Understanding Human Virome Biogeography Through Relationships with Bacterial Hosts

Geoffrey D Hannigan Melissa B Duhaime Patrick D Schloss

Contents

Introduction

The virome is a crucial component to microbial ecosystems and plays an important role in human health and disease. Altered viromes have been associated with numerous diseases and environmental perturbations including IBD[1], periodontal disease[2], and others[3–8]. The human virome has also been implicated in promoting antibiotic resistance throughout human-associated microbial communities[4,9]. These community shifts are not mere reflections of the bacterial communities, but rather act in concert with them as a single community[1,10]. Understanding the human virome is essential for more completely understanding the link between the microbiome and human health.

Communities of bacteria and their viruses (bacteriophages) are dynamic and complex. Bacteria and phages act in concert to maintain balanced, efficient ecosystems and their removal can disrupt or even collapse the system (cite isolated community work and modeling). Previous reports have forced bacterial community analyses concepts (e.g. poorly defined taxonomy, alpha diversity, beta diversity) onto phage communities in an attempt to identify community signatures associated with health and disease. This approach treats the phage community as an isolated system, which is inadequate given the metabolic and otherwise functional reliance of phages on their bacterial hosts. A more appropriate analysis will attempt to understand the virome as a subset of a greater interacting community by focusing on its relationships with its bacterial host community, rather than the phage entities alone.

Previous groups have benefited from more sophisticated analyses that focus on the relationships within ecosystems, including those between bacteria and phages. Such approaches allow for the creation of relationship networks that provide unique information into system biodiversity and functionality, including genetic material and resource transfer as well as ecosystem stability[11–18]. This work has shown that environmental conditions, including resource availability, impact ecological network structure. In isolated phage - bacteria systems, decreased resource availability has been shown to alter virome relationship structures, decreasing connectance and making those communities more vulnerable to network disintegration and extinction events[11]. Approaches such as this are valuable to our understanding of microbial ecology, but have yet to be applied to the human virome and the myriad landscapes the human virome inhabits. Here we focus our analysis on virome relationship with the aim of better understanding its ecology associated with different disease and resource availability states.

To investigate how networks of bacteria and phages change upon environmental disruptions, we leveraged three published microbiome datasets (one of which was published across two manuscripts) with paired virus and bacterial metagenomic sequence sets[4,5,19,20]. Sites included both the gut and skin. We built off of previous work on large-scale phage - bacteria ecological network analysis by inferring interactions using metagenomic datasets, instead of previous culture-based techniques[13,14]. Our metagenomic interaction inference model is powered beyond previous models by its inclusion of protein interaction data, inclusion of negative interactions as well as positive, and the use of a more sophisticated machine learning algorithm.

Just as the human microbiome field has benefited from an understanding of how different conditions impact microbial (primarily bacterial) community composition and diversity, we aim to provide an understanding of how the relationships between bacteria and phages change within the microbiome. Because our findings provide insight into how environmental conditions impact network communication, microbial hubs, and vulnerability, they will better inform microbiome-based therapeutics (e.g. pro-biotics), antibiotics, and horizontal gene transfer between elements.

Results

The Virome of the Human Microbiome Dataset

To provide a broad understanding of the human virome, we utilized all available high quality microbiome datasets that contained paired virus and and bacterial metagenomic samples. This resulted in the inclusion of four previously published datasets for three human virome studies. These studies included the impact of diet

on the gut virome[5], the impact of anatomical location on the skin virome[4], and the virome of monozygotic twins and their mothers[19,20]. The viromes associated with these datasets were subjected to virus-like particle (VLP) purification to eliminate other organism DNA including bacteria, fungi, and human. This approach is advantageous because it allows us to study the active virome because only VLPs will have been sequenced. Additionally, it provides confidence that the majority of the sequences within the virome are in fact viral and that we are not aligning bacterial genomic DNA to other bacteria in the metagenomic datasets.

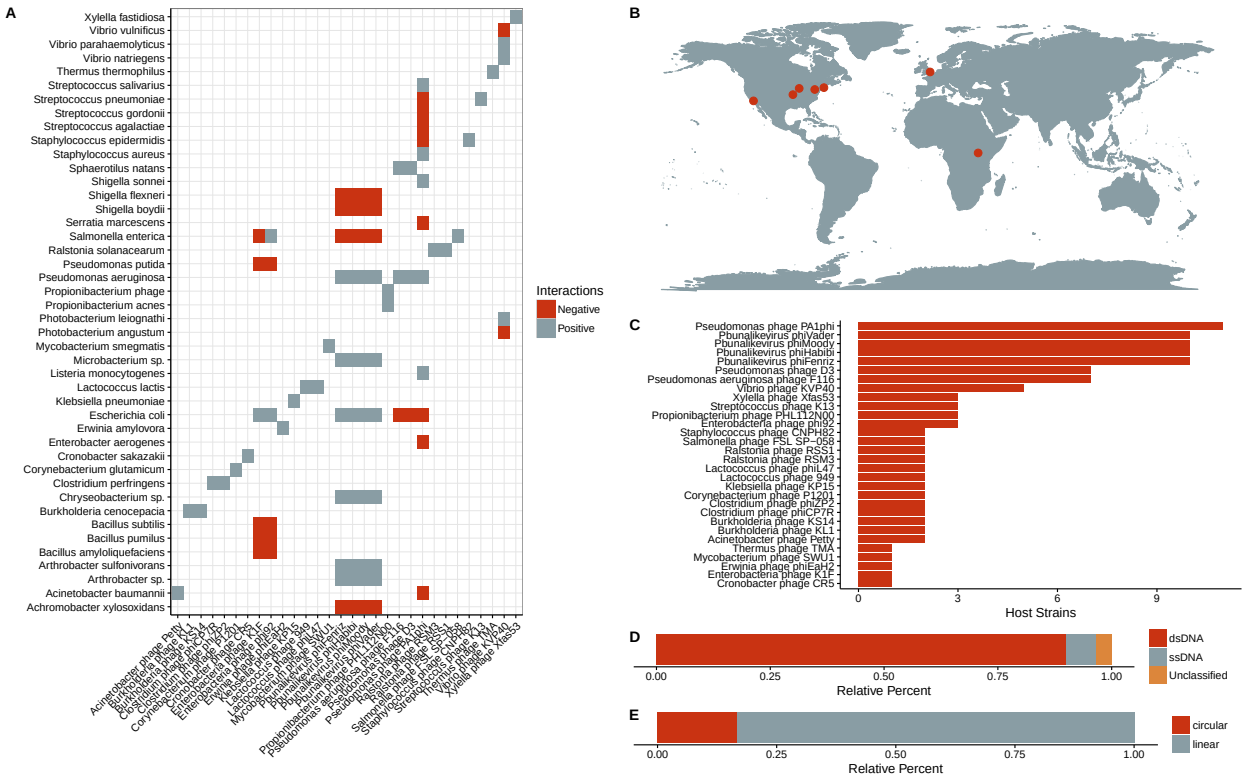

Figure 1: *Summary information of validation dataset used in the interaction predictive model. A) Categorical heatmap highlighting the experimentally validated positive and negative interactions. Only bacteria species are shown, which represent multiple reference strains. Phages are labeled on the x-axis and bacteria are labeled on the y-axis. B) World map illustrating the sampling locations used in the study (red dots). C) Quantificaiton of bacterial host strains known to exist for each phage. D) Genome strandedness and E) linearity of the phage reference genomes used for the dataset.*

The combined study raw sequences were quality filtered according to our high threshold and assembled into contigs that represent either complete viral genomes or genomic fragments. We assembled approximately 30,000 contigs whose sequencing depth ranged from ten to over ten thousand sequences. Contigs were tens of thousands of base pairs long. A large subset of contigs assembled as complete circles, suggesting complete coverage of a subset of viral genome sequences. I need to clean up this data for the manuscript.

Modeling Phage-Bacteria Interactions

We used Neo4J graph database software to construct a network of predicted interactions between bacteria and bacteriophages. Results from a variety of complementary interaction prediction approaches were layered into a single network. *In vitro*, experimentally validated interactive relationships were taken from the existing literature. Clustered Regularly Inter-spaced Short Palindromic Repeats (CRISPRs) are a sort of bacterial adaptive immune system that serves as a genomic record of phage infections by preserving genomic content from the infectious phage genome. These records were used to predict infectious relationships between

bacteria and phages. Infectious relationships were also predicted by identifying expected protein-protein interactions and known interacting protein domains between phages and their bacterial hosts. We finally used nucleotide blast to identify genomic similarity between bacteriophage genomes and sections of bacterial genomes. Such a match is a good predictor of an interaction between the phage and it's bacterial host.

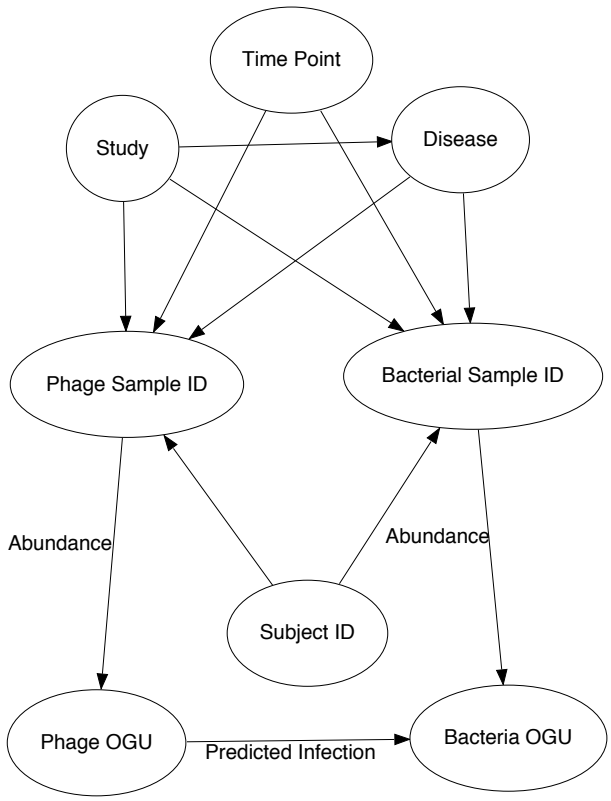

Figure 2: *Diagram illustrating the structure of the interactive network. Metadata relationships to samples (Phage Sample ID and Bacteria Sample ID) included the associated time point, the study, the subject the sample was taken from, and the associated disease. Infectious interactions were recorded between phage and bacteria operational genomic units (OGUs). Seuquence count abundance for each OGU within each sample was also recorded.*

We began by working in a controlled data environment in which the interactions and lack of interactions had been experimentally validated **(Figure 1 A)**. This dataset was extracted from manuscripts published between 1992 - 2015 and includes sampling representation from North America, Africa, and Europe **(Figure 1 B)**[21–26]. Many of the phages are known to target multiple bacterial hosts **(Figure 1 C)**. The majority of the reference phages used contained linear dsDNA genomes **(Figure 1 D-E)**. It is important to note the strength of our approach in that we used data of confirmed non-interactions as well as confirmed interactions. Previous approaches have claimed to perform tests of sensitivity and specificity, but assumed a lack of empirical evidence denoted a lack of interactions. Our approach circumvents this problematic assumption.

We used four predictive score categories of the controlled dataset with a tuned random forest model to classify each sample as an interaction or lack of interaction. The model was validated using repeated k-fold cross validation with k = 5 and ten repetitions. The model was optimized using the receiver operating characteristic (ROC) algorithm for the higher area under the curve (AUC) as implemented in R {caret}. The resulting model exhibited an AUC of 0.853, a sensitivity of 0.851, and a specificity of 0.774 **(Figure 3)**. These parameters describe only the interactions that were scored. Those that did not have scores were classified as having no interaction prior to predictive modeling. The most important predictor in the model was nucleotide similarity between genes, followed by nucleotide similarity of whole genomes. Protein family

(Pfam) interactions were moderately important to the model, while CRISPRs were minimally important (redundant with the blast information).

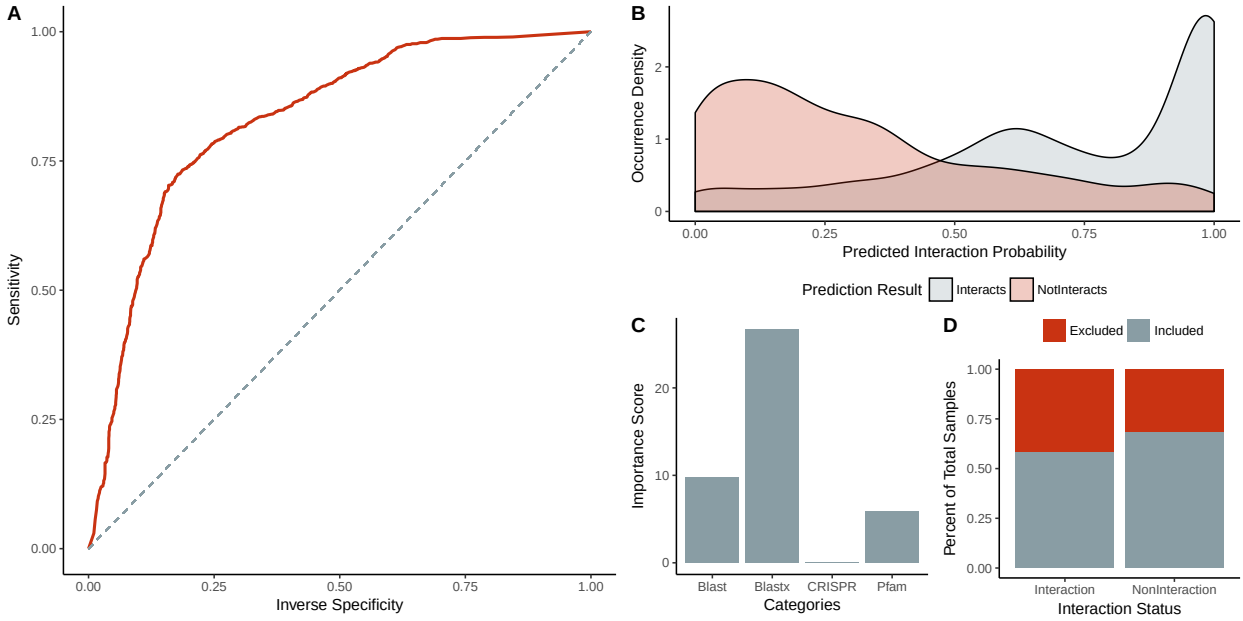

Figure 3: *Random forest model for bacteria - phage interactions. A) ROC curve of the ten iterations used to create the prediction model. B) Density plot of the distribution of sample interaction probability. Groups indicate whether the sample represented an interaction. C) Importance scores associated with the criteria used to create the random forest model. D) Proportions of samples excluded from model learning due to a lack of scoring. The true interaction status of the sample is noted on the x-axis and bars are colored by the proportion of sample excluded (red) and included (grey) in model training.*

Properties of the Global Virome Network

The total human virome network was highly connected and contained 578 nodes, 298 of which represented phage OGUs and 280 that represented bacterial OGUs **(Figure 4 A)**. There was a total of 72,287 infectious relationships between the bacterial and phage nodes, again supporting our finding that the bacteria and phages were overall highly connected throughout the human virome.

We additionally visualized all of the study subgraphs and found that they visually varied in size and distribution, likely due to the discrepancy in sequencing depth and sample size between the studies **(Figure 4 B - D)**.

Inter- and Intra-personal Diversity of Gut Microbial Networks

Previous work has show strong intra-personal diversity in the viromes and bacterial communities of the human gut and skin. Understanding this conservation of the microbiome has been important for establishing a basic understanding of microbiome individuality and even interpersonal transfer. We hypothesized that like other aspects of these microbial communities (e.g. diversity, community composition), there is a strong conservation of network structure in the skin and gut.

To test this hypothesis, we calculated the relationship dissimilarity (Hamming distance) among the same subjects over time, and the same subjects across different skin sites. When comparing gut microbial networks between the same subjects over time, we found that they were more similar to each other than to other

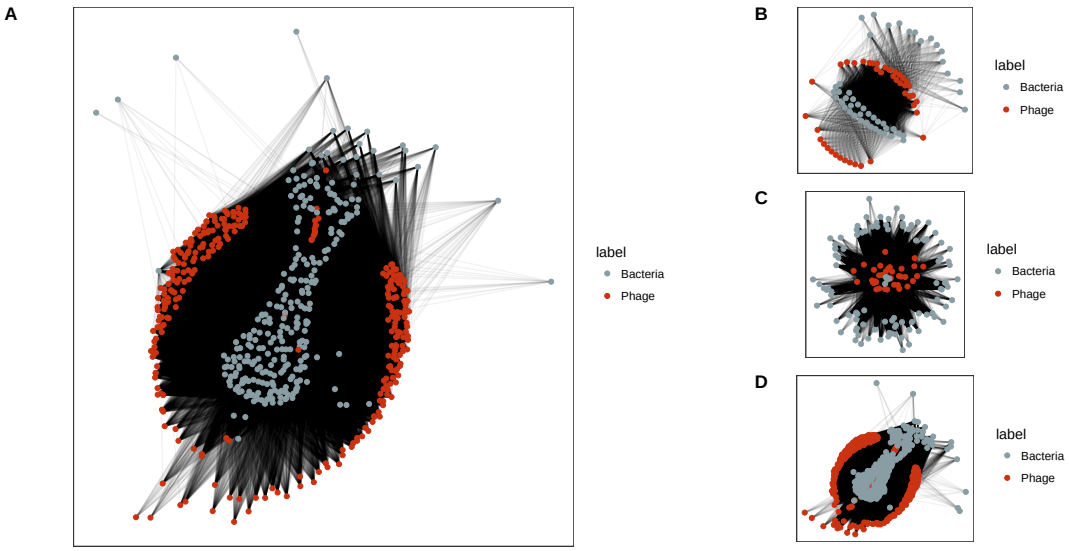

Figure 4: *Network diagram of the phage - bacteria relationships. A) The total network including all three studies. B) Subgraph representing all data associated with the diet gut virome study. C) Subgraph representing all data within the twin gut virome study. D) All data associated with the skin virome study.*

members of the study. The same trend was observed on the skin. The network relationship structure was more highly conserved within people across different anatomical locations than when compared to other people, all with different anatomical locations. Thus, much like bacterial diversity, microbial networks are conserved within individuals.

Role of Diet in Gut Network Structure

Diet is a major environmental factor that influences resource availability and gut microbiome composition, including bacteria and phages. To this end, we followed up on previous work that evaluated the role diet plays in modulating the gut virome by evaluating the role diet plays in virome host relationships within the gut. Instead of focusing on taxa and diversity, we are focusing on the relationships between bacteria and phages.

We evaluated the differences in gut phage-bacteria networks using three metrics: phage centrality, page rank, relationship Shannon entropy, and relationship dissimilarity (Hamming Distance). Using the alpha centrality metric, we found that diet had no significant impact of the degrees of phage centrality with the systems **(Figure 6 B)**. Likewise, we found a lack similarity between the Shannon entropy (alpha diversity) of the system between subjects fed with low or high fat diets **(Figure 6 C)**. This indicated a lack of difference in within-sample structure (centrality and relationship diversity) and led us to evaluate the differences in structure given shared composition between the networks.

We evaluated the differences between sample composition using the Hamming distances between the node edges. We first evaluated the degree of sample similarity between and within subjects over time. This allowed us to investigate whether gut virome networks are more similar within individuals compared to between individuals. We found that the gut microbial networks are more similar within the same subject over time compared to between subjects **(Figure 6 D)**. When comparing diet, we found that the networks were more highly similar within diet classes **(Figure 6 A)**. We also found that high fat diets were more consistent across subjects while the low fat diets were highly variable **(Figure 6 E)**.

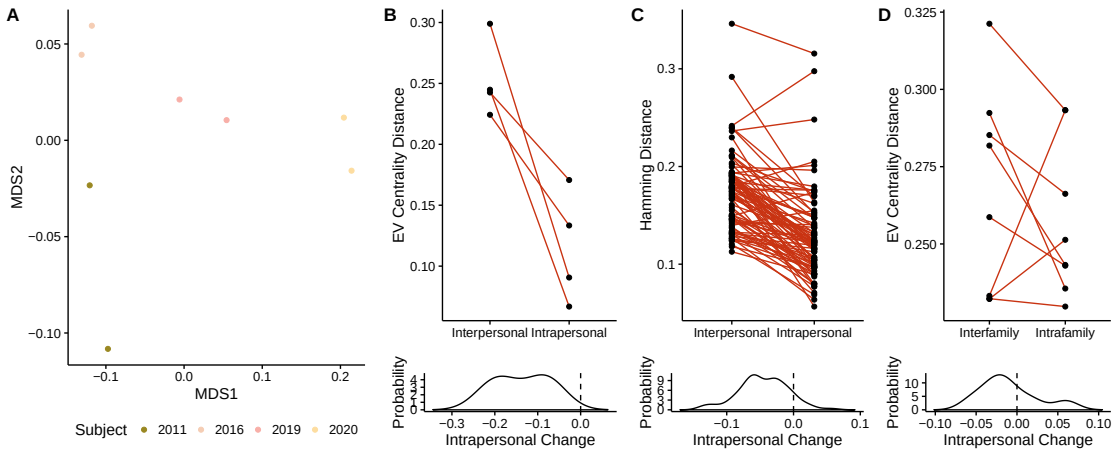

Figure 5: *Inter- vs intra-personal diversity of the human skin virome. A) NMDS ordination of gut virome diet samples by Hamming distances between networks. Samples are colored by subject ID with the distances within subjects representing 8 - 10 days between sampling. Patients were more significantly more similar to each other compared to other subjects. B) Quantification of dissimilarity within (intra-personal) and between (interpersonal) subjects. These were significantly different. C) Inter- and intra-personal distances of the skin virome within anatomical sites on each person and between people. These were significantly different. Dissimilarity between members within and between families while D) including and E) excluding mothers from the twin sets.*

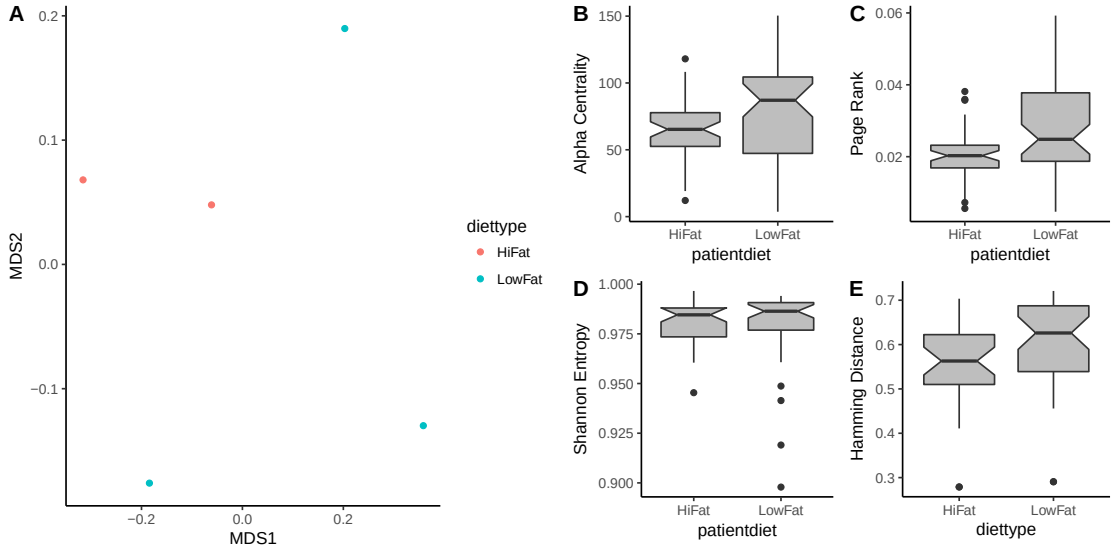

Figure 6: *Impact of diet on different aspects of the gut phage-bacteria ecological network. A) NMDS ordination visualizing the differences in networks between patients on either high or low fat diets. The results were statistically significant by ANOSIM ($p < 0.05$). Ordination based on Hamming distances. Lack of statistically significant difference in B) node centrality and C) Shannon Entropy between patients ingesting either low or high fat diet. D)Hamming distances between samples from D) the same (intra-personal) and different (interpersonal) subjects, as well as E) patients ingesting low or high fat diets. Observed differences were statistically significant ($p < 0.05$).*

Association Between Obesity and Virome - Host Relationships

The association between the microbiome and obesity remains a point of discussion among microbiome researchers and have seen conflicting evidence in past years. Although not a primary objective of our study, we were able to provide a preliminary observation of the link between the microbiome interactive network and obesity. The twin study incorporated into our analysis included three mothers, one of which was obese. Although the conclusions we can make are limited due to a low power, we found a lower degree of centrality in the obese network compared to the two non-obese networks **(Figure 7)**. While this is insufficient evidence for claiming a link between gut microbiome networks and obesity, it does support the utility of these techniques and warrants further, dedicated investigation.

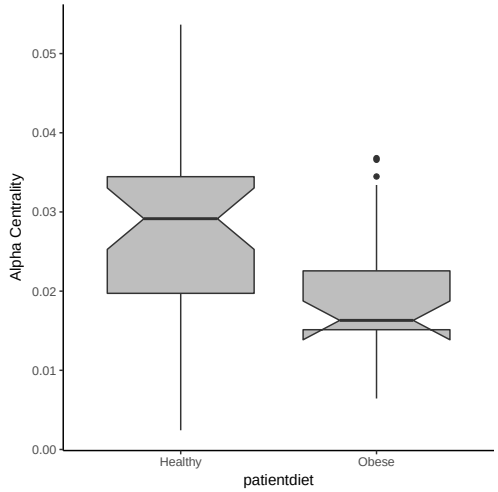

Figure 7: *Obesity is associated with decreased less network centrality. Page rank centrality for each node in the obese and non-obese networks. Difference is statistically significant (p = 0.01)*

Variation of Network Structure Across the Human Skin Landscape

Previous work has shown differences in microbial communities between anatomical sites. These differences include bacteria, viruses, and fungi, and vary by degree of moisture, oil, and occlusion from the external environment. We hypothesize that like microbial composition and diversity, microbial network structure differs between anatomical sites. We addressed this hypothesis by evaluating the changes in network structure between anatomical sites within the skin virome dataset.

The degree of centrality among nodes was significantly different across anatomical skin sites. The communities at the palm and atecubtal fossa were both highly connected and maintained diverse relationships. The high degree of alpha centrality suggests that the nodes at the sites are more highly connected to other highly connected nodes. The high Shannon entropy suggests that the relationships associated with the nodes are both numerous and their relative abundances evenly distributed. The Pagerank algorithm, which is weighted more heavily toward nodes connected to fewer nodes, revealed that the axilla and toe web had particularly high pageranks. This suggests that there are more nodes within those networks are more highly connected to bacteria without other predators. Dissimilarity analysis revealed that the only sites whose network structures were significantly different where the antecubital fossa and the axilla, the former having a high alpha centrality and the latter having a high page rank.

To evaluate the impact of skin environmental conditions, we assessed the differences in community relationship structure between levels of occlusion and moisture/oil. We found that the unstable sites (intermittently moist and intermittently occluded) maintained more diverse relationships and a higher degree of alpha centrality. As expected, these sites also had particularly low Pagerank values. All of the sites were highly and significantly

dissimilar by occlusion status. Only the intermittently moist site was significantly dissimilar from the moist site, although the other groups were nearly significantly different.

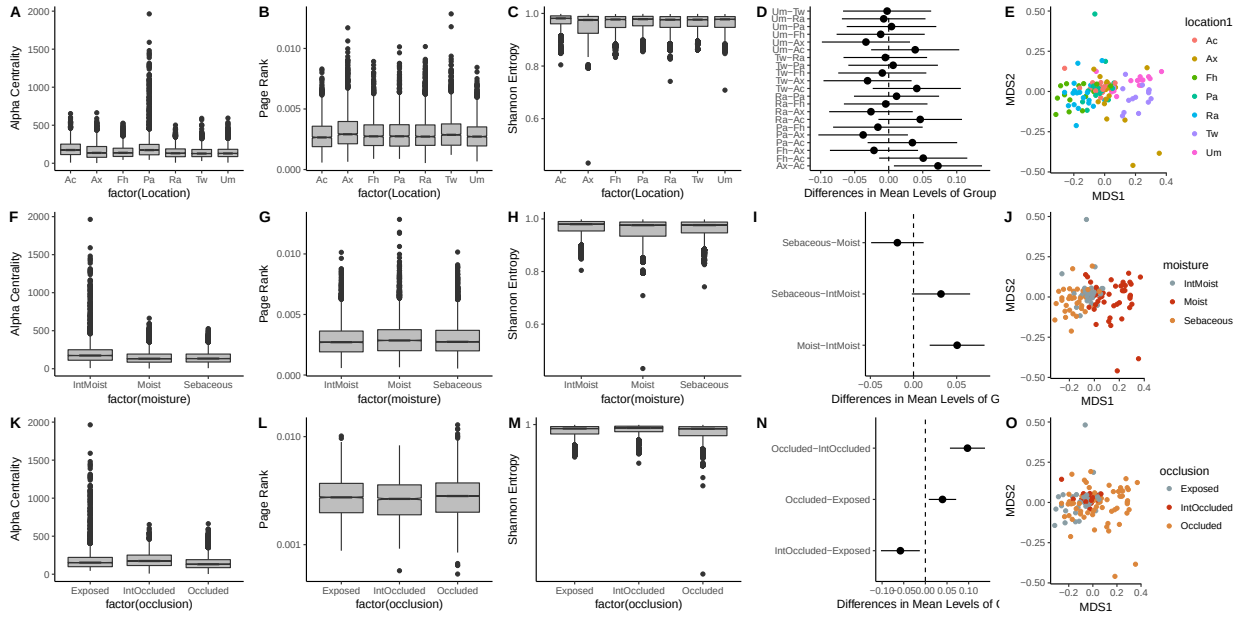

Figure 8: *Network properties of the human skin virome. A) Alpha centrality, B) Page Rank, C) Shannon entropy, and D) significance of Hamming dissimilarities which are E) visualized using NMDS ordination. Same metrics are shown for moisture of environments (F - J) and occlusion (K - O).*

Discussion

Here we build off of previous virome work to show that human virome relationship networks within the microbiome are altered by changing environmental conditions. The degree of node centrality, relationship entropy, and relationship dissimilarity was significantly altered upon changes in diet and anatomical skin site. Networks remained more similar with each individual over time, compared to all other individuals. Such network similarity was not observed within families or between mono-zygotic twin pairs. There may also be an association between network structure and obesity. These changes in network structure have important implications of the functionality of the ecological systems at large.

Previous work has shown that alterations in diet can impact virome diversity, but the functional implications of such a finding were unclear. We found that in addition to impacting virome diversity, alterations in diet had a significant impact on the structure of relationships the virome had with its host bacterial community. Phages were more highly central by the alpha centrality and Pagerank metrics in people being fed low fat diets, indicating a high degree of virome connectivity with bacterial hosts. The high alpha centrality indicates that low fat diets resulted in more phage infections of more bacteria that themselves were highly linked to other phages. This indicates a stronger potential for transmission of genetic elements between phages and/or bacterial hosts. Because these viruses are highly connected, this also indicates the low fat diets are associated with more resilient communities that can more efficiently compensate for removal of microbes. The high Pagerank suggests an even distribution of connectedness such that these phages are not only most highly connected to well connected bacteria, but also well connected to bacteria with minimally linked bacteria. While node centrality is enriched in low fat diets, the Shannon entropy (the degree of richness and evenness of communities) is unchanged between the different states. Together this indicates not only that the virome is different in guts subjected to different diets, but that low fat diets are more resilient to microbial community disruption and more capable of genomic material transmission.

In addition to reduced phage centrality in subjects with high fat diets, we also observed low centrality of phages within an obese human subject compared to two healthy individuals. While this is purely observational, it does correlate with our diet findings and suggests a reasonable possibility that obesity is linked to virome network structure. The association between obesity and the microbiome (e.g. bacterial diversity and community composition) remains a point of debate within the field. We believe this work supports the need for future studies to evaluate the link between the microbiome and obesity in new and creative ways, such as virome network structure. This new approach may provide more informative insights into the link between the microbiome and obesity, as well as disease in general.

We found that just as diet-based alterations to the gut environment influence the interactive potential of microbiome networks, anatomical location was also associated to different network structures of the skin virome. We found that there were significant differences in network structure between the seven skin sites evaluated, as well as differences between the overarching categories of occlusion and moist/oil levels. The palm and atecubtal fossa were enriched for phage nodes that were highly connected to bacteria that were in turn highly connected to other phages. Conversely, the axilla and toe web phages were more highly connected to bacteria that were less connected to other phages. Although they were different by their degree of connectivity, their relationship structures were overall not significantly dissimilar. Because the palms and atecubital fossa sites are more highly connected, we conclude that the virome structures of these sites are more resilient to ecological disturbances due to redundancy of interactions. The axilla and toe web sites are more susceptible to ecological disturbances.

While there were minimal differences in network structure between individual anatomical sites, the sites were completely dissimilar when grouped by occlusion status. There were also differences in relationship structure dissimilarity between sites categorized by their degree of moisture. The sites that were intermittently moist and intermittently occluded were more highly connected (alpha centrality) than the other associated sites. This link between transient environments and more robust network structures suggests that sites in a great state of flux are more highly connected. This supports previous work that suggests phages are needed to revive disrupted microbial communities.

In addition to evaluating the anatomical differences of the virome, we also evaluated the interpersonal dissimilarity of the virome relationship structures. Here we asked whether virome relationship structures are more similar in the same person over time compared to other people at the same time. We found that the gut viromes of individuals are more similar to each other over the course of approximately eight to ten days compared to other people, even when their diets were the same. We also found that skin virome relationship structures are more similar to each other over one month compared to other individuals at the same time. Interestingly, these similarities were strongly associated with individuals and were not preserved in families. Families of twins and their mothers were not more similar to each other compared to other families, and twins alone were not more similar to each other compared to other twins.

Together these findings present an initial look into the network structures of the human virome relationships with their host bacteria. We not only present the human virome as a network of its relationships with host bacteria, but also show that the structure of these relationships is dependent on the environmental conditions associated with the community, including diet, disease (obesity), and skin site.

This work is a first step on a long road toward understanding the virome through its relationships. Thus there are certainly caveats and many future directions associated with this work that are beyond the scope of this initial manuscript. While our model for classifying bacteria and phage relationships outperforms existing models, we recognize that, like most classification models, a lot of effort can be made to improve the model even further. This will include the use of new and creative metrics, as well as improved training and validation datasets. This is an area that we are excited to continue developing in future work.

These findings and methodologies are particularly exciting because they represent a new way of understanding the human microbiome. We hope this work will represent an initial step toward understanding the human microbiota as complex interacting communities instead of the reductionist approach often utilized today. Not only have we provided new insights into the biology of the human virome, but we do so while presenting sophisticated approaches toward studying the virome in parallel with the bacterial host community.

Materials & Methods

Data Availability

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

https://github.com/SchlossLab/Hannigan-2016-ConjunctisViribus.

Data Acquisition & Quality Control

Raw sequencing data and associated metadata was acquired from the NCBI sequence read archive. Supplementary metadata was acquired from the 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 `fastq-dump` within the NCBI SRA Toolkit (version). Sequences were quality trimmed using the Fastx toolkit to exclude bases with quality scores below 33 and shorter than 75 bp. Paired end reads were filtered to exclude and sequences missing their corresponding pair using the `get_trimmed_pairs.py` available in the source code.

Contig Assembly

Contigs were assembled using the Megahit assembly program (version). 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. This allowed us to detect lowly abundant or poorly assembled contigs within all samples. Sample sequences were aligned to all phage or bacteria contigs using the Bowtie2 global aligner (version). 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, due to insufficient sequencing depth. In order to reduce redundancy and artificially inflated genomic richness within our dataset, it was important to bin contigs based on their likelihood to be of the same phylogenetic groups. 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).

We clustered contigs using the CONCOCT program (version). Because of our large dataset and barriers 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.

Open Reading Frame Prediction

Open reading frames (ORFs) were identified using the Prodigal program (version) with the meta mode parameter and default settings.

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[21–26]. Associated reference genomes were downloaded from the European Bioinformatics Institute. The model was created based on the four metrics listed below.

The four scores were used as parameters in a random forest model for classify bacteria and bacteriophage pairs as either having infectious interactions or not. The classification model was built using the Caret R package. 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

CRISPRs were identified from bacterial genomes using the PilerCR program (version). 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. 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

Phages may share genes with their bacterial hosts, providing us with evidence of phage-host infectious pairs. We identified shared genes between bacterial and phage genomes by assessing amino acid similarity between the genes using the Diamond protein alignment algorithm. The mean alignment bitscores for each genome pair was recorded for use in our classification model.

Protein - Protein Interactions

The final method we used for predicting infectious interactions between bacteria and phages was by detecting 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 are predicted to interact using the Pfam interaction information within the Interact database (version). The mean bitscores of the matches between each pair were recorded for use in our classification model.

Virome 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 was saved as a graph database using Neo4j software. This network was used for downstream community analysis.

Network Diameter

The weighted diameters of the networks were calculated using the `diameter` function in the igraph database. Thus the network diameter was defined as the length of the longest weighted geodesic within that graph. Briefly, the weighted geodesic is defined as the path between nodes with the shortest sum of edge weights.

Centrality Analysis

Alpha centrality[27] and PageRank[28] were calculated using the associated functions within the igraph R package. Briefly, a vector of node alpha centrality values (`x`) was calculated as:

$$x = \left(I - \alpha A^T\right)^{-1} e$$

Where `I` is the associated identity matrix, `A` is the associated adjacency matrix, `e` is a the vector of external node importance values, and `alpa` is the constant alpha centrality parameter. Default paramters were used, thus providing an `alpha` and `e` value of 1.

Network Shannon Entropy

Shannon diversity was calculated according to Nathan Eagle *et al* and modified in concept to fit our environment[29]. Shannon entropy S for each node i with edges j was defined as:

$$S\left(i\right) = -\sum_{j=1}^{k} p_{ij} \log\left(p_{ij}\right)$$

Where k is the total number of edges associated with node i. We defined pij as:

$$p_{ij} = \quad \frac{V_{ij}}{\sum_{j=1}^{k} V_{ij}}$$

Where Vig is the difference between node weights between nodes i and j as:

$$V_{ij} = \quad \log_{10}\left(ij\right)$$

This was implemented using the igraph package available on CRAN.

Network Relationship Dissimilarity

Beta diversity was calculated using the Hamming distance implemented in R using functionality from the igraph package.

Membership between two network edge sets is defined as M such that:

$$M = \left[c = \left\|A \notin B\right\|, \quad b = \left\|B \notin A\right\|, \quad c = \left\|A \bigcap B\right\|\right]$$

The Hamming distance matrix between graphs A and B was also utilized and defined as the number of addition/deletion events required to to turn the edge set of A into B, normalized for the total number of edges within the system. This is defined as:

$$H(M) = \frac{c}{a+b+c}$$

Acknowledgments

We thank the members of the Schloss lab for their underlying contributions. GDH is supported in part by the University of Michigan Molecular Mechanisms of Microbial Pathogenesis Fellowship.

References

1. Norman, J. M. *et al.* Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160,** 447–460 (2015).

2. Ly, M. *et al.* Altered Oral Viral Ecology in Association with Periodontal Disease. *mBio* **5,** e01133–14–e01133–14 (2014).

3. Monaco, C. L. *et al.* Altered Virome and Bacterial Microbiome in Human Immunodeficiency Virus-Associated Acquired Immunodeficiency Syndrome. *Cell Host and Microbe* **19,** 311–322 (2016).

4. Hannigan, G. D. *et al.* The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *mBio* **6,** e01578–15 (2015).

5. Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* **21,** 1616–1625 (2011).

6. Santiago-Rodriguez, T. M., Ly, M., Bonilla, N. & Pride, D. T. The human urine virome in association with urinary tract infections. *Frontiers in Microbiology* **6,** 14 (2015).

7. Abeles, S. R., Ly, M., Santiago-Rodriguez, T. M. & Pride, D. T. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. *PLOS ONE* **10,** e0134941 (2015).

8. Abeles, S. R. *et al.* Human oral viruses are personal, persistent and gender-consistent. 1–15 (2014).

9. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499,** 219–222 (2013).

10. Haerter, J. O., Mitarai, N. & Sneppen, K. Phage and bacteria support mutual diversity in a narrowing staircase of coexistence. *The ISME Journal* **8,** 2317–2326 (2014).

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

12. Thompson, R. M. *et al.* Food webs: reconciling the structure and function of biodiversity. *Trends in ecology & evolution* **27,** 689–697 (2012).

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

14. Flores, C. O., Valverde, S. & Weitz, J. S. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal* **7,** 520–532 (2013).

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

16. Poisot, T. & Stouffer, D. How ecological networks evolve. *bioRxiv* (2016).

17. Flores, C. O., Meyer, J. R., Valverde, S., Farr, L. & Weitz, J. S. Statistical structure of host-phage interactions. *Proceedings of the National Academy of Sciences of the United States of America* **108,** E288–97 (2011).

18. Jover, L. F., Flores, C. O., Cortez, M. H. & Weitz, J. S. Multiple regimes of robust patterns between network structure and biodiversity. *Scientific Reports* **5,** 17856 (2015).

19. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466,** 334–338 (2010).

20. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).

21. Jensen, E. C. *et al.* Prevalence of broad-host-range lytic bacteriophages of Sphaerotilus natans, Escherichia

coli, and Pseudomonas aeruginosa. *Applied and Environmental Microbiology* **64,** 575–580 (1998).

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

23. Schwarzer, D. *et al.* A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis. *Journal of Virology* **86,** 10384–10398 (2012).

24. Kim, S., Rahman, M., Seol, S. Y., Yoon, S. S. & Kim, J. 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 (2012).

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

26. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews* **40,** 258–272 (2015).

27. Bonacich, P. & Lloyd, P. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks* **23,** 191–201 (2001).

28. Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* **30,** 107–117 (1998).

29. Eagle, N., Macy, M. & Claxton, R. Network Diversity and Economic Development. *Science* **328,** 1029–1031 (2010).