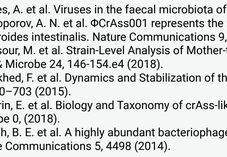
Acquisition, vertical transmission and strain diversity of crAssphage

Benjamin A. Siranosian[1], Fiona B. Tamburini[1], Gavin Sherlock[1], Ami S. Bhatt[1,2]
[1]Department of Genetics, Stanford University, Stanford, CA, USA.
[2]Department of Medicine, Division of Hematology, Stanford University, Stanford, CA, USA.

Mother-infant transmission of bacteria in the microbiome is well described for certain taxa. Infants often acquire their first microbes directly from the primary strain in the mother's microbiome. However, phages display a different pattern of inheritance and remain unique between mothers, infants and twins.

CrAssphage is a double-stranded DNA virus with a ~100 kb circular genome. The phage has been difficult to culture, but recent reports suggest it infects *Bacteroides* species. It's prevalent - found in the gut microbiome of up to 50% of individuals, and abundant - can represent up to 90% of viral sequencing reads in a stool metagenomic dataset. However, crAssphage has not been demonstrated to have an impact on host health or biology.

Given previously demonstrated vertical transmission of bacteria and the high prevalence of crAssphage, we hypothesized that mother-infant transmission could be responsible for crAssphage colonization in infants.



Electron micrograph of a representative crAssphage, from Shkoporov et al. (2018). This phage is a member of the *Podoviridae* and infects *Bacteroides Intestinalis*.

Methods

We examined two public stool shotgun metagenomic datasets:

Yassour et al. (2018)
- 44 families
- Mother sampling at -3, 0, 3 months relative to birth
- Infant sampling at birth, 2 weeks, 1, 2 and 3 months

Bäckhed et al. (2015)
- 98 families
- Mother sampling at birth
- Infant sampling at birth, 4 and 12 months

Metagenomic classification
We built a Kraken2 database containing all viral sequences of genome, chromosome or scaffold quality in GenBank. Samples were considered crAssphage positive if they had >1000 reads (roughly 1x coverage) classified to the crAssphage reference genome.

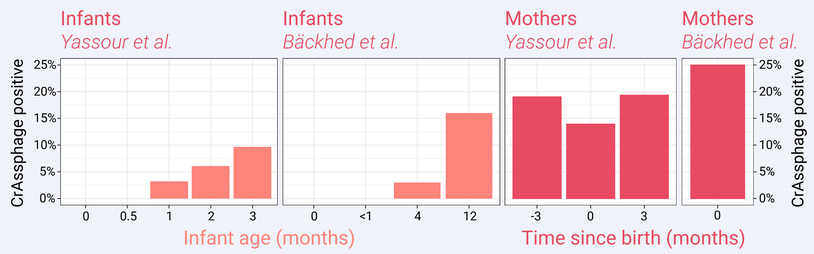
Genome assembly and comparison
Metagenomic assembly was performed with *metaSPAdes*, crAssphage contigs were pulled out by aligning to the reference genome with *BWA*, and pairwise alignments were performed with the *nucmer* suite of tools.

Strain diversity
Variants were called with *Snippy*, and multiallelic sites were classified as SNPs with multiple alleles and ≥5 reads supporting each allele. Variant effects were predicted with *SNPEff*.

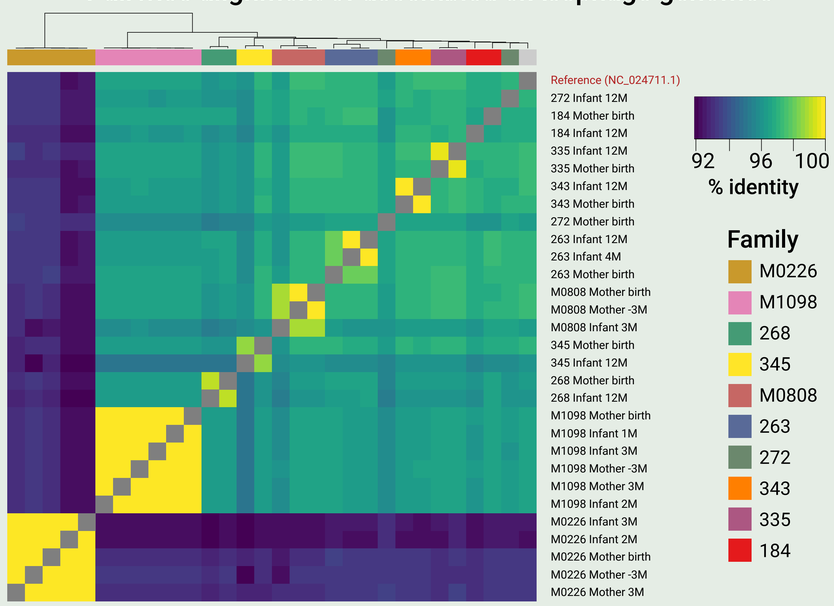
References

1.Reyes, A. et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466, 334–338 (2010).
2.Shkoporov, A. N. et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nature Communications 9, 4781 (2018).
3.Yassour, M. et al. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. Cell Host & Microbe 24, 146-154.e4 (2018).
4.Bäckhed, F. et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host & Microbe 17, 690–703 (2015).
5.Guerin, E. et al. Biology and Taxonomy of crAss-like Bacteriophages, the Most Abundant Virus in the Human Gut. Cell Host & Microbe 0, (2018).
6.Dutilh, B. E. et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nature Communications 5, 4498 (2014).

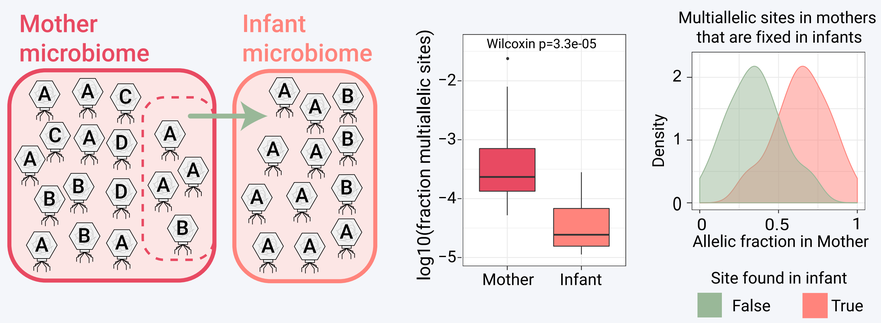
1 CrAssphage is not detected in infant microbiomes at birth, increases in prevalence with age, but doesn't reach the level of adults by 1 year



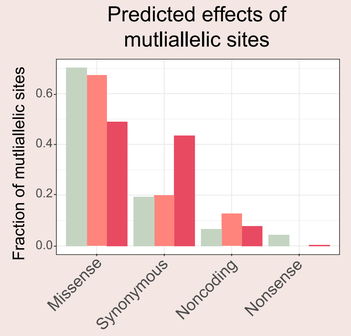
2 Mothers and infants share nearly identical crAssphage genomes in 40% of cases, suggesting vertical transmission



3 Infants have reduced crAssphage strain diversity and typically acquire the mother's dominant strain upon transmission



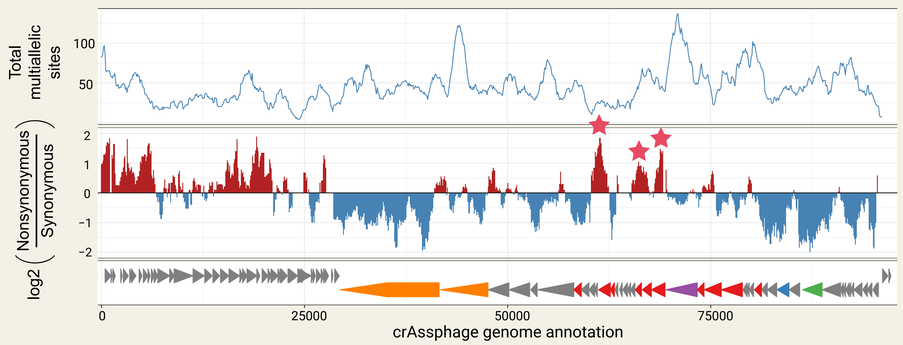
4 Strain diversity is mostly the result of neutral genetic variation, but infants are have more nonsynonymous changes than mothers



We used SNPs with multiple high quality alleles detected to compare strain diversity between samples. A background distribution (null model) is calculated by predicting the effect of every possible genomic change in the crAssphage reference genome.

Background distribution (null model)
Infants
Mothers

5 Strain diversity varies across the genome, and tail fiber genes are enriched for strain diversity with nonsynonymous variants



Does strain variation in tail fiber genes expand the host rage of crAssphage? Is variation suppressed in essential capsid, portal and RNAP genes?

6 These findings extend to crAss-like phages. Vaginally born infants are more likely to have crAss-like phages than those born via C-section

Guerin et al. (2018) proposed 4 subfamilies and 10 clusters for crAss-like phages. The crAssphage described here (also called prototypical crAssphage) falls into the Alpha subfamily, cluster 1. We observe a similar trend of increasing crAss-like phage prevalence with infant age, and a putative transmission rate of 29%. Additionally, presence of a crAss-like phage is strongly statistically associated with vaginal birth as opposed to cesarean section. Is vaginal birth responsible for crAss-like phage transmission, or for seeing microbes that crAss-like phages predate on? Future studies with more balanced cohorts are needed.

	Vaginal birth	*C-section*
crAss positive	54	2
crAss negative	59	19

Fisher's exact test: p=0.0012

Acknowledgements
This work was supported by the Stanford Genetics Department training grant and the Stanford Genomics Training Program (SGTP). This work utilized computing resources provided by the Stanford Genetics Bioinformatics Service Center. We thank the authors of the studies cited here for doing rigorous experiments and making their data publicly available for us to examine.

Get in touch!
bsiranosian.com
bsiranosian@gmail.com
twitter.com/bsiranosian
github.com/bsiranosian
github.com/bhattlab