

## **SUPPLEMENTARY MATERIAL for manuscript**

### Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks

Andrey Morgun<sup>1,8,\*</sup>, Amiran Dzutsev<sup>4</sup>, Xiaoxi Dong<sup>1</sup>, Renee L Greer<sup>2</sup>, D. Joseph Sexton<sup>3</sup>, Jacques Ravel<sup>5</sup>, Martin Schuster<sup>3</sup>, William Hsiao<sup>6</sup>, Polly Matzinger<sup>7</sup>, Natalia Shulzhenko<sup>1,8,\*</sup>

<sup>1</sup>College of Pharmacy, <sup>2</sup>College of Veterinary Medicine, <sup>3</sup>Department of Microbiology, Oregon State University, Corvallis, OR, USA; <sup>4</sup>Cancer and Inflammation Program, National Cancer Institute/Leidos Biomedical Research, Inc., Frederick, MD, USA; <sup>5</sup>Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA; <sup>6</sup>University of British Columbia, Vancouver, BC, Canada; <sup>7</sup>National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA; <sup>8</sup>equally contributed and corresponding authors

\*Correspondence to: Natalia Shulzhenko, 105 Dryden, Oregon State University, Corvallis, OR 97331, USA. [natalia.shulzhenko@oregonstate.edu](mailto:natalia.shulzhenko@oregonstate.edu) or [anemorgun@hotmail.com](mailto:anemorgun@hotmail.com)

### **Evaluation of mitochondria and cell viability in intestinal epithelium**

For isolation of intestinal epithelium, ~10 cm of the terminal ileum were open longitudinally, rinsed twice with ice-cold HBSS and cut into 5 mm pieces into HBSS with 0.2% BSA and 5 mM EDTA. The tissues were incubated at 37°C for 30 min with continuous stirring, after which the cell suspension was filtered and washed. Staining for flow cytometry was performed with fluorescently labeled antibodies for CD45 (to exclude hematopoietic cells) and Annexin V (Biolegend), in addition to 7AAD (BD) for cell viability and Mitotracker Green (Life Technologies) for mitochondria. BD FACSCanto II was used for data acquisition and FlowJo for data analysis.

As absolute fluorescence values for Mitotracker varied between different days, mean fluorescent intensity of each sample was normalized to the average value for that experiment so that the results are comparable across different experiments.

### **Immunofluorescent Confocal Microscopy**

To preserve fine cellular structure of the tissues, deeply anaesthetized mice were given intracardiac injection of 15 ml of 1% paraformaldehyde (Electron Microscopy Sciences) in phosphate buffer (2). Tissues were isolated and incubated in 1% PF for 12h, following by dehydration with 30% sucrose for another 6h. Tissues were then imbedded in OCT freezing media and snap frozen. 10µm sections were cut on Leica cryostat on to Superfrost Plus slides (VWR). Slides were blocked for 1h with PBS containing 0.2% Triton X-100 (Sigma), 0.5% BSA or 5% goat or 5% rat serum, depending on the antibodies used. After the blocking procedure slides were stained in the same buffer with directly-labeled (CD3, B220, IgA, all eBioscience) or indirectly-labeled antibodies (goat-anti-mouse Collagen IV, Abcam) and with DAPI (Invitrogen). Slides were mounted with Prolong Gold (Invitrogen). Confocal microscopy was performed on Leica SP2 or Leica SP5 instruments. Z-stacks (512x512p) were acquired with 0.3-1µm step. Image analysis was done using ImageJ software (NIH).

### **Laser Microdissection**

Protocol suggested by Leica Microsystems for preparing samples for laser microdissection was generally followed. Specifically, immediately after harvesting, 1-2 cm of terminal ileum were embedded in OCT media and frozen at -80°C. Sections of 12 µm were obtained using a cryostat at -18°C and placed on polyethylene naphthalate (PEN) coated slides (Leica). For microdissection, the slides were fixed in ethanol 70% (30 sec), washed in sterile water (10 dips), stained with hematoxylin (20 sec), washed again in water, fixed sequentially in ethanol 70%, 95% and 100% (1 min each). After a short air drying, the slides were positioned under the microscope of LMD6000 system (Leica) for microdissection. A video demonstrating the dissection of villi (red),

lamina propria (green) and crypts (blue) is provided as a supplemental file (Movie S1). In the movie, areas delineated with different colors were dissected and collected in different tubes containing RNA lysis buffer. The number of obtained cells was estimated from the dissected area size and varied from 2000 to 5000 cells. Dissected cells were directly harvested into RNA lysis buffer of RNeasy Micro kit (Qiagen) and the instructions for RNA isolation were followed. RNA was amplified and labeled using Aminoallyl RNA Amplification and Labeling System (NuGEN Technologies) following manufacturer's protocol. Cy5-labeled samples were co-hybridized to Agilent microarrays with Cy3-labeled common reference RNA from control mouse intestine.

### **Gene Expression Microarrays**

RNA samples were isolated, labeled, hybridized to Agilent arrays and scanned as described in (Shulzhenko et al., 2011). Individual samples labeled with Cy5 were co-hybridized with a common reference RNA labeled with Cy3. All microarray data that passed quality control were analyzed using BRB Array Tools developed by the Biometric Research Branch of the National Cancer Institute under the direction of R. Simon (<http://linus.nci.nih.gov/BRB-ArrayTools.html>). First, array data were filtered to limit analyses to probes with greater than 50% of samples showing spot intensities of >100 in at least one channel and spot sizes >10 pixels, thereafter, lowess smoother normalization was applied. Next, differentially expressed genes between groups of samples were identified by random variance t-test with adjustment for multiple hypotheses by setting the false discovery rate (FDR) below 10% in BRB Array Tools. To define overlapping and unique sets of genes, the expression of 1583 genes regulated by antibiotics were analyzed in the following set of comparisons: 1- control vs. germ-free; 2- germ-free vs. germ-free+ABx; 3- colonized with normal microbiota vs colonized with ABx-resistant microbiota. Gene expression that passed p-value of 0.05 with the direction of regulation (fold change) concordant with that in Control vs. ABx-treated mice was attributed to the respective factor (i.e. normal microbiota, direct effect of antibiotics, effect of ABx-resistant microbes). Raw and processed data files are available under GSE60568 super-series in the Gene Expression Omnibus (GEO) data repository.

### **Microbiome analysis**

#### *Shotgun sequencing data analysis*

DNA libraries were prepared using 50 ng of DNA from cecum content using Illumina Nextera sample preparation kit and Nextera Index kit (Illumina). Illumina HiSeq2000 was used for sequencing libraries with 4 to 12 samples per lane. The sequencing resulted in 2,816,120,546 sequences for 32 samples.

Reads with ambiguous character "N" were removed using prinseq (Schmieder and Edwards, 2011). Illumina adaptor and primer sequences were removed using trimmomatic with default option (Bolger et al., 2014). Reads that aligned to the mouse genome (NCBI version 37) (alignment with Bowtie2 (Langmead and Salzberg, 2012), using end-to-end --sensitive) were also removed. The set of 1,450,158,701 high-quality reads was then used for further analysis.

Alignment to reference genomes and taxonomical/functional analysis was performed using NCBI and Human Microbiome Project (HMP) non-redundant protein sequences (nr) using *rapsearch2* with options -e 0.001 -v 100 (Zhao et al., 2012). The alignment output was analyzed using MEGAN4 with default parameters to produce taxa and SEED function annotation for the reads (Huson et al., 2011). There were 276,849,551 reads with SEED function assignment and 164,941,710 with taxonomy assignment. The fraction of reads that were assigned to each taxa or SEED function among all annotated reads was used as a relative abundance of the taxa or function in subsequent analysis.

The data is available at NCBI (SRA), BioProject ID:PRJNA258166

### *Differential abundance analysis*

T-test of between group abundance fold changes are analyzed in BRB-ARRAY Tools. Taxas are considered with different abundance between ABx and control if T-test p-value <0.01 (FDR <0.02). SEED functions that are different are defined using p-value<0.05 (FDR<0.06).

### **Network analysis**

Gene-gene network is reconstructed using the steps described below and in more detail in Dong et al., <http://arxiv.org/abs/1411.0595>

1. For each pair of genes, four Pearson correlation coefficients and p-values are calculated from expression levels of each of the four groups of mouse samples on different genetic backgrounds (C57BL/6, Swiss-Webster, B10A and BALB/c; n=8-10 per strain)
2. Fisher combined p-value is calculated from four p-values using Fisher's combined probability test. FDR value is calculated from the combined pvalue. i.e. we calculate a test

statistic as 
$$X_{2k}^2 \sim -2 \sum_{i=1}^k \ln(p_i),$$
 where k is four and pi is the pvalue from one of the mouse strain background. A p-value (the combined pvalue) for  $X^2$  is calculated under the fact that it follows a chi-squared distribution with 2k degrees of freedom.

3. Network is generated including gene-gene pairs as has been previously described (Mine et al., 2013; Skinner et al., 2011). The step by step brief description is:
  - a. Individual p-value in each mouse strain sample group is <0.2
  - b. FDR calculated from combined fisher pvalue <0.05
  - c. The sign of correlation coefficients in four mouse strain groups should be consistent(all positive correlation or all negative correlation) and should be consistent with fold change relationship between the two genes (see methods in: (Yambartsev Anatoly, 2013) Unexpected links reflect the noise in networks [arXiv:1310.8341](https://arxiv.org/abs/1310.8341) [q-bio.MN])

The network was visualized in Cytoscape Software 2.6.3 (Shannon et al., 2003). We used the MCODE v1.2 (Molecular Complex Detection) (Bader and Hogue, 2003) plug-in for Cytoscape to identify clusters (subnetworks) of correlated genes. In the mouse gene expression network, we searched for subnetworks/clusters that had a K score above three, using the node score cut-off of 0.35. The two top ranked subnetworks were selected for further analyses, as all other subnetworks were representing parts of the first two.

Gene Ontology (biological process) enrichment analysis was performed using DAVID (Dennis et al., 2003) for each subnetwork. To illustrate functional enrichment analysis significant Gene Ontology (GO) terms and associated genes were used to create a gene-GO terms network to identify clusters of GO terms with overlapping common genes. Most common GO terms of the identified cluster were used to create the common name for the cluster.

Microbial SEED-host gene (**transkingdom**) network is reconstructed using the same procedure but using B6 and SW antibiotics-treated mice for which metagenomics sequencing was available. Because microarrays were carried out with two reference RNAs for each strain, the samples formed 4 groups and correlations were calculated within the 4 groups and meta-analysis was performed as described above.

*Taxa coverage analysis for SEED sub-networks:*

For each SEED protein, the corresponding reference sequences were retrieved from the SEED ftp: <ftp://ftp.theseed.org/>, reference sequence to taxonomy mapping is retrieved from NCBI ftp site. A taxon is considered as possessing a SEED function if it possessed the reference sequences annotation under the SEED function.

*SEED enrichment analysis:* The enrichment of SEED level1 categories is detected by calculating distribution of number of SEED categories in different subnetworks using chi-square test with p-value<0.05.

#### *Betweenness centrality analysis*

In this study we have adapted the classical betweenness centrality (Freeman, 1977). A standard betweenness centrality estimates the number of shortest paths between any two nodes in the network that go through a given node. Nodes with highest values in the network are candidates for being key drivers. We modified this approach. In order to find bacterial gene regulators of host functions we estimated the number of shortest paths from microbial to mouse compartments of the transkingdom network and selected microbial genes that are within the top 1% for this metric. Thus for each node in the network we measured a node's centrality between microbial SEED functions and host genes. The betweenness centrality of a node  $v$  is given by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where  $s$  belongs to SEED function and  $t$  belongs to host genes,  $\sigma_{st}$  is the total number of shortest paths from nodes  $s$  to node  $t$  and  $\sigma_{st}(v)$  is the number of those paths that pass through  $v$ .

#### *Annotation of cyanide synthase genes in shotgun reads:*

High quality shot-gun reads were searched against gene sequences for five cyanide synthase genes (algR, anR, hcnA, hcnB and hcnC) using rapsearch2 with different e-value thresholds. The numbers of reads that pass similarity e-value threshold are used to calculate the relative abundance of cyanide synthase genes.

#### **In vitro experiments**

MODE-K cells were maintained in DMEM growth media (10% FBS, 25mM glucose, 4mM L-glutamine, without Phenol red).  $1 \times 10^5$  cells per well in a 96-well plate were incubated in 100 $\mu$ l DMEM growth media for 24 hours prior to treatment with bacterial supernatant. 50 $\mu$ l supernatant was added to wells at time 0. For mitochondrial analysis MitoTracker green (Life Technologies) and TMRE (Sigma) were added to wells to final concentrations of 100nM each and incubated for 30 minutes. Media was collected; adherent cells were harvested by trypsin treatment and pooled with collected media. Dead cells were marked using Fixable viability dye eFluor660 (eBioscience) according to manufacturer's protocol. Cell suspensions were analyzed on a BD Accuri 6 flow cytometer. Wild type and mutant bacteria were previously generated and cultured under conditions that favor quorum sensing as described in (Wilder et al., 2011).

Cell free supernatant was prepared from *Pseudomonas aeruginosa* wild-type strain PAO1 and quorum-sensing negative mutant PAO1  $\Delta$ rhIR::GmR  $\Delta$ lasR::TetR (Rahim et al., 2001). Cultures were grown to stationary phase in MOPS buffered LB broth with incubation at 37°C with shaking for 14 hrs. Supernatant concentration was controlled by diluting cultures to an optical density at 600 nm (OD600) of 1. Cells were then pelleted and supernatant was passed through a 0.22 micron filter to remove remaining cells. Cell free supernatants were maintained at 4°C and used within one week of preparation.

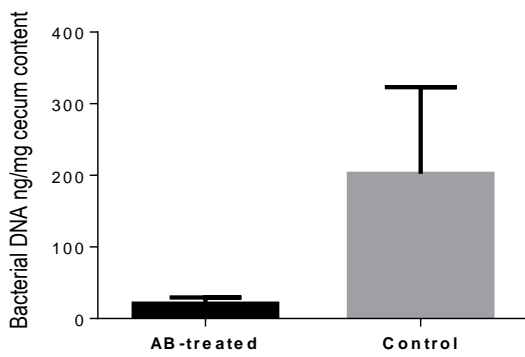


Figure S1. amount of bacterial DNA (ng) per mg of cecum content in control mice and mice given the cocktail of antibiotics for 5 weeks (AB-treated). N=8 in each group.

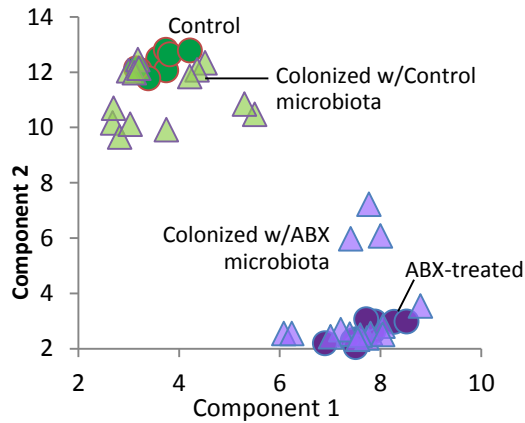


Figure S2. NMF (non-negative matrix factorization) clustering based on 16S rRNA showing the stability of microbial populations from Abx treated and control mice transplanted into germ free recipients. Each symbol represents an individual mouse; circles represent originally treated mice; triangles represent samples after colonization; green-control microbiota; purple, ABx-associated microbiota.

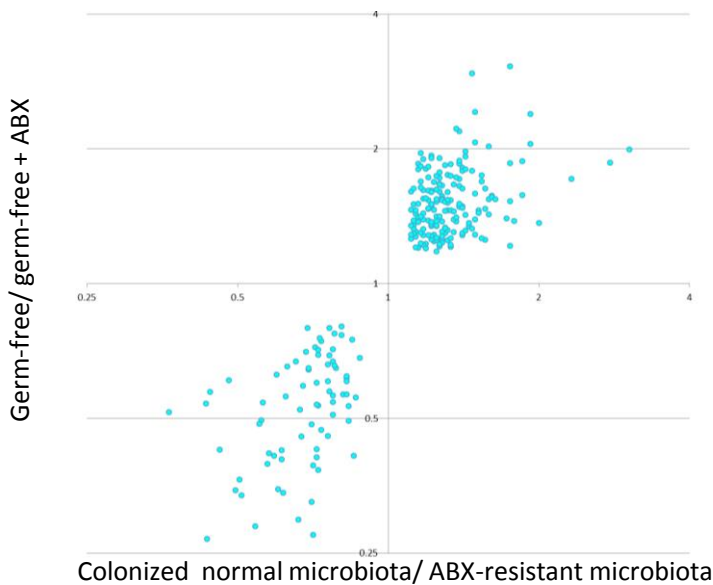
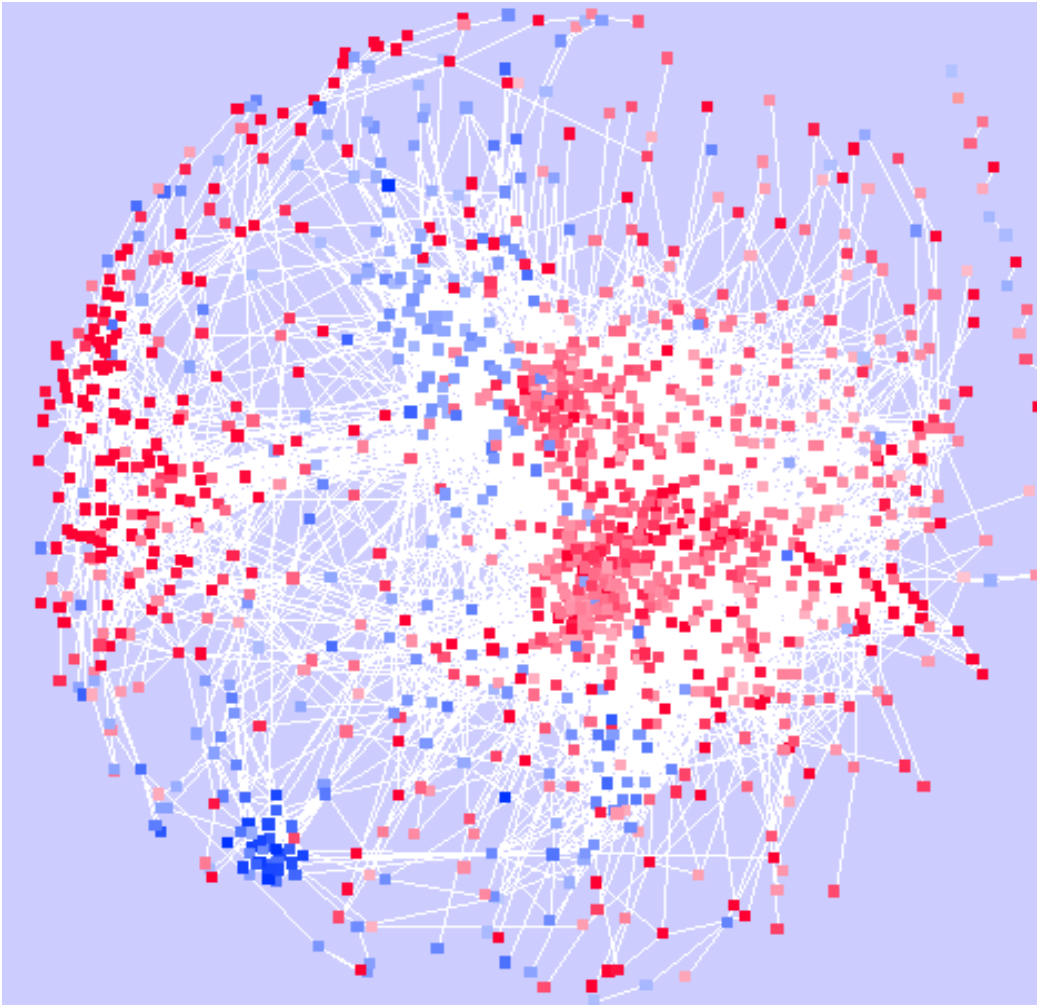


Figure S3. Genes from the 263-gene set that is influenced by both the direct effect of antibiotics (ABX) and of ABX-resistant microbes, showing that they are similarly regulated by ABX and by ABX-resistant microbes. Each symbol represents a gene from the set.

Figure S4

Gene expression network as in Fig. 2b but colored by fold change:

**red**, higher in control mice, i.e. lower in AB-treated; **blue**, lower in control mice; i.e. higher in AB-treated





## MCODE score

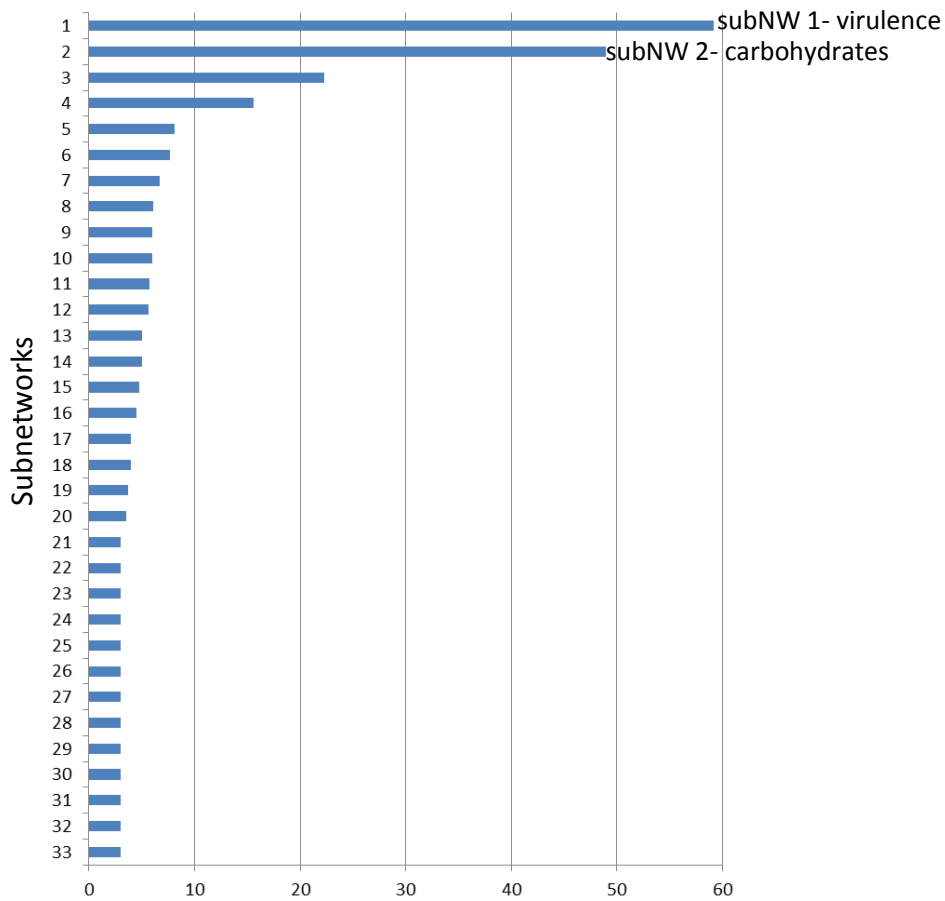


Figure S7. Revealing major microbial subnetworks (subNW) in the transkingdom network. MCODE (Molecular Complex Detection, <http://baderlab.org/Software/MCODE>) program implemented in Cytoscape was used to calculate the score of each subnetwork in the network on Fig. 5. We focused on the two top-ranked subnetworks.



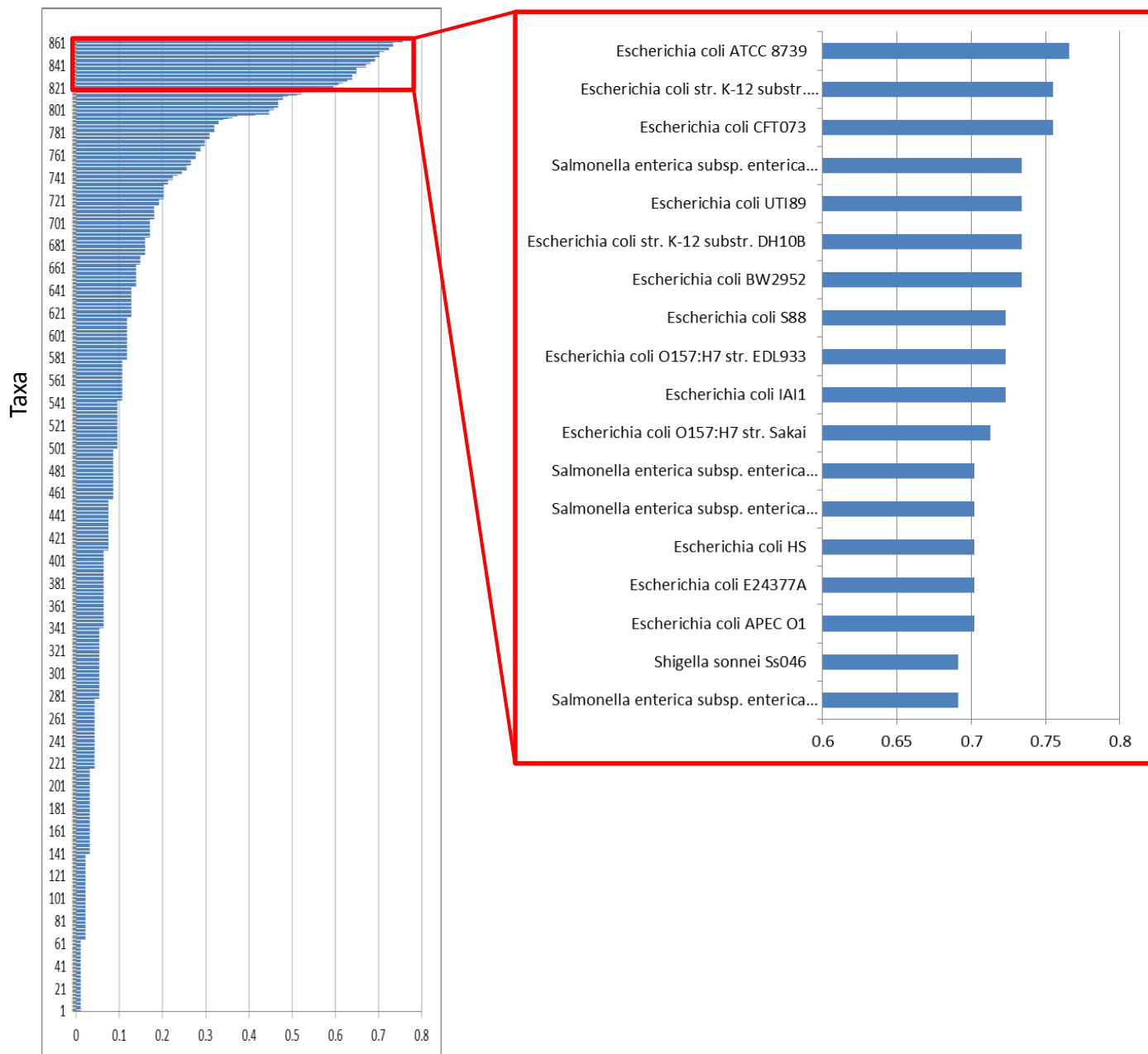


Figure S8. Taxa coverage for subNW2 from the transkingdom network (carbohydrates). Fraction of genes from this subnetwork that is found in the genome of a particular bacterial strain.

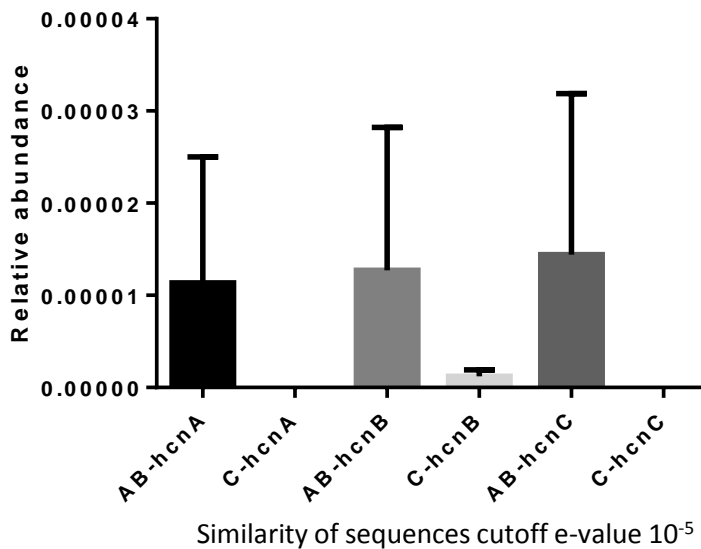
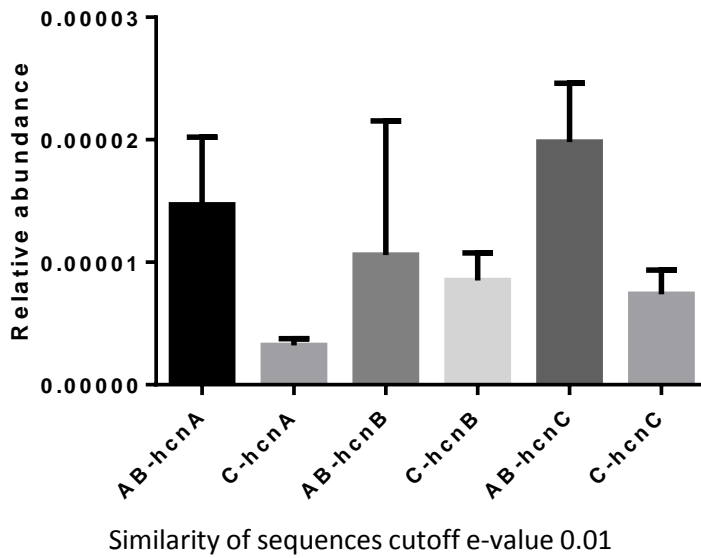


Fig. S9. Increased abundance of three cyanide synthase genes (hcnA, hcnB, hcnC) in microbiota from ABX treated B6 mice. Panels represent different similarity cut-offs (BLAST e-value) of 0.01 (upper panel) and 10<sup>-5</sup>(lower panel) for selecting sequences from metagenomic data similar to hcnA, hcnB and hcnC genes.

One of the many LasR-regulated virulence factors is cyanide synthase ([Pessi and Haas, 2000](#)), and cyanide is one of the most plausible mediators of mitochondria-dependent death after antibiotic treatment. Indeed, cyanide has been demonstrated to be one of the major mediators by which *P. aeruginosa* kills worms ([Gallagher and Manoil, 2001](#)) and neighboring bacteria ([Voggu et al., 2006](#)), and, in neural cells, cyanide-induced death involves mitochondrial inhibition ([Prabhakaran et al., 2007](#)). To see if cyanide could be involved in the mitochondrial dysfunction and apoptosis seen in ABx treated mice, we directly searched the metagenomics data for *P. aeruginosa* genes responsible for cyanide production, and found a definite increase in abundance of hydrogen cyanide synthase genes in the ABx associated microbiota (Figure S9). This increase was missed in our global analysis using SEED annotation, probably because of the very rigorous filtering we implemented, showing that global analysis can, and probably should, be complemented by hypothesis-driven investigations, because the global approaches may fail to detect some biologically meaningful, but not so strong, signals in the data.