43184
NZ_AAXF00000000	Bacteroides ovatus ATCC 8483
NZ_ABEY00000000	Coprococcus eutactus ATCC 27759
NZ_ACBY00000000	Subdoligranulum variabile DSM 15176

NZ_AAYG00000000	Ruminococcus gnavus ATCC 29149
NZ_ACEP00000000	Eubacterium hallii DSM 3353
NZ_ABFY00000000	Clostridium scindens ATCC 35704
NZ_AAXG00000000	Bacteroides capillosus ATCC 29799
NZ_AAYI00000000	Actinomyces odontolyticus ATCC 17982
NZ_ABEE00000000	Parvimonas micra ATCC 33270
NZ_ABXU00000000	Desulfovibrio piger ATCC 29098
NZ_AAVL00000000	Eubacterium ventriosum ATCC 27560
NZ_AAXB00000000	Dorea longicatena DSM 13814
NZ_AAXA00000000	Dorea formicigenerans ATCC 27755
NZ_AAVP00000000	Ruminococcus torques ATCC 27756
NZ_AAVO00000000	Ruminococcus obeum ATCC 29174
NC_008054	Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842
NC_008618	Bifidobacterium adolescentis ATCC 15703
NC_008086	Helicobacter pylori HPAG1
NC_010655	Akkermansia muciniphila ATCC BAA-835
NZ_ABDE00000000	Victivallis vadensis ATCC BAA-548
NC_007681	Methanosphaera stadtmanae DSM 3091
NC_008532	Streptococcus thermophilus LMD-9
NC_008526	Lactobacillus casei ATCC 334
NC_006449	Streptococcus thermophilus CNRZ1066
NC_006347	Bacteroides fragilis YCH46
NC_002662	Lactococcus lactis subsp. lactis Il1403
NC_008527	Lactococcus lactis subsp. cremoris SK11
NC_003228	Bacteroides fragilis NCTC 9343
NC_006448	Streptococcus thermophilus LMG 18311
NC_002505	Vibrio cholerae O1 biovar eltor str. N16961 chromosome I
NC_002506	Vibrio cholerae O1 biovar eltor str. N16961 chromosome II
NC_004663	Bacteroides thetaiotaomicron VPI-5482
NC_004307	Bifidobacterium longum NCC2705
NC_010816	Bifidobacterium longum DJO10A
NC_Bthetaiotaomicron3731[1]	Bacteroides thetaiotaomicron 3731
NC_Bthetaiotaomicron7330[1]	Bacteroides thetaiotaomicron 7330
NC_BWH2[1]	Bacteroides WH2
FP929033	Bacteroides xylanisolvens XB1A
NC_Bdentium[1]	Bifidobacterium dentium
NC_Csymbiosum[1]	Clostridium symbiosum
NC_DpigerGOR1[1]	Desulfovibrio piger GOR1
NC_Ecancerogenus[1]	Enterobacter cancerogenus
FP929042	Eubacterium rectale DSM17629
NC_M23A[2]	M23A
FP929051	Ruminococcus bromiiL263

1) Genome sequences available at The Genome Institute http://genome.wustl.edu/genomes/list/prokaryotic_microbes (Washington University, St. Louis).
2) Genome sequences obtained from Dr. Jeffrey Gordon (Washington University, St. Louis)

Table S13: AB95 antibiotic resistance genes with 100% identity to genes from one of three datasets isolated from the human intestinal microbiome. The 'Cultured Gut' dataset refers to a set of antibiotic resistant genes from a metagenomic library prepared with a set of isolates cultured from the fecal samples of two healthy adults *(20)*. The 'Rep. 128 Gut' dataset refers to a set of genome sequences from 128 representative organisms of the human gastrointestinal tract (table S12). The GI number from the best hit from each dataset is indicated. In the case of the representative 128 organisms, the ampC originated from Escherichia coli str. K-12 substr. MG1655 (NC000_913). The percentage covered refers to the length of the database gene covered by the queried AB95 antibiotic resistance gene.

Gene Name	GenBank ID	Annotation	Gut Dataset Hit	Best Hit GI#	% Hit Covered
AB95_TE_1.1	JX009365	tetA	Cultured Gut	254967249	100
AB95_AX_49.1	JX009213	ampC	Rep. 128 Gut	49175990 (NC000_913)	91.5

Table S14: Antibiotics used for functional selections of antibiotic resistance determinants from the fecal-derived metagenomic library HG1. Concentrations indicated are those used during functional selections and represent the minimum concentrations at which growth of *E. coli* TOP10 (Invitrogen) was 100% inhibited.

Antibiotic	Code	Selection Concentration	Antibiotic Class	Antibiotic Type
Amoxicillin	AX	50ug/mL	Beta-lactam	Semi-synthetic
Carbenicillin	CA	100ug/mL	Beta-lactam	Semi-synthetic
Cefdinir	CF	20ug/mL	Beta-lactam	Semi-synthetic
Chloramphenicol	CH	20ug/mL	Amphenicol	Natural
Minocycline	MN	20ug/mL	Tetracyclines	Semi-synthetic
Penicillin G	PE	100ug/mL	Beta-lactam	Natural
Piperacillin	PI	20ug/mL	Beta-lactam	Semi-synthetic
Sisomicin	SI	20ug/mL	Aminoglycoside	Natural

Table S15: Five pools, each comprised of *E. coli* clones transformed with one of six β-lactamase containing metagenomic fragments varying in length (fragment sizes: 1358bp, 1732bp, 2147bp, 2408bp, 2882bp, and 3292bp). In four of the five pools, one of these six fragments was spiked in at 10X the abundance of each of the other five while the fifth pool was an even mixture of each fragment. The number of 76bp paired-end reads assigned to the barcode for each pool is indicated in the final column.

Pool #	Composition	# PE Reads
1	10X 1358bp fragment	2,265,839
2	10X 2147bp fragment	1,743,311
3	10X 2882bp fragment	1,601,940
4	10X 3292bp fragment	1,502,033
5	Equal Abundance Pool	1,181,976

Table S16: Assembly statistics for each β-lactamase containing metagenomic fragment, by pool. The percentage of the original metagenomic fragment covered by the contigs generated using PARFuMS is indicated. The number of contigs required to achieve the reported coverage statistic are indicated in parentheses. For a pictorial representation of these values, please refer to fig. S6.

	Fragments Assembled					
Pool #	**1358bp**	**1732bp**	**2147bp**	**2408bp**	**2882bp**	**3292bp**
1	100%(1)	99.5%(1)	99.4%(1)	99.6%(2)	95.8%(2)	98.7%(2)
2	100%(1)	99.8%(1)	99.4%(1)	95.7%(2)	98.7%(1)	97.6%(2)
3	100%(1)	100%(1)	99.44%(1)	97.1%(2)	99.9%(1)	96.2%(1)
4	100%(1)	99.9%(1)	99.44%(1)	98.9%(1)	100%(1)	99.7%(1)
5	99.7%(1)	99.7%(1)	99.7%(1)	97.5%(2)	99.9%(1)	98.5%(1)

Table S17: Assembly and annotation statistics from sequencing eight selections of the fecal metagenomic library HG1 on various antibiotics. Four of the eight selections are listed in the top half of the table, while the remainder are listed in the lower half. Antibiotic abbreviations: AX-amoxicillin, CA-carbenicillin, CF-cefdinir, CH-chloramphenicol, MN-minocycline, PE-penicillin-G, PI-piperacillin, SI-sisomicin. ORF=Open Reading Frame.

Category	All HG1	HG1_AX	HG1_CA	HG1_CF	HG1_CH
Number of Contigs:	38	8	8	4	3
Average Assembled Contig Length (bp):	1203.55263	1140	1703.25	991.5	1905
n50 Length of Assembled Contigs (bp):	1659	1583	2060	2198	3415
Average Annotated ORF Length (bp):	610.71212	566.83333	727.21429	565.28571	476.16667
Antibiotic Resistance	20	4	4	2	3
Resistance-Related	19	3	4	4	0
Mobility Genes	0	0	0	0	0
Transporters	0	0	0	0	0
Other Known Annotations	14	4	5	1	3
Unknown	6	1	1	0	0
TOTAL ORFs:	59	12	14	7	6

Category	—	HG1_MN	HG1_PE	HG1_PI	HG1_SI
Number of Contigs:	—	4	5	4	2
Average Assembled Contig Length (bp):	—	589.25	1006.6	980.75	997.5
n50 Length of Assembled Contigs (bp):	—	837	3473	2169	1525
Average Annotated ORF Length (bp):	—	729.0909	557.875	586.16667	255
Antibiotic Resistance	—	3	2	1	0
Resistance-Related	—	0	3	4	1
Mobility Genes	—	0	0	0	0
Transporters	—	0	0	0	0
Other Known Annotations	—	0	0	0	1
Unknown	—	0	3	1	0
TOTAL ORFs:	—	3	8	6	2

Table S18: Genes from assemblies of HG1 antibiotic selections using short-read sequence data that were not present in the Sanger sequence data from the same selections. In two of three cases, partial genes were located at the end of an assembled contig and, thus, the entire gene sequence was not recovered. In these instances, however, the sequence was 100% identical to known antibiotic resistance genes, allowing classification of the resistant determinant.

Gene Length	Gene Annotation	Top Blast Hit [gbID--species]	Global BlastX %ID vs. NR Top Hit
648bp	Chloramphenicol-O-Acetyltransferase	302387060-- Clostridium saccharolyticum WM1	69.5
>489bp	TetQ: Tetracycline Resistance Protein	237713349-- Bacteroides sp. D1	100% over 489bp
>108bp	TetG: Tetracycline Resistance Protein	4633511-- Salmonella typhimurium DT104	100% over 108bp

Table S19: Primers used to amplify resistance genes.

Gene	Primer 1 (5' -> 3')	Primer 2 (5' -> 3')
AB95_PI_68.1	ATGCGCTCACGCAACTG	TCAGCGCGACTGTGATGTATAA
AB95_CH_13.1	ATGACCACCACACGCCC	CTAGACGACTGGCGACTTCTC
AB95_TE_2.2	ATGGGTCTCGGCCTCATC	TCACAATGAAGTTGCGAATGGT
AB95_TE_1.1	ATGTCCACCAACTTATCAGTG	TCAGCGATCGGCTCGT
AB95_GE_3.3	ATGCGCTCACGCAACT	TTAGGCCGCATATCGC
AB95_GE_3.1	ATGGTGACGGTGTTCGG	CTAACCCTCGGTCTCTGG
AB95_CH_21.1[1]	TTGGAGCAAATTAAAGAAGCAATG	GCATCACTGCGTGTTCG
AB95_CH_21.2	ATGAAGAACTATTTTGAGAGCCC	TTAGGCACTTGAACGCTGC
AB95_CH_33.1[2]	ATGACCACCACACGCCC	CTAGACGACTGGCGACTTCTC
AB95_CY_48.2	ATGACAATTACCCACAAGGAAAGC	CTATGCCTGCAGCCTGGAA

1) Primers used not flush with boundaries of open reading frame, rather slightly external, remaining within the region containing 100% nucleotide identity to pathogens.

2) Primers used identical to those of AB95_CH_13.1. The flanking regions of these genes are identical, whereas the internal sequence varies between the genes. Both variants were over 99% identical to human pathogens.

References and Notes

1. C. A. Arias, B. E. Murray, Antibiotic-resistant bugs in the 21st century--a clinical super-challenge. *N Engl J Med* **360**, 439 (Jan 29, 2009).
2. R. Benveniste, J. Davies, Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc Natl Acad Sci U S A* **70**, 2276 (Aug, 1973).
3. J. L. Martinez, Antibiotics and antibiotic resistance genes in natural environments. *Science* **321**, 365 (Jul 18, 2008).
4. L. Poirel, P. Kampfer, P. Nordmann, Chromosome-encoded Ambler class A beta-lactamase of Kluyvera georgiana, a probable progenitor of a subgroup of CTX-M extended-spectrum beta-lactamases. *Antimicrob Agents Chemother* **46**, 4038 (Dec, 2002).
5. L. Poirel, J. M. Rodriguez-Martinez, H. Mammeri, A. Liard, P. Nordmann, Origin of plasmid-mediated quinolone resistance determinant QnrA. *Antimicrob Agents Chemother* **49**, 3523 (Aug, 2005).
6. J. J. Farmer, 3rd *et al.*, Kluyvera, a new (redefined) genus in the family Enterobacteriaceae: identification of Kluyvera ascorbata sp. nov. and Kluyvera cryocrescens sp. nov. in clinical specimens. *J Clin Microbiol* **13**, 919 (May, 1981).
7. T. Stalder, O. Barraud, M. Casellas, C. Dagot, M. C. Ploy, Integron involvement in environmental spread of antibiotic resistance. *Frontiers in microbiology* **3**, 119 (2012).
8. B. M. Marshall, S. B. Levy, Food animals and antimicrobials: impacts on human health. *Clin Microbiol Rev* **24**, 718 (Oct, 2011).
9. V. M. D'Costa *et al.*, Antibiotic resistance is ancient. *Nature* **477**, 457 (Sep 22, 2011).
10. V. M. D'Costa, K. M. McGrann, D. W. Hughes, G. D. Wright, Sampling the antibiotic resistome. *Science* **311**, 374 (Jan 20, 2006).
11. H. K. Allen, L. A. Moe, J. Rodbumrer, A. Gaarder, J. Handelsman, Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME J* **3**, 243 (Feb, 2009).
12. J. J. Donato *et al.*, Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Appl Environ Microbiol* **76**, 4396 (Jul, 2010).
13. R. I. Aminov, R. I. Mackie, Evolution and ecology of antibiotic resistance genes. *FEMS Microbiol Lett* **271**, 147 (Jun, 2007).
14. H. Heuer, H. Schmitt, K. Smalla, Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr Opin Microbiol* **14**, 236 (Jun, 2011).
15. P. S. McManus, V. O. Stockwell, G. W. Sundin, A. L. Jones, Antibiotic use in plant agriculture. *Annu Rev Phytopathol* **40**, 443 (2002).
16. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood, *Practical Streptomyces Genetics*. (John Innes Foundation, Norwich, UK, ed. 1, 2000).
17. J. Davies, D. Davies, Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**, 417 (Sep, 2010).
18. H. W. Boucher *et al.*, Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis* **48**, 1 (Jan 1, 2009).
19. Materials and methods are available as supplementary materials on *Science* online
20. M. O. Sommer, G. Dantas, G. M. Church, Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128 (Aug 28, 2009).
21. G. Dantas, M. O. Sommer, R. D. Oluwasegun, G. M. Church, Bacteria subsisting on antibiotics. *Science* **320**, 100 (Apr 4, 2008).
22. V. L. Clark, F. E. Young, Inducible resistance to D-cycloserine in Bacillus subtilis 168. *Antimicrob Agents Chemother* **11**, 871 (May, 1977).

23. H. K. Allen *et al.*, Call of the wild: antibiotic resistance genes in natural environments. *Nat Rev Microbiol* **8**, 251 (Apr, 2010).
24. F. Rezzonico, G. Defago, Y. Moenne-Loccoz, Comparison of ATPase-encoding type III secretion system hrcN genes in biocontrol fluorescent Pseudomonads and in phytopathogenic proteobacteria. *Appl Environ Microbiol* **70**, 5119 (Sep, 2004).
25. M. W. Silby, C. Winstanley, S. A. Godfrey, S. B. Levy, R. W. Jackson, Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol Rev* **35**, 652 (Jul, 2011).
26. S. Romano *et al.*, Multilocus sequence typing supports the hypothesis that Ochrobactrum anthropi displays a human-associated subpopulation. *BMC Microbiology* **9**, (Dec 18, 2009).
27. P. S. G. Chain *et al.*, Genome of Ochrobactrum anthropi ATCC 49188(T), a Versatile Opportunistic Pathogen and Symbiont of Several Eukaryotic Hosts. *Journal of Bacteriology* **193**, 4274 (Aug, 2011).
28. C. S. Smillie *et al.*, Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**, 241 (Oct 30, 2011).
29. C. Walsh, Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**, 775 (Aug 17, 2000).
30. R. Lutz, H. Bujard, Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res **25**, 1203 (Mar 15, 1997).
31. A. Reyes *et al.*, IS-seq: a novel high throughput survey of in vivo IS6110 transposition in multiple Mycobacterium tuberculosis genomes. BMC genomics 13, 249 (Jun 15, 2012)
32. D. R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18**, 821 (May, 2008).
33. M. de la Bastide, W. R. McCombie, Assembling genomic DNA sequences with PHRAP. Curr Protoc Bioinformatics (John Wiley & Sons, Inc., Published Online, ed. 2008/04/23, 2007), vol. Chapter 11, pp. Unit11 4.
34. R. L. Tatusov, M. Y. Galperin, D. A. Natale, E. V. Koonin, The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res **28**, 33 (Jan 1, 2000).
35. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22**, 1658 (Jul 1, 2006).
36. B. Niu, Z. Zhu, L. Fu, S. Wu, W. Li, FR-HIT, a very fast program to recruit metagenomic reads to homologous reference genomes. Bioinformatics **27**, 1704 (Jun 15, 2011).
37. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**, 1792 (2004).
38. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One **5**, e9490 (2010).
39. J. J. Faith, A. J. Olson, T. S. Gardner, R. Sachidanandam, Lightweight genome viewer: portable software for browsing genomics data in its chromosomal context. BMC Bioinformatics **8**, 344 (2007).