# Biogeography and Environmental Conditions Shape Phage and Bacteria Interaction Networks Across the Healthy Human Microbiome

Geoffrey D Hannigan[1], Melissa B Duhaime[2], Danai Koutra[3], and Patrick D Schloss[1,*]

[1]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, 48109

[2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109

[3]Department of Computer Science, University of Michigan, Ann Arbor, Michigan, 48109

[*]To whom correspondence may be addressed.

***Corresponding Author Information***

Patrick D Schloss, PhD
1150 W Medical Center Dr. 1526 MSRB I
Ann Arbor, Michigan 48109
Phone: (734) 647-5801
Email: pschloss@umich.edu

***Running Title***: Network Diversity of the Healthy Human Microbiome

***Journal***: mSystems

***Keywords***: Virome, Microbiome, Graph Theory, Machine Learning

***Text Length***: 35,640 / 50,000 Characters

* *Figures at the end of the document for internal review only.*

## Abstract

Viruses and bacteria are critical components of the human microbiome and play important roles in health and disease. Most previous work has relied on studying microbes and viruses independently, thereby reducing them to two separate communities. Such approaches are unable to capture how these microbial communities interact, such as through processes that maintain community stability or allow phage-host populations to co-evolve. We developed and implemented a network-based analytical approach to describe phage-bacteria network diversity throughout the human body. We accomplished this by building a machine learning algorithm to predict which phages could infect which bacteria in a given microbiome. This algorithm was applied to paired viral and bacterial metagenomic sequence sets from three previously published human cohorts. We organized the predicted interactions into networks that allowed us to evaluate phage-bacteria connectedness across the human body. We found that gut and skin network structures were person-specific and were not conserved among cohabitating family members. High-fat diets and obesity were associated with less connected networks. There were significant differences in network structure between skin sites, with those exposed to the external environment being less connected and more prone to instability. This study quantified and contrasted the diversity of virome-microbiome networks across the human body and illustrated how environmental factors may influence phage-bacteria interactive dynamics. This work provides a baseline for future studies to better understand system perturbations, such as disease states, through ecological networks.

**Word Count**: 243 / 250

# Introduction

Viruses and bacteria are critical components to the human microbiome and play important roles in health and disease. Bacterial communities have been associated with diseases including a range of skin conditions (Hannigan and Grice 2013), acute and chronic wound healing conditions (Hannigan et al. 2014; Loesche et al. 2016), and gastrointestinal diseases including inflammatory bowel disease (He et al. 2016; Norman et al. 2015), *Clostridium difficile* infections (Seekatz et al. 2016), and colorectal cancer (Zackular et al. 2014; Baxter et al. 2014). Altered viromes (virus communities consisting primarily of bacteriophages) have been also associated with various diseases and perturbations including inflammatory bowel disease (Norman et al. 2015; Manrique et al. 2016), periodontal disease (Ly et al. 2014), spread of antibiotic resistance (Modi et al. 2013a), and others (Monaco et al. 2016; Hannigan et al. 2015; Minot et al. 2011; Santiago-Rodriguez et al. 2015; Abeles et al. 2015, 2014). The viruses act in concert with their microbial hosts as a single ecological community (Haerter et al. 2014). Viruses influence their living microbial host communities through processes including lysis, modulating host gene expression (Lindell et al. 2005, Tyler et al. (2013), Hargreaves et al. (2014)), influencing evolutionary processes, such as horizontal gene transfer (Moon et al. 2015, Modi et al. (2013b), Ogg et al. (1981), Frost et al. (2005)) or antagonistic co-evolution (Koskella and Brockhurst 2014), and by altering ecosystem processes and elemental stoichiometry (Jover et al. 2014).

Previous human microbiome work has focused on bacterial and viral communities, but have reduced them to two separate communities by studying them independently (Norman et al. 2015; Manrique et al. 2016; Ly et al. 2014; Monaco et al. 2016; Hannigan et al. 2015; Minot et al. 2011; Santiago-Rodriguez et al. 2015; Abeles et al. 2015, 2014). In reality, bacteria and phage communities are dynamic and complex. They frequently share genetic information and work together to maintain stable ecosystems. Removal of bacteria or phage can disrupt or even collapse those ecosystems (Haerter et al. 2014; Harcombe and Bull 2005; Middelboe et al. 2001; Poisot et al. 2011, 2012; Thompson et al. 2012; Moebus and Nattkemper 1981; Flores et al. 2013, 2011; Poisot and Stouffer 2016; Jover et al. 2015). Relationship-based network approaches allow us to capture this interaction information. We therefore leveraged machine learning and graph theory techniques to characterize human bacterial and phage communities by their inferred relationships. By doing so, we provide a foundation for further studies of disease network dynamics and gain broader insights into human microbiome network diversity across different body sites.

Human bacterial-phage network diversity was established and explored by utilizing three published microbiome datasets that contained paired virus and whole community metagenomic sequence sets dominated by bacteria

(Hannigan et al. 2015; Minot et al. 2011; Reyes et al. 2010; Turnbaugh et al. 2009a). Our approach built off of previous large-scale phage-bacteria microbiome network analyses by inferring interactions using metagenomic datasets, instead of using culture-based techniques, which limit the scale of the experiments and analyses (Flores et al. 2013). This metagenomic interaction inference model built on previous models, which have utilized a variety of techniques including taxonomic-based co-occurrence models using linear relationships between bacteria and phages (Lima-Mendez et al. 2015), as well as nucleotide similarity (including CRISPR targeting) models using OGUs (@ Roux et al. 2016) and whole genomes (Edwards et al. 2015). Our approach uniquely included protein interaction data, and was validated on experimentally determined negative and positive interactions, i.e., "who does and does not infect whom." Through this approach, we were able to provide a basic understanding of the network dynamics associated with phage and bacterial communities on the human body. By building and utilizing a microbiome network, we found that different people, body sites, and anatomical locations not only support distinct microbiome membership and diversity (Hannigan et al. 2015; Minot et al. 2011; Reyes et al. 2010; Turnbaugh et al. 2009a; Grice et al. 2009a; Findley et al. 2013; Costello et al. 2009, Consortium (2012)), but also support ecological communities with distinct communication structures and propensities toward community instability. Through an improved understanding of the healthy state of network structures across the human body, we empower future studies to investigate how these community structures are influenced by disease state and their overall impact on human health.

# Results

## Cohort Curation and Sample Processing

We studied the differences in virus-bacteria interaction networks across healthy human bodies by leveraging previously published sequence sets containing purified virome samples paired with bacterial metagenomes from whole metagenomic shotgun sequences. Our study contained three datasets that explored the impact of diet on the healthy human gut virome (Minot et al. 2011), the impact of anatomical location on the healthy human skin virome (Hannigan et al. 2015), and the viromes of monozygotic twins and their mothers (Reyes et al. 2010; Turnbaugh et al. 2009a). The twin and diet studies utilized multiple displacement amplification methods in their library preparations. These datasets were selected because they included virome samples subjected to virus-like particle (VLP) purification. To this end, they employed combinations of filtration, chloroform/DNase treatment, and cesium chloride gradients to eliminate organismal DNA and thereby allow for direct assessment of both the extracellular and fully-assembled intracellular virome **(Supplemental Figure S1 A-B)** (Minot et al. 2011, Hannigan et al. (2015), Reyes et al. (2010); Turnbaugh et al. 2009a). While the whole metagenomic shotgun sequence samples were not subjected to purification, they primarily consisted of bacteria (Minot et al. 2011, Hannigan et al. (2015), Reyes et al. (2010); Turnbaugh et al. 2009a).

The bacterial and viral sequences from these studies were quality filtered and assembled into contigs. We further grouped the related bacterial and phage contigs into operationally defined units based on their k-mer frequencies and co-abundance patterns, similar to previous reports **(Supplemental Figure S2 - S3)** (Roux et al. 2016). We referred to these operationally defined groups of related contigs as operational genomic units (OGUs). Each OGU represented a genomically similar sub-population of either bacteria or phages. Contig lengths within clusters ranged between $10^3$ and $10^{5.5}$ bp **(Supplemental Figure S2 - S3)**.

## Evaluating the Model to Predict Phage-Bacteria Interactions

We predicted which phage OGUs infected which bacterial OGUs using a random forest model trained on experimentally validated infectious relationships from six previous publications (Jensen et al. 1998; Malki et al. 2015; Schwarzer et al. 2012; Kim et al. 2012; Matsuzaki et al. 1992; Edwards et al. 2015). Only bacteria and phages were used in the model. The training set contained 43 diverse bacterial species and 30 diverse phage strains, including both broad and specific infectious ranges **(Supplemental Figure S4 A - B)**. Phages with linear and circular genomes, as well as ssDNA and dsDNA genomes, were included in the analysis. Because we used DNA sequencing studies, RNA phages were not considered **(Supplemental Figure S4 C-D)**. This training set included both positive relationships (a phage infects a bacterium) and negative relationships (a phage does not infect a bacterium). This allowed us to validate the false positive and false negative rates associated with our candidate models, thereby building upon previous work that only considered positive relationships (Edwards et al. 2015).

Four phage and bacterial genomic features were used to predict infectious relationships between bacteria and phages: 1) genome nucleotide similarities, 2) gene amino acid sequence similarities, 3) bacterial Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) spacer sequences that target phages, and 4) similarity of protein families associated with experimentally identified protein-protein interactions (Orchard et al. 2014). The resulting random forest model performed with an AUC of 0.846, a sensitivity of 0.829, and a specificity of 0.767 **(Figure 1 A)**. The most important predictor in the model was amino acid similarity between genes, followed by nucleotide similarity of whole genomes **(Figure 1 B)**. Protein family interactions were moderately important to the model, and CRISPRs were largely uninformative, due to the lack of identifiable CRISPRs in the dataset and their redundancy with the nucleotide similarity methods **(Figure 1 B)**. Approximately one third of the training set relationships yielded no score and therefore were unable to be assigned an interaction prediction **(Figure 1 C)**.

We used the random forest model to classify the relationships between bacteria and phage operational genomic units, which were then used to build the interactive network. The master network contained the three studies as sub-networks, which themselves each contained sub-networks for each sample **(Figure 1 D)**. Metadata including study, sample ID, disease, and OGU abundance within the community were stored in the master network for downstream analysis **(Supplemental Figure S5)**. The master network was highly connected and contained 72,287 infectious relationships among 578 nodes, 298 phages and 280 bacteria. Although the network was highly connected, not all relationships were present in all samples. As relationships were weighted by the relative abundances of their associated bacteria and phages, lowly abundant relationships could be present but not highly abundant. Like the master network, the skin network exhibited a diameter of 4 (measure of graph size; the greatest number of traversed vertices required between two vertices) and included 99.7% and 99.8% of the master network nodes and edges, respectively **(Figure 1 E - F)**. The phages and bacteria in the gut diet and twin sample sets were more sparsely related: each contained fewer than 150 vertices, fewer than 20,000 relationships, and diameters of 3 **(Figure 1 E - F)**.

## Role of Diet & Obesity in Gut Microbiome Connectivity

Diet is a major environmental factor that influences resource availability and gut microbiome composition and diversity, including bacteria and phages (Minot et al. 2011; Turnbaugh et al. 2009b; David et al. 2014). Previous work in isolated culture-based systems has suggested that changes in nutrient availability are associated with altered phage-bacteria network structures (Poisot et al. 2011), although this has yet to be tested in humans. We therefore hypothesized that a change in diet would also be associated with a change in virome-microbiome network structure in the human gut.

We evaluated the diet-associated differences in gut virome-microbiome network structure by quantifying how central each sample's network was on average. We accomplished this by utilizing two common centrality metrics: degree centrality and closeness centrality. Degree centrality, the simplest centrality metric, was defined as the number of connections each phage made with each bacterium. We supplemented measurements of degree centrality with measurements of closeness centrality. Closeness centrality is a metric of how close each phage or bacterium is to all of the other phages and bacteria in the network. A higher closeness centrality suggests that the effects of genetic information or altered abundance would be more impactful to all other microbes in the system. A network with higher average closeness centrality also indicates an overall greater degree of connections, which suggests a greater resilience against instability. We used this information to calculate the average connectedness per sample, which was corrected for the maximum potential degree of connectedness.

We found that the gut microbiome network structures associated with high-fat diets were less connected than those of low-fat diets **(Figure 2 A-B)**. Tests for statistical differences were not performed due to the small sample size. High-fat diets exhibited reduced degree centrality **(Figure 2 A)**, suggesting bacteria in high-fat environments were targeted by fewer phages and that phage tropism was more restricted. High-fat diets also exhibited decreased closeness centrality **(Figure 2 B)**, indicating that bacteria and phages were more distant from other bacteria and phages in the community. This would make genetic transfer and altered abundance of a given phage or bacterium less capable of impacting other bacteria and phages within the network.

In addition to diet, obesity was found to influence network structure. Obesity-associated networks demonstrated a higher degree centrality **(Figure 2 C)**, but less closeness centrality than the healthy-associated networks **(Figure 2 D)**. These results suggested that the obesity-associated networks are less connected, having microbes further from all other microbes within the community.

## Individuality of Microbial Networks

Skin and gut community membership and diversity are highly personal, with people remaining more similar to themselves than to other people over time (Grice et al. 2009b; Hannigan et al. 2015; Minot et al. 2013). We therefore hypothesized that this personal conservation extended to microbiome network structure. We addressed this hypothesis by calculating the degree of dissimilarity between each subject's network, based on phage and bacteria abundance and centrality. We quantified phage and bacteria centrality within each sample graph using the weighted eigenvector centrality metric. This metric defines central phages as those that are highly abundant ($A_O$ as defined in the methods) and infect many distinct bacteria which themselves are abundant and infected by many other phages. Similarly, bacterial centrality was defined as those bacteria that were both abundant and connected to numerous phages that were themselves connected to many bacteria. We then calculated the similarity of community networks using the weighted eigenvector centrality of all nodes between all samples. Samples with similar network structures were interpreted as having similar capacities for maintaining stability and transmitting genetic material.

We used this network dissimilarity metric to test whether microbiome network structures were more similar within people than between people over time. We found that gut microbiome network structures clustered by person (ANOSIM p-value = 0.005, R = 0.958, **Figure 3 A**). Network dissimilarity within each person over the 8-10 day sampling period was less than the average dissimilarity between that person and others, although this difference was not statistically significant (p-value = 0.125, **Figure 3 B**). The lack of statistical confidence was likely due to the small sample size of this dataset. Although there was evidence for gut network conservation among individuals, we found no evidence for conservation of gut network structures within families. The gut network structures were not more similar within families (twins and their mothers; intrafamily) compared to other families (inter-family) (p-value = 0.312, **Figure 3 C**).

Skin microbiome network structure was strongly conserved within individuals (p-value < 0.001, **Figure 3 D**). This distribution was similar when separated by anatomical sites. Most sites were statistically significantly more conserved within individuals **(Supplemental Figure S6)**.

## Association Between Environmental Stability and Network Structure Across the Human Skin Landscape

Extensive work has illustrated differences in diversity and composition of the healthy human skin microbiome between anatomical sites, including bacteria, virus, and fungal communities (Grice et al. 2009b; Findley et al. 2013; Hannigan et al. 2015). These communities vary by degree of skin moisture, oil, and environmental exposure. As viruses are known to influence microbial diversity and community composition, we hypothesized that microbe-virus network structure would be specific to anatomical sites, as well. To test this, we evaluated the changes in network structure between anatomical sites within the skin dataset.

The average centrality of each sample was quantified using the weighted eigenvector centrality metric. Intermittently moist skin sites (dynamic sites that fluctuate between being moist and dry) were significantly less connected than the more stable moist and sebaceous environments (p-value < 0.001, **Figure 4 A)**. Also, skin sites that were protected from the environment (occluded) were much more highly connected than those that were constantly exposed to the environment or only intermittently occluded (p-value < 0.001, **Figure 4 B)**.

To supplement this analysis, we compared the network signatures using the centrality dissimilarity approach described above. The dissimilarity between samples was a function of shared relationships, degree of centrality, and bacteria/phage abundance. When using this supplementary approach, we found that network structures significantly clustered by moisture, sebaceous, and intermittently moist status **(Figure 4 C,E)**. Occluded sites were significantly different from exposed and intermittently occluded sites, but there was no difference between exposed and intermittently occluded sites **(Figure 4 D,F)**. These findings provide further support that skin microbiome network structure differs significantly between skin sites.

## Discussion

Foundational work has provided a baseline understanding of the human microbiome by characterizing bacterial and viral diversity across the human body (Grice et al. 2009a; Findley et al. 2013; Hannigan et al. 2015; Costello et al. 2009, Consortium (2012); Schloss and Handelsman 2005; Minot et al. 2011). Here, we offer an approach to describe how phage-bacteria networks differ throughout the human body to provide a baseline for future studies of how and why microbiome networks differ in disease states. We developed and implemented a network-based analytical model to evaluate the basic properties of the human microbiome through bacteria and phage relationships, instead of membership or diversity alone. This enabled the application of network theory to provide a new perspective on complex ecological communities. We utilized metrics of connectivity to model the extent to which communities of bacteria and phages interact through mechanisms such as horizontal gene transfer, modulated bacterial gene expression, and alterations in abundance.

Just as gut microbiome and virome composition and diversity are conserved in individuals (Hannigan et al. 2015; Grice et al. 2009a; Findley et al. 2013; Minot et al. 2013), gut and skin microbiome network structures were conserved within individuals over time. Gut network structure was not conserved among family members. These findings suggested that the community properties inferred from microbiome interaction network structures, such as stability, the potential for horizontal gene transfer between members, co-evolution of populations, etc., were person-specific. These properties may be impacted by personal factors ranging from the body's immune system to external environmental conditions, such as climate and diet.

The ability of environmental conditions to shape gut and skin microbiome interaction network structure was further supported by our finding that diet and skin location were associated with altered network structures. We found evidence that diet was sufficient to alter gut microbiome network connectivity. Although our sample size was small, our findings provided evidence that high-fat diets were less connected than low-fat diets, and high-fat diets therefore may lead to less stable communities with a decreased ability for microbes to directly influence one another. We supported this finding with the observation that obesity may have been associated with decreased network connectivity. Together these findings suggest the food we eat may not only impact which microbes colonize our guts, but may also impact their infectious interactions. Further work will be required to characterize these relationships with a larger cohort.

In addition to diet, the skin environment influenced the microbiome interaction network structure as well. Network structure differed between environmentally exposed and occluded skin sites. The sites under greater environmental fluctuation and exposure (the exposed and intermittently exposed sites) were less connected and therefore were predicted to have a higher propensity for instability. Likewise, intermittently moist sites demonstrated less connectedness than the more stable moist and sebaceous sites. Together these data suggested that body sites under greater degrees of fluctuation harbored less connected, potentially less stable microbiomes. This points to a link between microbiome and environmental stability, and warrants further investigation.

While these findings take us an important step closer to understanding the microbiome through interspecies relationships, there are caveats to and considerations regarding the approach. First, as with most classification models, the infection classification model developed and applied is only as good as its training set – in this case, the collection of experimentally-verified positive and negative infection data, where genomes of all members are fully sequenced. Large-scale experimental screens for phage and bacteria infectious interactions that report high-confidence negative interactions (i.e., no infection) will provide more robust model training and improved model performance. Furthermore, just as we have improved on previous modeling efforts, we expect that new and creative scoring metrics will be integrated into this model to improve performance.

Second, although our analyses offer an informative proof of concept, this work was done retrospectively and relied on existing data up to seven years old. These archived datasets were limited by the technology and costs of the time. This resulted in small sequencing effort (as compared to today's dataset sizes) and thus datasets that were sub-optimally powered for the statistical analyses we strive for today. Further, two studies, the diet and twin studies, relied on multiple displacement amplification (MDA) in their library preparations–an approach used to overcome the large nucleic acids requirements, especially of older sequencing library generation protocols. MDA results in significant biases in microbial community composition (Yilmaz et al. 2010), as well as toward ssDNA viral genomes (Kim et al. 2008; Kim and Bae 2011), thus rendering the resulting microbial and viral metagenomes non-quantitative. Future work that employs larger sequence datasets and that avoids the use of MDA will build on and validate our findings, as well as inform the design and interpretation of further studies.

Finally, the networks in this study were built using operational genomic units, which represent groups of highly similar bacteria or phage genomes or genome fragments as clustered sub-populations. This operationally defined approach allows us to study whole community networks, but limits our ability to make conclusions about interactions among specific phage or bacterial species or populations. Although this approach lacks the resolution for drawing conclusions about specific bacteria or phages, future work could begin addressing these questions through improved binning methods and deeper metagenomic shotgun sequencing, as well as an improved understanding of what defines ecologically and evolutionarily cohesive units for both phage and bacteria, e.g., Polz et al., 2006 (Polz et al. 2006).

Such work will include more sophisticated approaches to defining operational genomic units and their taxonomic underpinnings (e.g. OGUs defined as clustering at the genus or species level). Previous work has implemented more sophisticated clustering and validation approaches that allow for such refined operational genomic unit clustering, such as the phylogenomic analyses performed to cluster cyanophage isolate genomes into informative groups using shared gene content, average nucleotide identity of shared genes, and pairwise differences between genomes (Gregory et al. 2016). Similar clustering definition and validation methods, both computational and experimental, have been implemented in some metagenomic sequencing studies (Roux et al. 2016; Brum et al. 2015; Deng et al. 2014; Minot et al. 2012) and could offer yet another level of sophistication to our network-based analyses, especially by allowing us to make conclusions about the specific virus ad bacterial interactions within the networks.

Together our work takes an initial step towards defining bacteria-virus interaction profiles as a characteristic of human-associated microbial communities. This approach revealed the impacts different human environments (e.g., the skin and gut) can have on microbiome connectivity. By focusing on relationships between bacterial and viral communities, they are studied as the interacting cohorts they are, rather than as independent entities. While our developed bacteria-phage interaction framework is a novel conceptual advance, the microbiome also consists of archaea and small eukaryotes, including fungi and *Demodex* mites (Hannigan and Grice 2013; Grice and Segre 2011). Further, the components of the microbiome can interact with human immune cells and other non-microbial community members (Round and Mazmanian 2009). Future work will build from our approach and include these additional community members and their diverse interactions and relationships (e.g., beyond phage-bacteria). This will result in a more robust network and a more holistic understanding of the human-associated microbiome and the evolutionary and ecological processes that drive its assembly and function.

# Materials & Methods

## Data Availability

All associated source code is available on GitHub at the following repository:

https://github.com/SchlossLab/Hannigan_ConjunctisViribus_GenRes_2017

## Data Acquisition & Quality Control

Raw sequencing data and associated metadata were acquired from the NCBI sequence read archive (SRA). Supplementary metadata were acquired from the same SRA repositories and their associated manuscripts. The gut virome diet study (SRA: `SRP002424`), twin virome studies (SRA: `SRP002523`; `SRP000319`), and skin virome study (SRA: `SRP049645`) were downloaded as `.sra` files. Sequencing files were converted to `fastq` format using the `fastq-dump` tool of the NCBI SRA Toolkit (v2.2.0). Sequences were quality trimmed using the Fastx toolkit (v0.0.14) to exclude bases with quality scores below 33 and shorter than 75 bp (Hannon). Paired end reads were filtered to exclude sequences missing their corresponding pair using the `get_trimmed_pairs.py` script available in the source code.

## Contig Assembly

Contigs were assembled using the Megahit assembly program (v1.0.6) (Li et al. 2016). A minimum contig length of 1 kb was used. Iterative k-mer stepping began at a minimum length of 21 and progressed by 20 until 101. All other default parameters were used.

## Contig Abundance Calculations

Contigs were concatenated into two master files prior to alignment, one for bacterial contigs and one for phage contigs. Sample sequences were aligned to phage or bacterial contigs using the Bowtie2 global aligner (v2.2.1) (Langmead and Salzberg 2012). We defined a mismatch threshold of 1 bp and seed length of 25 bp. Sequence abundance was calculated from the Bowtie2 output using the `calculate_abundance_from_sam.pl` script available in the source code.

## Operational Genomic Unit Binning

Contigs often represent large fragments of genomes. In order to reduce redundancy and the resulting artificially inflated genomic richness within our dataset, it was important to bin contigs into operational units based on their similarity. This approach is conceptually similar to the clustering of related 16S rRNA sequences into operational taxonomic units (OTUs), although here we are clustering contigs into operational genomic units (OGUs) (Schloss and Handelsman 2005).

Contigs were clustered using the CONCOCT algorithm (v0.4.0) (Alneberg et al. 2014). Because of our large dataset and limits in computational efficiency, we randomly subsampled the dataset to include 25% of all samples, and used these to inform contig abundance within the CONCOCT algorithm. CONCOCT was used with a maximum of 500 clusters, a k-mer length of four, a length threshold of 1 kb, 25 iterations, and exclusion of the total coverage variable.

OGU abundance ($A_O$) was obtained as the sum of the abundance of each contig ($A_j$) associated with that OGU. The abundance values were length corrected such that:

$$A_O = \frac{10^7 \sum_{j=1}^{k} A_j}{\sum_{j=1}^{k} L_j}$$

Where `L` is the length of each contig `j` within the OGU.

## Phage OGU Identification

To confirm a lack of phage sequences in the bacterial OGU dataset, we performed blast nucleotide alignment of the bacterial OGU representative sequences using an e-value < $10^{-25}$, which was stricter than the $10^{-10}$ threshold used in the random forest model below. We used a stricter threshold because we know there are genomic similarities between bacteria and phage OGUs from the interactive model, but we were interested in contigs with high enough similarity to references that they may indeed be from phages. 2% of the OGUs had nucleotide similarities to known bacteriophage genomes, although the alignments were short (a couple of kb) and represented a small fraction of the alignment sequences.

## Open Reading Frame Prediction

Open reading frames (ORFs) were identified using the Prodigal program (V2.6.2) with the meta mode parameter and default settings (Hyatt et al. 2012).

## Classification Model Creation and Validation

The classification model for predicting interactions was built using experimentally validated bacteria-phage infections or validated lack of infections from six studies (Jensen et al. 1998; Malki et al. 2015; Schwarzer et al. 2012; Kim et al. 2012; Matsuzaki et al. 1992; Edwards et al. 2015). Associated reference genomes were downloaded from the European Bioinformatics Institute (see details in source code). The model was created based on the four metrics listed below.

The four scores were used as parameters in a random forest model to classify bacteria and bacteriophage pairs as either having infectious interactions or not. The classification model was built using the Caret R package (v6.0.73) (Kuhn). The model was trained using five-fold cross validation with ten repeats. Pairs without scores were classified as not interacting. The model was optimized using the ROC value. The resulting model performance was plotted using the plotROC R package.

### Identify Bacterial CRISPRs Targeting Phages

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) were identified from bacterial genomes using the PilerCR program (v1.06) (Edgar 2007). Resulting spacer sequences were filtered to exclude spacers shorter than 20 bp and longer than 65 bp. Spacer sequences were aligned to the phage genomes using the nucleotide BLAST algorithm with default parameters (v2.4.0) (Camacho et al. 2009). The mean percent identity for each matching pair was recorded for use in our classification model.

### Detect Matching Prophages within Bacterial Genomes

Temperate bacteriophages infect and integrate into their bacterial host's genome. We detected integrated phage elements within bacterial genomes by aligning phage genomes to bacterial genomes using the nucleotide BLAST algorithm and a minimum e-value of 1e-10. The resulting bitscore of each alignment was recorded for use in our classification model.

### Identify Shared Genes Between Bacteria and Phages

As a result of gene transfer or phage genome integration during infection, phages may share genes with their bacterial hosts, providing us with evidence of phage-host pairing. We identified shared genes between bacterial and phage genomes by assessing amino acid similarity between the genes using the Diamond protein alignment algorithm (v0.7.11.60) (Buchfink et al. 2015). The mean alignment bitscores for each genome pair were recorded for use in our classification model.

### Protein - Protein Interactions

The final method used for predicting infectious interactions between bacteria and phages was the detection of pairs of genes whose proteins are known to interact. We assigned bacterial and phage genes to protein families by aligning them to the Pfam database using the Diamond protein alignment algorithm. We then identified which pairs of proteins were predicted to interact using the Pfam interaction information within the Intact database (Orchard et al. 2014). The mean bitscores of the matches between each pair were recorded for use in the classification model.

## Interaction Network Construction

The bacteria and phage operational genomic units (OGUs) were scored using the same approach as outlined above. The infectious pairings between bacteria and phage OGUs were classified using the random forest model described above. The predicted infectious pairings and all associated metadata were used to populate a graph database using Neo4j graph database software (v2.3.1) (). This network was used for downstream community analysis.

## Centrality Analysis

We quantified the centrality of graph vertices using three different metrics, each of which provided different information graph structure. When calculating these values, let $G(V, E)$ be an undirected, unweighted graph with $|V| = n$ nodes and $|E| = m$ edges. Also, let $\mathbf{A}$ be its corresponding adjacency matrix with entries $a_{ij} = 1$ if nodes $V_i$ and $V_j$ are connected via an edge, and $a_{ij} = 0$ otherwise.

Briefly, the **closeness centrality** of node $V_i$ is calculated taking the inverse of the average length of the shortest paths (d) between nodes $V_i$ and all the other nodes $V_j$. Mathematically, the closeness centrality of node $V_i$ is given as:

$$C_C\left(V_i\right)=\left(\sum_{j=1}^{n} d\left(V_i, V_j\right)\right)^{-1}$$

The distance between nodes $(\mathrm{d})$ was calculated as the shortest number of edges required to be traversed to move from one node to another.

Intuitively, the **degree centrality** of node $V_i$ is defined as the number of edges that are incident to that node:

$$C_D\left(V_i\right)=\sum_{j=1}^{n} a_{ij}$$

where $a_{ij}$ is the $ij^{th}$ entry in the adjacency matrix $\mathbf{A}$.

The eigenvector centrality of node $V_i$ is defined as the $i^{th}$ value in the first eigenvector of the associated adjacency matrix $\mathbf{A}$. Conceptually, this function results in a centrality value that reflects the connections of the vertex, as well as the centrality of its neighboring vertices.

The **centralization** metric was used to assess the average centrality of each sample graph $\mathbf{G}$. Centralization was calculated by taking the sum of each vertex $V_i$'s centrality from the graph maximum centrality $C_w$, such that:

$$C\left(G\right)=\frac{\sum_{i=1}^{n} Cw-c\left(V_i\right)}{T}$$

The values were corrected for uneven graph sizes by dividing the centralization score by the maximum theoretical centralization $(\mathrm{T})$ for a graph with the same number of vertices.

Degree and closeness centrality were calculated using the associated functions within the igraph R package (v1.0.1) (Csardi and Nepusz).

### Network Relationship Dissimilarity

We assessed similarity between graphs by evaluating the shared centrality of their vertices, as has been done previously. More specifically, we calculated the dissimilarity between graphs $G_i$ and $G_j$ using the Bray-Curtis dissimilarity metric and eigenvector centrality values such that:

$$B\left(G_i, G_j\right) = 1 - \frac{2C_{ij}}{C_i + C_j}$$

Where $C_{ij}$ is the sum of the lesser centrality values for those vertices shared between graphs, and $C_i$ and $C_j$ are the total number of vertices found in each graph. This allows us to calculate the dissimilarity between graphs based on the shared centrality values between the two graphs.

### Statistics and Comparisons

Differences in intrapersonal and interpersonal network structure diversity, based on multivariate data, were calculated using an analysis of similarity (ANOSIM). Statistical significance of univariate Eigenvector centrality differences were calculated using a paired Wilcoxon test.

Statistical significance of differences in univariate eigenvector centrality measurements of skin virome-microbiome networks were calculated using a pairwise Wilcoxon test, corrected for multiple hypothesis tests using the Holm correction method. Multivariate eigenvector centrality was measured as the mean differences between cluster centroids, with statistical significance measured using an ANOVA and post hoc Tukey test.

## Acknowledgments

We thank the members of the Schloss lab for their underlying contributions. GDH was supported in part by the Molecular Mechanisms in Microbial Pathogenesis Training Program (T32 AI007528). GDH and PDS were supported in part by funding from the NIH (P30DK034933, U19AI09087, and U01AI124255).

## Disclosure Declaration

The authors report no conflicts of interest.

# Figures

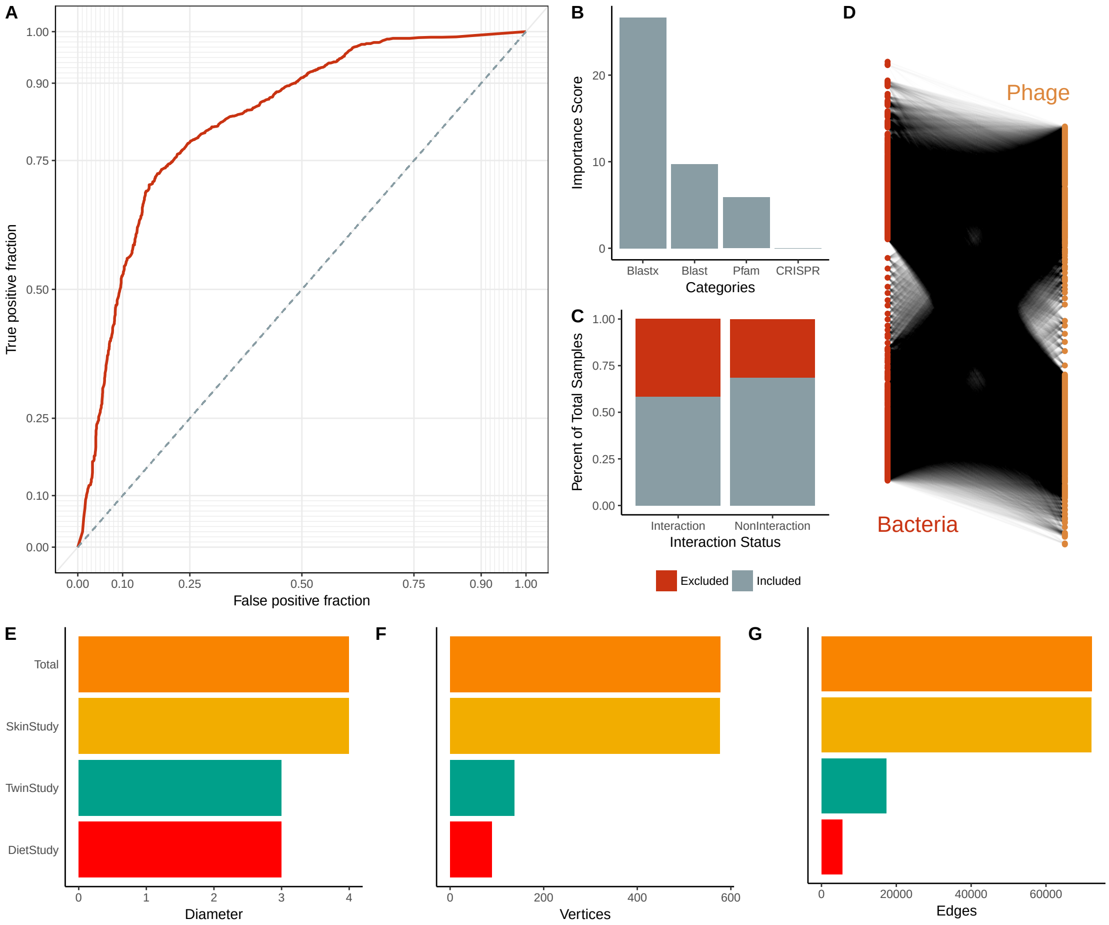

Figure 1: **Summary of Multi-Study Network Model.** *(A) Average ROC curve used to create the microbiome-virome infection prediction model. (B) Importance scores associated with the metrics used in the random forest model to predict relationships between bacteria and phages. The importance score is defined as the mean decrease in accuracy of the model when a feature (e.g. Pfam) is excluded. (C) Proportions of samples included (gray) and excluded (red) in the model. Samples were excluded from the model because they did not yield any scores. Those interactions without scores were defined as not having interactions. (D) Bipartite visualization of the resulting phage-bacteria network. This network includes information from all three published studies. (E) Network diameter (measure of graph size; the greatest number of traversed vertices required between two vertices), (F) number of vertices, and (G) number of edges (relationships) for the total network (yellow) and the individual study sub-networks (diet study = red, skin study = green, twin study = orange).*

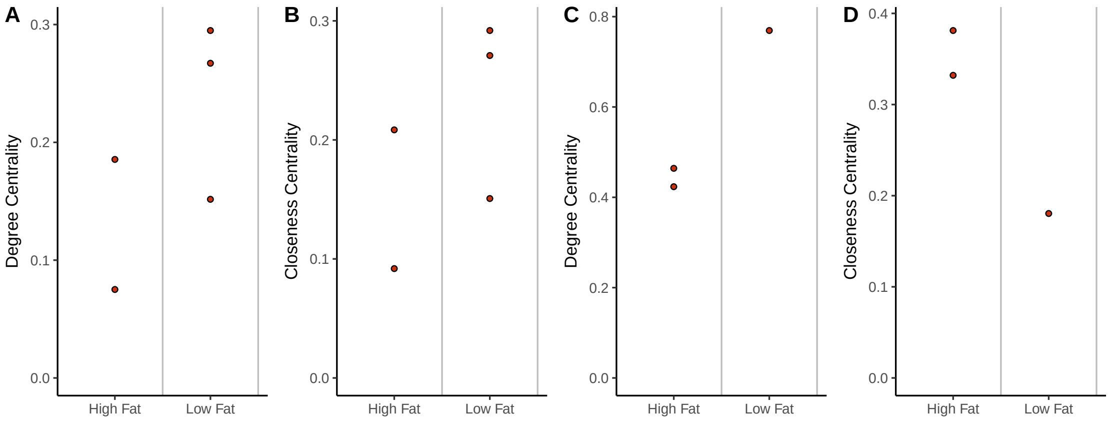

Figure 2: **Impact of Diet and Obesity on Gut Network Structure.** *(A) Quantification of average degree centrality (number of edges per node) and (B) closeness centrality (average distance from each node to every other node) of gut microbiome networks of subjects limited to exclusively high-fat or low-fat diets. Lines represent the mean degree of centrality for each diet. (C) Quantification of average degree centrality and (D) closeness centrality between obese and healthy adult women.*

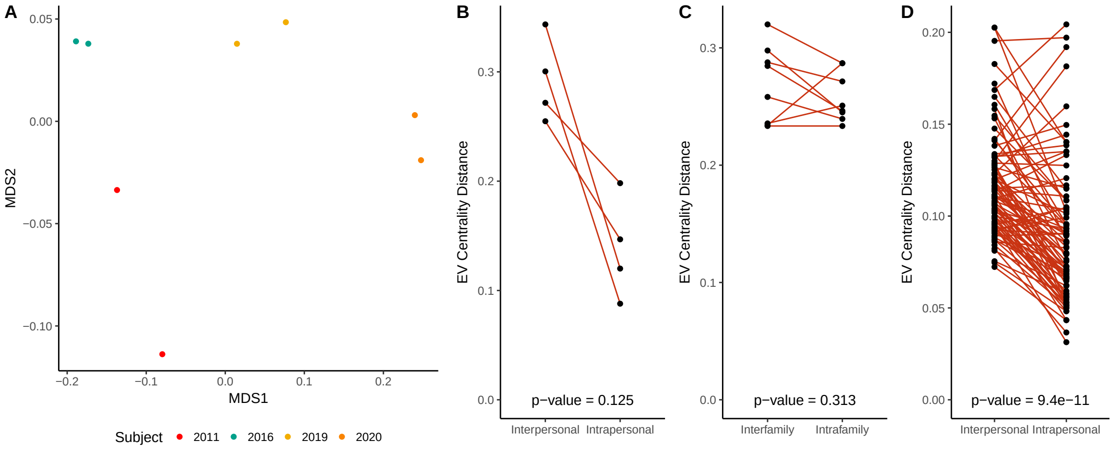

Figure 3: **Intrapersonal vs Interpersonal Network Dissimilarity Across Different Human Systems.** *(A) NMDS ordination illustrating network dissimilarity between subjects over time. Each sample is colored by subject, with each sample pair collected 8-10 days apart. Dissimilarity was calculated using the Bray-Curtis metric based on abundance weighted eigenvector centrality signatures, with a greater distance representing greater dissimilarity in bacteria and phage centrality and abundance. (B) Quantification of gut network dissimilarity within the same subject over time (intrapersonal) and the mean dissimilarity between the subject of interest and all other subjects (interpersonal). The p-value is also provided. (C) Quantification of gut network dissimilarity within subjects from the same family (intrafamily) and the mean dissimilarity between subjects within a family and those of other families (interfamily). The p-value is also provided. (D) Quantification of skin network dissimilarity within the same subject and anatomical location over time (intrapersonal) and the mean dissimilarity between the subject of interest and all other subjects at the same time and the same anatomical location (interpersonal). P-value was calculated using a paired Wilcoxon test.*

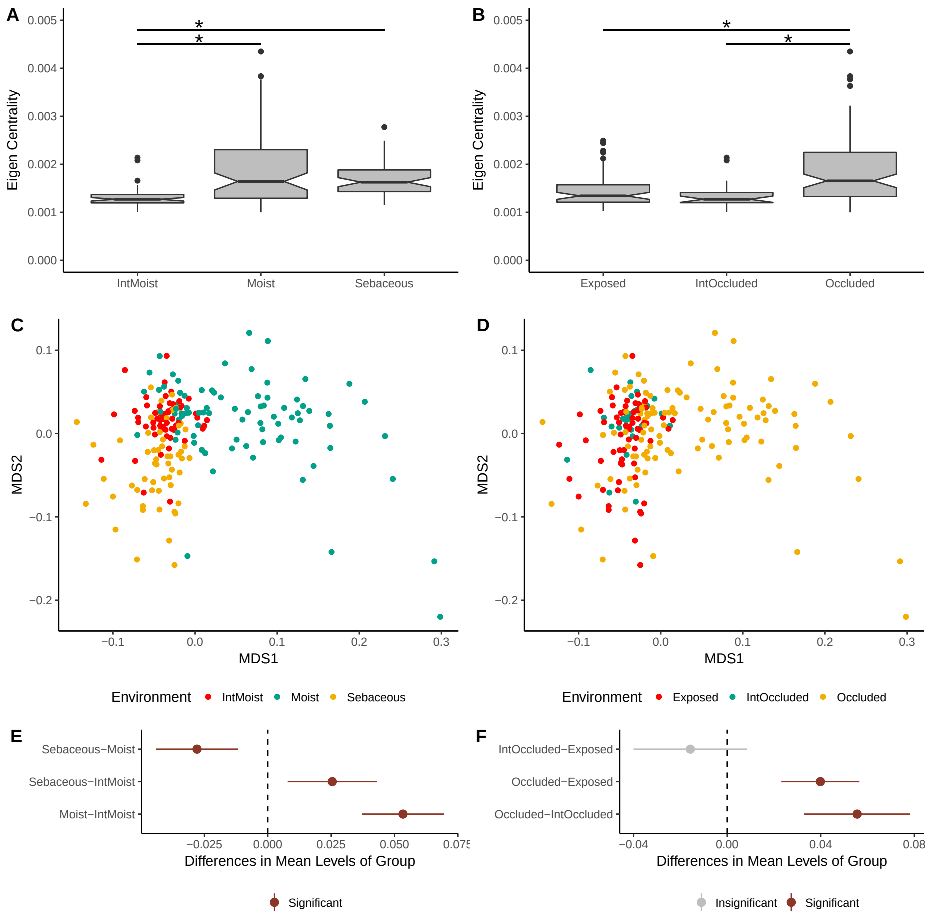

Figure 4: **Impact of Skin Micro-Environment on Microbiome Network Structure.** *(A) Notched box-plot depicting differences in average eigenvector centrality between moist, intermittently moist, and sebaceous skin sites and (B) occluded, intermittently occluded, and exposed sites. Notched box-plots were created using ggplot2 and show the median (center line), the inter-quartile range (IQR; upper and lower boxes), the highest and lowest value within 1.5 * IQR (whiskers), outliers (dots), and the notch which provides an approximate 95% confidence interval as defined by 1.58 * IQR / sqrt(n). (C) NMDS ordination depicting the differences in skin microbiome network structure between skin moisture levels and (D) occlusion. Samples are colored by their environment and their dissimilarity to other samples was calculated as described in figure 3. (E) The statistical differences of networks between moisture and (F) occlusion status were quantified with an anova and post hoc Tukey test. Cluster centroids are represented by dots and the extended lines represent the associated 95% confidence intervals. Significant comparisons (p-value < 0.05) are colored in red, and non-significant comparisons are gray.*

411 ## Supplemental Figures

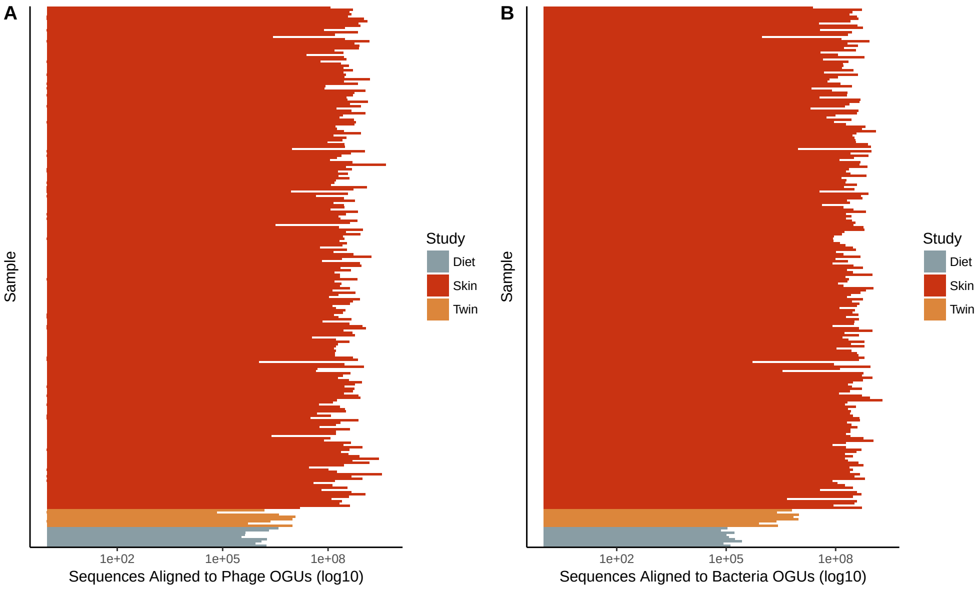

Figure S1: **Sequencing Depth Summary.** *Number of sequences that aligned to (A) Phage and (B) Bacteria operational genomic units per sample and colored by study.*

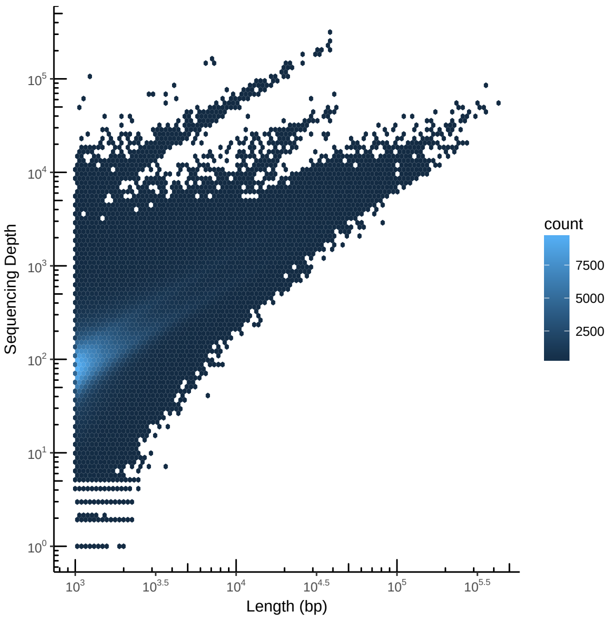

Figure S2: **Contig Summary Statistics.** *Scatter plot heat map with each hexagon representing the abundance of contigs. Contigs are organized by length on the x-axis and the number of aligned sequences on the y-axis.*

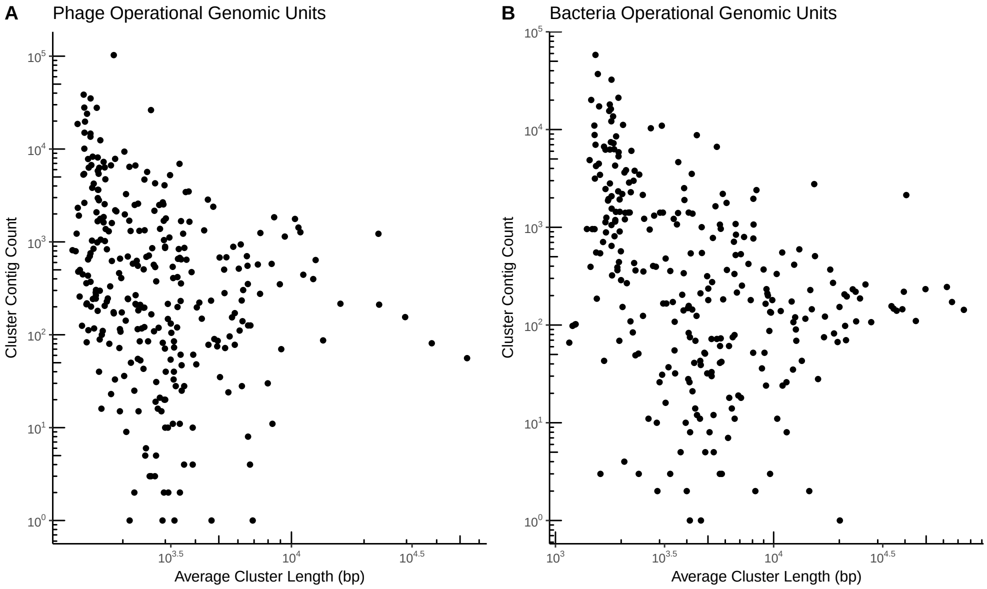

Figure S3: **Operational Genomic Unit Summary Statistics.** *Scatter plot with operational genomic unit clusters organized by average contig length within the cluster on the x-axis and the number of contigs in the cluster on the y-axis. Operational genomic units of (A) bacteriophages and (B) bacteria are shown.*

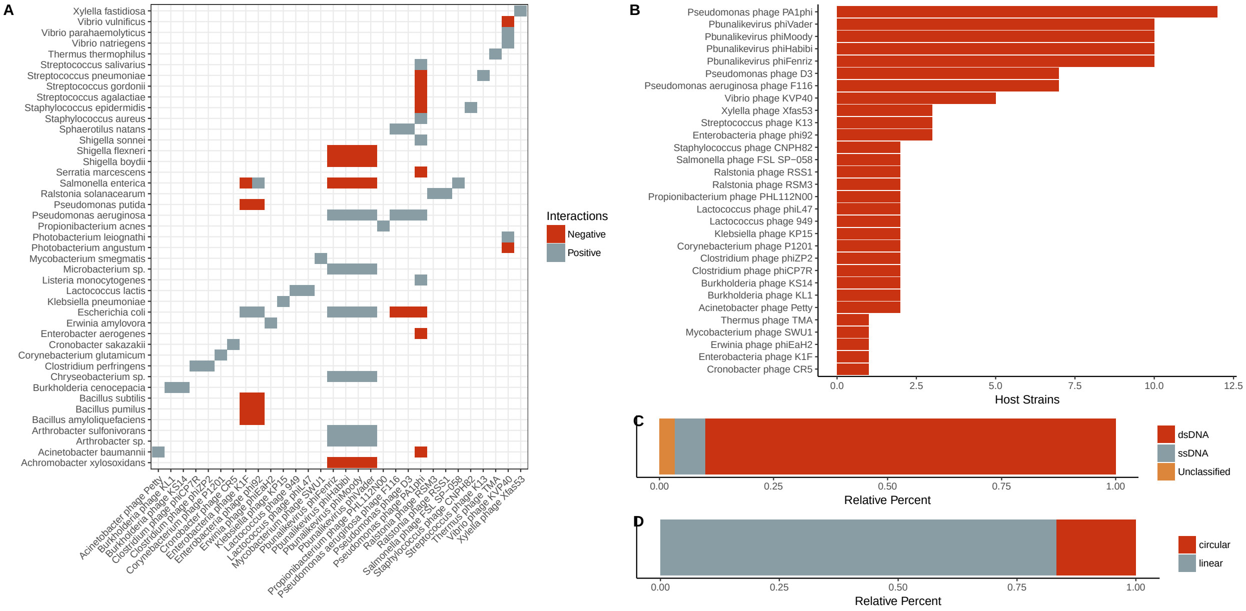

Figure S4: **Summary information of validation dataset used in the interaction predictive model.** *A) Categorical heat-map highlighting the experimentally validated positive and negative interactions. Only bacteria species are shown, which represent multiple reference strains. Phages are labeled on the x-axis and bacteria are labeled on the y-axis. B) Quantification of bacterial host strains known to exist for each phage. C) Genome strandedness and D) linearity of the phage reference genomes used for the dataset.*

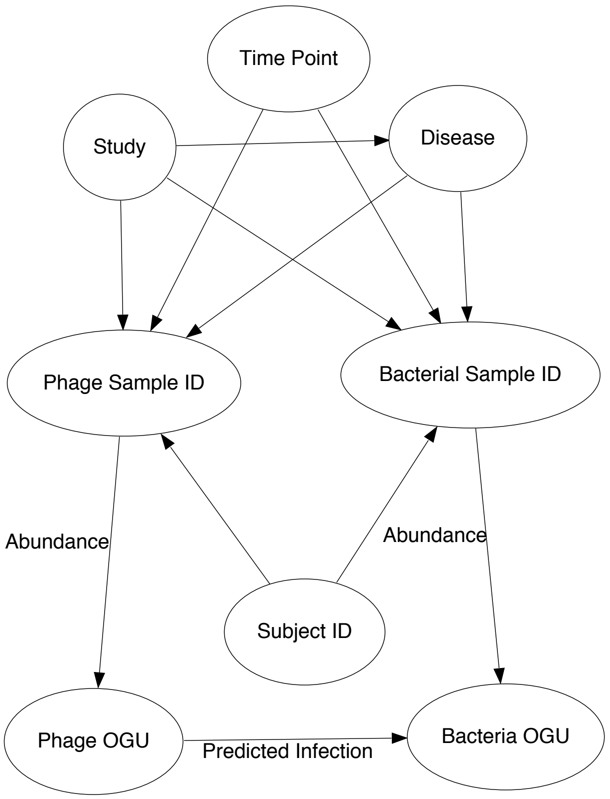

Figure S5: **Structure of the interactive network.** *Metadata relationships to samples (Phage Sample ID and Bacteria Sample ID) included the associated time point, the study, the subject the sample was taken from, and the associated disease. Infectious interactions were recorded between phage and bacteria operational genomic units (OGUs). Sequence count abundance for each OGU within each sample was also recorded.*

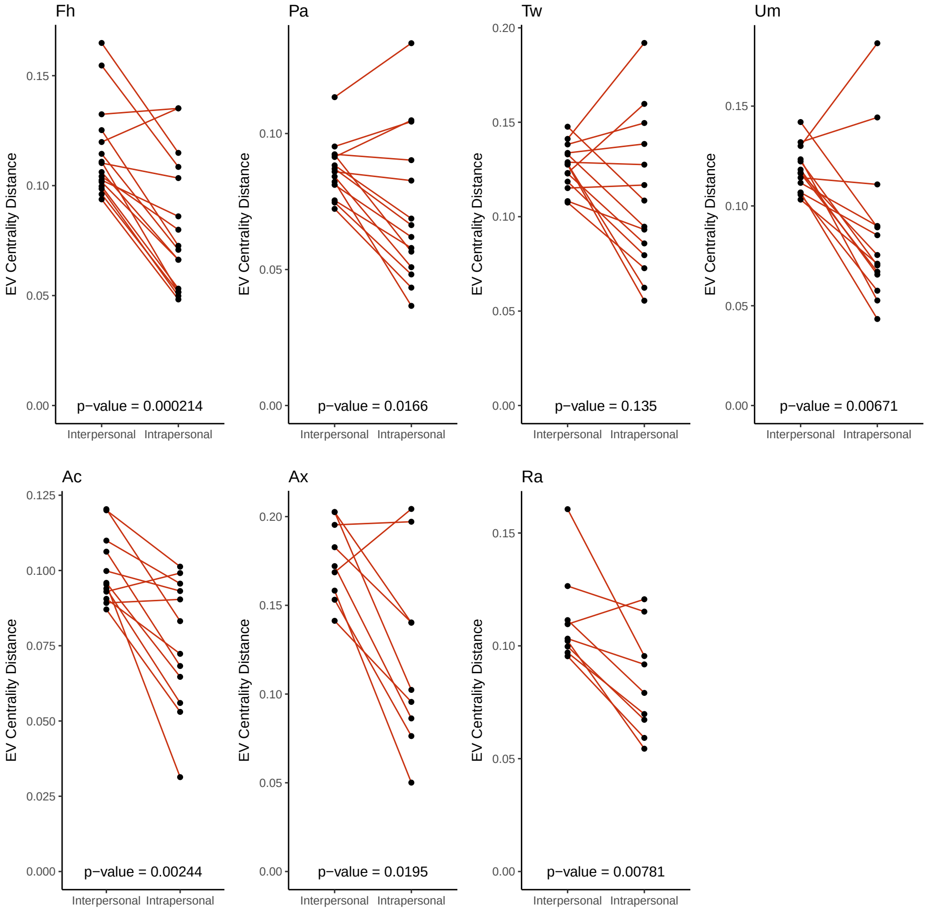

Figure S6: **Intrapersonal vs Interpersonal Dissimilarity of the Skin.** *Quantification of skin network dissimilarity within the same subject and anatomical location over time (intrapersonal) and the mean dissimilarity between the subject of interest and all other subjects at the same time and the same anatomical location (interpersonal), separated by each anatomical site (forehead [Fh], palm [Pa], toe web [Tw], umbilicus [Um], antecubital fossa [Ac], axilla [Ax], and retroauricular crease [Ra]). P-value was calculated using a paired Wilcoxon test.*

## References

Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT. 2015. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLOS ONE* **10**: e0134941.

Abeles SR, Robles-Sikisaka R, Ly M, Lum AG, Salzman J, Boehm TK, Pride DT. 2014. Human oral viruses are personal, persistent and gender-consistent. 1–15.

Alneberg J, Bjarnason BS aacute ri, Bruijn I de, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods* 1–7.

Baxter NT, Zackular JP, Chen GY, Schloss PD. 2014. Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. *Microbiome* **2**: 20.

Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, Vargas C de, Gasol JM, et al. 2015. Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* **348**: 1261498–1261498.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**: 59–60.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 1.

Consortium THMP. 2012. A framework for human microbiome research. *Nature* **486**: 215–221.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694–1697.

Csardi G, Nepusz T. The igraph software package for complex network research.

David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al. 2014. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**: 559–563.

Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, Sullivan MB. 2014. Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. *Nature* **513**: 242–245.

Edgar RC. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**: 18.

Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2015. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews* **40**: 258–272.

Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, NIH Intramural Sequencing Center Comparative Sequencing Program, et al. 2013. Topographic diversity of fungal and bacterial communities in human skin. *Nature* 1–6.

Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. 2011. Statistical structure of host-phage interactions. *Proceedings of the National Academy of Sciences of the United States of America* **108**: E288–97.

Flores CO, Valverde S, Weitz JS. 2013. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal* **7**: 520–532.

Frost LS, Leplae R, Summers AO, Toussaint A. 2005. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* **3**: 722–732.

Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, Maitland A, Chittick L, Dos Santos F, Weitz JS, et al. 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* **17**: 930.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, et al. 2009a. Topographical and Temporal Diversity of the Human Skin Microbiome. *Science* **324**: 1190–1192.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, et al. 2009b. Topographical and Temporal Diversity of the Human Skin Microbiome. *Science* **324**: 1190–1192.

Grice EA, Segre JA. 2011. The skin microbiome. *Nature Reviews Microbiology* **9**: 244–253.

Haerter JO, Mitarai N, Sneppen K. 2014. Phage and bacteria support mutual diversity in a narrowing staircase of coexistence. *The ISME Journal* **8**: 2317–2326.

Hannigan GD, Grice EA. 2013. Microbial Ecology of the Skin in the Era of Metagenomics and Molecular Microbiology. *Cold Spring Harbor Perspectives in Medicine* **3**: a015362–a015362.

Hannigan GD, Hodkinson BP, McGinnis K, Tyldsley AS, Anari JB, Horan AD, Grice EA, Mehta S. 2014. Culture-independent pilot study of microbiota colonizing open fractures and association with severity, mechanism, location, and complication from presentation to early outpatient follow-up. *Journal of Orthopaedic Research* **32**: 597–605.

Hannigan GD, Meisel JS, Tyldsley AS, Zheng Q, Hodkinson BP, SanMiguel AJ, Minot S, Bushman FD, Grice EA. 2015. The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *mBio* **6**: e01578–15.

Hannon GJ. FASTX-Toolkit. GNU Affero General Public License.

Harcombe WR, Bull JJ. 2005. Impact of phages on two-species bacterial communities. *Applied and Environmental Microbiology* **71**: 5254–5259.

Hargreaves KR, Kropinski AM, Clokie MR. 2014. Bacteriophage behavioral ecology: How phages alter their bacterial host's habits. *Bacteriophage* **4**: e29866.

He Q, Li X, Liu C, Su L, Xia Z, Li X, Li Y, Li L, Yan T, Feng Q, et al. 2016. Dysbiosis of the fecal microbiota in the TNBS-induced Crohn's disease mouse model. *Applied Microbiology and Biotechnology* 1–10.

Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. 2012. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**: 2223–2230.

Jensen EC, Schrader HS, Rieland B, Thompson TL, Lee KW, Nickerson KW, Kokjohn TA. 1998. Prevalence of broad-host-range lytic bacteriophages of Sphaerotilus natans, Escherichia coli, and Pseudomonas aeruginosa. *Applied and Environmental Microbiology* **64**: 575–580.

Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. 2014. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nature Reviews Microbiology* **12**: 519–528.

Jover LF, Flores CO, Cortez MH, Weitz JS. 2015. Multiple regimes of robust patterns between network structure and biodiversity. *Scientific Reports* **5**: 17856.

Kim K-H, Bae J-W. 2011. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Applied and Environmental Microbiology* **77**: 7663–7668.

Kim KH, Chang HW, Nam YD, Roh SW. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Applied and ...*.

Kim S, Rahman M, Seol SY, Yoon SS, Kim J. 2012. Pseudomonas aeruginosa bacteriophage PA1Ø requires type IV pili for infection and shows broad bactericidal and biofilm removal activities. *Applied and Environmental Microbiology* **78**: 6380–6385.

Koskella B, Brockhurst MA. 2014. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiology Reviews* **38**: 916–931.

Kuhn M. caret: Classification and Regression Training.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *METHODS* **102**: 3–11.

Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, et al. 2015. Ocean plankton. Determinants of community structure in the global plankton interactome. *Science* **348**: 1262073–1262073.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–89.

Loesche M, Gardner SE, Kalan L, Horwinski J, Zheng Q, Hodkinson BP, Tyldsley AS, Franciscus CL, Hillis SL, Mehta S, et al. 2016. Temporal stability in chronic wound microbiota is associated with poor healing. *Journal of Investigative Dermatology*.

Ly M, Abeles SR, Boehm TK, Robles-Sikisaka R, Naidu M, Santiago-Rodriguez T, Pride DT. 2014. Altered Oral Viral Ecology in Association with Periodontal Disease. *mBio* **5**: e01133–14–e01133–14.

Malki K, Kula A, Bruder K, Sible E. 2015. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. *Virology*.

Manrique P, Bolduc B, Walk ST, Oost J van der, Vos WM de, Young MJ. 2016. Healthy human gut phageome. *Proceedings of the National Academy of Sciences of the United States of America* 201601060.

Matsuzaki S, Tanaka S, Koga T, Kawata T. 1992. A Broad-Host-Range Vibriophage, KVP40, Isolated from Sea Water. *Microbiology and Immunology* **36**: 93–97.

Middelboe M, Hagström A, Blackburn N, Sinn B, Fischer U, Borch NH, Pinhassi J, Simu K, Lorenz MG. 2001. Effects of Bacteriophages on the Population Dynamics of Four Strains of Pelagic Marine Bacteria. *Microbial Ecology* **42**: 395–406.

Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. 2013. Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 12450–12455.

Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. 2011. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* **21**: 1616–1625.

Minot S, Wu GD, Lewis JD, Bushman FD. 2012. Conservation of gene cassettes among diverse viruses of the human gut. *PLOS ONE* **7**: e42342.

Modi SR, Lee HH, Spina CS, Collins JJ. 2013a. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**: 219–222.

Modi SR, Lee HH, Spina CS, Collins JJ. 2013b. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**: 219–222.

Moebus K, Nattkemper H. 1981. Bacteriophage sensitivity patterns among bacteria isolated from marine waters. *Helgoländer Meeresuntersuchungen* **34**: 375–385.

Monaco CL, Gootenberg DB, Zhao G, Handley SA, Ghebremichael MS, Lim ES, Lankowski A, Baldridge MT, Wilen CB, Flagg M, et al. 2016. Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host and Microbe* **19**: 311–322.

Moon BY, Park JY, Hwang SY, Robinson DA, Thomas JC, Fitzgerald JR, Park YH, Seo KS. 2015. Phage-mediated horizontal transfer of a Staphylococcus aureus virulence-associated genomic island. *Scientific Reports* **5**: 9784.

Norman JM, Handley SA, Baldridge MT, Droit L, Liu CY, Keller BC, Kambal A, Monaco CL, Zhao G, Fleshner P, et al. 2015. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**: 447–460.

Ogg JE, Timme TL, Alemohammad MM. 1981. General Transduction in Vibrio cholerae. *Infection and Immunity* **31**: 737–741.

Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, et al. 2014. The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research* **42**: D358–63.

Poisot T, Canard E, Mouillot D, Mouquet N, Gravel D. 2012. The dissimilarity of species interaction networks. *Ecology letters* **15**: 1353–1361.

Poisot T, Lepennetier G, Martinez E, Ramsayer J, Hochberg ME. 2011. Resource availability affects the structure of a natural bacteriabacteriophage community. *Biology letters* **7**: 201–204.

Poisot T, Stouffer D. 2016. How ecological networks evolve. *bioRxiv*.

Polz MF, Hunt DE, Preheim SP, Weinreich DM. 2006. Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**: 2009–2021.

Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**: 334–338.

Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nature reviews Immunology* **9**: 313–323.

Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, et al. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**: 689–693.

Santiago-Rodriguez TM, Ly M, Bonilla N, Pride DT. 2015. The human urine virome in association with urinary tract infections. *Frontiers in Microbiology* **6**: 14.

Schloss PD, Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology* **71**: 1501–1506.

Schwarzer D, Buettner FFR, Browning C, Nazarov S, Rabsch W, Bethe A, Oberbeck A, Bowman VD, Stummeyer K, Mühlenhoff M, et al. 2012. A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. *Journal of Virology* **86**: 10384–10398.

Seekatz AM, Rao K, Santhosh K, Young VB. 2016. Dynamics of the fecal microbiome in patients with recurrent and nonrecurrent Clostridium difficile infection. *Genome medicine* **8**: 47.

Thompson RM, Brose U, Dunne JA, Hall RO, Hladyz S, Kitching RL, Martinez ND, Rantala H, Romanuk TN, Stouffer DB, et al. 2012. Food webs: reconciling the structure and function of biodiversity. *Trends in ecology & evolution* **27**: 689–697.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, et al. 2009a. A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.

Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. 2009b. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine* **1**: 6ra14–6ra14.

Tyler JS, Beeri K, Reynolds JL, Alteri CJ, Skinner KG, Friedman JH, Eaton KA, Friedman DI. 2013. Prophage induction is enhanced and required for renal disease and lethality in an EHEC mouse model. *PLoS Pathogens* **9**: e1003236.

Yilmaz S, Allgaier M, Hugenholtz P. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* **7**: 943–944.

Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. 2014. The human gut microbiome as a screening tool for colorectal cancer. *Cancer prevention research (Philadelphia, Pa)* **7**: 1112–1121.

Neo4j.