Resolving gut microbiome networks within Chiropterans

Timothy J. Rogers, Laurel R. Yohe, Richard A. White †

Department of Bioinformatics, University of North Carolina, Charlotte, NC 28223; Department of Bioinformatics and Genomics, North Carolina Research Center (NCRC), Kannapolis, NC

† To whom correspondence should be addressed

Abstract (250 words)

Introduction

- Bats are known carriers of human associated pathogens
- The reason bats are bioreactors is not understood
- The diet of bats may contribute to the gut microbiota makeup
- Phage associated with these microbiota can benifit and hinder microbial populations and have an impack on bat immune responses
- Some ideas suggest viral tolerance is linked to
    - Uniqueness of bats and their variation in Diets
        * Diversity
        * Ecological role
        * Pathogenic role
- Methods for characterizing microbiome and virome. As well as methods for linking the two
    - Culture dependant vs independent
        * What has been discovered with these methods
        * How these methods have been applied to bats
            · What has been found
            · What has yet to be discribed
- Goal of this study

Materials and Methods

- Sample location discription
- Sample collection
- Sample processing ### Phase genomics portion
- (AND)

In house analysis of MAGs, vcontigs, and viral-host pairing

vOTU curration

- (AND) Genomad (v x.x.x) was used to varify contigs identified as viral by phase genomics and for taxonomic identification.

- (AND) We also ran all assembly files from phase genome through genomad to identify potential viral sequences that the phase genome pipeline may have missed
- (AND) Quality filtered viral sequences were then clustered into species-level equivalent viral operational taxanomic units (vOTUs) at 95% average nucleatide identity over 85% of the alignment fraction of the shorter sequence using the greedy centroid algorithm (anicalc.py and aniclust.py) from checkV (v x.x.x).

MAG curration

- (AND) MAGs from phase metagenomic were quality filtered using checkM (v x.x.x).
- (AND) Quality filtered MAGs were then dereplicated using dRep (v x.x.x) at 99% ANI with the following settings (–S_algorithm fastANI, -comp 50, –SkipMash)
- (AND) MAGs representatives were taxonomically identified using the Genome Taxonomy Database Toolkit (GTDB-Tk v x.x.x)

vOTU and MAG coverage

- (AND) Quality controled reads were mapped to the vOTUs and representative MAGs using bowtie (v x.x.x) and sam files were converted to indexed and sorted bam files using samtools (v x.x.x).
- (AND) Bam files were then fed into anvi'o (v x.x.x) to calculate the Q2Q3 coverage of vOTUs and MAGs across bat species.
- (AND) Q2Q3 coverages for both vOTUs and MAGs were then normalized by sequence length.
- (AND) For alpha diveristy analyzes, normalized Q2Q3 coverages for both vOTUs and MAGs were rarefied using the rrarify function of the vengan R package.
- (AND) For beta diversity analyzes, normalized Q2Q3 coverages for both vOTUs and MAGs were used to create rarefied Hellinger distance matrices using the vegan (v x.x.x) and labdsv (v x.x.x) R packages.
- (AND) For abundance pattern analyzes, normalized Q2Q3 coverages were coverted into units of GCPM (Genome Copy Per Million Reads) for MAGs, as laid out in Rogers et al. 2022 (citation), and TPM (Transcripts Per Million Reads) for vOTUs.
- (AND) Abundance correlations of vOTUs and MAGs across and within samples were determined using hierarchical clustering of Spearman rank correlation distance matrice in base R (v x.x.x).

Metabolic predictions

- The program metacerberus (v x.x.x) was used for gene anotation of both viral sequences and MAGs using the following data bases: Functional Ontology Assignments for Metagenomes (FOAM), KEGG, CAZy/dbCAN, VOG, pVOG, PHROG, and COG.

AMR and CAZY gene abundance patterns

- (AND) Read coverage and abundance for AMR and CAZY genes were calculated using a custom ihhouse script.

Linking virus to host

- (AND) IPHoP (v x.x.x) was used to link all viral sequences to all medium to high quality MAGs, not just to the representative vOTUs and MAGs.
- (AND) The viral and MAG IDs of those within the viral-host predictions from both IPHoP and Phase genomes were replaced with the corresponding representative vOTU and MAG IDs.

Statistical analysis

- (AND) Differences in MAG and vOTU composition among bat species were analyzed via distance based redundancy analysis (db-RDA) on a quntitative Hellinger distance matirx.
- (AND) The richness and evenness of MAGs and vOTUs were compared using the R package micro-biome for richness (chao1) and the package hillR (v x.x.x) for evenness (Pielou's index).
- (AND) Statistical significants was determined based on 9999 permutations of the data in the vegan R package.

Results

MAGs

- (AND) We recovered 5 Archaeal and 182 Bacterial medium to high quality MAGs with $\geq$ 50 % and $\leq$ 9 % redundancy.
- (AND) Of these 187 MAGs, 106 were $\geq$ 90% complete and $\leq$ 10% redundant, while YY are of high quality as determined by parameters layed out in XXX et al. (** TABLE **).

- (AND) *Bacillota*, *Actinomycetota*, and *Pseudomonadata* were the predominate phylum across the gut microbiome within all 3 of the bat species (**Figure**).

- (AND) No significant differnces were found in the richness (Choa1) or evenness (Pielou's Index) of prokaryotic MAGs across the 3 bat species.

- (AND) db-RDA analysis revealed no significant differnce in the B-diversity of the prokaryotic MAGs across the 3 bat species.
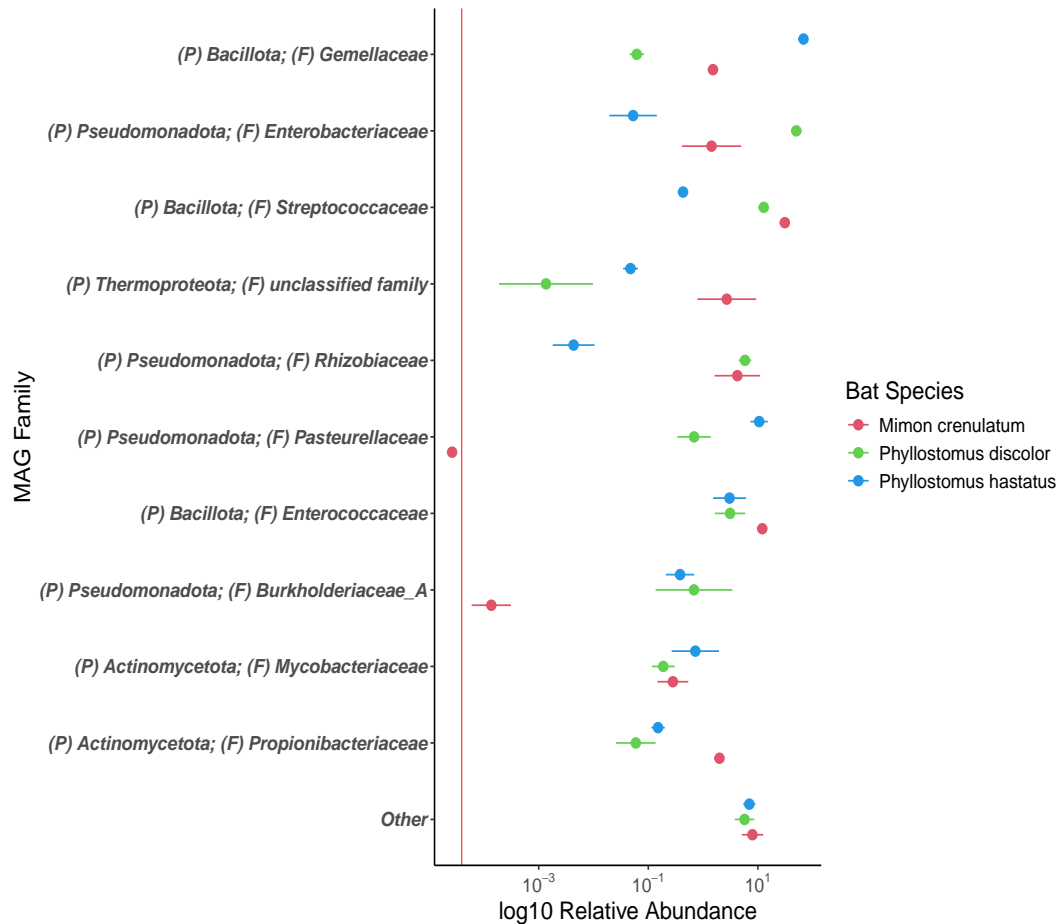

Figure 1: Distribution of the top 10 most abundant prokaryotic families across the gut microbiome of the three bat species

vOTUs

- (AND) A total of 17702 potential viral sequences were ID from the 6 individual assemblies.
- (AND) Clustering at 95% ANI over 85% of the shortest sequence identified 16289 viral operational taxanomic units (vOTUs) that were $\geq$ 1 kbp in length.

- (AND) Further filtering for sequences $\geq$ 2.5 kpb resulted in a final set of 6235 vOTUs.

- (AND) After rarefaction, 5582 vOTUs were retained.

- (AND) CheckV was used to assess the quality of these sequences, revealing that 361 (6%) were $\geq$ 50% complete including 39 complete vOTUs that identified on the bases of direct terminal repeats (DTR), 74 high quality vOTUs that were identified on the bases of AAI (54 vOTUs) and HMM (20 vOTUs), 198 medium quality vOTUs that were idenified on the bases on AAI (136 vOTUs) and HMM (62 vOTUs), and 50 low quality vOTUs that were identified based on AAI (41 vOTUs) and HMM (9 vOTUs). The reset of the vOTUs (5221) were of low quality (3298) or the quality was undetermined (1923).

- (AND) An unclassified order of the class *Caudoviricetes*, families *Retroviridae*, *Adintoviridae*, and *Iridoviridae*, an unclassified family of *Kyanoviridae*, families *Inoviridae* and *Bornaviridae*, an unclassified family of *Herelleviridae*, families *Mimiviridae* and *Parvoviridae* were the most predominate viral taxa across the gut virome with all 3 of the bat species (**Figure**).

- (But) No statistically significant differences were found in the viral richness, envenness, or B-diversity nor did indicator analysis reveal any indicator viral spices for the bat species.

Metabolic capabilities

- (AND)

Virus-host predicitons

- (AND) A total of 1308 viru-host prodictions were made using phase HiC and IPHoP methods.

- (AND) These predictions included 186 MAGs and 953 vOTUs.

- (AND) Of these, 352 were made by only phase, 867 were made by only IPHoP, and 80 were agreed matches between both methods (**Figure**).

- (AND) Of those predictions made by IPHoP, 565 are based on blastn and 391 are based on iPHoP-RF

- (AND) MAGs within the bacteria phyla *Actinomycetota*, *Bacillota*, *Pseudomonadota*, and *Desulfobacterota* had the highest frequency of being targeted by viruses (84, 62, 18, and 13 MAGs respectively).

- (AND) Host families most often targeted within *Actinomycetota* include *Mycobacteriaceae* (20 MAGs) and *Nocardiaceae* (10 MAGs).

- (AND) The host family most often targeted within *Bacillota* is *Enterococcaceae* (15 MAGs).

- (AND) The host family most often targeted within *Pseudomonadota* is *Enterobacteriaceae* (6 MAGs).

- (AND) The host family most often targeted within *Desulfovibrionaceae* is *Desulfobacterota* (9 MAGs).
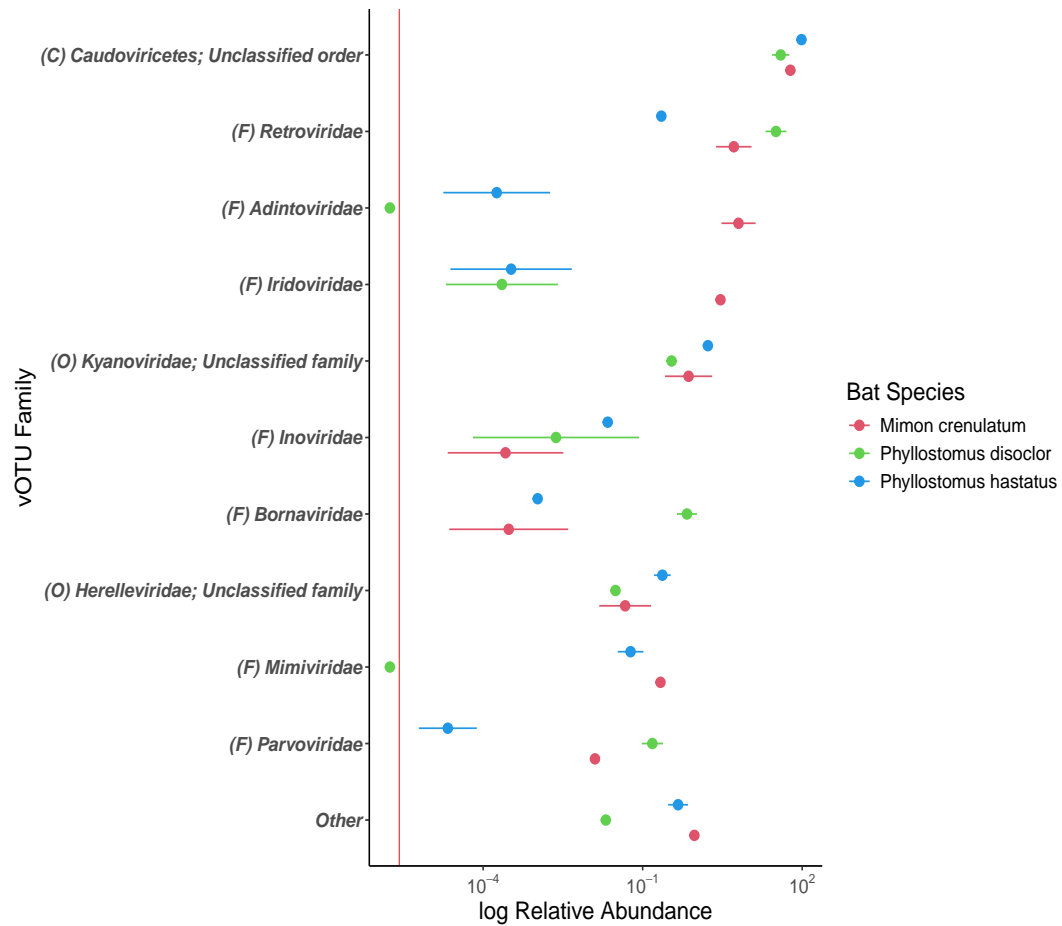
Figure 2: Distribution of the top 10 most abundant viral families across the gut virome of the three bat species
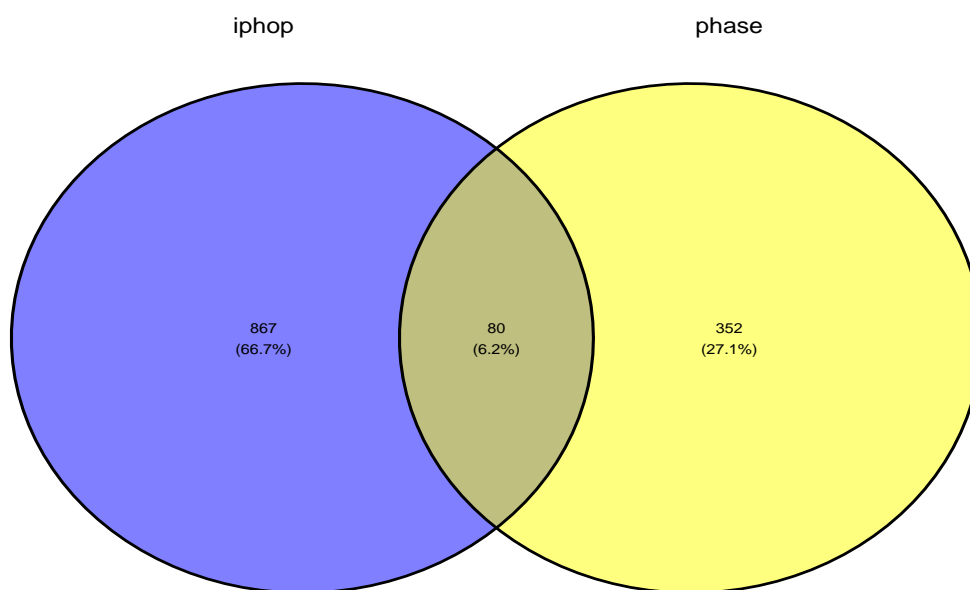
Figure 3: Venn diagram showing the number of virus-host predictions made by each method used
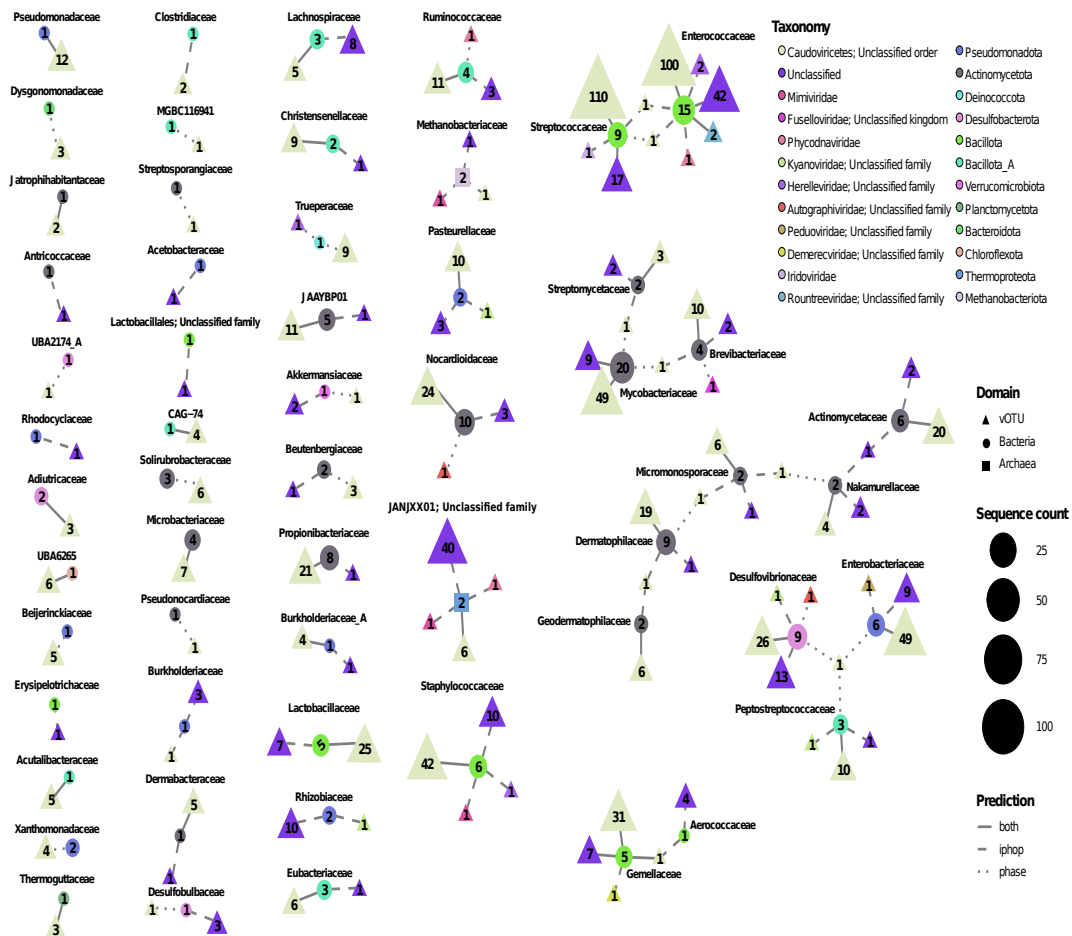
Figure 4: Virus-host network analysis

Descusion

- (AND) The families *Mycobacteriaceae* and *Nocardiaceae* within the bacteria phylum *Actinomycetota* contain known pathogenic members (**CITATION**).
- (AND) The family *Enterococcaceae* within the bacteria phylum *Bacillota* contains known pathogenic members (**CITATION**).
- (AND) The family *Enterobacteriaceae* within the bacteria phylum *Pseudomonadota* contain known pathogenic memebers (**CITATION**).
- (AND) The family *Desulfovibrionaceae* within the bacteria phylum *Desulfobacterota* contain known opportunistic pathogenic members (**CITATION**).

Macro Eukarotic * *Iridoviridae* Bornaviridae _Paroviruses_

Micro Eukarotic **Mimiviridae*

Bacterial phage * an unclassified family of *Kyanoviridae* * *Inoviridae* * an unclassified family of *Herelleviridae*
Conclusion

Acknowledgement