Schneider-Crease India A (Orcid ID: 0000-0002-2699-5304)
Snyder-Mackler Noah (Orcid ID: 0000-0003-3026-6160)

Author Manuscript

Intestivirid acquisition increases across infancy in a wild primate population

India A. Schneider-Crease[1,2,3*], Isabella L. Moya[1,4], Kenneth L. Chiou[1,4], Alice Baniel[5], Abebaw Azanaw Haile[6], Fanuel Kebede[6], Belayneh Abebe[7], Amy Lu[8], Thore J. Bergman[9,10], Noah Snyder-Mackler[1,2,4], Arvind Varsani[1,4,11,12]

[1]Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, USA
[2]School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA
[3]Department of Anthropology, Wollo University, Dessie, Ethiopia
[4]School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[5]Institut des Sciences de l'Évolution de Montpellier UMR5554, CNRS, IRD, EPHE, Université de Montpellier, Montpellier, France
[6]Ethiopian Wildlife Conservation Authority, Addis Ababa, Ethiopia
[7]African Wildlife Foundation, Debark, Ethiopia
[8]Department of Anthropology, Stony Brook University, Stony Brook, New York, USA
[9]Department of Anthropology, University of Michigan, Ann Arbor, Michigan, USA
[10]Department of Ecology and Evolution, University of Michigan, Ann Arbor, Michigan, USA
[11]The Biodesign Center of Fundamental and Applied Microbiomics, Arizona State University, Tempe, Arizona, USA
[12]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, 7925, South Africa

Corresponding author email address: indiaaugusta@gmail.com

Abstract

Intestivirids (order Crassvirales, family Intestiviridae), viruses that infect Bacteroidales bacteria in the mammalian gastrointestinal tract, have been identified as a highly abundant component of the healthy human virome that may shape patterns of human health and disease through direct action on the microbiome. While double-stranded DNA bacteriophages called crAssphages (*Carjivirus communis*) that infect bacteria in the Bacteroidales order have been identified in humans within the first month of life, the enormous variation in post-parturition infant environments and diets has inhibited a

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.17801

This article is protected by copyright. All rights reserved.

Author Manuscript

robust understanding of the physiological and environmental factors that govern acquisition patterns. We turned to a wild population of graminivorous nonhuman primates (geladas, *Theropithecus gelada*) under long-term study in the Simien Mountains National Park, Ethiopia. We analyzed fecal samples from infants and mothers in this population across the infancy period for richness and presence of crAssphage-like viruses (family *Intestiviridae*). Eight intestivirid genomes were identified based on terminal redundancy representing six unique variants (<98% intergenomic similarity) closely related to the human crAssphage. The prevalence of intestivirids in gelada fecal samples begins to rise at about 10 months of age, peaks in the months surrounding weaning (~18 months), and somewhat decrease but maintain high levels into adulthood. We found a strong association between cumulative rainfall and intestivirid detection, with a higher likelihood accompanying wetter seasons with higher grass availability. In this population, the months prior to weaning have been found to be accompanied by shift in the microbiome characterized by a decrease in glycan degrader Bacteroidales taxa and an increase in fermentative Bacteroidales taxa, and wetter seasons when the vast majority of the gelada diet comprises grasses are associated with an increase in fermentative Bacteroidales taxa. In the context of these microbiome shifts, our results suggest that the intestivirid-bacterial host relationship may interact with major developmental and seasonal dietary shifts in the mammalian host.

Keywords: *Crassvirales, Intestiviridae,* geladas, *Theropithecus gelada, bacteriophage, primate virome*

Sequence data: Sequences have been deposited in GenBank with accession #s PP965493 - PP965500 and the raw reads mapping to these have been deposited under Bioproject # PRJNA795171, BioSamples #s SAMN42063139 - SAMN42063146 and SRA #s SRR29560746 - SRR29560753.

Introduction

The viruses that infect bacteria and archaea– generally referred to as ‘bacteriophages’ or ‘phages’– are by far the most ubiquitous and diverse organisms on earth, with approximately one trillion bacteriophages for each grain of sand on earth [1]. Phages infect every known type of bacteria, shaping the population dynamics and evolution of bacterial communities in nearly every described ecosystem by lysing or lysogenizing their bacterial hosts [1,2]. The ability of phages to reduce the abundance of certain bacteria through predator-prey interactions (i.e., lysis) has produced antagonistic coevolution among certain phages and their microbial hosts [3,4]. Beyond reducing abundance, phage genomes can also integrate into the bacterial genome (i.e., lysogenesis) and affect bacterial virulence [5,6], among other dimensions of bacterial architecture [3,7–9]. These properties make phages widely considered as major drivers

Author Manuscript

in the shape of bacterial communities [10]. Bacterial communities, in turn, have repeatedly been shown to be major players in health and disease in humans and other species [11–20]. If phages can shape bacterial communities, and bacterial communities can shape health, then it stands to reason that phages may contribute to health via actions on the microbiome [21–23]. Indeed, an emerging body of research positions phages as possible therapeutic alternatives in the face of multi-drug resistant bacterial infections [1]. Among phages, the human crAssphage (cross-Assembly phage; family *Intestiviridae*, subfamily *Crudevirinae*, genus *Carjivirus*, species *Carjivirus communi*s, accession # BK010471) has been identified as a highly abundant and prevalent component of the healthy virome that may play a critical role in regulating the microbiome [1,24,25].

CrAssphage and other intestivirids in the four *Crassvirales* families (*Crevaviridae*, *Intestiviridae*, *Steigviridae*, *Suoliviridae*) are predicted to infect bacteria in the order Bacteroidales [25–28], which are broadly implicated in the digestion of milk oligosaccharides [29] and plant polysaccharides [16,30], with evidence building that crAssphages can infect a number of species including *Bacteriodes intestinalis* [28,31,32], *B. xylanisolvens* [33], and *B. thetaiotaomicron* [34]. Intestivirids appear to have coevolved with primates, with evidence of phage-primate codivergence across millions of years (Gogarten et al. 2021, [3]). While strains exhibit temporal and geographic clustering [3] and comparisons between captive and wild primates show plasticity in phage communities (Gogarten et al. 2021), phages show strong conservation of genomic structure across evolutionarily and geographically dispersed primate species [24]. This long evolutionary history suggests that crAssphage and other intestivirids, which have not yet been shown to be associated with metrics of human health, might play at least an indirect role in the maintenance of health throughout the lifespan as regulators of important bacterial communities [24,35]. Early work on the dynamics of human gut-associated crAssphage has suggested that it appears in infants within the first month of life, with some evidence pointing to a role of delivery route (i.e., vaginal delivery versus cesarean section) in shaping observed patterns [35,36]. Infants sometimes share crAssphage strains with their mothers, and exhibit increasing strain diversity as they age [35]. Together, these patterns suggest the possibility of multiple pathways for crAssphage transmission and acquisition; however, identifying these pathways in humans– particularly using metagenomic data– is challenging because of the difficulty of accounting for the enormous variation in infant environments.

Wild primates can serve as powerful models for understanding the natural dynamics of viral and bacterial communities in humans because of their lower within-population and within-species variability in early life environments and experience, paired with their close phylogenetic proximity to and shared evolutionary history with humans. Where

Author Manuscript

human infants can be born vaginally or by cesarean-section and fed formula or breastmilk diets, both of which can influence viral and bacterial communities [23,37,38], wild primates are uniformly born vaginally and feed exclusively on breast milk until transitioning to their species-specific adult diet. In addition, wild primate populations can be easily sampled over time. Because crAssphages and other intestivirids can alter the microbiome through lysing and lysogenizing their bacterial hosts, understanding the factors that influence presence and patterning of these phages in natural systems—particularly during infancy—is a critical step towards a fuller understanding of the ecological and evolutionary dynamics of the phageome and microbiome and their impacts on health and disease. To characterize the patterns and dynamics of natural intestivirid acquisition, we turned to a wild population of geladas (*Theropithecus gelada*), graminivorous primates related to baboons, under long-term study in the Simien Mountains National Park, Ethiopia.

The patterning of microbial communities in geladas parallels that of humans: alpha diversity increases rapidly after birth and evens out at weaning, when infants transition from breastmilk to an exclusively graminivorous diet at a population average of ~18 months [39]. In the 5-7 months prior to weaning, infant gelada microbiomes shift from high abundances of glycan degrader Bacteroidales taxa (e.g., *Bacteroides fragilis*) relevant for milk digestion to high abundances of fermentative taxa (e.g., *Prevotella* sp.) relevant for plant digestion [39]. This likely reflects the weaning process, in which infants begin supplementing breastmilk with plants prior to the complete cessation of breastmilk consumption and exclusive consumption of plants. In adults, annual shifts in the composition of the gut microbiome are associated with seasonal changes in diet; the microbiome is dominated by amylolytic taxa that break starches into sugars (e.g., *Lactobacillus, Streptococcus*) during the dry season when geladas feed primarily on starchy underground plant parts [40], and by fermentative Bacteroidales taxa (e.g., *Prevotella*) that break down plant polysaccharides into simple sugars after the primary action of cellulolytic degraders [24]. While seasonal patterns of the gastrointestinal microbiome that correspond to changes in food availability have been seen in other mammals [41–43], it should be noted that explanations for the relationship between rainfall and microbiome such as changes host physiology or environmental abundance of microbes cannot be ruled out [44,45]. Because intestivirids infect taxa in the Bacteroidales order, intestivirids should recapitulate the seasonal and developmental patterning of the gelada microbiome.

Here, we describe patterns of intestivirid acquisition for a group of crAssphages in the subfamily *Obtuvirinae* (within the *Intestiviridae* family) that are phylogenetically closely related to the human crAssphage (*Carjivirus communis*, family *Intestiviridae*, subfamily *Crudevirinae*). We first assemble intestivirid genomes, assess gene synteny and

Author Manuscript

sequence homologies, and perform phylogenetic analyses to classify these genomes and root them in relation to already classified intestivirid genomes. We then use metagenomic data generated from fecal samples taken from our focal gelada population, including sequential samples taken from infants and mothers, to assess the impact of developmental stage, sex, and cumulative rainfall (as a proxy of dietary shifts in the gut microbiome as described above) on intestivirid richness and presence.

Methods

Sample collection

Fecal samples were collected from geladas in the Simien Mountains National Park (SMNP), Ethiopia, between October 2015 and March 2020. The Simien Mountains Gelada Research Project has collected near-daily behavioral, demographic, and life history data on a population of ~300 geladas in the SMNP since 2006, and geladas in this population are individually identifiable with nearly 100% accuracy by observers. Fecal samples were collected from known individuals immediately following observed defecation (n = 248 from 43 individuals) and stored in RNAlater. Samples were stored in a -20'C freezer in the field until shipment and in a -80'C freezer in the lab until analysis to further prevent DNA degradation.

Samples were collected from adult individuals, defined as post-sexual maturation (n = 52 from 22 individuals ranging in age from ~7.2 years to ~20.5 years), and from young individuals, defined as pre-sexual maturation (n = 196 from 21 individuals ranging in age from 2 days to ~3.8 years). Among young individuals, 120 samples were from 21 weaned individuals while 76 samples were from 19 unweaned individuals (all but two young individuals were sampled pre- and post-weaning). Weaning was determined either based on observation or based on the population average of 18 months at weaning (for the four individuals known to be independent but missing a weaning date). Seventy samples were collected from eleven known mother-offspring pairs within a one-to-three-day window. Climatic data (season, rainfall, and temperature) were recorded continuously throughout the study period. We calculated cumulative rainfall over the 90 days prior to each sample as a proxy for diet composition. Cumulative rainfall over the past 90 days is a good predictor of green grass availability and concordant grass consumption by geladas [40,49], and is also associated with changes in the gelada gut microbiome that include increases in fermentative families in the *Bacteroidales* order that are likely to serve as intestivirid hosts (Baniel et al. 2021).

Author Manuscript

Library preparation and sequencing

Fecal samples were first partially processed using the Qiagen Dneasy PowerLyzer PowerSoil DNA Isolation Kit (cat. No. 12855). We performed lysis, bead-beating, and inhibitor removal following manufacturer instructions up to step 10 (manual v. 07272016), terminating with the successful preparation of fecal crude lysates, which we used as input for downstream library preparation. We prepared metagenomic libraries using a modified version of the Illumina DNA prep kit (cat. No. 20060059) protocol (https://doi.org/10.17504/protocols.io.rm7vzbob4vx1/v1). We used 2 µl of crude lysate as input, which we diluted to 20 µl with nuclease-free water prior to tagmentation. We used a reduced volume of tagmentation mix (5 µl each of BLT and TB1). For PCR, we reduced reaction volumes by adding to samples 10 µl of EPM PCR mix with 2.5 µl of each 10 µM primer. We ran 12 cycles of PCR. The remainder of our protocol followed manufacturer instructions for maximizing library yield and quality. We sequenced libraries on an Illumina Novaseq 6000 S4 flow cell (2 x 150 bp).

Metagenome assembly

Cutadapt [50] was used to trim adapters and low-quality bases from the raw reads and reads mapping to the gelada genome [51] were removed using BWA. MetaSPAdes v 3.15.2 [52] was used to *de novo* assemble the remaining non-host reads into contiguous sequences (“contigs”).

Intestivirid identification and annotation

We selected only circular *de novo* assembled contigs with terminal redundancy (denoting complete circular genomes), which we analyzed using Diamond BLASTx [53] against a RefSeq viral protein database (release 220) to identify viral-like sequences. We further filtered the data for intestivirids based on blast similarity The complete intestivirid genomes were annotated using Cenote-Taker3 [54] and manually checked. The genomes were reoriented ending in the terminase open reading frame. The intestivirid genomes have been deposited in GenBank with accession #s PP965493 – PP965500 and the raw reads mapping to these have been deposited under Bioproject # PRJNA795171, BioSamples #s SAMN42063139 – SAMN42063146 and SRA #s SRR29560746 – SRR29560753.

Representative sequences of classified intestivirid species (n=18) were downloaded from GenBank and analyzed with the ones from this study. Intergenomic similarities between all the genomes of intestivirids were inferred using VIRIDIC [55]. We used VipTree [56] to infer a proteomic phylogenetic tree to show the relationship of various

intestivirids. We used Clinker [57] to generate a genome comparison plot of intestivirid to illustrate gene synteny and sequence homologies.

Per sample viral operational taxonomic unit (vOTU) mapping

Viral reads from each fecal sample were mapped using Bbmap [58] to seven variant sequences (sequences that share <98% intergenomic similarity as determined using VIRIDIC [55]. We assigned presence or absence of variants per sample based on reads mapped to genomes with at least 10% coverage across the genome. We chose to use presence/absence in our downstream analyses rather than relative abundance because of the methodological constraints of working with metagenomic data.

Statistical analyses

We used linear mixed models (LMMs) and generalized linear mixed models (GLMMs) to assess the drivers of variation of intestivirid richness per sample and the presence of any intestivirid in a sample [37]. We first modeled predictors of intestivirid richness in each sample across our entire dataset (n = 248 samples) using an LMM, with richness (i.e., number of intestivirid strains) as the dependent variable and developmental stage (unweaned, weaned, adult), sex (female, male), and scaled cumulative 90-day rainfall as the independent variables. We included logged number of reads per sample (i.e., sequencing depth) and individual ID as random effects. We then modeled predictors of presence/absence (0/1) of any intestivirid in each sample across our entire dataset using a GLMM fitted with a binomial distribution (n = 248 samples), with developmental stage (unweaned, weaned, adult), sex (female, male), and scaled cumulative 90-day rainfall as the independent variables and logged number of reads and individual ID included as random effects. We determined that time from collection to analysis did not affect results or improve model fit and thus did not include this variable in our final models. Finally, we tested whether mother-offspring pairs shared intestivirids with Fisher's exact tests. We matched samples taken from infants to samples taken from their mothers within a 30-day window. We performed a Fisher's exact test for each intestivirid to assess whether samples were more likely to share an intestivirid (0/1 indicating absence or presence of each intestivirid in both samples) if the samples were mother-offspring pairs. All models were run using the lme4 package in R (v. 2022.12.0+353).

Author Manuscript

Author Manuscript

Results

Genomes of intestivirids

In the *de novo* assemblies from 247 fecal sample sequencing libraries, we identified eight complete genomes of intestivirids from eight samples (two genomes from two adults; six genomes from four infants; Table 1). These eight genomes range in size from 85045 to 96711 nts and GC content of 27.3 to 31.3%. Three of the genomes (PP965497-PP965499) share >98% intergenomic similarity whereas the other five share <91.7% intergenomic similarity with all intestivirids including those from this study. The intestivirid genomes with accession #s PP965498 and PP965499 share 99.2% similarity. For the purpose of this study and more clear read mapping analyses, we used a 98% intergenomic similarity to define vOTUs. Thus, in this study we identify seven vOTUs (Table 1) within the eight assembled genomes.

Based on the 95% intergenomic similarity threshold of species classification of bacteriophages, the eight intestivirid genomes identified here represent six species. We thus named the eight intestivirids from this study geladintestivirus 1 through 6 (Table 1), with the name geladintestivirus derived from **gelada** **intestivirus**. Geladintestivirus 5 (PP965495) and 6 (PP965496) share 91.7% intergenomic similarity and geladintestivirus 3 (PP965493) and 4 (PP965494) share 87.0% similarity (Figure 1). All the intestivirids identified in this study are distantly related to those already classified except geladintestivirus 5 (PP965495) and 6 (PP965496) which share ~75% similarity with uncultured phage cr271_1 (MT774394) from porcine feces.

Table 1: Summary of the eight intestivirid genomes identified with GenBank and SRA accession numbers. These eight genomes contain seven vOTUs and represent six species (under “virus name”).

Virus name	Isolate	GenBank accession	vOTU	Genome length (nts)	%GC	Collection date	Latitude / Longitude	SRA
Geladintestivirus 1	Int_RNL_2016_0117_DIX	PP965497	1	89398	27.40%	4/11/2016	13.2027 N 37.8876 E	SRR29560751
Geladintestivirus 1	Int_RNL_2018_0945_COW	PP965498	2	89153	27.30%	7/3/2018	13.2027 N 37.8876 E	SRR29560752
Geladintestivirus 1	Int_RNL_2017_0546_COW	PP965499	2	87261	27.30%	12/2/2017	13.2027 N 37.8876 E	SRR29560753
Geladintestivirus 2	Int_RNL_2018_0288_CRY	PP965500	3	85045	27.90%	3/3/2018	13.2027 N 37.8876 E	SRR29560748
Geladintestivirus 3	Int_RNL_2017_0019_DDA	PP965493	4	96711	31.20%	6/2/2017	13.2027 N 37.8876 E	SRR29560749
Geladintestivirus 4	Int_RNL_2017_0055_MCB	PP965494	5	94950	31.30%	6/9/2017	13.2027 N 37.8876 E	SRR29560750
Geladintestivirus 5	Int_RNL_2018_1252_VOL	PP965495	6	90043	29.40%	9/12/2018	13.2027 N 37.8876 E	SRR29560746
Geladintestivirus 6	Int_RNL_2018_1178_PEE	PP965496	7	89831	29.60%	8/14/2018	13.2027 N 37.8876 E	SRR29560747

Author Manuscript

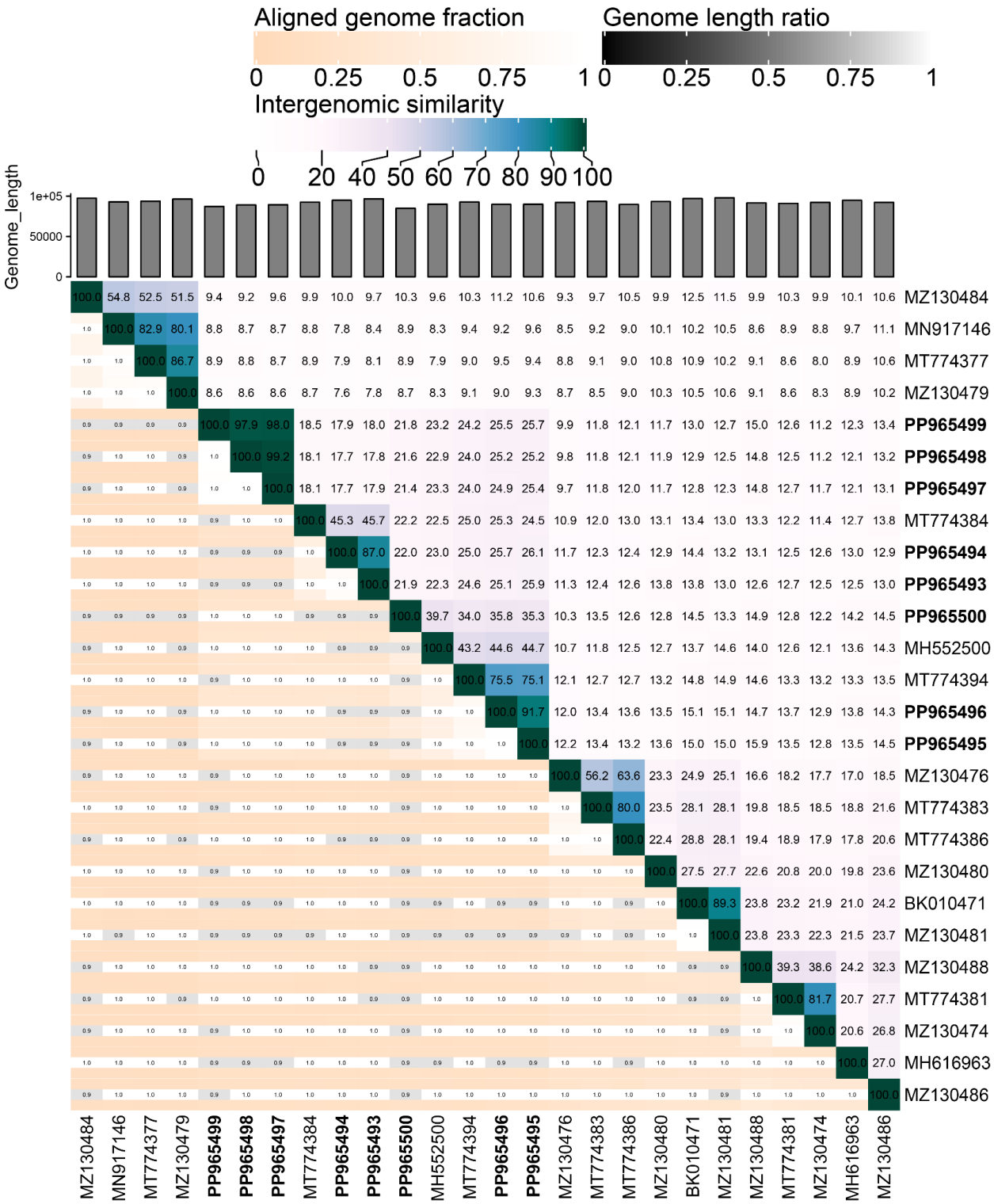


Figure 1: Intergenomic similarities of classified intestivirids together with the eight identified in this study determined using VIRIDIC [55]. The GenBank accession numbers of the intestivirid genomes from this study are in bold font.

We used Clinker [57] to analyze the annotated genomes of geladintestivirus 1-6 and three of the most closely related intestivirids, uncultured phage cr8_1 (species *Fohxhuevirus gastrointestinalis*; MT774384) from human feces, crAssphage sp. Ctcc615 (species *Wotdevirus murinus*; MH552500) from mouse tissue and uncultured phage cr271_1 (species *Hacihdavirus animalis*; MT774394) from porcine feces (Figure 2). We note that there is significant gene synteny across all these genomes, especially in the core genes.

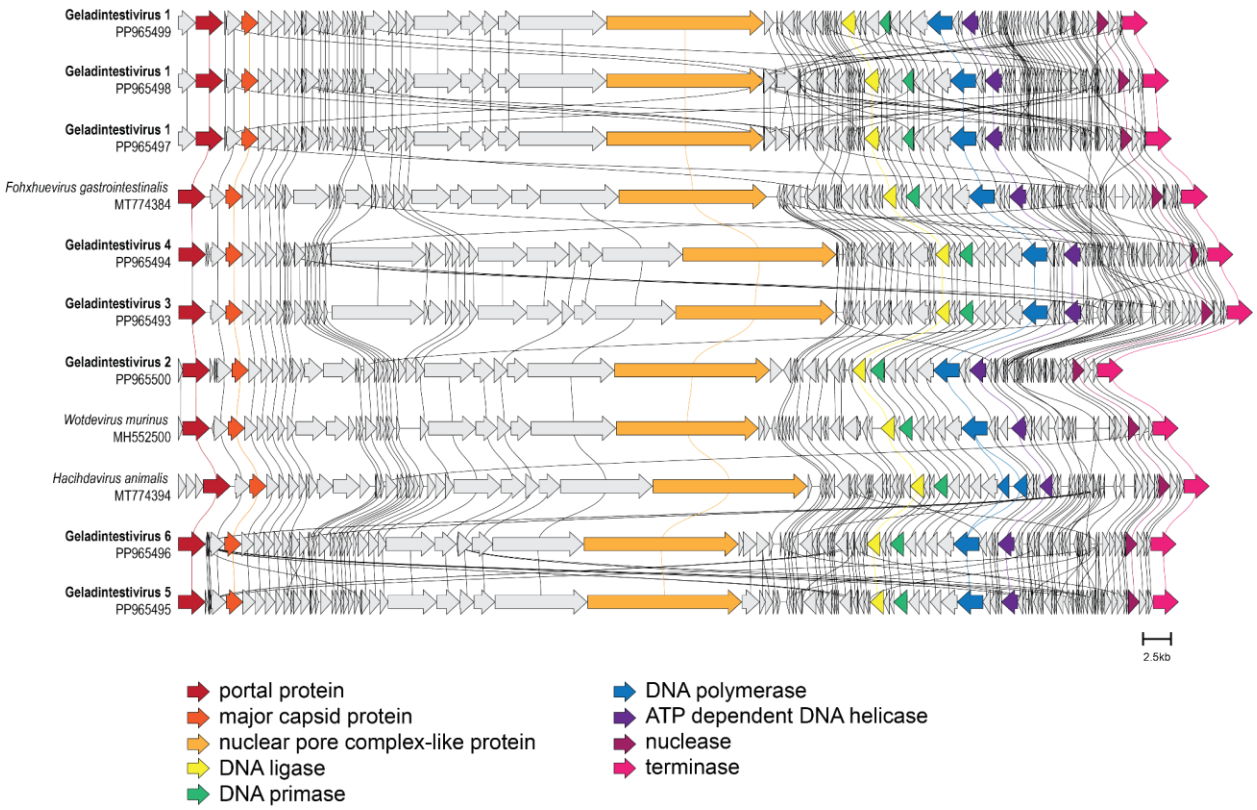


Figure 2: Genome organization of the intestivirids in the subfamily *Obtuvirinae*. Some of the core coding regions are highlighted. Representatives of the classified members of the subfamily *Obtuvirinae* are shown with their species names and accession numbers. Geladintestivirus 1-6 are highlighted in bold font.

Author Manuscript

ViPTree proteomic phylogeny analyses of the geladintestivirus 1 – 6 shows that they are part of the subfamily *Obtuvirinae* in the family *Intestiviridae* (Figure 3). Furthermore, the proteomic phylogeny coupled with intergenomic similarities provide strong support for the classification of the eight intestivirids from this study into six new species in the subfamily *Obtuvirinae*.

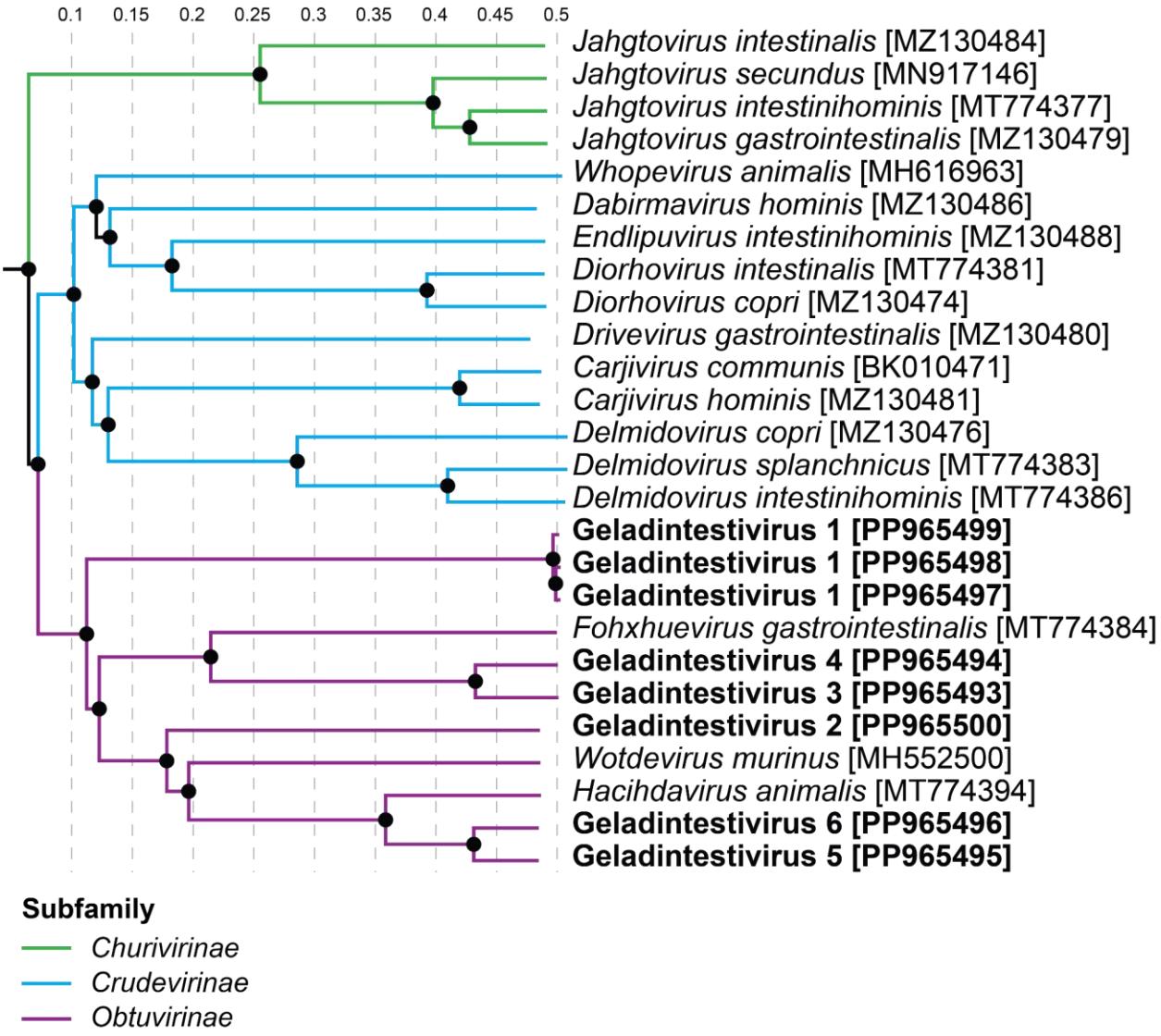


Figure 3: ViPTree proteomic phylogeny of classified intestivirids together with the eight identified in this study. Representatives of the classified intestivirids are shown with their species names and accession numbers. Geladintestivirus 1 – 6 are highlighted in bold font.

Author Manuscript

Author Manuscript

Intestivirid distribution

We identified at least one intestivirid sequence (geladintestivirus 1 – 6) in 169 of 248 fecal samples (68.2%), of which 131 were from young individuals (66.8% of 196 samples from young individuals) and 38 were from adults (73.1% of 38 samples from adults). A single geladintestivirus 1 – 6 sequence was found in 66 samples (26.6%), two sequences were found in 44 samples (17.7%), three sequences were found in 30 samples (12.1%), four sequences were found in 18 samples (7.3%), five sequences were found in nine samples (3.6%), and six sequences were found in two samples (0.8%). Among all individuals with at least one geladintestivirus 1 – 6 sequences, 95 were females (66 young individuals, 29 adults) and 74 were males (65 young individuals, 9 adults). Within young individuals, at least one geladintestivirus was identified in 40 out of 76 samples from unweaned individuals (52.6%; Figure 4) and in 91 out of 120 samples from weaned individuals (75.8%; Figure 4). Geladintestivirus 6 was the most prevalent in our dataset (identified in 75 samples), followed by geladintestivirus 2 (67 samples), geladintestivirus 4 (66 samples), geladintestivirus 3 (65 samples), geladintestivirus 5 (54 samples), and geladintestivirus 1 (46 samples). Geladintestiviruses 4 and 6 were the most prevalent among adults (24.5% prevalence for both), and geladintestiviruses 6 and 2 were the most prevalent among young individuals (29.1%, 28.1%, respectively).

Author Manuscript

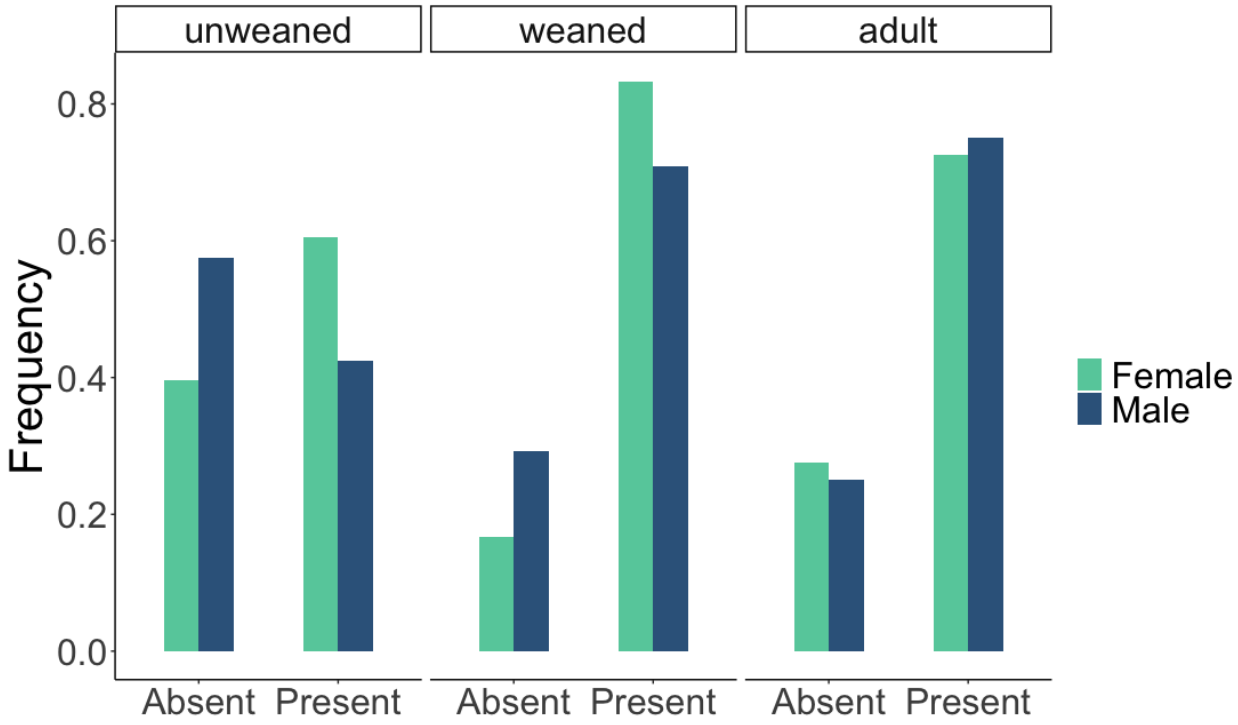


Figure 4: Frequencies of presence/absence of any of the geladintestiviruses (1 – 6) within unweaned infants, weaned infants, and adults. Frequencies are presented for both females (n =131) and males (n = 117).

Intestivirid incidence in infants increases over time until weaning

Intestivirids (geladintestivirus 1 - 6) first appeared in infant fecal samples beginning at 0.4 months of age, and the incidence of positive samples increased over time for non-adult individuals. The highest concentration of geladintestivirus-positive fecal samples was in infants around 18 months of age (Figure 5), which is the average age at weaning for this population [39]. Following this peak, incidence somewhat declined and maintained similar levels throughout the rest of juvenility and adulthood. Weaned and adult individuals had significantly higher intestivirid richness than unweaned individuals in the LMM (weaned estimate = 0.76, $p < 0.001$; adult estimate = 0.65, $p = 0.005$; Table S1). Relative to unweaned individuals, both weaned individuals and adults had significantly higher likelihoods of being identified with any intestivirid in the GLMM (weaned OR = 2.84, $p = 0.002$; adult OR = 2.76, $p = 0.016$; Table S2). In the analyses of individual geladintestivirus presence, weaned individuals had a higher likelihood of being identified with geladintestivirus 3 (OR = 3.31, $p = 0.004$), geladintestivirus 4 (OR = 2.93, $p = 0.010$), geladintestivirus 5 (OR = 2.74, $p = 0.021$), and geladintestivirus 6 (OR = 2.39, $p = 0.028$). Adults had a higher likelihood of being identified with geladintestivirus 4 (OR = 4.07, $p = 0.003$) and geladintestivirus 6 (OR = 3.51, $p = 0.009$)

relative to unweaned individuals. Mother-offspring pairs were more likely to share a single geladintestivirus than non-pairs (geladintestivirus 3; OR = 1.5, $p < 0.001$, Figure 6, Table S4), and no more likely than non-pairs to share the remaining five (Table S4).

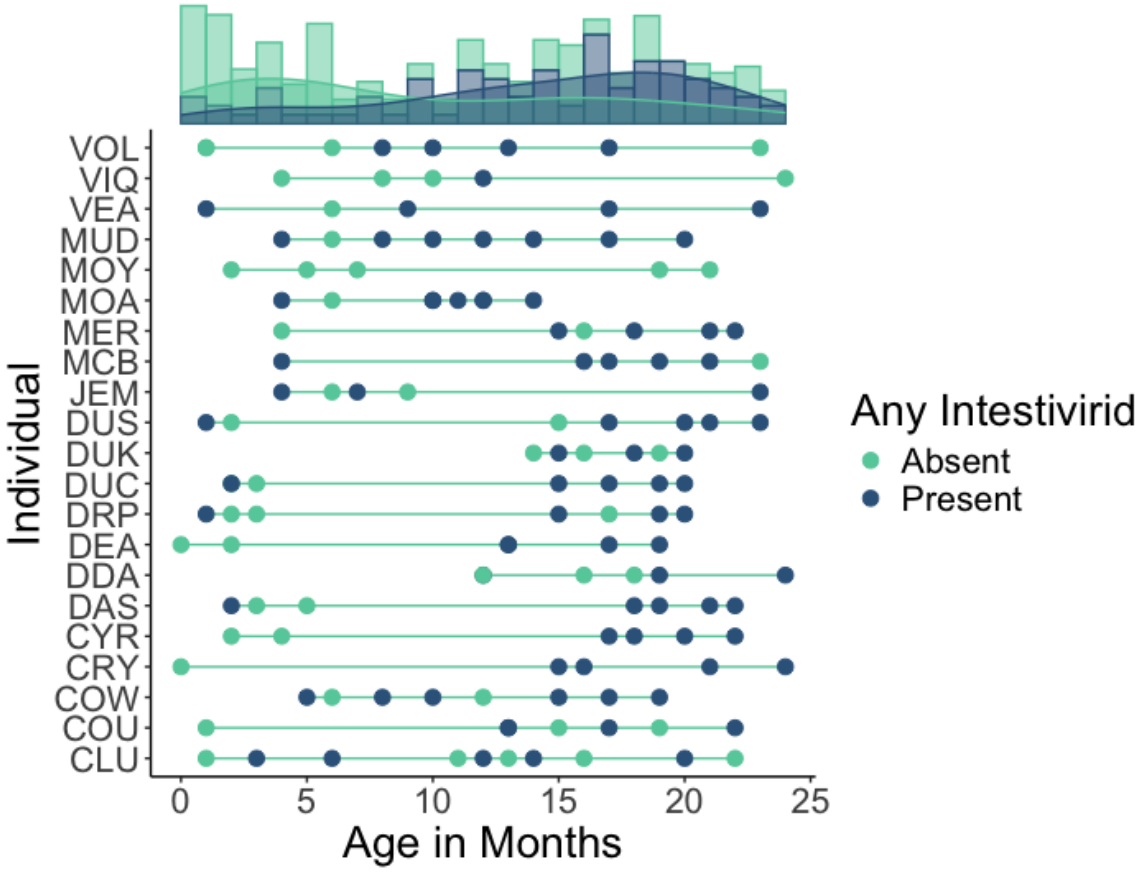


Figure 5: Geladintestivirus incidence increases over infancy and peaks around the population average age at weaning. Horizontal lines follow fecal sampling (dots) for each young individual (names on y-axis) over time (age in months on y-axis). Light blue dots indicate fecal samples without geladintestivirus; dark blue lines indicate fecal samples with geladintestivirus. The histograms represent the density of geladintestivirus-absent (aquamarine) and geladintestivirus-present (dark blue) samples over time, with a peak in geladintestivirus-present samples at ~18 months (the average age at weaning for this population).

Author Manuscript

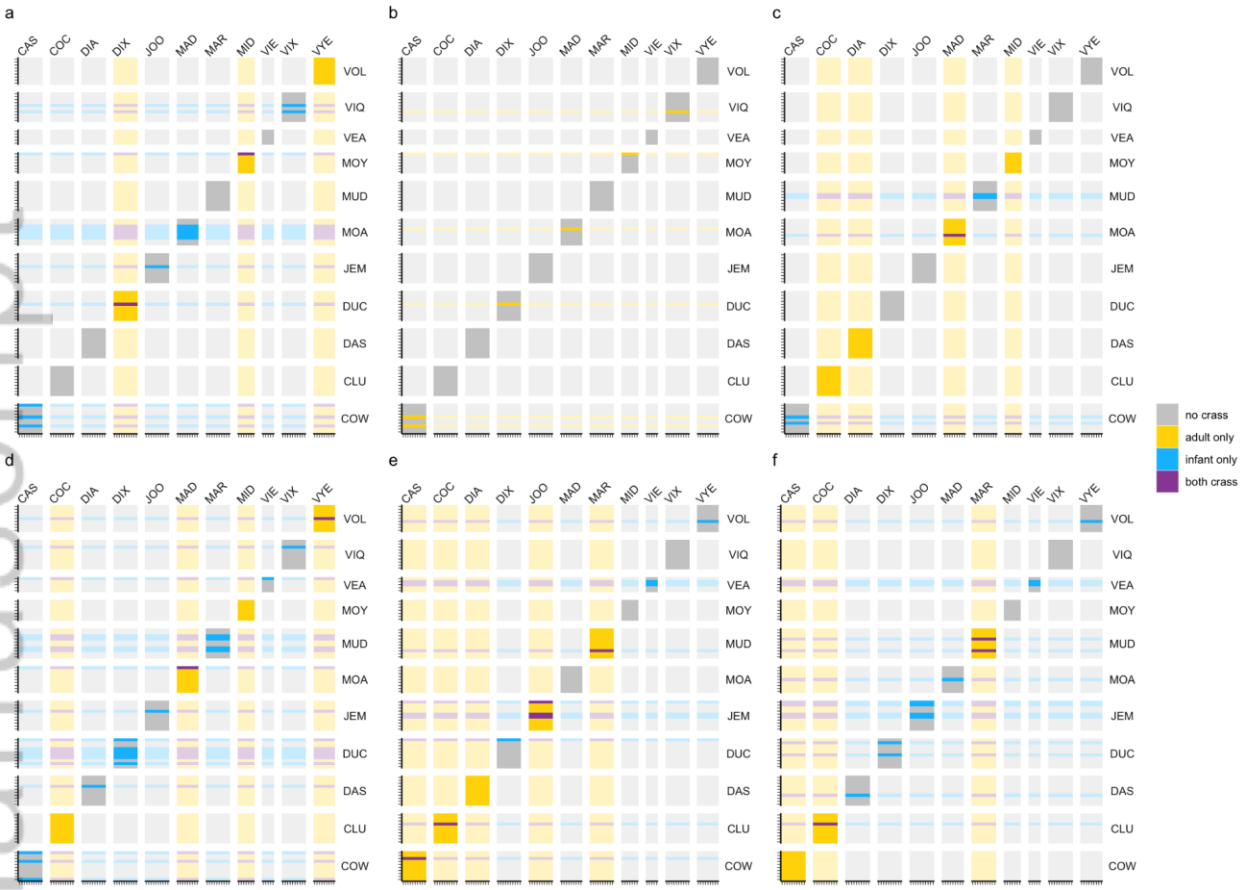


Figure 6: Geladintestiviruses are generally not shared by mother-offspring pairs, with the exception of geladaintestivirus 5 (panel e). Each panel depicts the presence/absence of each geladintestivirus (a-f representing 1-6, respectively) in all possible pairs of adult females (x-axis) and infants (y-axis) that were sampled within six months of each other. The diagonal represents a true mother-offspring pair. Lines are gray if neither adult female nor infant had that geladintestivirus during that window, yellow if only the adult female had geladintestivirus during that window, blue if only the infant had geladintestivirus during that window, and purple if both had geladintestivirus during that window. Purple lines in the boxes along the diagonal thus indicate a mother-offspring pair with a shared geladintestivirus.

Rainfall predicts Intestivirid richness and presence

Cumulative rainfall was associated with increased likelihood of higher geladintestivirus richness (estimate = 0.36, $p < 0.001$, Table S1), any geladintestivirus presence (OR = 1.37, $p < 0.047$, Figure 7, Table S2), and the presence of four geladintestiviruses (geladintestivirus 3, OR = 1.52, $p = 0.008$; geladintestivirus 4, OR = 1.44, $p = 0.019$; geladintestivirus 5, OR = 1.62, $p = 0.004$; geladintestivirus 6, OR = 1.8, $p < 0.001$, Table S3).

Author Manuscript

Author Manuscript

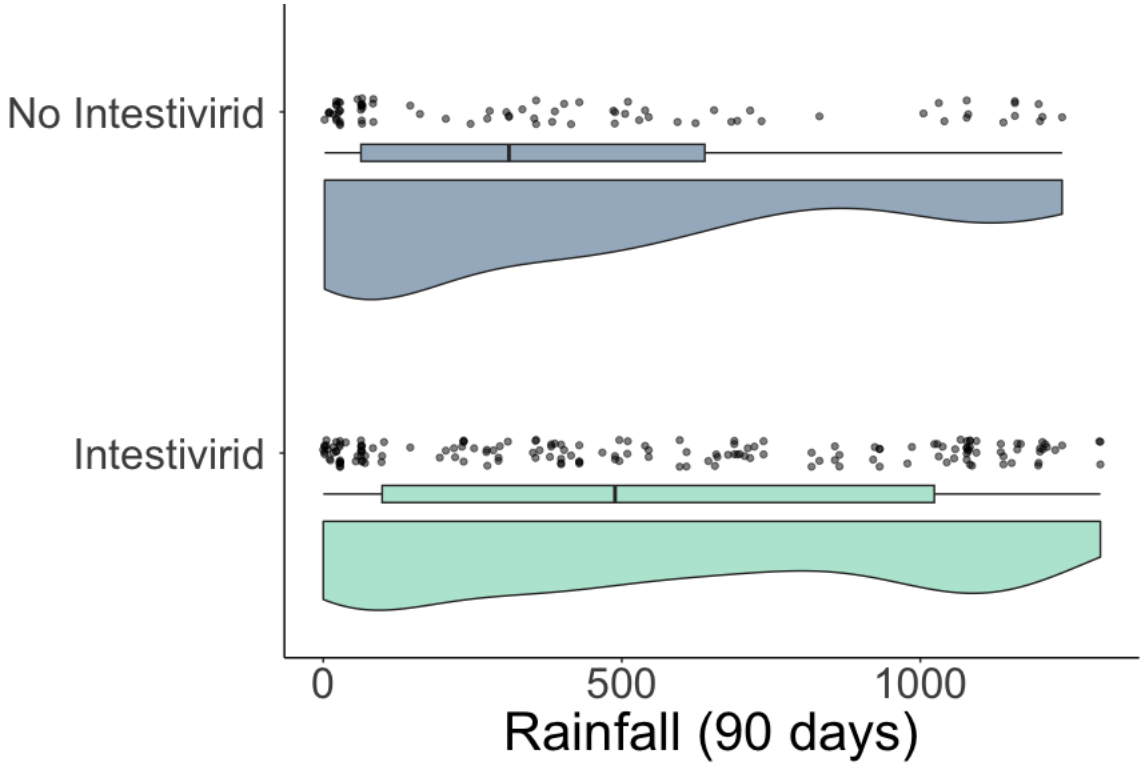


Figure 7: Higher cumulative rainfall (in mm) over the 90 days prior to sampling was associated with an increased likelihood of intestivirid presence in gelada samples.

Discussion

We found that detection of geladintestiviruses in this wild primate population peaked in the months surrounding the average age at weaning and was strongly associated with cumulative rainfall over the 90 days preceding sample collection. Weaning in geladas reflects a transition from breast milk to a plant-based diet and is accompanied by shifts in the gut microbiome characterized by increases in fermentative Bacteroidales taxa [23], which also increase in the adult gelada microbiome during the wet season when geladas shift to a grass-dominant diet. Paired with the lack of detection of shared geladintestiviruses between mother and offspring pairs, this suggests that intestivirid acquisition is not tied to maternally mediated exposure such as vertical transmission through breastfeeding. Rather, our results suggest that intestivirid acquisition follows developmental, dietary, and seasonal patterns of its Bacteroidales hosts in the meta-host gut microbiome.

Author Manuscript

The gradual acquisition of intestivirids throughout infancy in geladas recapitulates the pattern observed in humans, in which crAssphage is not detected at birth but becomes increasingly prevalent over the first year of life [35,36]. The role of vertical transmission in crAssphage acquisition in human remains unclear, with one study finding no difference in crAssphage incidence between infants delivered vaginally and those delivered via cesarean section [36] and another finding a higher incidence of crAssphage in infants delivered vaginally, but only when considering multiple studies [35]. In wild geladas, which are all delivered vaginally, the concentration of intestivirid detection around the average age of weaning suggests that vertical transmission around parturition may not contribute significantly to acquisition in this setting. Weaning in this species reflects an enormous shift in diet, moving from breast milk to grasses, roots, and tubers, and in social behavior, moving from maternal dependence to independence and greater social interactions with non-maternal individuals [39,59]. Furthermore, where some studies have found nearly identical intestivirid sequences in a subset of mother-offspring pairs [35], we found that mother-offspring pairs were only more likely to have one geladintestivirus than non-pairs. We did not explicitly assess the role of vertical transmission, and the lack of matching mother-offspring geladintestiviruses could also be due to detection challenges leading to false negatives. However, when paired with the timing of infection, our results suggest that gelada infants may largely be acquiring intestivirids from an external social or ecological source around weaning rather than directly from their mothers.

Indeed, the microbiome composition of this species shifts notably during the period surrounding weaning. Bacteroidales taxa related to lactose fermentation (e.g., *Bacteroides fragilis*) are highly abundant early in life, and drop significantly in the months leading up to weaning (~10-15 months of age) as Bacteroidales taxa related to plant fermentation (e.g., *Prevotella*) rise [39]. Our data show that intestivirids begin rising in incidence at around the age window, peaking around the average age at weaning in this population (~18 months). This may suggest that intestivirids infect *Bacteroides fragilis* or other amylolytic taxa and play a regulatory role in the microbiome by driving down the abundance of those taxa. Alternatively, and perhaps more likely, intestivirids may infect one or more fermentative taxa, and their increased incidence reflects increasing bacterial host availability during the transition to plant-based diets during infancy.

We found that cumulative rainfall was a strong predictor of multiple metrics of intestivirid infection. Cumulative rainfall over 90 days predicts the availability of green grasses, and is thus not only an indicator of seasonal shifts overall but also of seasonal shifts in the gelada diet since geladas preferentially consume green grasses when available [40]. To digest the grasses that comprise up to 93% of the gelada diet in the rainy season, primary fermenters break down plant cell walls into soluble polysaccharides that are further processed by secondary fermenters—many in the Bacteroidales order (e.g., *Prevotella*)—into simple sugars [24]. These taxa become less abundant during the dry season, when geladas feed on starchy underground plant parts in the absence of grasses and the gut microbiome is dominated by amylolytic taxa that break starches into sugars (e.g., *Lactobacillus, Streptococcus*) [60]. The higher overall probability of

Author Manuscript

intestivirid detection and higher intestivirid richness associated with greater cumulative rainfall thus suggests that intestivirid hosts in this population include one or more fermentative taxa and that intestivirid patterning follows bacterial host composition, which is shaped by dietary changes that are both developmental (as above) and environmental. Alternate explanations for this relationship would be the impact of seasonal changes that co-occur with changes in rainfall, such as changes in gelada behavior other than diet that shape their gastrointestinal microbiome communities and changes in variables such as temperature and water distribution that change the environmental abundances of bacteria including intestivirid hosts. These alternate explanations cannot be ruled out with our dataset, although changes in the gastrointestinal microbiome that are associated with rainfall and corresponding changes in food availability have also been observed in other wild primates [41,42] in addition to geladas [60] and in wild rodents with seasonally impacted diets [43].

Overall, our results demonstrate the importance of extrinsic factors in shaping the patterning of intestivirids in this population through compositional shifts in the gut microbiome. While vertical transmission may still be at play as a driver of intestivirid acquisition, the later timing of acquisition in this wild primate indicates that breastfeeding–as opposed to parturition– would be the pathway of such vertical transmission. In addition, we were only able to address questions based on intestivirid presence rather than relative abundance due to the constraints of our dataset—further work here should use qPCR to quantify abundance and ask more fine-grained questions. Additional work should characterize the exact pathways of either vertical or horizontal intestivirid transmission, and comparisons with other less seasonal primate species and cross-culturally in humans will permit an assessment of whether the results from geladas can be extended to understand crAssphage acquisition in humans. In the context of human crAssphages, our results suggest that the impact of timing and nature of solid food and social introductions during infancy might thus contribute to the observed wide variation in intestivirid (i.e., crAssphage) appearance and strain-sharing across infancy [35,36] and position wild primate populations as models for understanding the natural dynamics of intestivirids. Further work with viral communities in this population will work on elaborating the transmission, acquisition, and patterning of other viruses and bacteriophages and contextualizing intestivirid transmission within these broader viral and bacterial dynamics.

Ethical approval: All data collection was conducted with permission from the Ethiopian Wildlife Conservation Authority and was approved by the Institutional Animal Care and Use Committees at Arizona State University (20-1754 R) and Stony Brook University (773805_TR002).

Conflicts of interest: The authors declare that there are no conflicts of interest.

Acknowledgements: We dedicate this paper to the memory of Ambaye Fenta, and thank him, Eshete Jejaw, and Setegne Girmay for their dedication to nearly two decades of research with the Simien Mountains Gelada Research Project. We thank the

Author Manuscript

Ethiopian Wildlife Conservation Authority and all park officials who have permitted, facilitated, and aided our research.

Funding Information: Long-term gelada research was supported by the National Science Foundation (IOS-1255974, BCS-0715179, BCS-1723237, BCS-1723228), the National Institutes of Health (5R00AG051764-04), the Leakey Foundation, and the National Geographic Society (Gr. #8989-11, Gr. #8100-06). ISC acknowledges funding from the Arizona State University Center for Evolution and Medicine.

Data Access: Sequences have been deposited in GenBank with accession #s PP965493 - PP965500 and the raw reads mapping to these have been deposited under Bioproject # PRJNA795171, BioSamples #s SAMN42063139 - SAMN42063146 and SRA #s SRR29560746 - SRR29560753.

1. Keen EC. A century of phage research: bacteriophages and the shaping of modern biology. Bioessays. 2015;37: 6–9.

2. Wagner PL, Waldor MK. Bacteriophage control of bacterial virulence. Infect Immun. 2002;70: 3985–3993.

3. De Sordi L, Lourenço M, Debarbieux L. The battle within: Interactions of bacteriophages and bacteria in the gastrointestinal tract. Cell Host Microbe. 2019;25: 210–218.

4. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, et al. Phage-host coevolution in natural populations. Nat Microbiol. 2022;7: 1075–1086.

5. Bille E, Meyer J, Jamet A, Euphrasie D, Barnier J-P, Brissac T, et al. A virulence-associated filamentous bacteriophage of Neisseria meningitidis increases host-cell colonisation. PLoS Pathog. 2017;13: e1006495.

6. Schroven K, Aertsen A, Lavigne R. Bacteriophages as drivers of bacterial virulence and their potential for biotechnological exploitation. FEMS Microbiol Rev. 2020. doi:10.1093/femsre/fuaa041

7. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev. 2004;68: 560–602, table of contents.

8. Harrison E, Brockhurst MA. Ecological and evolutionary benefits of temperate phage: What does or doesn't kill you makes you stronger. Bioessays. 2017;39. doi:10.1002/bies.201700112

9. Pinto G, Sampaio M, Dias O, Almeida C, Azeredo J, Oliveira H. Insights into the genome architecture and evolution of Shiga toxin encoding bacteriophages of Escherichia coli. BMC Genomics. 2021;22: 366.

Author Manuscript

10. Chevallereau A, Pons BJ, van Houte S, Westra ER. Interactions between bacterial and phage communities in natural environments. Nat Rev Microbiol. 2022;20: 49–62.

11. Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. N Engl J Med. 2016;375: 2369–2379.

12. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. Nat Rev Microbiol. 2021;19: 55–71.

13. Dupont HL, Jiang Z-D, Dupont AW, Utay NS. THE INTESTINAL MICROBIOME IN HUMAN HEALTH AND DISEASE. Trans Am Clin Climatol Assoc. 2020;131: 178–197.

14. Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. BMJ. 2017;356: j831.

15. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Curr Opin Gastroenterol. 2015;31: 69–75.

16. Zafar H, Saier MH Jr. Gut Bacteroides species in health and disease. Gut Microbes. 2021;13: 1–20.

17. Belizário JE, Napolitano M. Human microbiomes and their roles in dysbiosis, common diseases, and novel therapeutic approaches. Front Microbiol. 2015;6: 1050.

18. Bull MJ, Plummer NT. Part 1: The human gut microbiome in health and disease. Integr Med (Encinitas). 2014;13: 17–22.

19. Mezouar S, Chantran Y, Michel J, Fabre A, Dubus J-C, Leone M, et al. Microbiome and the immune system: From a healthy steady-state to allergy associated disruption. Human Microbiome Journal. 2018;10: 11–20.

20. Ehrlich SD. The human gut microbiome impacts health and disease. C R Biol. 2016;339: 319–323.

21. Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. Healthy human gut phageome. Proc Natl Acad Sci U S A. 2016;113: 10400–10405.

22. Federici S, Nobs SP, Elinav E. Phages and their potential to modulate the microbiome and immunity. Cell Mol Immunol. 2021;18: 889–904.

23. Cao Z, Sugimura N, Burgermeister E, Ebert MP, Zuo T, Lan P. The gut virome: A new microbiome component in health and disease. EBioMedicine. 2022;81: 104113.

24. Edwards RA, Vega AA, Norman HM, Ohaeri M, Levi K, Dinsdale EA, et al. Global

Author Manuscript

phylogeography and ancient evolution of the widespread human gut virus crAssphage. Nat Microbiol. 2019;4: 1727–1736.

25. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun. 2014;5: 4498.

26. Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, et al. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. Nat Microbiol. 2018;3: 38–46.

27. Smith L, Goldobina E, Govi B, Shkoporov AN. Bacteriophages of the order *Crassvirales*: what do we currently know about this keystone component of the human gut virome? Biomolecules. 2023;13: 584.

28. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. Nat Commun. 2018;9: 4781.

29. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, et al. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe. 2011;10: 507–514.

30. Salyers AA, Vercellotti JR, West SE, Wilkins TD. Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. Appl Environ Microbiol. 1977;33: 319–322.

31. Shkoporov AN, Khokhlova EV, Stephens N, Hueston C, Seymour S, Hryckowian AJ, et al. Long-term persistence of crAss-like phage crAss001 is associated with phase variation in Bacteroides intestinalis. BMC Biol. 2021;19: 163.

32. Li Y, Gordon E, Shean RC, Idle A, Deng X, Greninger AL, et al. CrAssphage and its bacterial host in cat feces. Sci Rep. 2021;11: 815.

33. Guerin E, Shkoporov AN, Stockdale SR, Comas JC, Khokhlova EV, Clooney AG, et al. Isolation and characterisation of ΦcrAss002, a crAss-like phage from the human gut that infects Bacteroides xylanisolvens. Microbiome. 2021;9: 89.

34. Hryckowian AJ, Merrill BD, Porter NT, Van Treuren W, Nelson EJ, Garlena RA, et al. Bacteroides thetaiotaomicron-Infecting Bacteriophage Isolates Inform Sequence-Based Host Range Predictions. Cell Host Microbe. 2020;28: 371-379.e5.

35. Siranosian BA, Tamburini FB, Sherlock G, Bhatt AS. Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages. Nat Commun. 2020;11: 280.

36. McCann A, Ryan FJ, Stockdale SR, Dalmasso M, Blake T, Ryan CA, et al. Viromes of one year old infants reveal the impact of birth mode on microbiome diversity.

Author Manuscript

PeerJ. 2018;6: e4694.

37. Howe A, Ringus DL, Williams RJ, Choo Z-N, Greenwald SM, Owens SM, et al. Divergent responses of viral and bacterial communities in the gut microbiome to dietary disturbances in mice. ISME J. 2016;10: 1217–1227.

38. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. 2011;21: 1616–1625.

39. Baniel A, Petrullo L, Mercer A, Reitsema L, Sams S, Beehner JC, et al. Maternal effects on early-life gut microbiome maturation in a wild nonhuman primate. bioRxiv. 2021. p. 2021.11.06.467515. doi:10.1101/2021.11.06.467515

40. Jarvey JC, Low BS, Pappano DJ, Bergman TJ. Graminivory and fallback foods: annual diet profile of geladas (Theropithecus gelada) living in the Simien Mountains National Park, Ethiopia. International Journal of. 2018. Available: https://link.springer.com/article/10.1007/s10764-018-0018-x

41. Hicks AL, Lee KJ, Couto-Rodriguez M, Patel J, Sinha R, Guo C, et al. Gut microbiomes of wild great apes fluctuate seasonally in response to diet. Nat Commun. 2018;9: 1786.

42. Orkin JD, Campos FA, Myers MS, Cheves Hernandez SE, Guadamuz A, Melin AD. Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. ISME J. 2019;13: 183–196.

43. Brown BRP, Khasoha LM, Lokeny P, Jakopak RP, Reed CG, Dyck M, et al. Spatiotemporal variation in the gut microbiomes of co-occurring wild rodent species. Ecosphere. 2024;15. doi:10.1002/ecs2.4854

44. Litchman E. Climate change effects on the human gut microbiome: complex mechanisms and global inequities. Lancet Planet Health. 2025;9: e134–e144.

45. Williams CE, Williams CL, Logan ML. Climate change is not just global warming: Multidimensional impacts on animal gut microbiota. Microb Biotechnol. 2023;16: 1736–1744.

46. Santiago-Rodriguez TM, Tremblay RL, Toledo-Hernandez C, Gonzalez-Nieves JE, Ryu H, Santo Domingo JW, et al. Microbial quality of tropical inland waters and effects of rainfall events. Appl Environ Microbiol. 2012;78: 5160–5169.

47. Sato K, Hara-Chikuma M, Yasui M, Inoue J, Kim Y-G. Sufficient water intake maintains the gut microbiota and immune homeostasis and promotes pathogen elimination. iScience. 2024;27: 109903.

48. Rocca JD, Simonin M, Blaszczak JR, Ernakovich JG, Gibbons SM, Midani FS, et al. The Microbiome Stress Project: Toward a global meta-analysis of environmental

Author Manuscript

stressors and their effects on microbial communities. Front Microbiol. 2018;9: 3272.

49. Jarvey JC, Low BS, Haile AA, Chiou KL, Snyder-Mackler N, Lu A, et al. Aggression rates increase around seasonally exploited resources in a primarily grass-eating primate. Behav Ecol. 2023. doi:10.1093/beheco/arad079

50. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10–12.

51. Chiou KL, Janiak MC, Schneider-Crease IA, Sen S, Ayele F, Chuma IS, et al. Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. Nat Ecol Evol. 2022;6: 630–643.

52. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27: 824–834.

53. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18: 366–368.

54. Tisza MJ, Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. Cenote-Taker 2 democratizes virus discovery and sequence annotation. Virus Evol. 2021;7: veaa100.

55. Moraru C, Varsani A, Kropinski AM. VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. Viruses. 2020;12: 1268.

56. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33: 2379–2380.

57. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics. 2021;37: 2473–2475.

58. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States); 2014 Mar. Report No.: LBNL-7065E. Available: https://www.osti.gov/biblio/1241166

59. Barale CL, Rubenstein DI, Beehner JC. Juvenile social relationships reflect adult patterns of behavior in wild geladas. Am J Primatol. 2015;77: 1086–1096.

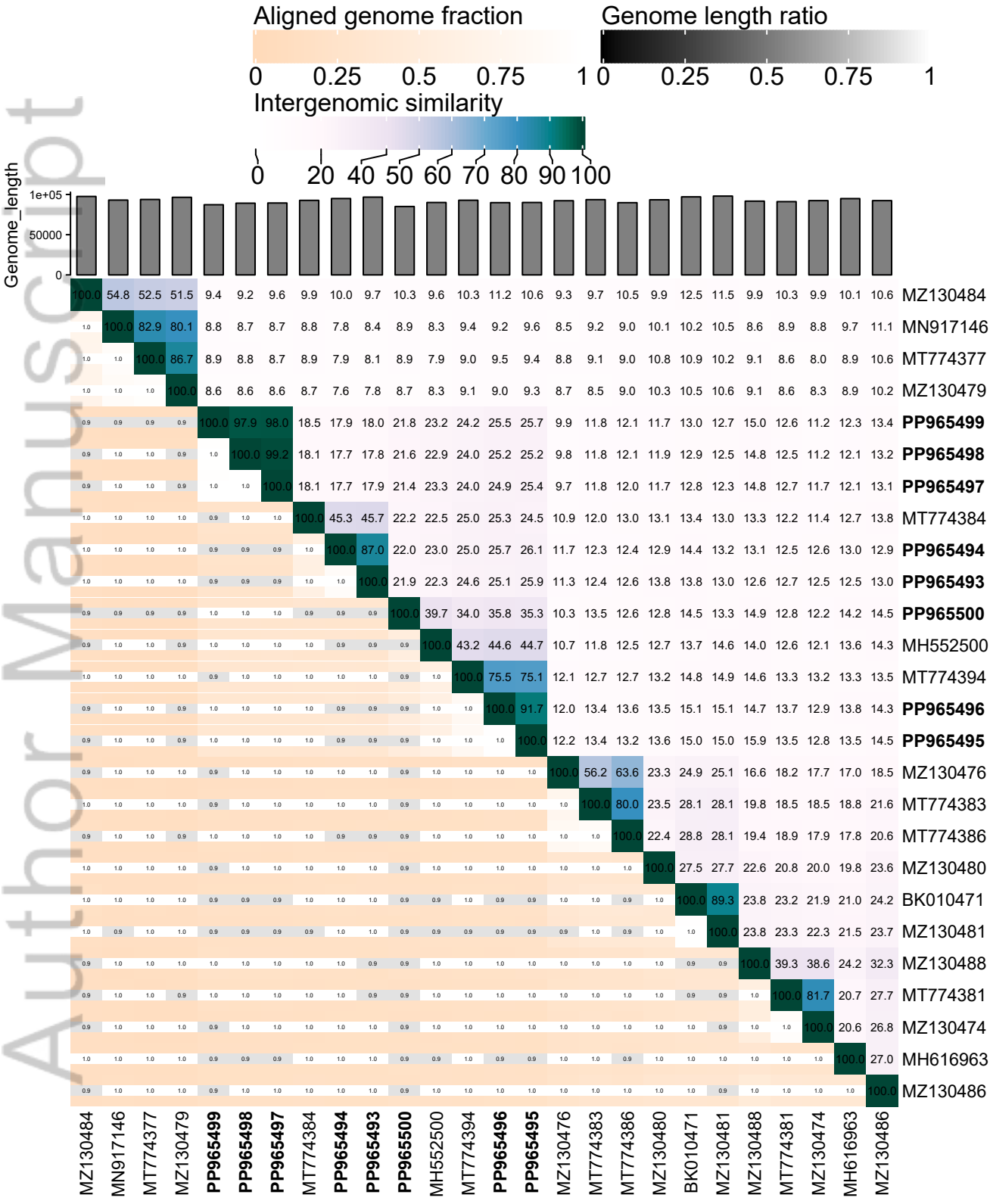
60. Baniel A, Amato KR, Beehner JC, Bergman TJ, Mercer A, Perlman RF, et al. Seasonal shifts in the gut microbiome indicate plastic responses to diet in wild geladas. Microbiome. 2021;9: 26.

37. Schneider Crease, I., Moya, I., Chiou, K. L., Azanaw, A., Kebede, F., Abebe, B., Lu, A., Bergman, T. J., Snyder-Mackler, N., & Varsani, A. Intestivirid acquisition increases across infancy in a wild primate population [Data set]. https://github.com/GeladaResearchProject/Schneider-Crease_et_al_intestivirids_2024

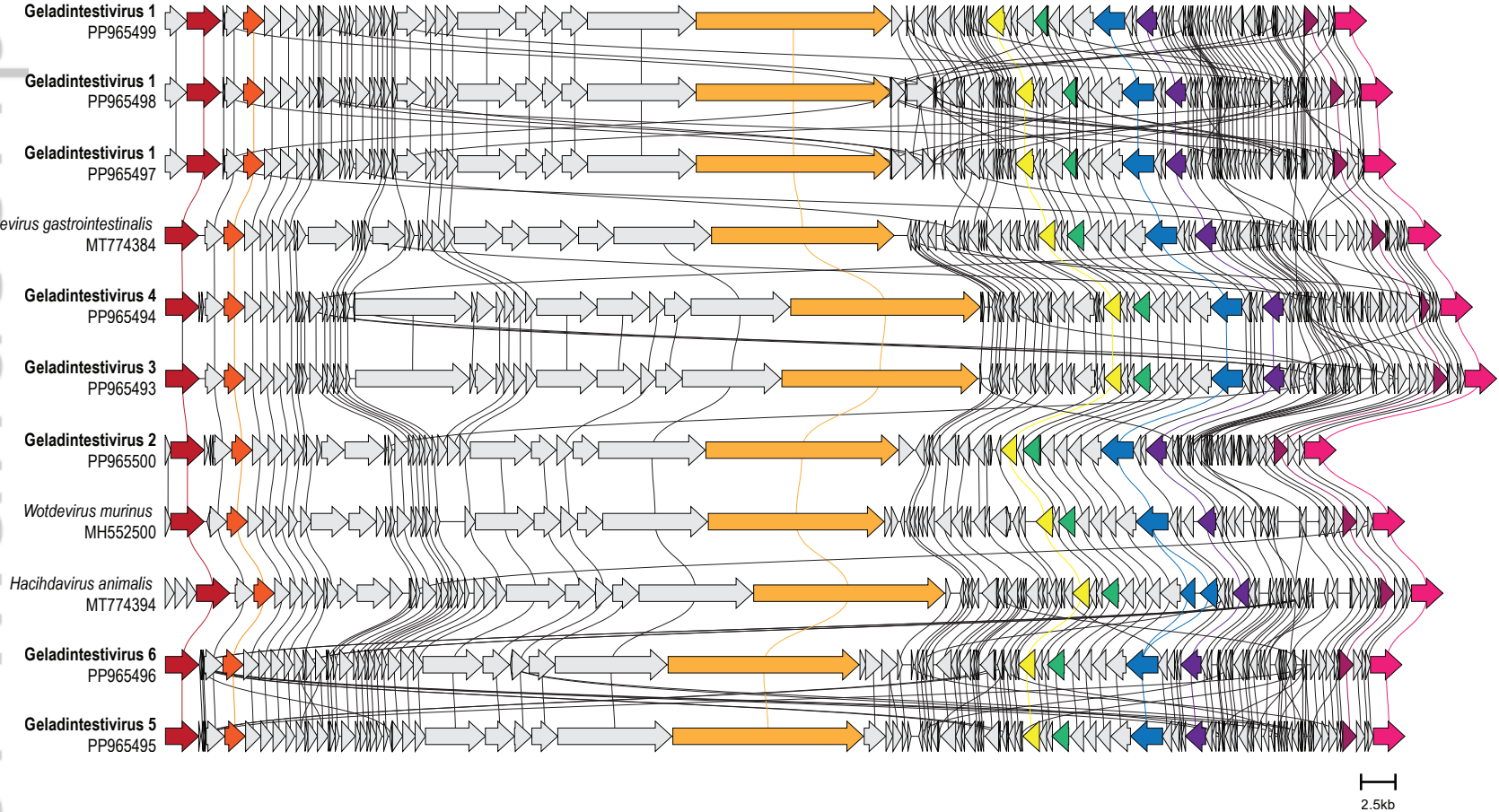
Author Manuscript

Data Accessibility and Benefit-Sharing

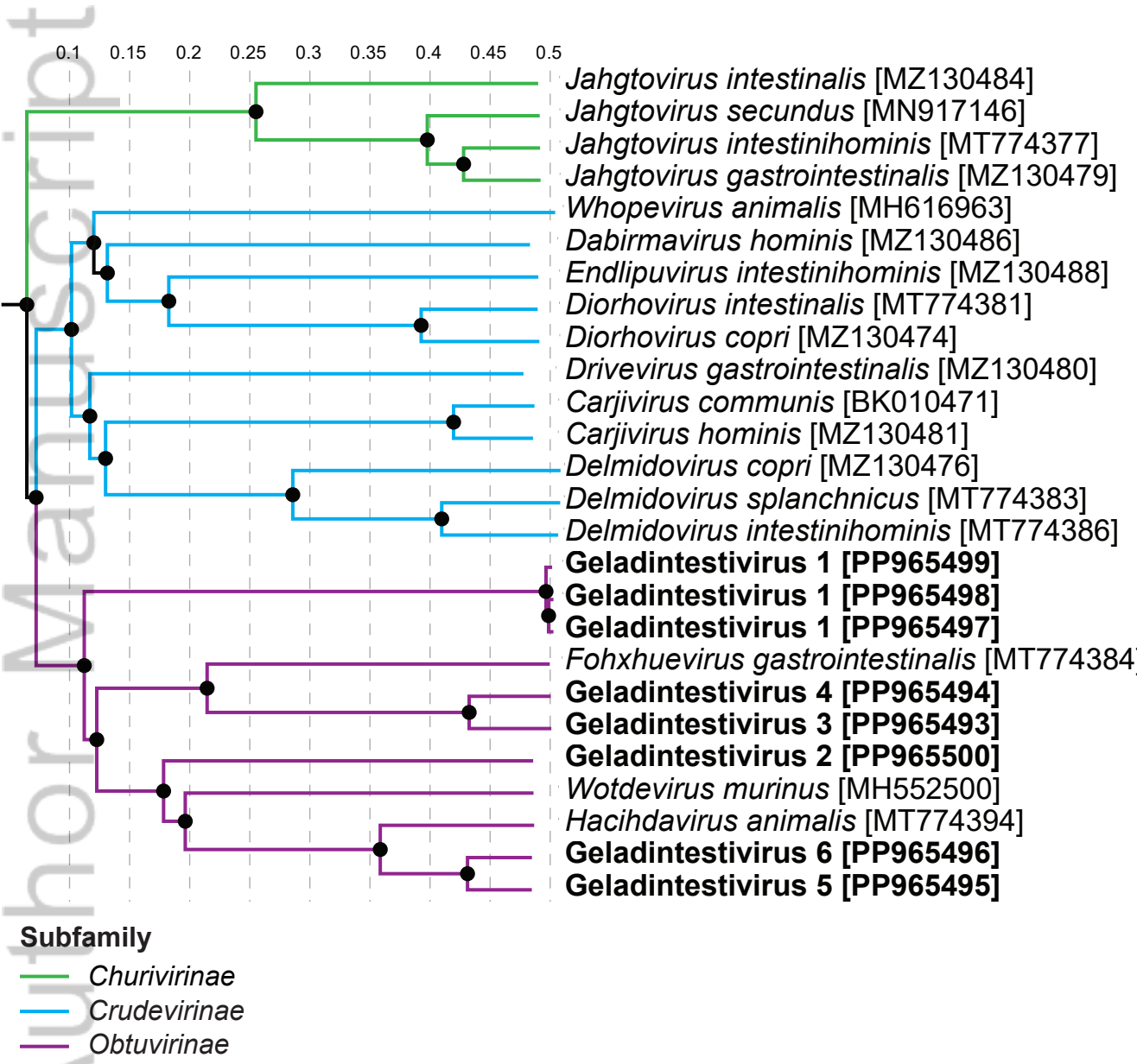
Sequences have been deposited in GenBank with accession #s PP965493 - PP965500 and the raw reads mapping to these have been deposited under Bioproject # PRJNA795171, BioSamples #s SAMN42063139 - SAMN42063146 and SRA #s SRR29560746 - SRR29560753. Code and data to replicate analyses can be found at: https://github.com/GeladaResearchProject/Schneider-Crease_et_al_intestivirids_2024/



Author Manuscript

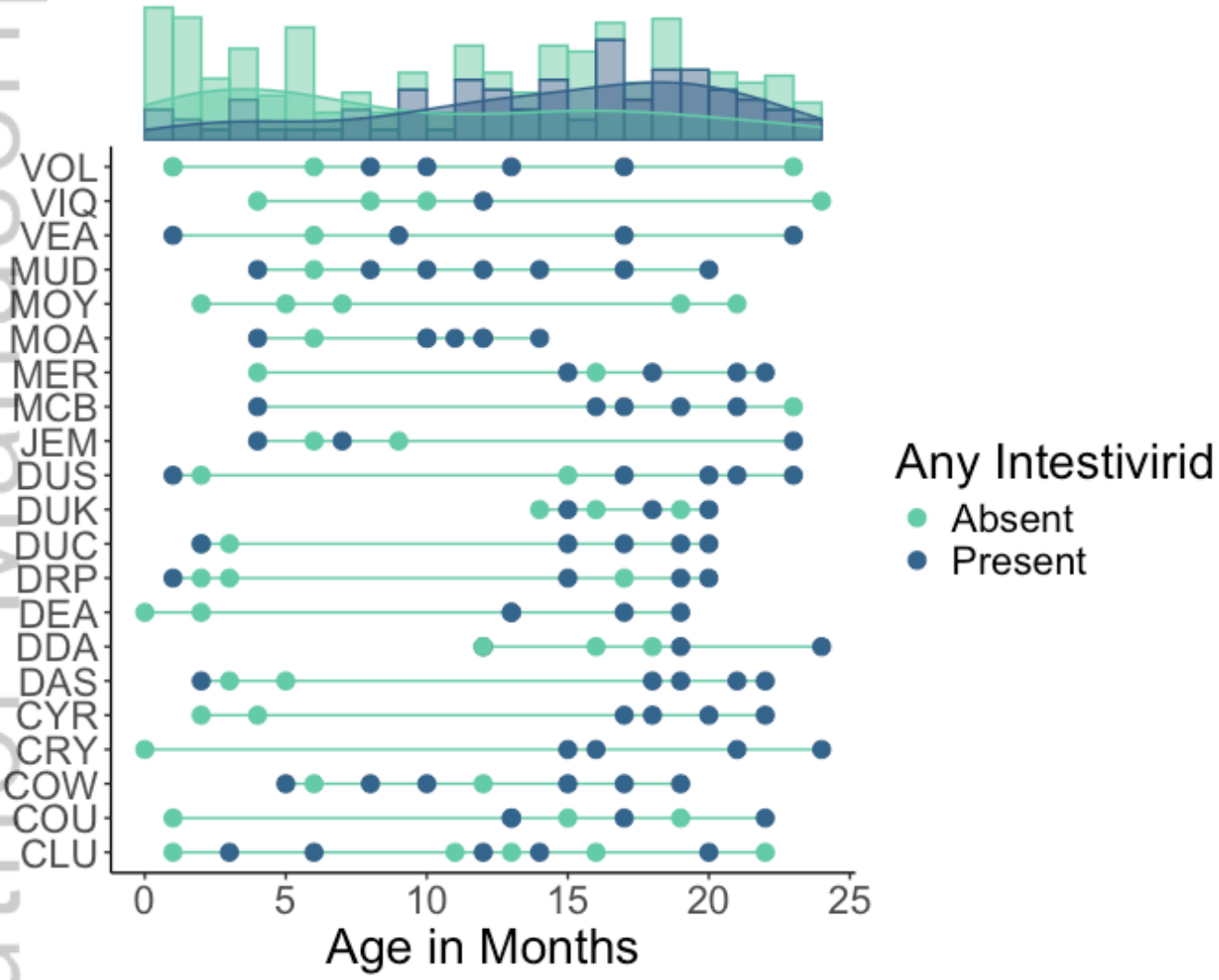


Author Manuscript



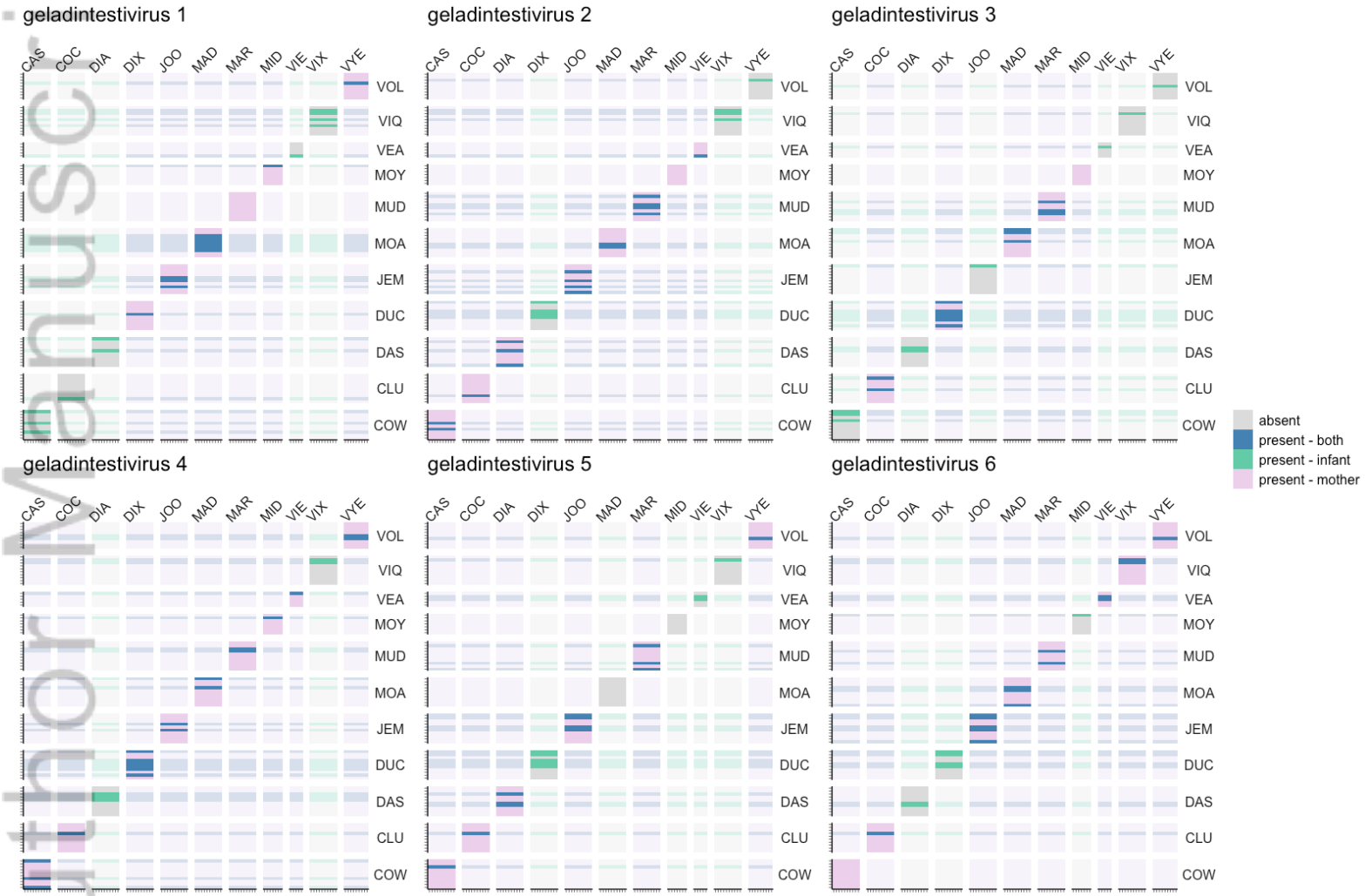
Author Manuscript

Author Manuscript



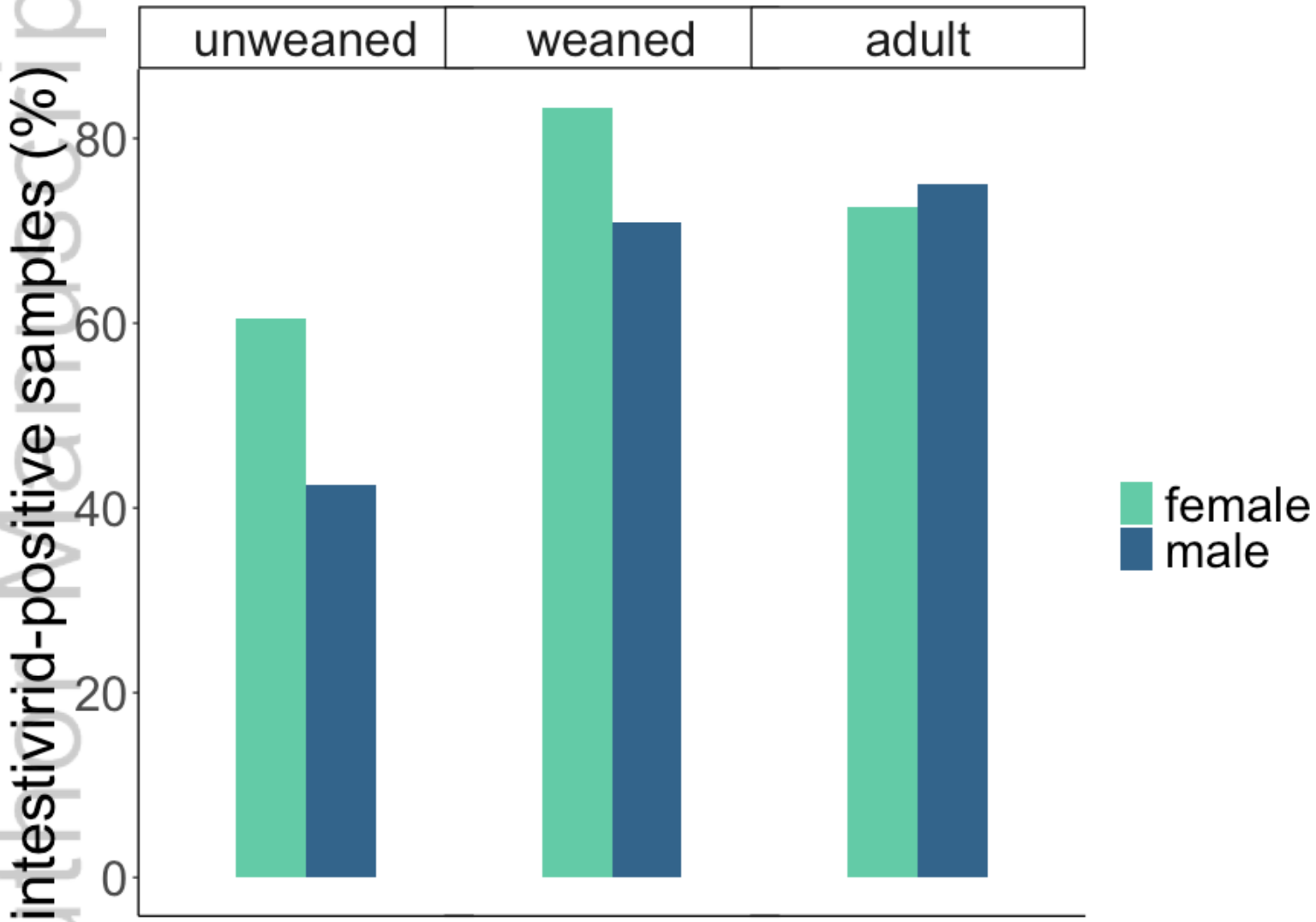
infant_trajectories_10_legend.png

Author Manuscript



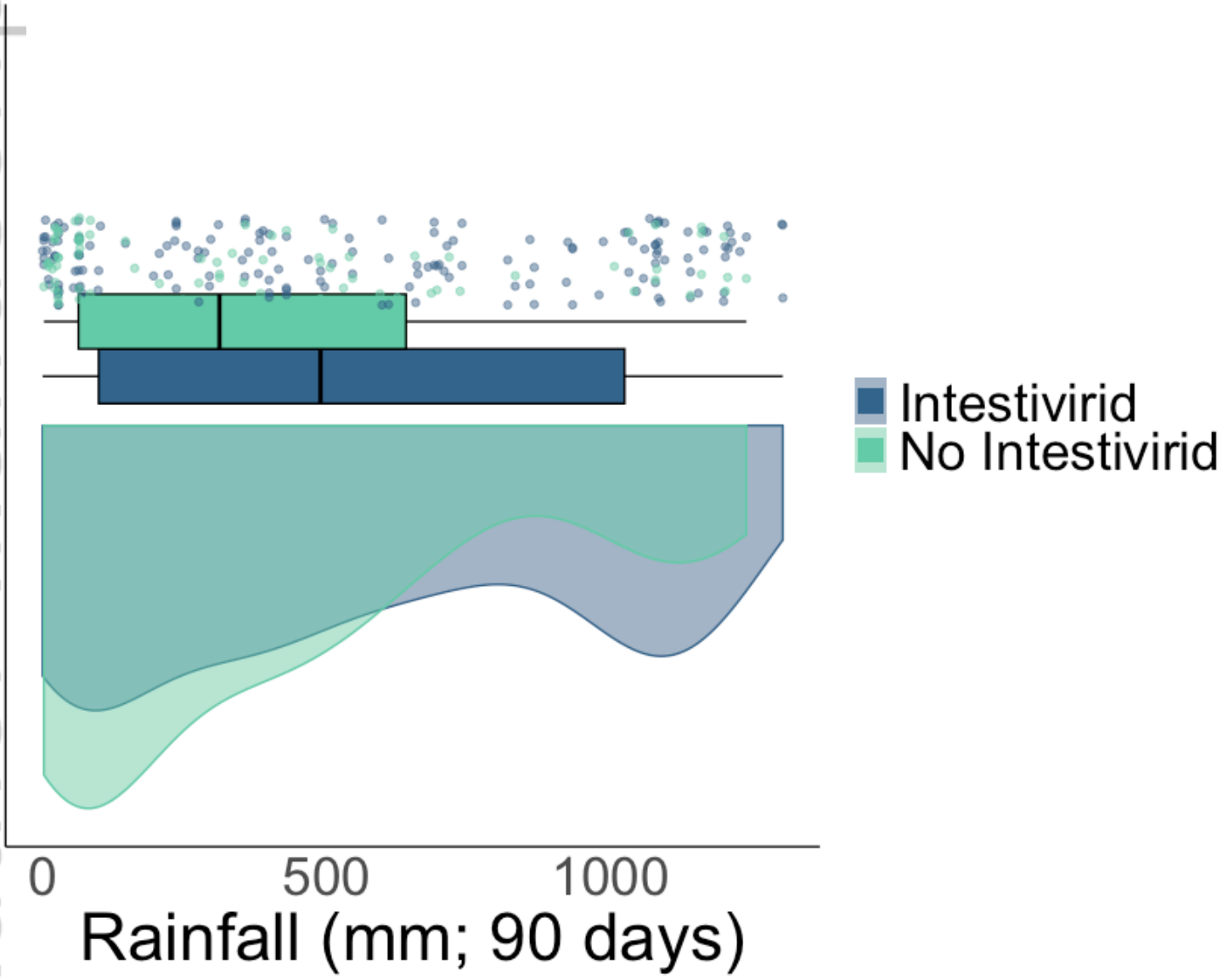
matched_10_fig.png

Author Manuscript



percentages.png

Author Manuscript



rainfall_raincloud.png