Investigating the Population and Genetic Dynamics of CrAssphage

Sanath Angalakudati, Benjamin H. Good, Oskar Hallatscheck
Department of Physics and Bioengineering, UC Berkeley

Introduction

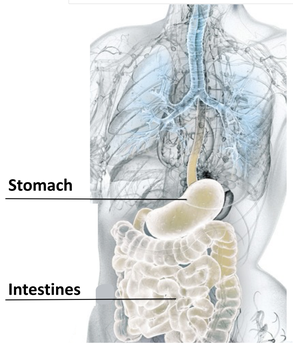

Fig 1. The Human Gut Microbiome The human gut microbiome represents a highly diverse population of microbes that habitat the human digestive tract, predominantly in the stomach and the intestines. Each healthy human gut contains over 100 trillion bacteria, a population that consists of over 1000 different species.

The microbes of the human gut microbiome are essential to human digestive. These microbes break down food into nutrients that are then absorbed by the human body. These bacteria are regulated by the Human Gut Virome, or the population of bacteriophages that habitat the human digestive tract. These bacteriophages ensure no one species of bacteria begins to overpopulate the microbiome, which could lead to various digestive conditions.

In order to maintain a healthy microbial community, the human body highly regulates the conditions in the gut microbiome. Due to the precise conditions of the gut, creating a culture of gut microbes in incredibly difficult. A In order to study the gut microbiome, we use a set of techniques called metagenomics. Metagenomics allows us to sequence the DNA of microbes inhabiting one human's microbiome at one time point. By analyzing these DNA sequences, we generate data on not only microbial populations, but also of gene and SNP data for a specific species.

Fig 2. Metagenomics. A metagenomics study entails analyzing a stool sample from a single subject, which is then filtered for microbial DNA strands. These strands are then spliced into exact 100 base pair fragments. These fragments are then sequenced, producing individual "reads". Reads are then aggregated into a single metagenomic file, and are published in publicly available databases online.

We used metagenomic files to analyze a recently discovered bacteriophage found to exist in the gut microbiome: crAssphage. A recent paper by Dutilh et. al (2014) finds that crAssphage, when present in a microbiome, can compose up to 76% of the reads in its metagenomic file. crAssphage's abundance in some microbiomes indicates a potentially crucial role in the function and evolution of the gut.

Materials and Methods

We analyzed a metagonomic dataset published by Voight et.al (2015), which contains rich sampling of seven subjects taken over two years. We identified two subjects with both sufficient data and evidence of existing crAssphage, "Bugkiller" and "Peacemaker."

We analyzed Bugkiller's and Peacemaker's metagenomic data using the Metagemonic Intraspecies Diversity Analysis System (MIDAS). MIDAS is an analytical tool that aligns the reads from the metagenomic file to a database of reference genomes. MIDAS alignment offers analysis on the population, gene, and SNP levels. MIDAS allows us to quantify the abundance of various microbiota as well as isolate loci along the reference genome where reads diverged from the reference. Results from MIDAS were analyzed with Microsoft Excel and various Python packages, and visualizations of metagenomic files were done using the Integrative Genomics Viewer (IGV).

Results

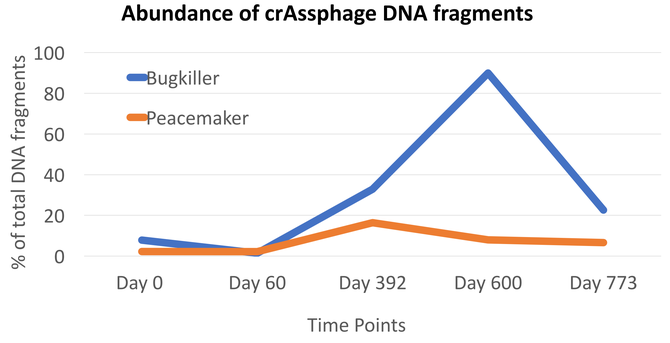

Fig 3. Population Data In both Peacemaker and Bugkiller, crAssphage abundances are highly variable over time. crAssphage reaches a peak abundance of 18% of all reads in peacemaker, while in bugkiller it reaches 88% Both populations crash by Day 773.

Our population analysis reinforces the findings of Dutilh. In Bugkiller, we see that crAssphage abundance is startlingly high. Interestingly, our analysis shows that crAssphage is not always highly abundant in the microbiome, especially evident in regards to Bugkiller. In order to investigate why the crAssphage population fluctuates so much, we analyzed SNP level data from Bugkiller.

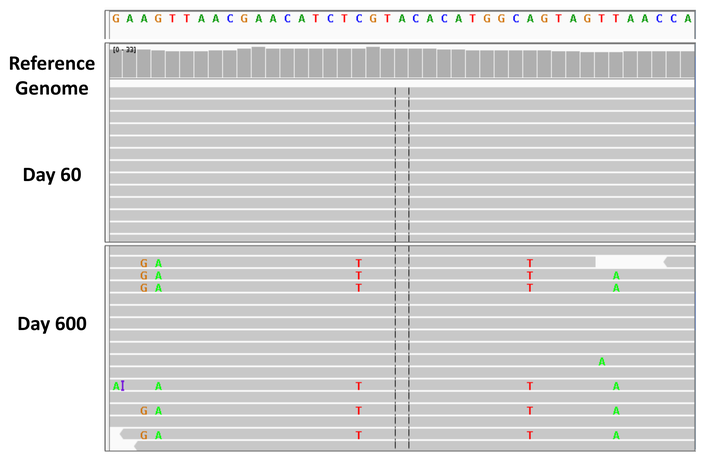

Fig 4. Visualization of Reads This screesnhot from the IGV program highlights a particular locus along the crAssphage genome. At Day 60, all reads found in Bugkiller's metagenomic data at this locus perfectly align to the reference genome.

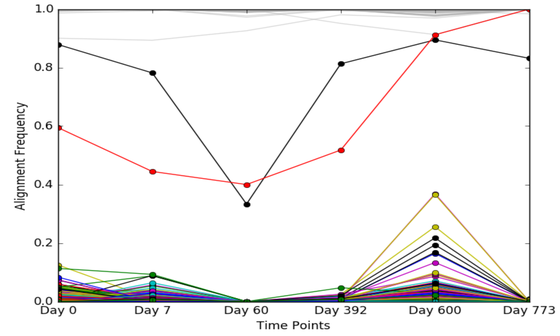

Fig 4. Alignment Frequency We tracked alignment frequency of SNPS to the reference genome. A SNP with an alignment frequency of 1.0 always matches with the reference, while one with 0.5 aligns to the reference half the time. s

Corresponding to the population bloom observed, we see a spike in SNP alignment frequencies at Day 600. Previously, at Day 392, these SNPS aligned with frequencies of less than 5%. At Day 773, these frequencies crashed.

Conclusions

Our findings suggest that while crAssphage can compose a large portion of flora found in the gut microbiome, its population is subject to fluctuation. Currently, our SNP level data is insufficient to fully explain these population blooms. However, the alignment frequency spike observed at Day 600 in Bugkiller's microbiome represents a promising lead. Further studies of time series metagenomic files would further shed light on the population and genetic dynamics of crAssphage.

Citations

Bull, Matthew J., and Nigel T. Plummer. "Part 1: The Human Gut Microbiome in Health and Disease." *Integrative Medicine: A Clinician's Journal* 13.6 (2014): 17–22. Print.

Dutilh, B. E. et al. "Unknown sequences in faecal metagenomes reveal a widely distributed and highly abundant bacteriophage." Nat. Commun. 5:4498 doi: 10.1038/ncomms5498 (2014).

Voigt, Anita Y et al. "Temporal and technical variability of human gut metagemones." BioMed Central 16:73 doi: 10.1186/s13059-015-0639-8

Acknowledgements

I'd like to thank my mentor Benjamin H. Good for being wonderful and providing guidance throughout my summer. Further thanks are owed to the Hallatscheck Lab for being willing lab mates. I'd like to thank Eva Campodonico and the STEM class of 2017 for supporting me throughout my adventure.