The Laboratory for
RNA-Based Lifeforms
https://RNAlab.ca | @RNA_Life

RNAlab preprint, 2023, pp. 1–29

doi: pending
Online Date: September 18, 2023
Research Article - preprint

RESEARCH ARTICLE - PREPRINT

A Parasite Odyssey: An RNA virus concealed in *Toxoplasma gondii*

Purav Gupta,[1,2,*] Aiden Hiller,[1,2,*] Jawad Chowdhury,[1,2,*] Declan Lim,[1,2,*] Dillon Lim,[3,*] Artem Babaian,[1,2,*,†] Luke Pereira[1,2,*] and Alex Morales[4,*]

[1]Department of Molecular Genetics, University of Toronto, Toronto, Canada, [2]The Donnelly Centre for Cellular + Biomolecular Research, University of Toronto, Toronto, Canada, [3]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK and [4]MOTA SOFTWARE CONSULTING LLC, Deleware, USA
[*]Authors contributed equally, order was determined by Starfox64 tournament.[†]Corresponding author: a.babaian@utoronto.ca

Abstract

We are in the midst of the "Platinum Age of Virus Discovery", an era characterized by the exponential growth in the discovery of virus biodiversity. In humans (or any animal) there are more viruses than simply those directly infecting the animal cells. Viruses infect all species constituting the microbiome: including bacteria, fungi, and unicellular parasites. Thus the complexity of possible interactions between host, microbe, and viruses is unfathomable. To understand this network we must employ computationally assisted virology as a means of analyzing and interpreting the millions of available samples to decipher the myriad of ways in which viruses may intersect human health.

From a global screen for human neuron-associated viruses, we identify a novel narnavirus *Achaeavirus odysseus* (Ao) which likely infects the neurotropic parasite *Toxoplasma gondii*. Previously, several parasitic protozoan viruses (PPVs) have been mechanistically established as triggers of host innate responses, and here we present *in silico* evidence that Ao is a plausible pro-inflammatory factor in mouse and human cells infected by *T. gondii*. *T. gondii* infects billions of people worldwide, yet the prognosis of toxoplasmosis disease is highly variable, and PPVs like Ao could function as a hitherto undescribed hypervirulence factor. More broadly, we explore phylogenetically proximal viruses to Ao and discover 18 new species of *Achaeavirus*, all found in vertebrate transcriptome and metatranscriptomes. While the Narnavirus samples making up this genus-like clade are derived from sheep, goat, bat, rabbit, chicken, and pigeon, the presence of virus is strongly predictive of parasitic (*Apicomplexa*) nucleic acid co-occurrence, supporting that these are a clade of parasite-infecting viruses.

This is a computational proof-of-concept study in which we rapidly analyze millions of datasets from which we distil a mechanistically, ecologically, and phylogenetically refined hypothesis. We predict this highly diverged Ao RNA virus is biologically a *T. gondii* infection, and that Ao, and other viruses like it, will modulate this disease which afflicts billions worldwide.

Key words: RNA Virus Discovery, Narnaviruses, Toxoplasma gondii, Hypervirulence, Computational Virology

Introduction

RNA virus discovery is undergoing a revolutionary expansion in the characterization of virus diversity (Shi et al., 2016, 2018; Wolf et al., 2020; Edgar et al., 2022; Neri et al., 2022). Of the 10^8 to 10^{12} virus species on Earth (Koonin et al., 2020), $\approx 300,000$ mammalian viruses are thought to have human-infecting potential (Anthony et al., 2013), of which we know ≈ 160 RNA viruses (Woolhouse and Adair, 2013). This bulk of unknown wildlife-originating (zoonotic) viruses is expected to cause the majority of emerging infectious diseases in humans (Jones et al., 2008), with precedent set by the 1918 Spanish influenza, AIDS, SARS, Ebola, and more recently COVID-19. This establishes the real and continued threat that viral zoonoses pose to global health.

Most established relationships between a disease and its causal RNA virus are direct and proximal to infection, such relationships are tractable to reductionist interrogation. Yet it is evident that confounding variables can impede linking cause and effect, for example, virus genetic heterogeneity, chronic infections, asymptomatic carriers, prolonged latency periods, and microbiome interactions all add complexity in establishing the link between causal virus and disease. An indirect, yet causal, relationship is well exemplified by the Epstein-Barr virus and multiple sclerosis (EBV-MS). The EBV-MS association has long been statistically postulated, yet a causality was only established last year (Bjornevik et al., 2022; Soldan and Lieberman, 2023). Thus, statistical and computational inference, while not sufficient to establish causation, do allow for a radical simplification of the space of plausible hypotheses, thereby accelerating the time to discovery. We opine that in addition to virus discovery, virus effect inference should be a primary objective of the emerging field of computational virology.

Parasitic protozoan viruses, nested invaders

The diversity of viral niches within the human ecosystem encompasses more than human cells themselves. Viruses interact with, infect, and alter the biology of all constituent organisms in the human holobiont. One functional grouping of such viruses is the parasitic protozoan viruses (PPVs), a subset of which have been called Matryoshka viruses (Russian nested-doll) since they infect unicellular parasites, which in turn infect animal host cells. Notably PPVs are found within the parasitic phylum *Apicomplexa*, in the genera *Cryptosporidia*, *Trichomonas*, *Leishmania*, *Giardia*, *Entamoeba*, and possibly *Plasmodium* (Charon et al., 2019; Kinsella et al., 2020; Graves et al., 2019; de Carvalho et al., 2019).

More than mere passengers, PPVs enter animal host cells via their carrier parasite and can inadvertently modulate host immune responses, which alters the parasite's pathogenicity (Gómez-Arreaza et al., 2017; Zhao et al., 2023). Since viral nucleic acids, including viral RNA (vRNA), are recognized as pathogen-associated molecular patterns, they are sufficient to initiate an innate immune response via nucleic acid sensors (NAS). Specifically, some toll-like receptors (TLRs) survey endosomes for dsRNA (TLR3), ssRNA (TLR7,8), or unmethylated CpG motifs in ssDNA (TLR9) (Fitzgerald and Kagan, 2020). Additionally, the three RIG-I-like receptors (RLRs) - the signalling RIG-I and MDA5, and the regulatory LGP2 - survey the cytosol for vRNA transcripts lacking a 5' cap, or for misprocessed cellular RNA (Rehwinkel and Gack, 2020). Collectively, NAS orchestrate an antiviral type I interferon (IFN) response (Lee and Ashkar, 2018).

For instance, the dsRNA virus Cryptosporidium parvum virus 1 (CSpV1) which infects *Cryptosporidium parvum*, activates a type I IFN inflammatory cascade in mouse models. Interestingly, this response undermines host defences against *C. parvum*, as evidenced by the enhanced parasitic immunity observed in mice lacking type I IFN receptor (Deng et al., 2023). While the underlying mechanism for the inflammation remains elusive, PPV presence appears to be necessary for parasite pathogenicity, potentially by diverting the host's innate immune system towards the activation of antiviral immunity and away from antiparasitic immunity. In a related context, Trichomonas vaginalis virus (TVV) infects *Trichomonas vaginalis*, the causative agent for a prevalent sexually transmitted infection (Narayanasamy et al., 2022). The RNA genome of this virus activates a TLR3-mediated type I IFN response, thereby exacerbating the inflammatory pathogenesis of the disease (Fichorova et al., 2012). Conversely, PPVs, such as Giardia lamblia virus 1 (GlV1) can mitigate parasitic pathogenesis. In the case of GlV1, the virus inhibits the growth of *G. lamblia* (Miller et al., 1988), highlighting the complexity of this tripartite relationship.

In general, PPVs are poorly characterized. The majority of PPVs identified to date are dsRNA Totiviruses and Partitiviruses (Gómez-Arreaza et al., 2017). However, ssRNA Narnaviruses have been recently linked with *Apicomplexa* and *Euglenozoa* parasites. Notably, the Matryoshka viruses are associated with *Plasmodium vivax*, a causative agent of malaria, and species in the genus *Leucocytozoon* (Charon et al., 2019; Rodrigues et al., 2022).

Toxoplasma gondii, a global pathogen

Surprisingly, there are no known viruses associated with *Toxoplasma gondii*, an apicomplexan parasite infecting up to 60% of the world's population and tropic for all animal tissues, including the brain (Pappas et al., 2009). While *T. gondii* is not commonly thought of as a pathogenic parasite, sporadic *T. gondii* strains are hypervirulent, causing severe disease (Dardé et al., 2020). Thus the full global health burden of Toxoplasmosis remains unclear, especially its possible role in chronic neuroinflammation and subsequent neurodegenerative disease.

Felids (cats) are the definitive host for *T. gondii*, where the parasite can sexually reproduce. The felids then excrete oocysts which contaminate soil and water supplies. These oocysts can be ingested, infecting a remarkably broad host range of warm-blooded animals wherein the parasite replicates asexually (Dubey, 2021a). During initial infection, there is rapid proliferation of *Toxoplasma* as tachyzoites. Subsequently, *T. gondii* transitions to a slow proliferation bradyzoite phase, characterized by the

formation of cysts in the muscle, heart, lung, liver, eye, or central nervous system (CNS) tissues (Johnson and Johnson, 2021). The consumption of raw or poorly cooked meat containing these cysts serves as the second significant route of transmission for *T. gondii* (Johnson and Johnson, 2021). The majority of *T. gondii* is asymptomatic or self-limiting, but transplacental vertical transmission can cause severe congenital toxoplasmosis (McAuley, 2014). Likewise, transplantation-associated toxoplasmosis poses a serious risk, particularly in immunosuppressed transplant recipients (Khurana and Batra, 2016).

Independent of host factors, *T. gondii* pathogenicity has high variance. *Toxoplasma* can actively invade any mammalian cell, such as intestinal enterocytes of the gut lumen upon ingestion (Gregg et al., 2013), or they can be taken up by tissue-resident phagocytes (Werk, 1985). The latter immune tropism is an important niche, for it supports the dissemination of the parasite across body sites, potentially with particular importance for CNS infiltration (Matta et al., 2021; Schneider et al., 2022). Intracellularly, the tachyzoites reside in a parasitophorous vacuole masking them from host immune recognition. Asexual reproduction eventually leads to cell rupture and the spread of the parasite. This produces the non-specific symptoms of toxoplasmosis which are fatigue, headache, mild fever, myalgia, arthralgia, and isolated cervical lymphadenopathy indicative of spread to the head and neck (Montoya and Liesenfeld, 2004; Dubey, 2021a). Occasionally, symptoms are also produced by lytic replication in peripheral tissues - myocarditis reflecting cardiac spread, polymyositis reflecting skeletal muscle spread, encephalitis reflecting CNS spread, and so on(Montoya and Liesenfeld, 2004; Dubey, 2021a). Subsequently, *T. gondii* can transition to a slow proliferation bradyzoite phase. A severe potential complication is ocular toxoplasmosis, the leading cause of infectious posterior uveitis which presents as foggy vision and is associated with white focal lesions on the retina and inflammation in the vitreous humour (Montoya and Liesenfeld, 2004; Matta et al., 2021; Kalogeropoulos et al., 2022). Given this invasive potential of *T. gondii*, it would serve as an effective Trojan horse for disseminating a virus.

While CNS-symptomatic toxoplasmosis is more common in immunocompromised patients (Montoya and Liesenfeld, 2004), mounting evidence links latent neuronal infection with altered host behaviour (Johnson and Johnson, 2021; Matta et al., 2021). Biochemically, mouse infection models undergo altered neurotransmitter metabolism with increased dopamine and decreased noradrenaline levels (Johnson and Johnson, 2021). Cellularly, reduced dendritic spine and synapse counts are associated with sustained *Toxoplasma* neuroinflammation (Matta et al., 2021). In both rodents and hominids, infection leads to a suppressed aversion to felid urine, which is predicted to increase the likelihood of felid predation and infection of the definitive host. Other serious changes, including cognitive impairment, prolonged reaction times, associations with suicidal behaviour, and prevalence of several psychiatric conditions, have also been observed in humans (Johnson and Johnson, 2021). There have also been suggestions that chronic *Toxoplasma* infection of the CNS serves as a risk factor in neurodegenerative conditions such as Alzheimer's disease (AD) where there is a known neuroinflammatory component (Nayeri et al., 2021; Yang et al., 2021), although others have noted potentially strain-dependent neuroprotective effects in AD mouse models (Jung et al., 2012; Cabral et al., 2017), and others still find no significant effect on murine age-related cognitive decline (McGovern et al., 2020). This underscores the importance of a better understanding of factors which influence *Toxoplasma* pathogenicity factors, why some patients develop CNS pathology, while others are seemingly unaffected, and even protected.

Asking an old question in a new way

The volume of freely available sequencing data in the Sequence Read Archive (SRA) has grown explosively for over a decade (Katz et al., 2021). Currently, there are in excess of 29 petabases (Pbp) of freely available sequencing data, capturing $>$ 10 million biological samples. The emerging field of petabase-scale computational biology strives to analyze the totality of this data and enable expansive meta-analyses. The SRA-STAT project has processed over 10.8 million datasets (27.9 Pbp) using a k-mer hashing algorithm to create an index of reads matching known taxa genomes. (Katz et al., 2021). Likewise, Serratus which is aimed at uncovering known and novel RNA viruses using a translated-nucleotide search for the RNA-dependent RNA polymerase (RdRp) RNA virus hallmark gene, has processed over 7.5 million RNA-sequencing datasets (18.97 Pbp) (Edgar et al., 2022). Our group focuses on leveraging this critical mass of freely available data to re-interrogate fundamental questions in virology in a data-driven philosophy. The objective is to minimize a priori bias and maximize the discovery of unexpected biology. Our immediate focus is to characterize highly divergent neuroinflammatory RNA viruses. We hypothesize that human neuroinflammatory diseases of unknown etiology, such as AD, can be caused by highly-divergent viral infections.

Results

Identification of T. gondii associated viral sequences

To screen for novel human neuroinflammatory RNA viruses, we searched the BioSample metadata associated with the SRA annotated as the human and central nervous system. From the 7,675,502 sequencing runs or 18.97 Pbp processed in Serratus, 483,173 (6.29%) runs are annotated with a central nervous system term, of which 82,454 (17.07%) are human. Manually inspecting the putative palmprints in this data, we identified a divergent RdRp fragment (palmprint: `u380516_cranialChivalry`) common to 3 human neuronal datasets (`SRR1205923`, `SRR1204654`, and `SRR1205190`) from one BioProject (`PRJNA241125`). We have termed the BioProject's associated paper the Ngô study, wherein the authors infected three human cell types (neuronal stem cells, neuronal differentiated cells, and monocytes) *in vitro* with nine different strains of the apicomplexan parasite *T. gondii*, or a mock control (Ngô et al., 2017). We *de novo* assembled all 117 available Ngô mRNA-seq runs and identified a 3,177 nucleotide (nt) putative viral contig with a coding-complete 989 amino acid RdRp. A BLASTp search on the translated protein revealed highly divergent

homologs, with the highest percent identity among these being an unclassified *Riboviria* RdRp (accession: `QIM73983.1`).

The RdRp contig lacked structural genes, and we therefore postulated the existence of additional segments. To screen for co-occurring RdRp contigs, we depleted contigs mapping to the genomes of either *T. gondii* or *H. sapiens*. This produced a second 1,283 nt contig correlated with the RdRp (Pearson correlation $R^2 = 0.90$). This contig encodes four open reading frames (ORFs) with no identifiable homologs (Figure 1). Given the well-conserved RdRp sequence is already highly divergent, it is not unexpected that a viral accessory or structural gene would prove more difficult to identify.

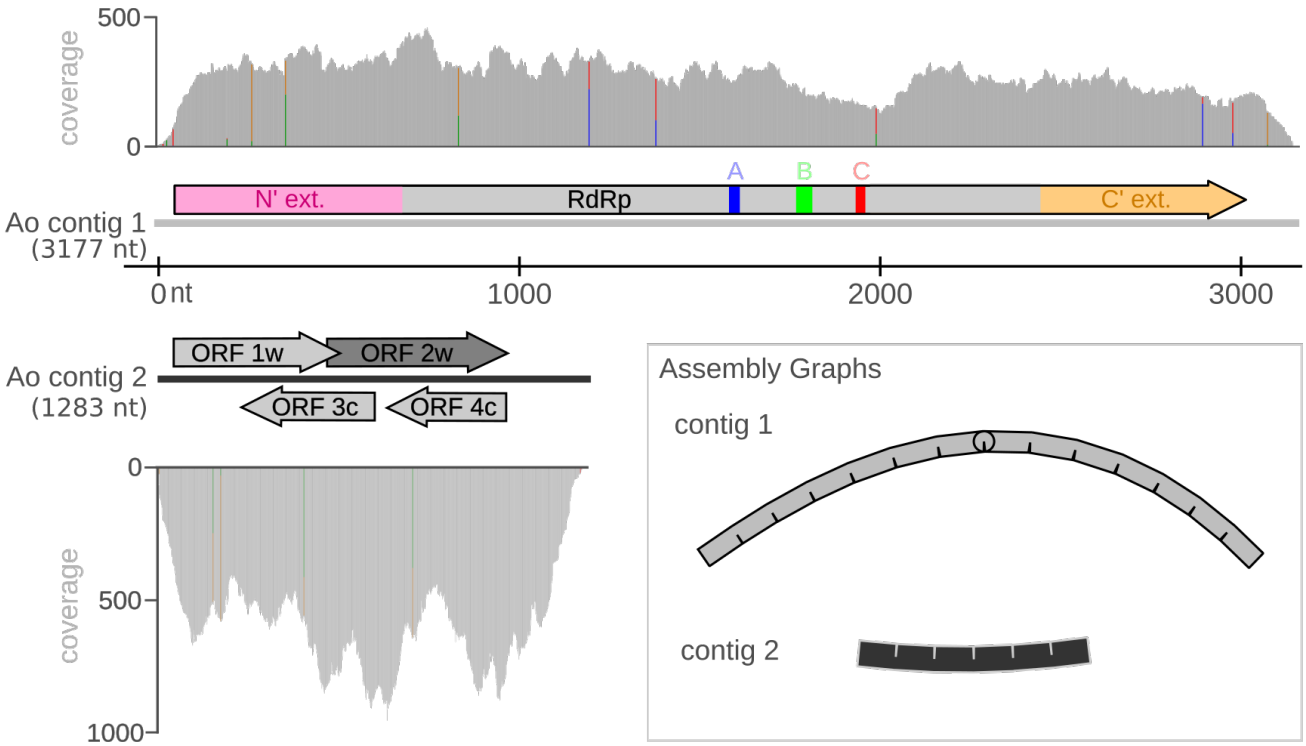


Figure 1. *Achaeavirus odysseus* (Ao) genome. Assembling (rnaSPAdes) the Ngô RNA sequencing datasets, we recovered a high-coverage (123x coverage, `SRR1205923`) complete RdRp-encoding contig and identified a second, correlated, contig of likely viral origin. Contig 2 (279x coverage) contains four ORFs: ORF1w (Watson), ORF2w, ORF3c (Crick), and ORF4c but none of these ORFs have identifiable homologs via BLASTp in the non-redundant and transcriptome shotgun assembly databases, TBLASTN with the nucleotide database, INTERPROSCAN, or Foldseek.

To quantify the relative abundance of human, parasite, and virus nucleic acids across each library, we aligned the reads against each respective genome and observed that high expression of the putative viral RdRp contig (RPKM ≥ 10) was specific to samples experimentally infected by the *T. gondii* - RUB strain (mRNAseq = 4/4, miRNAseq 5/5), and was mostly absent (RPKM ≤ 1.0) from other strains or controls (mRNAseq = 112/113, miRNAseq 115/115, Fisher's Exact Test, $p < 1e-4$) (Figure 2A). We rule out that the RdRp is an endogenous viral element since zero of six DNAseq replicates show any mapping reads in either *T. gondii* - RUB or *T. gondii* - COUGAR strains (Figure 2B). Trace RdRp expression (RPKM 0.5 - 2.0) was observed in three *T. gondii*-infected samples, all annotated as other strains: RH, GT1, and RH-GRA10[KO]. The relatively low abundance of sequencing reads is suggestive of sequencing cross-contamination, although early stages of viral cross-infection between samples cannot be ruled out, in which case the RUB strain would be the likely source (Figure 2C).

We assert that these two contigs constitute the genome of a bi-segmented RNA virus infecting *T. gondii* - RUB. We name the virus *Achaeavirus odysseus* (Ao), after the Greek mythical warrior who built a wooden horse to sneak his soldiers into the besieged city of Troy.

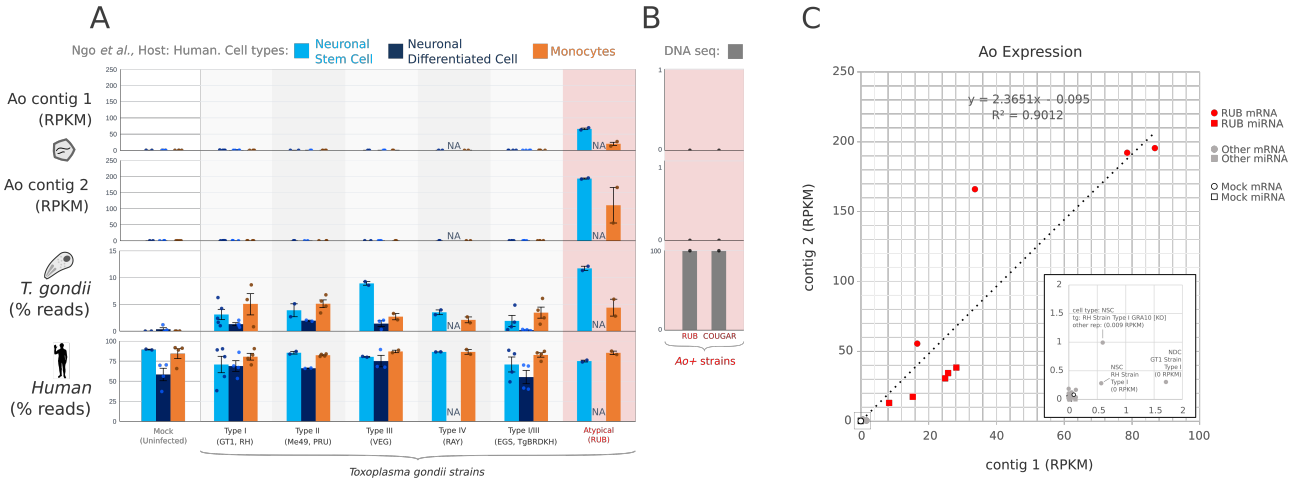


Figure 2. Quantification of *A. odysseus* (Ao) genome across Toxoplasma-infected human cell lines in the Ngô study. A. The expression (reads per million kilobase of mapped reads, RPKM) of contig 1 and contig 2 of *A. odysseus* (Ao) was quantified in three cell types (n = 37 neuronal stem cell, NSC; n = 24 neuronal differentiated cell, NDC; and n = 40 monocytes), infected by one of nine strains (n = 117 runs total) of *T. gondii*, in addition to quantifying toxoplasma or human genome-mapped reads (percentage total) from the same datasets. Ao was found at high expression exclusively in *T. gondii* - RUB strain samples within Ngô *et al.* mRNAseq and miRNAseq (not shown). Data points and 2 SD Error Bars are shown. **B.** DNAseq quantification of Ao in four replicates of *T. gondii* - RUB (0/40,151,652 reads) and two *T. gondii* - COUGAR (0/31,562,801 reads) in BioProject: `PRJNA61119` fails to identify any virus mapping reads. **C.** Ao contig 1 and contig 2 expression (reads per kilobase of million mapped reads, RPKM) and Pearson correlation between contigs across all mRNA and miRNAseq datasets in the Ngô study, N = 237.

A. odysseus infects geographically distinct strains of T. gondii

30-60% of the world's population will be infected by *T. gondii*, and while the majority of infections will resolve asymptomatically, there are sporadic cases of severe toxoplasmosis. Next, we investigated the global distribution of *T. gondii* and more broadly apicomplexa-containing sequencing data and their association, if any, with Achaeaviruses.

We hypothesized that Ao would only be observed if its parasite host was present. To test this, we re-queried all 7.5 million sequencing runs processed by Serratus for Ao and identified one additional Ao+ BioProject (0.00052%, or 1/191,678 BioProjects): `PRJNA114693`, henceforth "the Melo Study" (Melo et al., 2013). Reviewing the Melo study, they infected murine macrophages *in vitro* with 30 distinct strains of *T. gondii*, or a mock control (Supplementary Table S1). Following *de novo* assembly of all 32 Melo samples, we identified Ao in 2 samples: *T. gondii* - RUB strain (again), and *T. gondii* - COUGAR strain. From the RUB library (`SRR446933`), we uncovered a 3,202 nucleotide contig encoding a 989 amino acid RdRp. This RdRp protein had a 100% alignment match to the Ngô Ao RdRp. Thus we isolated near identical Ao from two independent studies and established that *T. gondii* - RUB is a common source of virus. Moreover, the Ao+ *T. gondii* - COUGAR strain (`SRR446909`) contains a 3,184 nucleotide contig encoding a 987 amino acid RNA dependent RNA polymerase (RdRp). The Ao COUGAR RdRp has 95.64% amino acid alignment identity to the Ao RUB RdRp, demonstrating the existence of multiple viral strains of Ao amongst *T. gondii* strains. Further, the second Ao contig was found in both RUB and COUGAR Melo libraries spanning 1,286 nucleotides (99.4% aa-identity to Ao RUB Ngô) and 1,322 nucleotides (92.26% aa-identity to Ao RUB Ngô), respectively. Similar to the Ngô study, Ao contig 1 and contig 2 exhibited a high expression (RPKM $\geq$ 10) in RUB and COUGAR libraries, and are absent (RPKM $\leq$ 1) from 28 other *T. gondii* strains or the mock control (Figure 3 and Supplementary Table S2). To identify additional RUB or COUGAR RNA sequencing data, we performed a metadata search but failed to find any additional datasets in the SRA from these strains. Given the available data, Ao appears to be fully penetrant amongst *T. gondii* RUB and COUGAR RNA.

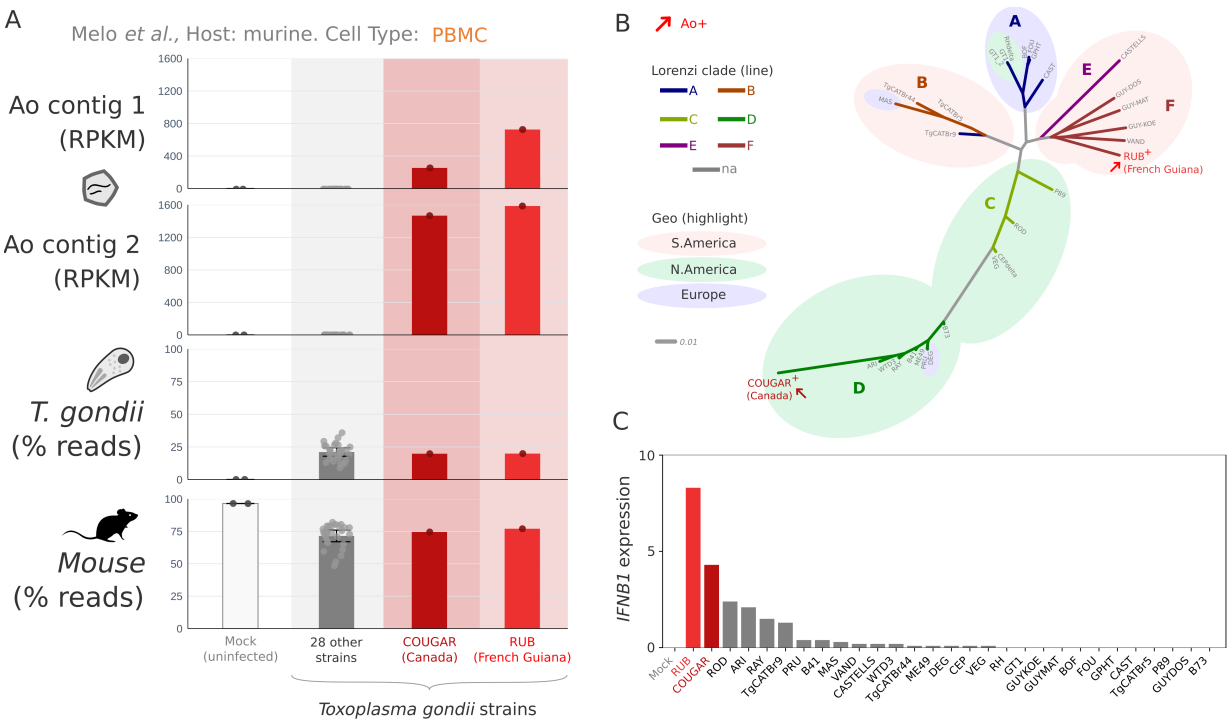


Figure 3. Quantification of *A. odysseus* in Melo *et al.* A. Quantification of Ao-aligning reads amongst 31 strains of *T. gondii* strains and mock control, error bars indicate 2 standard deviations. **B.** Unrooted *T. gondii* strain phylogeny (IQ-TREE) from transcriptome in the Melo study (149,777 measured sites; nodes have 100% bootstrap support, see also Supplementary Figure S1). The original geographic isolation of strains is highlighted as North American (green), European (blue), or South American (pink), and lines are coloured by the genomic *T. gondii* clades defined by Lorenzi et al (Supplementary Table S2). *A. odysseus* (Ao, red arrow) was detected at high levels in two geographically and phylogenetically unrelated strains of *T. gondii*; RUB isolated from a symptomatic soldier in French Guiana in 1991 and COUGAR isolated from a cougar in British Columbia in 1996. **C.** RNA sequencing expression data obtained from Supplementary Table 1 of the Melo *et al.* study showing normalized expression levels of *Ifnb1* in murine macrophages infected with strains of *T. gondii*.

Over the estimated 1 million years since *T. gondii* arose in South America (Bertranpetit et al., 2017), the parasite has radiated into phylogeographically distinct clades A through F (Lorenzi et al., 2016). Hypervirulence is polyphyletic amongst *T. gondii* strains, and not isolated to one geographical locale. *T. gondii* - RUB and COUGAR strains are phylogenetically and geographically divergent from one another, RUB is a clade F strain isolated from a human in French Guiana (Dardé et al., 1998), while COUGAR is a clade D strain isolated from a mountain lion associated with a toxoplasmosis outbreak in British Columbia, Canada (Aramini et al., 1998) (Supplementary Figure S1). However, all available samples of RUB and COUGAR are experimental/laboratory-associated, so it remains undetermined if Ao is found in wild isolates. Regardless, given the evolutionary, geographic, and host differences (both wild source, and experimental differences between projects) between the Ao RUB and COUGAR strains, and the absence of virus in all intermediate parasite strains: it is plausible that Ao is a horizontally infectious virus circulating in *T. gondii*.

Achaeaviruses are a diverse clade of parasite-associated Narnaviruses

To elucidate the evolutionary and ecological context of the *T. gondii* virus in the landscape of viral diversity, we interrogated the viruses phylogenetically proximal to Ao. Although hidden Markov model (HMM) homology search via PFAM and InterProScan (Finn et al., 2014; Jones et al., 2014) failed to recognize Ao's RdRp, remote homologs were identified with BLASTp (Altschul et al., 1990) within the Narnavirus lineage, this discrepancy is due to a lack of adequate Narnavirus representation in standard HMMs. Narnaviruses are mostly capsid-less single-stranded RNA viruses, the exemplar species being 20S and 23S Narnavirus which infect the laboratory model fungus, *Saccharomyces cerevisiae*, to which Ao shares 25% and 28% amino acid identity, respectively. The highest similarity match to Ao RdRp was an unclassified ribovirus partial RdRp (52% aa identity, query coverage 40%, accession: `QIM73983.1`) isolated in 2016 from the lung of a Ryukyu mouse (*Mus caroli*) in Thailand (Wu et al., 2021). The remaining Narnaviruses matches were distal, at sub-40% amino acid identity.

Due to a paucity of data within the family *Narnaviridae*, and the divergent nature of Ao, we re-queried Serratus for Ao-like RdRp palmprints. We retrieved 166 matching runs from 46 BioProjects with 32 species-like operational taxonomic units (sOTUs) (186 distinct virus-run observations, Supplementary Table S3). For each sOTU, we selected the top 3 libraries ranked by contig-coverage where available, to a total of 60 assemblies. We recovered relevant Narnavirus RdRp contigs for all 32/32 (100%) sOTU, of which 22/32 (75%) contained a sufficiently long RdRp sequence ($\approx$ 390 amino acids, motif F-E and thumb) for reliable phylogenetic reconstruction. In addition to the criterion requiring motif F-E of RdRp alignment, we enforced that RdRps must be $<$ 90% identity

from one another to be considered distinct viruses. Thus, two pigeon-associated sOTUs (*Columba livia*) were merged and four domestic animal-associated sOTUs (two *Ovis aries*, *Capra hircus*, and *Sus scrofa*) were merged, yielding 18 *Achaeavirus sp.* (Table 1).

Using both the novel RdRps recovered from the SRA and alignable RdRp hits in GenBank, we constructed a maximum-likelihood phylogenetic tree (Figure 4). Ao and the Ao-like Serratus hits, as well as the rodent lung virome Narnavirus, form a monophyletic clade (bootstrap support 100%). These RdRp share at least 70% amino acid identity with another member in the clade, and less than 70% identity with all members outside the clade (Supplementary Figure S2). We propose these viruses constitute a novel genus within *Narnaviridae*, which we have dubbed *Achaeavirus*. Achaeaviruses are situated within a larger Achaeavirus-proximal clade that also comprises the Matryoshka RNA viruses (Charon et al., 2019), all of which are in *Narnaviridae* with high confidence (Figure 4). The *Achaevirus* is named after the soldiers who hid in the Trojan Horse along with Odysseus.

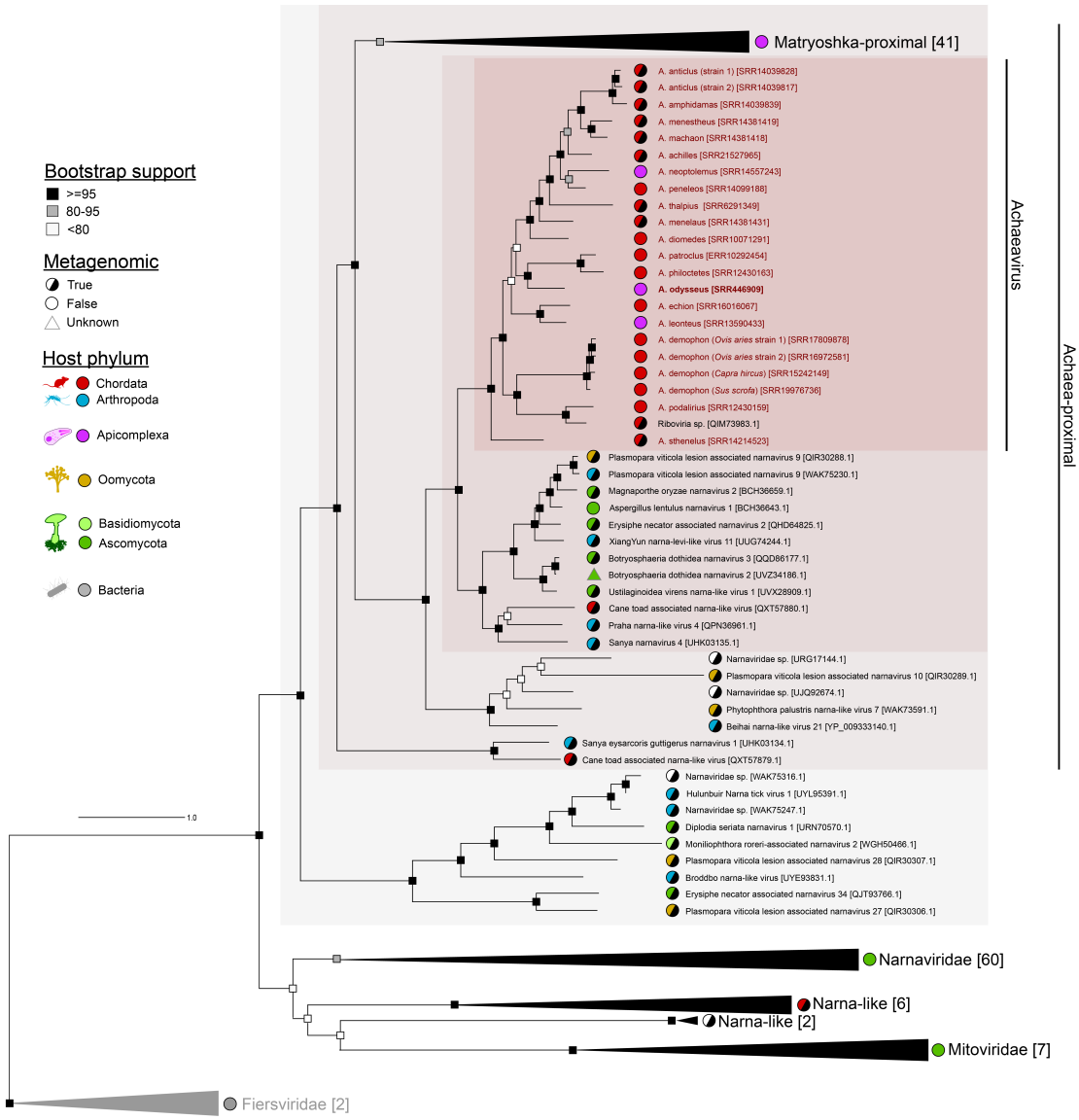


Figure 4. RdRp Phylogeny Maximum-likelihood phylogenetic tree of *A. odysseus* (Ao) and related Achaeaviruses. Phylogeny is estimated based on RdRp palm (motifs F-E) and thumb subdomains. The tree is outgroup-rooted on two members of the family *Fiersviridae*: bacteriophages Qbeta [P14647.2] and MS2 [P00585.3]. The scale bar represents 1.0 amino acid substitutions per site and high-confidence bootstraps are demarcated with black squares. Operational taxonomic units are annotated by their predicted hosts based on available GenBank/BioSample/SRA metadata. For collapsed nodes, the number of clade members is denoted in square brackets.

Table 1. Achaeavirus run data

Achaeavirus	Abbrv.	SRA_run	Source	Host	Parasite_(STAT)	Geography	Metagenome	Contig1	Contig2	Reference
A. anticlus 1	Aanti 1	SRR14039828	feces	Columba livia	-	China: Zhejiang	T	3158	-	(Chen et al., 2022)
A. anticlus 2	Aanti 2	SRR14039817	feces	Columba livia	Cryptosporidium parvum	China: Zhejiang	T	4448	-	(Chen et al., 2022)
A. amphidamas	Aampd	SRR14039839	feces	Columba livia	Cryptosporidium parvum	China: Zhejiang	T	3133	-	(Chen et al., 2022)
A. menestheus	Amens	SRR14381419	feces	Myotis muricola	-	China:Yunnan	T	3180	1291	(Zhou et al., 2021)
A. machaon	Amacha	SRR14381418	feces	Rhinolopus malayanus	-	China:Yunnan	T	3147	1249	(Zhou et al., 2021)
A. achilles	Aachi	SRR21527965	feces	Branta canadensis	Eimeria sp.; Babesia sp.; Blastocystis sp.	USA: Connecticut	T	3193	1193	(Gil and Hird, 2022)
A. neoptolemus	Aneop	SRR14557243	feces	Gallus gallus	Eimeria sp.	China: Yangzhou	F	3267	-	(Gao et al., 2021)
A. peneleos	Apene	SRR14099188	liver	Typhlomys cinereus	-	China: Fujian, Wuyishan,Guadun	F	2743	1144	(He et al., 2021)
A. thalpius	Athal	SRR6291349	gut, liver, gill	Okamejei acutispina	-	China: Eastern China Sea	T	3175	1116	(Chen et al., 2021b)
A. menelaus	Amene	SRR14381431	feces	Myotis laniger	Blastocystis hominis	China:Yunnan	T	3136	1126	(Zhou et al., 2021)
A. diomedes	Adiom	SRR10071291	intestine	Sus scrofa	Sarcocystidae sp.	China: Chengyang	F	3194	1037	-
A. patroclus	Apatr	ERR10292454	kidney	Acomys russatus	-	-	F	3494	2009	(Rhie et al., 2021)
A. philoctetes	Aphil	SRR12430163	kidney	Mus musculus	-	USA: Oakland, CA	F	3022	1909	(Bittner et al., 2022)
A. odysseus - COUGAR	Ao COUGAR	SRR446909	blood	Mus musculus	Toxoplasma gondii - COUGAR	Canada: British Columbia	F	3184	1193	(Melo et al., 2013)
A. echion	Aechio	SRR16016067	intestine	Oryctolagus cuniculus	Eimeria sp.	China: Sichuan	F	3202	1422	(Chen et al., 2021a)
A. leonteus	Aleon	SRR13590433	liver	Oryctolagus cuniculus	Cyclospora cayetanensis	China:Sichuan	F	3848	1341	(Xie et al., 2021)
A. demophon - Ovis 1	Ademo Ovis 1	SRR17809878	muscle	Ovis aries	Eimeria sp., Sacrocytis sp.	Tibetan Plateau	F	3257	-	(Cui et al., 2022)
A. demophon - Ovis 2	Ademo Ovis 2	SRR16972581	muscle	Ovis aries	Eimeria sp.; Babesia sp.	Tibetan Plateau	F	3228	-	(Wen et al., 2022)
A. demophon - Capra	Ademo Capra	SRR15242149	muscle	Capra hircus	Eimeria sp.; Babesia sp.	China: Hebei	F	3208	-	(Zhou et al., 2022b)
A. demophon - Sus	Admop Sus	SRR19976736	muscle	Sus scrofa	Eimeria sp.; Ichthyophthirius sp.	China: Harbin	F	3252	-	(Sun et al., 2022)
-	RtMc-NcV/Tu2016	QIM73983.1	lung	Mus caroli	-	Thailand	T	1238	-	(Wu et al., 2021)
A. podalirius	Apoda	SRR12430159	kidney	Mus musculus	Plasmodium sp.	USA: Berkeley, CA	F	4302	-	(Bittner et al., 2022)
A. sthenelus	Asthe	SRR14214523	lung	Hoplobatrachus rugulosus	-	China: Zhejiang	T	3132	-	(Chen et al., 2021b)

Treating the SRA as a sampling of nature, the prevalence of Ao and the Achaeaviruses is rare. Only 6/23,530 (0.025%) runs annotated as *T. gondii* are Ao-positive, and zero runs containing $\geq$ 1024 *Toxoplasma sp.* reads measured by STAT (Katz et al., 2021), and not annotated as *T. gondii*, are Ao-positive (0/9,071) (Figure 5A). By definition, if *T. gondii* was annotated by the authors, then the parasite is a known entity in that experiment and these runs would be biased towards lab strains. In contrast, in the absence of *Toxoplasma sp.* annotations, these are more likely to be "wild" observations of *T. gondii*. Regardless, the rate of Achaeavirus positivity is estimated to be well below 1% in Toxoplasma.

Predicting virus host range based on metagenomic data is an ongoing challenge in viromics. The overwhelming majority of *Narnaviridae* are metagenomic or from complex samples such as leaf lesions. The closest identified Narnavirus with a plausible host assignment is Aspergillus lentulus Narnavirus 1 (AleNV1, BCH36643.1), isolated from cultured *Aspergillus lentulus* mycelium (Chiba et al., 2021). The next closest is the Matryoshka RNA viruses isolated from human blood co-infected with *Plasmodium* or *Leucocytozoon* parasites (Charon et al., 2019).

Nominally, Ao was identified in libraries annotated as *H. sapiens* (Ngô et al., 2017) and *M. musculus* (Melo et al., 2013), yet through metadata and nucleic acids imputed a specific *T. gondii* strain as the probable host due to the specificity of the virus association (Ao is found in multiple replicates specific to 1/9 strains in Ngô, and 2/31 strains in Melo). Generalizing this rationale, we measured for apicomplexa reads associated with Achaeavirus-positive libraries; and 10/18 (55%) of the viruses are associated with apicomplexa in at least one dataset (Table 1). To test if Achaeavirus are significant predictors for apicomplexa nucleic acids in the same library, we compared the rate of detecting apicomplexa reads in Achaeavirus-positives datasets relative to a background rate of detecting apicomplexa reads in the remaining datasets, controlling for the animal host. The Achaeaviruses are highly predictive (via Fisher's exact test) for apicomplexa in, pig (p = 9.38E-11), chicken (p = 5.58E-17), sheep (p = 1.28E-42), goat (p = 2.38E-16), rabbit (p = 6.58E-19), and sparrow (p = 6.09E-05) (Figure 5B). Conversely, we can measure the rate of Achaeavirus-positivity (prevalence) amongst apicomplexa-positive samples in the SRA. Achaeavirus detection rate varies within apicomplexa-positive samples in, pig (6/260, 2.30%), chicken (8/211, 3.79%), sheep (33/191, 17.28%), goat (9/69, 13.04%), rabbit (11/31, 35.48%), and sparrow (5/13, 38.46%).

We cannot rule out that Achaeaviruses are infecting vertebrate cells, perhaps made more susceptible by parasitic infection, but parsimony suggests that such repeated and significant enrichment of parasitic reads in association with Achaeaviruses are caused (simply) by *Apicomplexa sp* infecting viruses, and the parasites infect vertebrates. Worth noting, the viruses A. neoptolemus, A. diomedes, and specifically A. demophon show a relatively high prevalence in parasites (mainly *Eimeria sp.*) associated with livestock, implicating the viruses in agriculturally relevant pathogens. *Eimeria sp.* are the causal agents of coccidiosis in poultry, a high mortality disease causing billions of dollars of economic loss for farmers (Blake and Tomley, 2014).

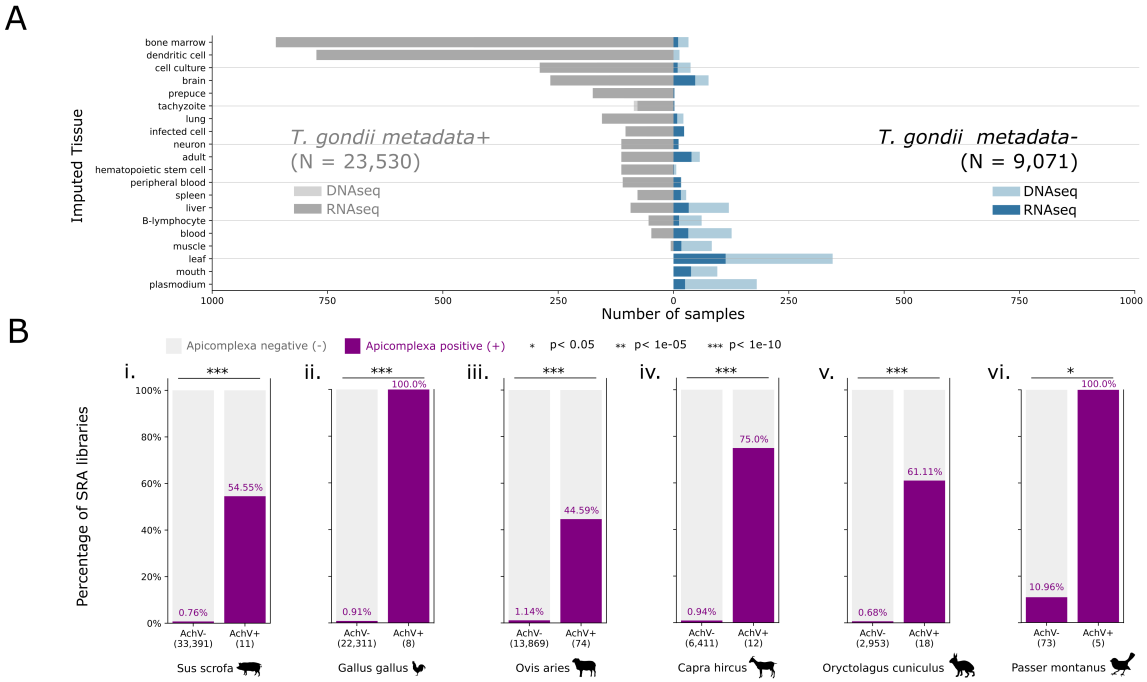


Figure 5. A. T. gondii metadata analysis reveals ≈32k sequence datasets marked *T. gondii*-positive (72% by metadata, 62% by sequence-search). "Unannotated" *T. gondii* samples allow sampling of thousands of "wild" parasites for virus co-occurrence. Samples labelled as "`TRANSCRIPTOMIC`", "`TRANSCRIPTOMIC SINGLE CELL`", "`METATRANSCRIPTOMIC`" or "`VIRAL RNA`" classified as RNA sequencing while all other values (`GENOMIC`, `GENOMIC SINGLE CELL`, `METAGENOMIC`, `OTHER`, `SYNTHETIC`) classified as DNA sequencing. Tissue values were extracted from the associated NCBI BioSample database and linked to the BRENDA Tissue Ontology (BTO). Analysis of the 20 most occurring tissues shows contrasting skews between metadata-positive and "unannotated" datasets. **B.** Analysis of host and apicomplexa co-association. For each of i. pig, ii. chicken, iii. sheep, iv. goat, v. rabbit and vi. tree sparrow, runs were categorized as Achaeavirus-negative (AchV-) or Achaeavirus-positive (AchV+) and we measured for the absence (grey) or presence (purple, $\geq$ 1024 Apicomplexa-classified reads by SRA-STAT (Katz et al., 2021)). Significance was evaluated with a Fisher's Exact Test.

Achaean-proximal viruses have conserved N' and C' terminal RdRp extensions

The efficacy of antiviral drugs is highly dependent on RdRp structure. If it is indeed the case that PPVs cause exacerbated immune response in parasitic infection, it may be worthwhile to investigate antivirals as a form of treatment. This has already been suggested in the case of TVV, wherein the canonical treatment for the parasite (metronidazole) counterintuitively worsens patient outcomes as vRNA is released from lysed cells (Narayanasamy et al., 2022; Fichorova et al., 2012). However, no previous studies have analyzed the structure of narna-like PPV RdRp in depth. We noted that the RdRp-coding contig of Achaeaviruses has long N'- and C'-extensions—220 and 224 amino acids, respectively—that are beyond core RdRp (motifs F-E). To investigate what activity these extensions may provide, we first sought to analyze sequence motifs conserved in the Achaean-proximal clade, then apply these to structure. We realigned more closely related (Achaea-proximal) RdRp, which allowed us to create a higher confidence multiple sequence alignment (MSA) of the full coding sequence that was inclusive of the extensions. We identify 12 high occupancy and high conservation regions from the MSA and designate them as Achaea-proximal motifs I through XII. These motifs had similar conservation to the RdRp core motifs against a background of "non-motifs" — high occupancy inter-motif sequences (Figure 6). The 12 novel motifs range in length from 10 amino acids to 32, and 5 of these motifs fall within the N'- and C'-extensions.

Next, we investigated how these motifs may intersect with RdRp structural evolution. Therefore, using Colabfold (Mirdita et al., 2022), we generated predicted structures for phylogenetically select RdRps and aligned these with TM-align (Zhang and Skolnick, 2005). Comparing the predicted structure to the essential catalytic motifs from the experimentally-solved Poliovirus RdRp structure, we validate that the Achaeavirus RdRp encodes a biochemically competent RdRp, on the basis of core RdRp motif conservation. Moreover, it is likely some of these motifs play a role in catalysis and may modulate RNA replication. For instance, motif III encodes an arginine at position 9 that is 100% conserved in our MSA. This residue is predicted to sit directly parallel to the highly-conserved lysine and arginine at positions 5 and 7, respectively, in motif VII. Together, the three-point directly into the catalytic site above motif C. It is likely that the three positively charged residues are interacting with the template RNA backbone, possibly acting as a stabilizing mechanism. Such considerations would need to be explored in further depth, possibly through the use of molecular docking software or wet lab experimentation, for the deployment of effective antivirals.

Based on topology, the N'- and C'-extensions appear to be unique domains distinct from core RdRp. The motifs identified in these regions are structurally conserved within the Achaean-proximal RdRps, where the N'-motifs (I-III) are more strongly

conserved than the C'-ones (XI and XII). Moreover, some of these (e.g. motif II) may be conserved across all of *Narnaviridae* and *Mitoviridae*, but it is difficult to say for certain as structural alignments may be capturing similarity, and not homology at these distances (Figure 7).

We also attempted to construct an MSA for the open readings frames from the second Achaeavirus contig. However, even within the Achaeavirus space, local HMMs were required to build a plausible alignment (see Methods in subsection 4.11), and the results demonstrated what appears to be a conserved protein for ORF1 and ORF2, albeit rapidly evolving. It should be noted that related viruses have also been shown to contain a second segment, notably the Matryoshka viruses, another PPV *Leptomonas seymouri* Narna-like virus 1 (Sukla et al., 2017), and *Aspergillus lentulus* Narnavirus 1 (Chiba et al., 2021). All of these landed in the Achaea-proximal clade on our phylogeny, yet their second segments were too far diverged to align confidently or to identify a sequence (InterProScan (Jones et al., 2014)) or structural (Foldseek (van Kempen et al., 2023)) homolog.

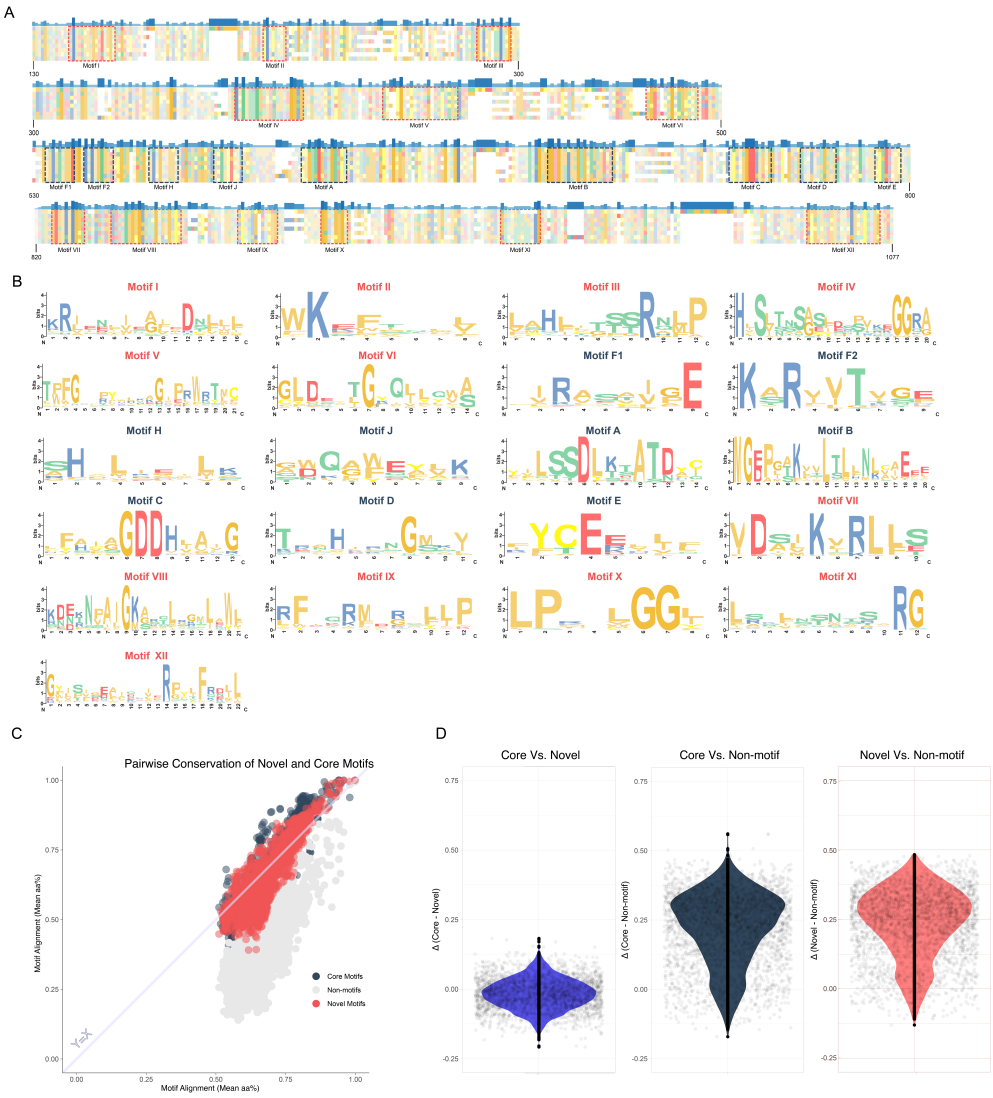


Figure 6. Novel RdRp Motif Sequence Analysis A. Conservation map of selected RdRps from "Ao-proximal" RdRp MSA. Sequences as follows (from top to bottom): Broddbo narna-like virus [UYE93831.1], Aspergillus lentulus Narnavirus 1 [BCH36645.1], Matryoshka RNA virus 3 [DAZ89878.1], Ao [SRR446909], Matryoshka RNA virus 4 [DAZ89879.1], Ao-like sp. [SRR14557243], Cane toad associated narna-like virus, [QXT57879.1], Beihai narna-like virus 21 [YP_009333140.1]. Novel and core motifs are highlighted, residues colored by chemical properties and degree of conservation. **B.** Sequence logos for novel and core motifs. **C.** Scatterplot of mean pairwise percent conservation between novel and core motifs. Pairwise percent conservation was calculated within each motif for every sequence selected. The mean conservation was calculated across the core motifs (F-E) and the novel motifs (I-XII) and plotted for each pairwise comparison. Inter-motif regions with high occupancy are also plotted in the same manner, serving as a background conservation rate. **D.** Violin plots for the difference in mean conservation for groups specified. Strongly positive/negative values indicate a large difference in conservation. Values close to zero suggest both sets of motifs are equivalently conserved.

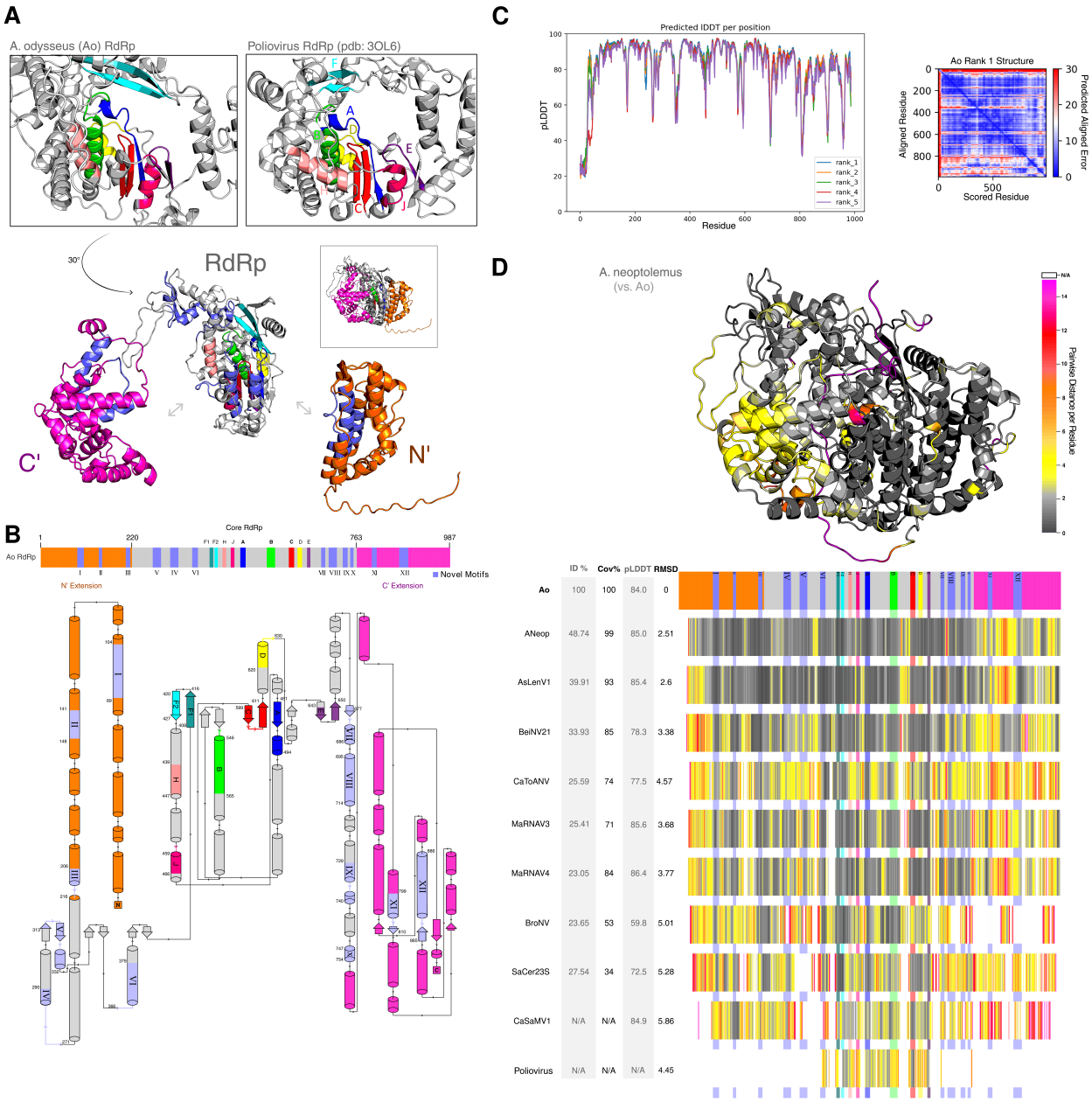


Figure 7. Structural Analysis A. ALPHAFOLD predicted structures of Ao RdRp and X-ray diffraction of poliovirus RdRp (PDB accession 3OL6), at their respective catalytic cores, with motifs F-E highlighted. Ao RdRp is shown in an exploded view, with the core rotated 30 degrees such that motif C is facing the view head-on, with distinct N' and C' terminal extensions separate from the catalytic core RdRp. **B.** Secondary structure topology map for Ao core RdRp and extension domains with canonical and novel motifs highlighted. **C.** Plots for the predicted local difference distance test (pLDDT), a per-residue estimate of accuracy, for the five predicted structures and predicted alignment error for the rank one model. **D.** Distance maps for pairwise alignments of select RdRp structures against Ao. The structure for *A. Neoptolemus* is shown, with each residue colored by its distance to the corresponding residue in the structural alignment against Ao RdRp. Residues that did not align are shown in white.

Is A. odysseus a T. gondii hypervirulence factor?

T. gondii RUB and COUGAR are atypically hypervirulent in humans. RUB was isolated from an immunocompetent soldier in French Guyana (1991) presenting with fever, myalgia, and leukopenia which developed into rales, respiratory failure, and renal deterioration (Dardé et al., 1998). While the COUGAR strain was isolated from a mountain lion (Dubey et al., 2008), it is believed to be identical to the strain which infected 3000-7000 people in a water-borne 1994/1995 toxoplasmosis outbreak in Victoria, BC (Bowie et al., 1997). Notably, the incidence of ocular inflammation (retinitis) toxoplasmosis amongst these immunocompetent patients was high in the Victoria outbreak (Bowie et al., 1997). The Melo *et al.* publication notes that exactly these two "atypical" *T. gondii* strains are outliers by virtue of their capacity to induce a type I interferon inflammatory response in murine cells (see

also: Figure 3C). Noting this immunological response, we sought to re-analyse the Ngô *et al.* human dataset, hypothesizing that the presence of Ao in both RUB and COUGAR provides a plausible mechanism for an immune-mediated hypervirulence of *T. gondii*.

Using the Ngô *et al.* data, we performed differential gene expression analysis (DGE) and gene set enrichment analysis (GSEA) to compare the Ngô study (human) on the immune response to the described Melo (murine macrophage) immune response. The Ngô *et al. T.gondii* - RUB infection RNAseq data was available in two cell types, macrophage (n = 2) and neuronal stem cells (NSCs, n =2). Initially, we quantified *IFNB1* induction in neuronal (Figure 8) and macrophage cell types (Supplementary Figure S3). We noted a large variation in *IFNB1* gene expression, especially across the macrophage datasets including a second set of mock controls, which could be explained by a sequencing batch-effect in the data (Supplementary Figure S3, Supplementary Table S1). Samples segregated by batch in a principal component analysis (Supplementary Figure S4A) are indicative of global profile differences, thus we limit differential expression analysis to intra-batch comparisons.

Amongst the comparable *T. gondii*-infected neuronal samples, the RUB strain (n = 2) had the highest *IFNB1* expression (26.1 fold increase vs. Mock; n = 2, p = 0.029) relative to TgBRDKH (n = 2), GT1 (n = 2), ME49 (n = 2), RAY (n = 2) or VEG (n = 2) (Figure 8B), in agreement with *IFNB1* expression in the Melo study (see also Figure 3C). Next, we sought to contextualize the type of *IFNB1* response, asking which of the five gene sets containing *IFNB1*, if any, are differentially regulated in RUB versus mock and the rank of this difference amongst other *T. gondii* strains. While both RUB and RAY strains show upregulation of the gene signature for "Interferon mediated signalling pathway" (p-value 0.009 and 0.049) and "Receptor signalling via STAT", (p-value 0.006 and 0.039) a known downstream effector of interferon, the RUB strain was specifically enriched for "cellular response to virus" (p-value 0.019), supporting the existence of specific host biological viral-response against Ao.

To characterize the broader host response of *T. gondii* RUB, we performed GSEA using the well-defined Hallmark gene set (Liberzon et al., 2015). As expected, all *T. gondii* infections induce an "Inflammatory Response" with the involvement of the "TNFA signalling via NFkB" pathway relative to mock controls. Yet the magnitude of these responses in neurons is the highest in RUB, even when compared to other (non-viral) hypervirulent strains such as RAY or VEG (Figure 8D). In addition, RUB was the only *T. gondii* strain inducing "Apoptosis", "Glycolysis", "Peroxisome", and the "PI3K AKT MTOR Signalling" axis, which altogether is consistent with the Melo study conclusions that *T. gondii* RUB is exceptionally pro-inflammatory, even amongst strains of *T. gondii*.

A similar inflammatory response trend is observed in the *T. gondii* RUB macrophage-infection experiments, but the response is less specific/exceptional (Supplementary Figure S3). A higher overall level of experimental variation in the macrophages (n = 2 each) is evident when comparing the MA plots across cell types for statistically differentially regulated genes (compare NSC Figure 8A and macrophages Supplementary Figure S3A).

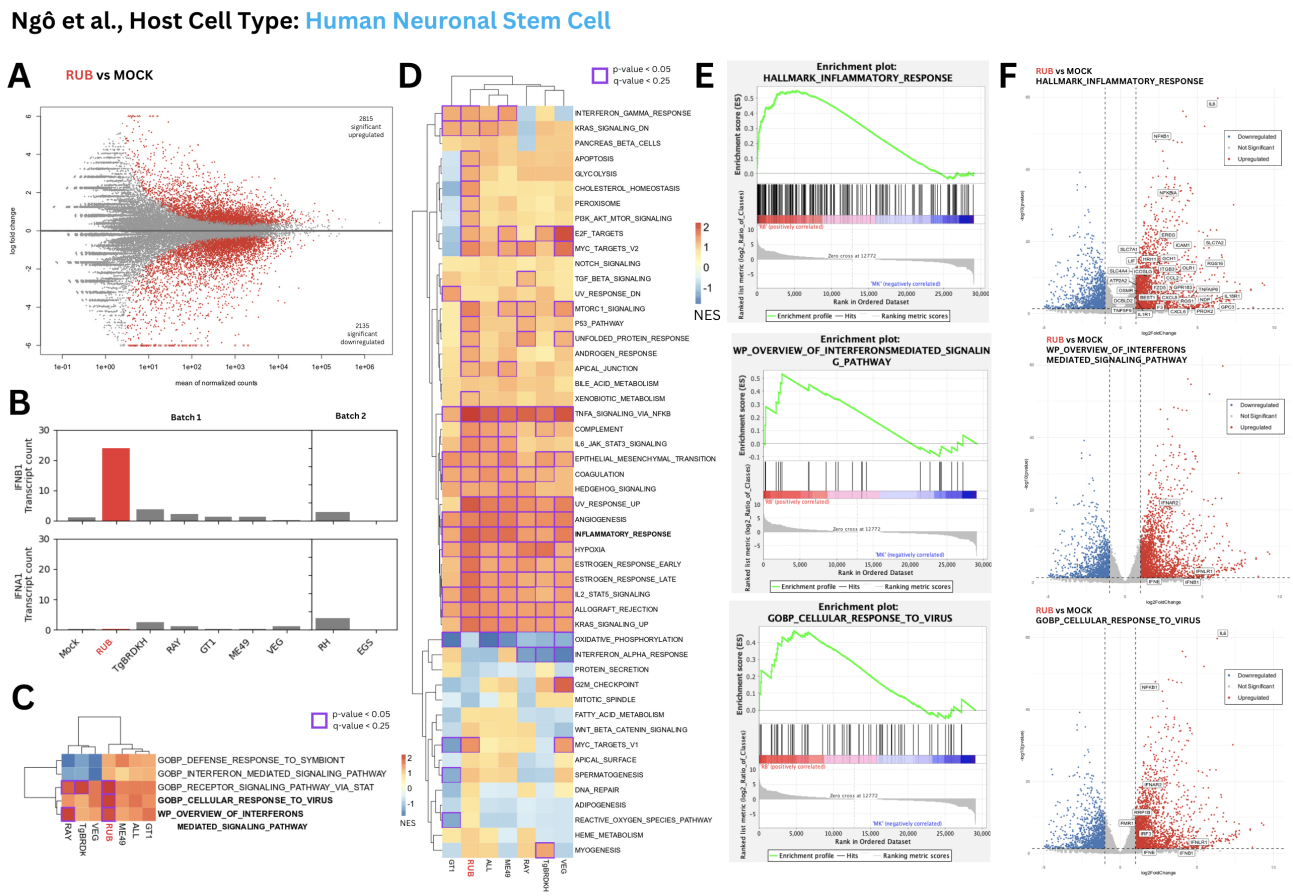


Figure 8. Differential Gene Expression (DGE) of human neuronal stem cells infected with various *T. gondii* strains. **A.** MA plot of *T. gondii* - RUB vs. Mock genes (highlighted: Benjamini-Hochberg adjusted p-value < 0.05). **B.** Bar plot of normalized transcription counts of *IFNB1* and *IFNA1* across *T. gondii* strains and Mock sequence in Ngô *et al.* experiments, separated by batch. **C.** Heat map of Normalized Enrichment Scores (NES) from Gene Set Enrichment Analysis (GSEA) using gene sets possessing interferon-specific genes, namely *IFNA1* and *IFNB1*, applied to the *T. gondii* strains. **D.** Heat map of NES values from GSEA using the Hallmark gene sets. **E.** GSEA curves comparing RUB vs. Mock strain using gene sets of inflammatory response, cellular response to virus, and interferon-mediated signalling pathway. **F.** Volcano plot of differentially regulated genes with significant members of notable gene sets being labelled.

A major limitation of these *in silico* analyses is that we cannot establish a causal relationship between Ao and the host inflammatory responses. Notably, *T. gondii* RUB genetic factors are likely to confound this analysis. Yet we are able to recapitulate a statistically significant association of the Ao+ RUB strain with a robust inflammatory response in human cells. In the Melo *et al.* study, this inflammatory response was experimentally demonstrated to be nucleic-acid mediated which, taken together with the discovery of Ao, allows us to propose a virus-mediated hypervirulence model (Figure 9).

Discussion

We discover a novel Narnavirus, *Achaeavirus odysseus*, which tightly associates with two distinct yet hypervirulent strains of *T. gondii* - RUB and COUGAR. *A. odysseus* is a member of a broadly distributed clade (putatively at the genus level) of Narnaviruses, the *Achaeaviruses*. We also describe 18 additional *Achaeaviruses*, which are likely to infect apicomplexans which in turn infect chordates. Further, we provide initial (*in silico*) evidence to assert that *A. odysseus*, upon entering cells inside *T. gondii*, may act upon the innate immune system of human and mouse cells and therefore may be a hitherto uncharacterized hypervirulence factor for this ubiquitous parasite.

While PPVs have been described for decades, their entanglement in the parasite-host relationship and pathogenesis is more recent (Gómez-Arreaza et al., 2017; Zhao et al., 2023). Multiple causal studies establish a molecular mechanism by which parasite-infecting viruses trigger chordate innate immune responses. Examples of this menage à trois include: Trichomonas vaginalis virus (TVV) in *Trichomonas vaginalis* (Fichorova et al., 2012), Leishmania RNA virus 1 (LRV1) in *Leishmania Viannia sp.* (Ives et al., 2011; Eren et al., 2016), Giardia lamblia virus (GLV) in *Giardia lamblia* (Pu et al., 2021) and, Cryptosporidium parvum virus 1 (CSpV1) in *Cryptosporidia sp.* (Deng et al., 2023). Unlike *A. odysseus*, the other PPVs are double-stranded RNA viruses, and

a narnavirus contributing to disease in vertebrates has not yet been characterized. A recurrent theme amongst these studies is enhanced host pro-inflammatory cytokine production, specifically interferon, which worsens parasite virulence. While it has been suggested that the PPV may promote parasite fitness via host immune modulation, such as recently suggested in *Cryptosporidium* (Deng et al., 2023), we do not expect this to hold true across all parasite strains, species and hosts; the virus-parasite-host fitness relationships are by definition complex.

Given the recurrent examples of PPV-mediated pathogenesis, it is parsimonious to suggest that *A. odysseus* is a sufficient trigger of host IFN-driven immune responses (Figure 9). Melo *et al.* found that induction of type I IFN in the COUGAR strain is dependent upon the RNA-sensor RIG-I, and the signal cascade proceeds through MyD88/TRIF and ultimately IRF3, an inducer of type I IFN (Melo et al., 2013). IRF3 phosphorylation and IFN-induction are also observed in other PPV infections, albeit with a double-stranded RNA ligand (Ives et al., 2011; Fichorova et al., 2012; Eren et al., 2016; Deng et al., 2023). Not knowing of the existence of *A. odysseus* within their experiment, the authors ascribe COUGAR/RUB hypervirulence as mediated by parasite genomic DNA/RNA, which is plausible but does not explain the specificity of response in these two unrelated strains and absence from the remaining 29 other (virus-negative) strains that also possess parasite nucleic acids (Melo et al., 2013).

We do note differences between existing pathways previously elucidated for innate response to PPVs and a potential contribution of *A. odysseus* to pathogenicity. The other cases involve dsRNA viruses via TLR3/IRF3, whereas a ssRNA Narnavirus genome would likely trigger TLR7,9 (ssRNA ligand) or RIG-like receptor (uncapped RNA), and downstream of this, some dependence on *IRF7* (and, had it been investigated, *IRF5*) (Jensen and Thomsen, 2012; Schoenemeyer et al., 2005). Melo *et al.* however find that in murine macrophages infected by COUGAR strain tachyzoites, *IRF7[KO]* in fact increases *IFNB1* expression (Melo et al., 2013). It is anticipated that differences between their experimental set-up and the pathophysiological scenario, particularly in humans, could explain this discrepancy. First, the constitutive knockout of IRF genes and subsequent long-term depression of innate immune pathways are likely to perturb macrophage biology. Second, the presence of *TLR11* and *TLR12* in murine cells which detects *Toxoplasma*'s profilin and signals downstream via *IRF8*, means that murine macrophages have an additional pathway (that humans lack) for innate sensing of *Toxoplasma*-associated antigens (Raetz et al., 2013). This will affect immune cell responses at baseline and the way in which they compensate for IRF knockouts. Others have observed that in the presence of functional *TLR11/TLR12*, a *TLR7/TLR8* KO has no significant effect on the production of IL-12 in response to ME49 tachyzoite infection, while the knock-out of either *TLR3* or *TLR11* (or both) impairs the response (Andrade et al., 2013). Together, these differences may explain the unexpected finding of Melo *et al.* in the context of our hypothesis of *A. odysseus*-mediated hypervirulence.

While *T. gondii* pathogenicity is certainly multi-variate, with variance in disease partially explained by known components of host immune status, age, and parasite genetic variation, there is still a lack of reliable prognostic indicators. A viral hypervirulence factor can explain, at least in part, some of the stochastic nature of parasitic disease burden in humans and other animals. Given the enormous infectious burden of *T. gondii* and other parasites (impacting multiple billions of people), it is prudent to improve monitoring for viruses within the niche of the human biome, as it interlinks human health with the health of microbial flora. In general, if the presence of PPV is prognostic (as they are in TTV/CSpV1), then determining Ao and the Achaeavirus status can impact treatment and outbreak-management decisions to the burden of toxoplasmosis.

There is good evidence to suggest that the type I IFN response plays an important role in the control of *Toxoplasma*, even if debate continues around the direction of this response (i.e. whether type I IFN limits or supports infection). Knockout of *IFNAR* appears protective and supplementation of type I IFN detrimental after intraperitoneal injection of PRU strain *Toxoplasma* in mice (Hu et al., 2022). Similarly, an intestinal epithelial cell-specific knockout of *IFNAR* is protective in mice after oral infection of *Cryptosporidium* (another apicomplexan pathogen) carrying CSpV1 (Deng et al., 2023). In *Leishmania donovani* infection, $IFNAR^{-/-}$ mice have significantly lower parasite burden compared to wild-type (Kumar et al., 2020). The authors propose that type I IFNs stimulate dendritic cells to attenuate anti-parasitic defence by affecting the balance between Th1 helper and Treg1 regulatory cell subsets which impairs the type II IFN response; they also demonstrate that ruxolitinib, the monoclonal against JAK kinases involved in interferon transduction, exerts anti-parasitic effects that are not observed in $IFNAR^{-/-}$ mice, suggesting a dominant effect via inhibition of type I IFN signalling (Kumar et al., 2020).

On the other hand, others have found that after oral infection of ME49 cysts, $IFNAR^{-/-}$ mice had significantly poorer survival rates and increased numbers of brain cysts (Matta et al., 2019). Differences between strains, route of infection, and multiplicity of infection, with effects on the timing, duration, and magnitude of the type I IFN signal, could explain these contrasting findings (Silva-Barrios and Stäger, 2017; Lee and Ashkar, 2018). Regardless of the eventual outcomes on parasite burden and host survival, we propose that the additional type I IFN response, brought on by the presence of vRNA in Ao+ strains, will at least acutely contribute to more severe disease. PPV-mediated inflammation has also initiated the investigation of the efficacy of antivirals Indinavir and Ribavirin, individually or in combination with the anti-parasitic paromomycin against CSpV1+ *Cryptosporidium* in a 2D monolayer model of human enteric epithelia (Deng et al., 2023).

Ultimately, these proposed mechanisms and models require experimental validation, for which we outline an efficient set of critical experiments. First, necessity and sufficiency of *A. odysseus*-mediated IFN-production/pathogenicity needs to be established. For necessity, Ao-infection of a naive *T. gondii* strain (ME49) will promote *IFNB1* expression relative to an uninfected control. Curing of COUGAR/RUB *T. gondii* of the virus will cause a loss of IFN production/pathogenicity (sufficiency), and this effect can be rescued by subsequent re-infection of the cured strain. The gain and loss of the *IFNb* will be dependent on ssRNA receptors TLR7,9 or RLRs, and not TLR3 or TLR11,12. This minimal set of experiments would establish a causal relationship between Ao

and proinflammatory induction, while the same set of experiments in an animal model can formalize the relationship between Ao and parasite pathogenicity.

Hypervirulence in *Toxoplasma* can present with different symptoms. Broadly, however, we note that in humans, both infections with COUGAR or RUB are associated with ocular involvement and overt acute pathology even in immunocompetent patients, respectively (Bowie et al., 1997; Dubey, 2021b; Dardé et al., 1998). In mice, both strains are associated with increased mortality (Khan et al., 2009; Niedelman et al., 2012; Hassan et al., 2019). COUGAR is associated with enhanced oral infectivity when compared to other strains, although whether RUB shares this phenotype is unknown (Su et al., 2003). A recent report of COUGAR strain infection in four southern sea otters (Enhydra lutris nereis) observed severe myocarditis and marked subcutaneous and peritoneal steatitis (Miller et al., 2023). Type I IFN signalling is thought to play an important role in driving adipocyte inflammation (Chan et al., 2020), so steatitis could present an unusual consequence of the vRNA contribution to toxoplasmosis in strains infected by Ao.

A proposed reason for type I IFN production in response to COUGAR and not PRU strain tachyzoites is enhanced killing, which releases greater quantities of *Toxoplasma* nucleic acids (Melo et al., 2013). While this is not necessarily inconsistent with a contribution of vRNA to the innate response, we find that in the absence of similar enhanced killing data for the RUB strain, the presence of Ao in both strains remains the single most persuasive explanation for the differential host immune response. Furthermore, infection of a fibroblast, during which the *Toxoplasma* mediates its own entry into a parasitophorous vacuole, is a fundamentally different process to phagocytosis by a macrophage, which is followed by phagolysosome maturation. Even if PRU strain tachyzoites are better able to proliferate in fibroblasts, they are unlikely to be able to survive macrophage phagocytosis in the same way. Therefore, regardless of strain, we expect a generic exposure of macrophage NASs to toxoplasma nucleic acid common in pathogenic and non-pathogenic strains – the key distinguishing factor in COUGAR and RUB is the viral contribution.

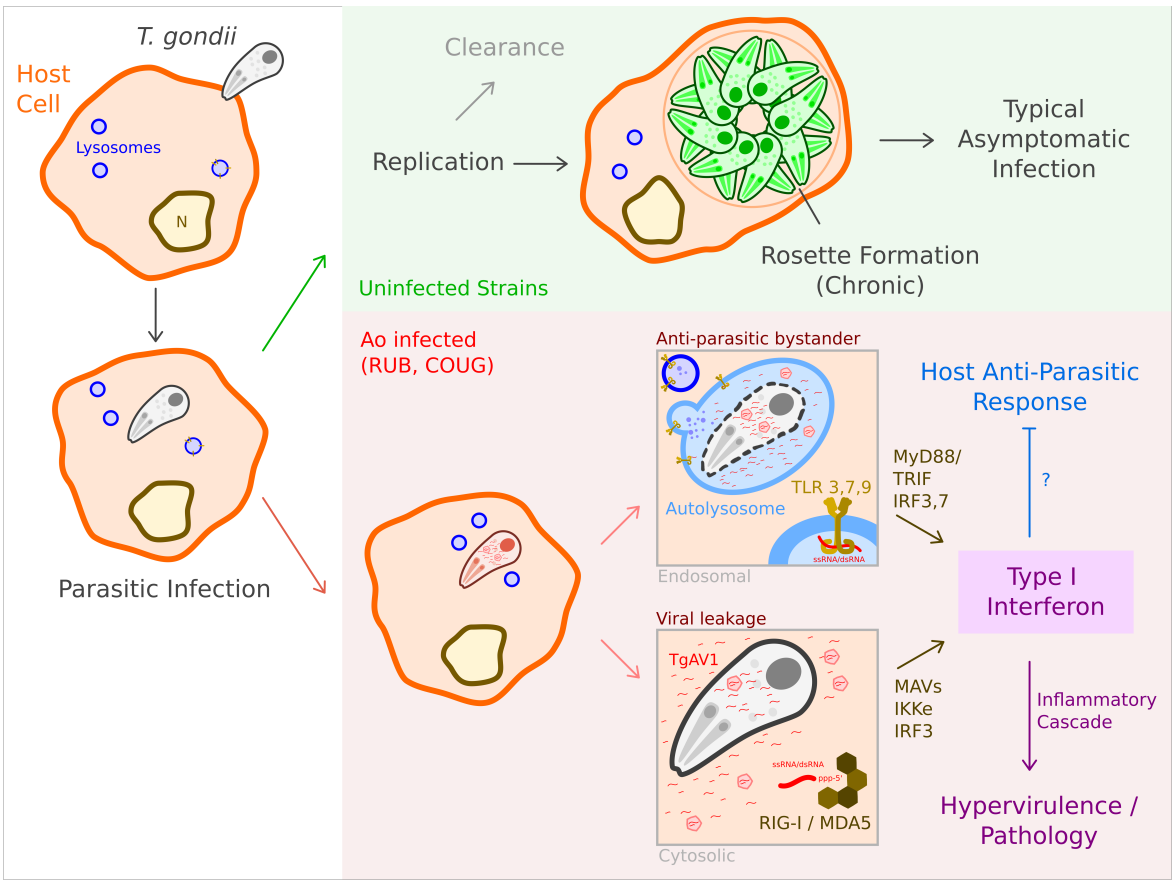


Figure 9. Hypervirulence model. *T. gondii* infects host cells. In uninfected strains, this either leads to successful clearance by the host or the establishment of chronic infection, typically asymptomatic. In virus+ strains, such as RUB and COUGAR, detection of viral RNA (vRNA) - either by endosomal TLRs or by RLRs surveying the cytosol - leads to the induction of type I IFN. The effect of this is twofold - the viral pathogen-assoicated molecular pattern drives additional inflammation which leads to pathology and overt symptoms, but potentially also impairs or diverts host anti-parasitic immunity.

The objective of this study was to rapidly screen for a candidate highly divergent neuroinflammatory virus *in silico*. Here we discovered *Achaeavirus odysseus*, a narnavirus likely infecting *T. gondii* which through the innate immune sensing of vRNA, could drive a type I IFN-mediated inflammatory response. While we cannot establish this mechanism causally, it is plausible and supported by the body of available mechanistic and gene-response evidence. A link between neurotropic toxoplasmosis and neuroinflammatory and degenerative conditions, such as Alzheimer's disease has been proposed (Sanford and McEwan, 2022; Jin et al., 2022). In AD, CNS inflammation is balanced between beneficial (e.g., priming microglia for amyloid plaque clearance) and detrimental (e.g., synaptic loss, immune infiltration) function. An *Achaeavirus*, or similar PPV, could skew this equilibrium in chronic toxoplasmosis towards neuro-pathogenicity.

The strength of computational virology is in its capacity for searching massive swathes of data and identifying novel and unexpected relationships between viruses and diseases. Ao and neuroinflammation is such a case study and should serve to demonstrate the need for and merit of the comprehensive characterization of Earth's RNA virome.

Materials and methods

Querying the BioSample database

In the initial screen for novel and highly divergent neuro-tropic viruses, we searched the BioSamples SQL table in the Serratus SQL database (`https://github.com/ababaian/serratus/wiki/SQL-Schema`) with the query `SELECT * FROM tismap_summary WHERE biosample_tags LIKE '%neuron%' ORDER BY coverage desc;` and identified SRA run `SRR1205923` in BioProject `PRJNA290363` through manual curation. BLASTp and BLASTx were performed with a query of the `u150420` palmprint identified in three separate libraries to measure the percent amino acid identity to known viral proteins using the non-redundant proteins database (nr). Furthermore, BLASTn against the nr/nt database was performed.

Genomic analyses

The three libraries (`SRR1205923`, `SRR1204654`, `SRR1205190`) identified through a Serratus SQL search were *de novo* assembled using rnaSPAdes (Bushmanova et al., 2019). rnaSPAdes was run with parameters '`--rna --s1 -t 64`'. Palmscan (Babaian and Edgar, 2022)—an RNA-dependent RNA Polymerase detector—was used to identify the viral RdRp contig in the assembled transcript file. Parameters used were '`--threads 64 --seqtype nt`'. For each library, the contigs identified by Palmscan were put through ORFfinder and the longest Open Reading Frame (ORF) was extracted. This amino acid sequence was identical across both mRNA libraries `SRR1205923` and `SRR1204654` (`SRR1205923` had a sequencing error, so manually fixing the frameshift deletion was performed resulting in the identical amino acid sequence). The ORF from library `SRR1205923` was subsequently put through Colabfold (Mirdita et al., 2022) and a prediction of the RdRp structure was generated. BWA-sw (Li and Durbin, 2010) was used for long-read alignments of `SRR1205923` transcripts generated by the assembler back to the reads upon which they were assembled. The resulting SAM file was converted to a BAM file and indexed with SAMtools (Danecek et al., 2021), and the alignment view was exported from IGV (Robinson et al., 2011). The entire Open Reading Frame (ORF) from the `SRR1205923` RdRp contigs were queried against the non-redundant proteins database (nr) using BLASTp. RdRp proteins from the top 50 results were used to create an MSA using MUSCLE (Edgar, 2004). To create the phylogeny of these related viral RdRps, IQ-TREE (Minh et al., 2020) was used. The phylogeny of these related viruses was viewed in Dendroscope (Huson et al., 2007).

Endogenous virus check

The reference *T. gondii* RUB and reference *T. gondii* COUGAR DNA-seq data was downloaded from their respective BioProjects (`PRJNA61119` and `PRJNA71479`). Bowtie2 '`-build`' (Langmead and Salzberg, 2012) created a dataset with both Ao contig 1 and contig 2. Afterwards, Bowtie2 with parameters '`--local --very-sensitive-local --threads 64 -q -x -U`' was used to align the DNA-seq data to the FASTQ files for reference *T. gondii* RUB and reference *T. gondii* COUGAR. SAMtools was then used to convert the SAM file to BAM and the SAMtools `flagstat` command was used to see the number of primary mapped reads in the DNA-seq data.

Depletion experiment (finding contig 2)

A copy of the reference human genome: `GRCh38` was downloaded from NCBI (`https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/` and the reference *Toxoplasma gondii* Me49 genome was acquired from ToxoDB (Gajria et al., 2008). The Human genome and the *T. gondii* genome was combined into one file and Bowtie2 '`-build`' was used to create a 'background' dataset. Bowtie2 with parameters '`--local --very-sensitive-local --threads 64 -f -x -U`' was used to align the 4 RUB mRNA Ngô libraries (`SRR1205923`, `SRR1204654`, `SRR1204653`, `SRR1204652`) to the background file. SAMtools view with parameters '`-S -b -@64`' converted the SAM to a BAM file and another SAMtools view command with parameters '`-f 4 -b -@ 64`' removed all the mapped reads and kept only the unmapped reads. All 4 unmapped mRNA RUB libraries were aligned against NCBI's nr proteins database using DIAMOND (Buchfink et al., 2015). BLASTx was run with parameters '`--masking 0 --unal 1 --mid-sensitive -l 1 -p14 -k1 --threads 64 -f 6 qseqid qstart qend qlen qstrand sseqid sstart send slen pident evalue full_qseq staxids sscinames`'. The contig with the highest coverage was classified as E. coli by DIAMOND, the 3rd highest by coverage was the Ao RdRp contig and the 2nd highest across all 4 RUB libraries was a $\approx$ 1280 nucleotide long contig with no known associations through BLAST. This unknown contig was common throughout all 4 RUB libraries and was called the putative contig 2. We validate contig 2 is viral in origin when a co-evolving homolog was identified

in 12 related *Achaeavirus sp.* Finally, with both contig 1 and contig 2 a genome map representation was created for both contigs with gggenes (Wilkins, 2020), an R package for drawing gene arrow maps and GGPLOT2 (Wickham, 2016).

Assembling and quantifying of Ngô et al. Libraries

117 out of the 237 libraries in the Ngô et al. BioProject are labelled as mRNA-seq data, the other 120 libraries are micro RNA (miRNA). The 117 mRNAseq libraries were pre-fetched and the FASTQ files were extracted using SRATOOLKIT. Next, all 117 Ngô mRNA libraries were *de novo* assembled with RNASPADES with parameters '`--rna --s1 -t 64`'. Using BOWTIE2 '`-build`' a dataset was created with all 4 genomes/sequences: Human (HgCh38), *T. gondii* Me49 (Tg64), Ao contig 1 RdRp, Ao contig 2. Next, the reads for all 237 libraries in Ngô *et al.* were aligned against this dataset using BOWTIE2 with parameters '`--local --very-sensitive-local --threads 64 -q -x -U`'. To convert the SAM file to a BAM file, SAMTOOLS with parameters '`-S -b -@ 64 -F 260`' were used. Flag `-F 260` removes all unmapped reads and non-primary alignments. The output of SAMTOOLS was then piped into SEQKIT (Shen et al., 2016) with the '`-c`' parameter which counts the mapped reads per chromosome/sequence. In Figure 2, all the metadata for the entire Ngô *et al.* BioProject (`PRJNA241125`) was collected with Google's BIGQUERY tool. Using this metadata and the output of SEQKIT, Figure 2 was generated in R with DPLYR and GGPLOT2.

Assembling and quantifying Melo **et al.** Libraries

Similarly to subsection 4.5, for the BioProject from Melo *et al.*, the FASTQ files for all 32 libraries were downloaded using SRATOOLKIT. All 32 Melo libraries were de novo assembled with RNASPADES parameters '`--rna --pe -t 64`'. BOWTIE2 was used to create a dataset with all 4 genomes/sequences: Murine (GRCm39), Tg64, Ao contig 1 RdRp, Ao contig 2. Next, the reads for all 32 libraries in Melo *et al.* were aligned against this dataset using BOWTIE2. To convert the SAM file to a BAM file, SAMTOOLS was used (see subsection 4.5 for parameters used). All the metadata for the entire Melo *et al.* BioProject (`PRJNA241125`) was collected with Google's BIGQUERY tool. Using this metadata and the data parsed from PANDAS, Figure 3 was generated in R with DPLYR and GGPLOT2.

Screening for and assembling the Achaeaviruses

To identify RdRp related to Ao, we aligned the sOTU in palmDB (version 2023-04) Babaian and Edgar (2022) to the Ao palmprint (`usearch -usearch_global palmDB.palmprint.faa -db ao.palmprint.faa -id 0.3`), yielding matches in '`u380516`', '`u476932`', '`u706419`', '`u492272`', '`u602981`', '`u1051092`', '`u845773`', '`u857620`', '`u665910`', '`u819619`', '`u584295`', '`u691934`', '`u71279`', '`u145522`', '`u626963`', '`u150420`', '`u592253`', '`u828652`', '`u617666`', '`u964003`', '`u643849`', '`u993146`', '`u770301`', '`u533578`', '`u460145`', '`u419915`', '`u849389`', '`u1004674`', '`u942391`', '`u10732`', '`u761722`', '`u599206`' and '`u1009116`'. To query which SERRATUS-processed SRA runs contained microcontigs matching these palmprints, we performed the SQL query '`SELECT * FROM palm_sra2 WHERE qc_pass = 'true' AND sotu IN ('u380516', 'u476932', 'u706419', 'u492272', 'u602981', 'u1051092', 'u845773', 'u857620', 'u665910', 'u819619', 'u584295', 'u691934', 'u71279', 'u145522', 'u626963', 'u150420', 'u592253', 'u828652', 'u617666', 'u964003', 'u643849', 'u993146', 'u770301', 'u533578', 'u460145', 'u419915', 'u849389', 'u1004674', 'u942391', 'u10732', 'u761722', 'u599206', 'u1009116')`', yielding the 185 matching libraries (Supplementary Table S3). Sorting by microcontig coverage for each palmprint-library, FASTQ files from the top three libraries were downloaded using SRATOOLKIT and assembled using RNASPADES with parameters '`--rna -t 64 --pe`'. RdRp-contigs were identified as described above in each library.

MSA and phylogenetic tree of Ao and related Narnaviruses

In order to sample Ao-related viruses, we first searched for available sequences in GenBank's non-redundant protein database, using PSI-BLAST and default parameters with the `BLOSUM62` substitution matrix. We queried the database with the entire amino acid sequence for the Ao RdRp recovered from the RUB strain of *T. gondii*. PSI-BLAST iterated until 500 protein sequences were found (2 iterations); all sequences had E-values $< 7e-28$. Notably, this approach failed to capture more distantly related Narnaviruses, such as those previously described in arthropod metagenomic samples, yeast, and nematodes. Thus, we supplemented the Achaea-proximal viruses with an additional BLASTP search on *Coquillettidia venezuelensis* Narnavirus 1 (`QBA55488.1`). For a more distant clade, we selected 7 ICTV-recognized *Mitoviridae* species, two from each genus except *Kvaramitovirus*, for which there is only one accepted species. As an outgroup, we added two species from the family *Fiersviridae*: bacteriophages MS2 and Qbeta. The sequence with the highest percent identity relative to Ao RUB among these was `QIM73983.1` at 53.53% identity, but only 40% coverage. For this reason, after aligning the sequence with MUSCLE v3.8.1551, we decided to manually trim the MSA to the RdRp palm (motifs F-E) and thumb subdomains. This ensured partial reads could be fairly represented in the phylogeny. Some BLAST hits, however, were too incomplete to satisfy this requirement and were therefore removed.

With the closest relative in GenBank at such a low identity, we re-queried the SRA for Ao-like sequences. Using DIAMOND a dataset of all Ao-like palmprint sOTUs was created. This dataset was then aligned against the assembled transcripts for each Ao-like library with BLASTX using parameters '`--very-sensitive --threads 64 -f 6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalue bitscore full_qseq`'. The output of this alignment was sorted and cleaned up by PANDAS (pandas development team, 2020). With PANDAS, each TSV file was combined into one large file. Manually each transcript outputted from DIAMOND was evaluated and only contigs with 100% identity to the palmprint, high coverage and ≈ 3000 nt length were chosen to be the representative contig for that virus. 8 of the 55 libraries did not have a quality representative contig, so they were discarded. The remaining libraries each had a representative contig and were all put through the ORFFINDER command line

to find the ORFs. The longest ORF was chosen and was confirmed to be the RdRp using BLASTp. Those ORFs were combined into a FASTA file creating a dataset of RdRps of all Ao-like viruses. The final MSA contains 169 RdRps, a full list of the selected sequences, their GenBank/SRA accessions, and associated metadata is available in the supplementary materials. Based on the RdRp palm, we generated a maximum-likelihood phylogenetic tree using IQ-TREE v2.2.2.7, where the substitution matrix was selected as VT+F+R8 by ModelFinder (Kalyaanamoorthy et al., 2017). Bootstrap values on 1000 replicates were generated via UFBoot (Hoang et al., 2018).

Novel Ao-proximal RdRp Motif Identification

Based on the phylogeny, sequences displaying high relatedness with Ao were selected. The full sequences (as opposed to just the palm and thumb) were retrieved and aligned with MUSCLE. A conservation map showing alignments for selected sequences in the MSA was generated using ggmsa (Zhou et al., 2022a). Subsequences with high occupancy and conservation were selected as novel motifs I-XII. We also selected "non-motif" subsequences to serve as a background. These constituted the high occupancy regions between motifs. Sequence logos for the novel and canonical RdRp motifs were generated via WebLogo (Crooks et al., 2004). Pairwise conservation was calculated based on the percentage of residues with similar chemical properties between two aligned subsequences. No penalties were given to mismatched residues or gaps. A matrix of pairwise comparisons between all motifs across all selected sequences was calculated with Pandas. All FASTA files and the Jupyter notebook used for calculation are available in the supplementary data. Mean conservation across the novel, core, and non-motifs was calculated and used as the basis for a scatterplot in R using ggplot2. Deltas (differences between mean values for the novel, core, and non-motif sets) were also plotted with ggplot2 in the form of a violin plot.

RdRp structure analysis

RdRp structures were predicted using Colabfold on the full-length RdRp-containing contigs for the Achaeaviruses and select Achaea-proximal RdRps. Structures were manipulated and colored in PyMol (Schrödinger, LLC, 2015). The topology map was generated with PDBSum (Laskowski et al., 2018) and further edited with Adobe Illustrator. TM-align (Zhang and Skolnick, 2005) was used for structural alignments, and the ColorByRMSD plugin for PyMol was used to generate the distance maps.

Contig 2 recovery

To recover contig 2 for each Ao-like virus, a DIAMOND database of all 4 mRNA RUB Ngô Ao contig 2's was created. Using BLASTx, the transcripts for all Ao-like viruses were aligned against this database. Pandas was used to sort and combine all the alignment data. For each library, the contig with the best alignment identity, e-value, coverage, and length was chosen. These contigs were combined into one FASTA file. This combined file was then put through the ORFfinder command line and all the ORFs were extracted. An MSA of the extracted ORFs was created using MUSCLE. Upon viewing the alignment, 2 profiles were generated: one for contig 2's ORF 1 and the other for ORF 2. Based on these profiles, 2 HMM models (one for ORF 1 and one for ORF 2) were created with HMMER3 version 3.4 Using these HMM models, HMMER's `hmmsearch` was used to search all Ao-like virus libraries. Any ORFs that hit against the HMM model were labelled either ORF 1 or ORF 2 according to which HMM model they aligned against.

Differential gene expression

To begin, the human reference genome (`GRCh38.p14`), its corresponding gene annotation, as well as the FASTQ files of the Ngô dataset accessions for Macrophage (MonoMac6) and Neuronal Stem Cell (NSC) were downloaded. Initially, all experiments were performed on macrophages, and then replicated with the NSC dataset. HISAT2 (Kim et al., 2019) was used to extract splice sites and exons, and then used to index the human genome to generate SAM files for each accession. SAM files were converted to BAM files using SAMtools (See subsection 4.5). A matrix of counts per human gene symbol was generated using the `featureCounts` tool in the SubRead package (Liao et al., 2014). DESeq2 (Love et al., 2014) was used to analyze the count matrix and test for differential gene expression. DESeq2 and ggplot2 were used to create MA plots, PCA plots, and volcano plots with highlighted values having Benjamini-Hochberg adjusted p-value < 0.05. Upon further analysis, it was noted that of the four control Mock strains used, two appeared to be pro-inflammatory (Supplementary Figure S4). To investigate this, a spreadsheet consisting of all of the metadata from the Ngô *et al.* experiment was created. From this, it was observed that Ngô *et al.* performed their experiments in separate runs. This batch effect explains the variation in the immune system response for the control Mock strains. To account for this batch effect, any samples that were used in batches that presented hyper-inflammatory symptoms were not used for further analysis. To determine whether the defined set of genes showed statistically significant differences between phenotypes, we used the Gene Set Enrichment Analysis (GSEA) software (Subramanian et al., 2005). Using the Molecular Signatures Database (MsigDB), we focused on the Hallmark collection of gene sets, as well as five non-hallmark gene sets. The five non-hallmark gene sets were chosen on the basis of containing the genes *IFNB1*, *IFNW1*, *IFNA13*, *IFNA1*, *IFNE*, *IFNL3*, *IFNL4*, all of which were the interferon genes being upregulated amongst *T.gondii* strain infection. Using all of these gene sets, GSEA data matrices were then used to create both enrichment plots and heat maps of NES scores.

Acknowledgments

We'd like to thank R. Valencia, J. Shen, J. Parkinson, J. Charon, and A. Reinke for their helpful comments on the manuscript. This work is supported by a Project Grant from the Canadian Institutes for Health Research (CIHR PJT - 190150). AH is supported by the University of Toronto's Department of Molecular Genetics undergraduate research opportunity program (UROP). JC is

supported by the University of Toronto's Department of Molecular Genetics undergraduate research opportunity program (UROP). We are grateful to the entire team managing the NCBI SRA and the biology community for openly sharing their data. Computing resources were provided by the University of British Columbia Community Health and Wellbeing Cloud Innovation Centre, powered by AWS.

Data availability and supplementary data

Achaeavirus sequences will be available in GenBank under BioProject `XXXXXXXXXX` (pending submission). All sequences and supporting data are available at `https://github.com/ababaian/serratus/wiki/Achaeavirus`. Project notebooks and code are available via S3 (`https://mtnemo.s3.amazonaws.com/README.md`).

References

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990. ISSN 0022-2836. doi: https://doi.org/10.1016/S0022-2836(05)80360-2.

W. A. Andrade, M. d. C. Souza, E. Ramos-Martinez, K. Nagpal, M. S. Dutra, M. B. Melo, D. C. Bartholomeu, S. Ghosh, D. T. Golenbock, and R. T. Gazzinelli. Combined action of nucleic acid-sensing toll-like receptors and TLR11/TLR12 heterodimers imparts resistance to toxoplasma gondii in mice. *Cell Host Microbe*, 13(1):42–53, Jan. 2013.

S. J. Anthony, J. H. Epstein, K. A. Murray, I. Navarrete-Macias, C. M. Zambrana-Torrelio, A. Solovyov, R. Ojeda-Flores, N. C. Arrigo, A. Islam, S. Ali Khan, P. Hosseini, T. L. Bogich, K. J. Olival, M. D. Sanchez-Leon, W. B. Karesh, T. Goldstein, S. P. Luby, S. S. Morse, J. A. K. Mazet, P. Daszak, and W. I. Lipkin. A strategy to estimate unknown viral diversity in mammals. *MBio*, 4(5):e00598–13, Sept. 2013.

J. J. Aramini, C. Stephen, and J. P. Dubey. Toxoplasma gondii in vancouver island cougars (felis concolor vancouverensis): serology and oocyst shedding. *J. Parasitol.*, 84(2):438–440, Apr. 1998.

A. Babaian and R. Edgar. Ribovirus classification by a polymerase barcode sequence. *PeerJ*, 10(e14055):e14055, Oct. 2022.

E. Bertranpetit, T. Jombart, E. Paradis, H. Pena, J. Dubey, C. Su, A. Mercier, S. Devillard, and D. Ajzenberg. Phylogeography of Toxoplasma gondii points to a South American origin. *Infect Genet Evol*, 48:150–155, Mar 2017.

N. K. J. Bittner, K. L. Mack, and M. W. Nachman. Shared patterns of gene expression and protein evolution associated with adaptation to desert environments in rodents. *Genome Biol. Evol.*, 14(11), Nov. 2022.

K. Bjornevik, M. Cortese, B. C. Healy, J. Kuhle, M. J. Mina, Y. Leng, S. J. Elledge, D. W. Niebuhr, A. I. Scher, K. L. Munger, and A. Ascherio. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*, 375(6578):296–301, Jan. 2022.

D. P. Blake and F. M. Tomley. Securing poultry production from the ever-present eimeria challenge. *Trends Parasitol.*, 30(1): 12–19, Jan. 2014.

W. R. Bowie, A. S. King, D. H. Werker, J. L. Isaac-Renton, A. Bell, S. B. Eng, and S. A. Marion. Outbreak of toxoplasmosis associated with municipal drinking water. the BC toxoplasma investigation team. *Lancet*, 350(9072):173–177, July 1997.

B. Buchfink, C. Xie, and D. H. Huson. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 12(1):59–60, Jan. 2015.

E. Bushmanova, D. Antipov, A. Lapidus, and A. D. Prjibelski. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, 8(9):giz100, 09 2019. ISSN 2047-217X. doi: 10.1093/gigascience/giz100.

C. M. Cabral, K. E. McGovern, W. R. MacDonald, J. Franco, and A. A. Koshy. Dissecting amyloid beta deposition using distinct strains of the neurotropic parasite toxoplasma gondii as a novel tool. *ASN Neuro*, 9(4):1759091417724915, July 2017.

C. C. Chan, M. S. M. A. Damen, M. E. Moreno-Fernandez, T. E. Stankiewicz, M. Cappelletti, P. C. Alarcon, J. R. Oates, J. R. Doll, R. Mukherjee, X. Chen, R. Karns, M. T. Weirauch, M. A. Helmrath, T. H. Inge, and S. Divanovic. Type I interferon sensing unlocks dormant adipocyte inflammatory potential. *Nat. Commun.*, 11(1):2745, June 2020.

J. Charon, M. J. Grigg, J.-S. Eden, K. A. Piera, H. Rana, T. William, K. Rose, M. P. Davenport, N. M. Anstey, and E. C. Holmes. Novel RNA viruses associated with plasmodium vivax in human malaria and leucocytozoon parasites in avian disease. *PLoS Pathog.*, 15(12):e1008216, Dec. 2019.

L. Chen, K. Du, X. Bai, J. Shao, T. Tang, S. Xia, H. Fan, J. Wang, X. Jia, and S. Lai. Transcriptomics analysis reveals the immune response mechanism of rabbits with diarrhea fed an antibiotic-free diet. *Animals (Basel)*, 11(10), Oct. 2021a.

X.-X. Chen, W.-C. Wu, and M. Shi. Discovery and characterization of actively replicating DNA and retro-transcribing viruses in lower vertebrate hosts based on RNA sequencing. *Viruses*, 13(6):1042, May 2021b.

Y.-M. Chen, S. Sadiq, J.-H. Tian, X. Chen, X.-D. Lin, J.-J. Shen, H. Chen, Z.-Y. Hao, M. Wille, Z.-C. Zhou, J. Wu, F. Li, H.-W. Wang, W.-D. Yang, Q.-Y. Xu, W. Wang, W.-H. Gao, E. C. Holmes, and Y.-Z. Zhang. RNA viromes from terrestrial sites across china expand environmental viral diversity. *Nat. Microbiol.*, 7(8):1312–1323, Aug. 2022.

Y. Chiba, S. Oiki, T. Yaguchi, S.-I. Urayama, and D. Hagiwara. Discovery of divided RdRp sequences and a hitherto unknown genomic complexity in fungal viruses. *Virus Evol.*, 7(1):veaa101, Jan. 2021.

G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner. WebLogo: a sequence logo generator. *Genome Res.*, 14(6):1188–1190, June 2004.

R. Cui, X. Kang, Y. Liu, X. Liu, S. Chan, Y. Wang, Z. Li, Y. Ling, D. Feng, M. Li, F. Lv, and M. Fang. Integrated analysis of the whole transcriptome of skeletal muscle reveals the ceRNA regulatory network related to the formation of muscle fibers in tan sheep. *Front. Genet.*, 13:991606, Oct. 2022.

P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 02 2021. ISSN 2047-217X. doi: 10.1093/gigascience/giab008. URL `https://doi.org/10.1093/gigascience/giab008`. giab008.

M. L. Dardé, I. Villena, J. M. Pinon, and I. Beguinot. Severe toxoplasmosis caused by a toxoplasma gondii strain with a new isoenzyme type acquired in french guyana. *J. Clin. Microbiol.*, 36(1):324, Jan. 1998.

M.-L. Dardé, A. Mercier, C. Su, A. Khan, and M. E. Grigg. Chapter 3 - molecular epidemiology and population structure of toxoplasma gondii. In L. M. Weiss and K. Kim, editors, *Toxoplasma gondii (Third Edition)*, pages 63–116. Academic Press, third edition edition, 2020. ISBN 978-0-12-815041-2. doi: https://doi.org/10.1016/B978-0-12-815041-2.00003-7.

R. V. H. de Carvalho, D. S. Lima-Junior, M. V. G. da Silva, M. Dilucca, T. S. Rodrigues, C. V. Horta, A. L. N. Silva, P. F. da Silva, F. G. Frantz, L. B. Lorenzon, M. M. Souza, F. Almeida, L. M. Cantanhêde, R. d. G. M. Ferreira, A. K. Cruz, and D. S. Zamboni. Leishmania RNA virus exacerbates leishmaniasis by subverting innate immunity via TLR3-mediated NLRP3 inflammasome inhibition. *Nat. Commun.*, 10(1):5273, Nov. 2019.

S. Deng, W. He, A.-Y. Gong, M. Li, Y. Wang, Z. Xia, X.-T. Zhang, A. S. Huang Pacheco, A. Naqib, M. Jenkins, P. C. Swanson, K. M. Drescher, J. K. Strauss-Soukup, M. Belshan, and X.-M. Chen. Cryptosporidium uses CSpV1 to activate host type I interferon and attenuate antiparasitic defenses. *Nat. Commun.*, 14(1):1456, Mar. 2023.

J. P. Dubey. *Toxoplasmosis of animals and humans*. CRC Press, London, England, 3 edition, Dec. 2021a.

J. P. Dubey. Outbreaks of clinical toxoplasmosis in humans: five decades of personal experience, perspectives and lessons learned. *Parasit. Vectors*, 14(1):263, May 2021b.

J. P. Dubey, T. Quirk, J. A. Pittt, N. Sundar, G. V. Velmurugan, O. C. H. Kwok, D. Leclair, R. Hill, and C. Su. Isolation and genetic characterization of toxoplasma gondii from raccoons (procyon lotor), cats (felis domesticus), striped skunk (mephitis mephitis), black bear (ursus americanus), and cougar (puma concolor) from canada. *J. Parasitol.*, 94(1):42–45, Feb. 2008.

R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5):1792–1797, Mar. 2004.

R. C. Edgar, J. Taylor, V. Lin, T. Altman, P. Barbera, D. Meleshko, D. Lohr, G. Novakovsky, B. Buchfink, B. Al-Shayeb, J. F. Banfield, M. de la Peña, A. Korobeynikov, R. Chikhi, and A. Babaian. Petabase-scale sequence alignment catalyses viral discovery. *Nature*, 602(7895):142–147, Feb. 2022.

R. O. Eren, M. Reverte, M. Rossi, M.-A. Hartley, P. Castiglioni, F. Prevel, R. Martin, C. Desponds, L.-F. Lye, S. K. Drexler, W. Reith, S. M. Beverley, C. Ronet, and N. Fasel. Mammalian innate immune response to a leishmania-resident RNA virus increases macrophage survival to promote parasite persistence. *Cell Host Microbe*, 20(3):318–328, Sept. 2016.

R. N. Fichorova, Y. Lee, H. S. Yamamoto, Y. Takagi, G. R. Hayes, R. P. Goodman, X. Chepa-Lotrea, O. R. Buck, R. Murray, T. Kula, D. H. Beach, B. N. Singh, and M. L. Nibert. Endobiont viruses sensed by the human host - beyond conventional antiparasitic therapy. *PLoS One*, 7(11):e48418, Nov. 2012.

R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta. Pfam: the protein families database. *Nucleic Acids Res.*, 42(Database issue):D222–30, Jan. 2014.

K. A. Fitzgerald and J. C. Kagan. Toll-like receptors and the control of immunity. *Cell*, 180(6):1044–1066, Mar. 2020.

B. Gajria, A. Bahl, J. Brestelli, J. Dommer, S. Fischer, X. Gao, M. Heiges, J. Iodice, J. C. Kissinger, A. J. Mackey, D. F. Pinney, D. S. Roos, C. J. Stoeckert, Jr, H. Wang, and B. P. Brunk. ToxoDB: an integrated toxoplasma gondii database resource. *Nucleic Acids Res.*, 36(Database issue):D553–6, Jan. 2008.

Y. Gao, Z. Suding, L. Wang, D. Liu, S. Su, J. Xu, J. Hu, and J. Tao. Full-length transcriptome sequence analysis of eimeria necatrix unsporulated oocysts and sporozoites identifies genes involved in cellular invasion. *Vet. Parasitol.*, 296(109480):109480, Aug. 2021.

J. C. Gil and S. M. Hird. Multiomics characterization of the canada goose fecal microbiome reveals selective efficacy of simulated metagenomes. *Microbiol. Spectr.*, 10(6):e0238422, Dec. 2022.

A. Gómez-Arreaza, A.-L. Haenni, I. Dunia, and L. Avilán. Viruses of parasites as actors in the parasite-host relationship: A "ménage à trois". *Acta Trop.*, 166:126–132, Feb. 2017.

K. J. Graves, A. P. Ghosh, P. J. Kissinger, and C. A. Muzny. Trichomonas vaginalis virus: a review of the literature. *Int. J. STD AIDS*, 30(5):496–504, Apr. 2019.

B. Gregg, B. C. Taylor, B. John, E. D. Tait-Wojno, N. M. Girgis, N. Miller, S. Wagage, D. S. Roos, and C. A. Hunter. Replication and distribution of toxoplasma gondii in the small intestine after oral infection with tissue cysts. *Infect. Immun.*, 81(5):1635–1643, May 2013.

M. A. Hassan, A.-A. Olijnik, E.-M. Frickel, and J. P. Saeij. Clonal and atypical toxoplasma strain differences in virulence vary with mouse sub-species. *Int. J. Parasitol.*, 49(1):63–70, Jan. 2019.

K. He, Q. Liu, D.-M. Xu, F.-Y. Qi, J. Bai, S.-W. He, P. Chen, X. Zhou, W.-Z. Cai, Z.-Z. Chen, Z. Liu, X.-L. Jiang, and P. Shi. Echolocation in soft-furred tree mice. *Science*, 372(6548):eaay1513, June 2021.

D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, 35(2):518–522, Feb. 2018.

Z. Hu, D. Wu, J. Lu, Y. Zhang, S.-M. Yu, Y. Xie, H. Li, J. Yang, D.-H. Lai, K. Zeng, H. Jiang, Z.-R. Lun, and X. Yu. Inflammasome activation dampens type I IFN signaling to strengthen anti-toxoplasma immunity. *MBio*, 13(6):e0236122, Dec. 2022.

D. H. Huson, D. C. Richter, C. Rausch, T. Dezulian, M. Franz, and R. Rupp. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, 8(1):460, Nov. 2007.

A. Ives, C. Ronet, F. Prevel, G. Ruzzante, S. Fuertes-Marraco, F. Schutz, H. Zangger, M. Revaz-Breton, L.-F. Lye, S. M. Hickerson, S. M. Beverley, H. Acha-Orbea, P. Launois, N. Fasel, and S. Masina. Leishmania RNA virus controls the severity of mucocutaneous leishmaniasis. *Science*, 331(6018):775–778, Feb. 2011.

S. Jensen and A. R. Thomsen. Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *J. Virol.*, 86(6):2900–2910, Mar. 2012.

M. Jin, R. Xu, L. Wang, M. M. Alam, Z. Ma, S. Zhu, A. C. Martini, A. Jadali, M. Bernabucci, P. Xie, K. Y. Kwan, Z. P. Pang, E. Head, Y. Liu, R. P. Hart, and P. Jiang. Type-I-interferon signaling drives microglial dysfunction and senescence in human iPSC models of down syndrome and alzheimer's disease. *Cell Stem Cell*, 29(7):1135–1153.e8, July 2022.

S. K. Johnson and P. T. J. Johnson. Toxoplasmosis: Recent advances in understanding the link between infection and host behavior. *Annu. Rev. Anim. Biosci.*, 9(1):249–264, Feb. 2021.

K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993, Feb. 2008.

P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240, May 2014.

B.-K. Jung, K.-H. Pyo, K. Y. Shin, Y. S. Hwang, H. Lim, S. J. Lee, J.-H. Moon, S. H. Lee, Y.-H. Suh, J.-Y. Chai, and E.-H. Shin. Toxoplasma gondii infection in the brain inhibits neuronal degeneration and learning and memory impairments in a murine model of alzheimer's disease. *PLoS One*, 7(3):e33312, Mar. 2012.

D. Kalogeropoulos, H. Sakkas, B. Mohammed, G. Vartholomatos, K. Malamos, S. Sreekantam, P. Kanavaros, and C. Kalogeropoulos. Ocular toxoplasmosis: a review of the current diagnostic and therapeutic approaches. *Int. Ophthalmol.*, 42 (1):295–321, Jan. 2022.

S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*, 14(6):587–589, June 2017.

K. S. Katz, O. Shutov, R. Lapoint, M. Kimelman, J. R. Brister, and C. O'Sullivan. STAT: a fast, scalable, MinHash-based k-mer tool to assess sequence read archive next-generation sequence submissions. *Genome Biol.*, 22(1):270, Sept. 2021.

A. Khan, S. Taylor, J. W. Ajioka, B. M. Rosenthal, and L. D. Sibley. Selection at a single locus leads to widespread expansion of toxoplasma gondii lineages that are virulent in mice. *PLoS Genet.*, 5(3):e1000404, Mar. 2009.

S. Khurana and N. Batra. Toxoplasmosis in organ transplant recipients: Evaluation, implication, and prevention. *Trop. Parasitol.*, 6(2):123–128, July 2016.

D. Kim, J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, 37(8):907–915, Aug. 2019.

C. M. Kinsella, A. Bart, M. Deijs, P. Broekhuizen, J. Kaczorowska, M. F. Jebbink, T. van Gool, M. Cotten, and L. van der Hoek. Entamoeba and giardia parasites implicated as hosts of CRESS viruses. *Nat. Commun.*, 11(1):4620, Sept. 2020.

E. V. Koonin, V. V. Dolja, M. Krupovic, A. Varsani, Y. I. Wolf, N. Yutin, F. M. Zerbini, and J. H. Kuhn. Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.*, 84(2), May 2020.

R. Kumar, P. T. Bunn, S. S. Singh, S. S. Ng, M. Montes de Oca, F. De Labastida Rivera, S. B. Chauhan, N. Singh, R. J. Faleiro, C. L. Edwards, T. C. M. Frame, M. Sheel, R. J. Austin, S. W. Lane, T. Bald, M. J. Smyth, G. R. Hill, S. E. Best, A. Haque, D. Corvino, N. Waddell, L. Koufariotis, P. Mukhopadhay, M. Rai, J. Chakravarty, O. P. Singh, D. Sacks, S. Nylen, J. Uzonna, S. Sundar, and C. R. Engwerda. Type I interferons suppress anti-parasitic immunity and can be targeted to improve treatment of visceral leishmaniasis. *Cell Rep.*, 30(8):2512–2525.e9, Feb. 2020.

B. Langmead and S. L. Salzberg. Fast gapped-read alignment with bowtie 2. *Nat. Methods*, 9(4):357–359, Mar. 2012.

R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, and J. M. Thornton. PDBsum: Structural summaries of PDB entries. *Protein Sci.*, 27(1):129–134, Jan. 2018.

A. J. Lee and A. A. Ashkar. The dual nature of type I and type II interferons. *Front. Immunol.*, 9:2061, Sept. 2018.

H. Li and R. Durbin. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5):589–595, Mar. 2010.

Y. Liao, G. K. Smyth, and W. Shi. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, Apr. 2014.

A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, and P. Tamayo. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.*, 1(6):417–425, Dec. 2015.

H. Lorenzi, A. Khan, M. S. Behnke, S. Namasivayam, L. S. Swapna, M. Hadjithomas, S. Karamycheva, D. Pinney, B. P. Brunk, J. W. Ajioka, D. Ajzenberg, J. C. Boothroyd, J. P. Boyle, M. L. Dardé, M. A. Diaz-Miranda, J. P. Dubey, H. M. Fritz, S. M. Gennari, B. D. Gregory, K. Kim, J. P. J. Saeij, C. Su, M. W. White, X.-Q. Zhu, D. K. Howe, B. M. Rosenthal, M. E. Grigg, J. Parkinson, L. Liu, J. C. Kissinger, D. S. Roos, and L. D. Sibley. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic toxoplasma gondii genomes. *Nat. Commun.*, 7:10147, Jan. 2016.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12):550, 2014.

S. K. Matta, P. Olias, Z. Huang, Q. Wang, E. Park, W. M. Yokoyama, and L. D. Sibley. Toxoplasma gondii effector TgIST blocks type I interferon signaling to promote infection. *Proc. Natl. Acad. Sci. U. S. A.*, 116(35):17480–17491, Aug. 2019.

S. K. Matta, N. Rinkenberger, I. R. Dunay, and L. D. Sibley. Toxoplasma gondii infection and its implications within the central nervous system. *Nat. Rev. Microbiol.*, 19(7):467–480, July 2021.

J. B. McAuley. Congenital toxoplasmosis. *J. Pediatric Infect. Dis. Soc.*, 3 Suppl 1(suppl_1):S30–5, Sept. 2014.

K. E. McGovern, C. M. Cabral, H. W. Morrison, and A. A. Koshy. Aging with toxoplasma gondii results in pathogen clearance, resolution of inflammation, and minimal consequences to learning and memory. *Sci. Rep.*, 10(1):7979, May 2020.

M. B. Melo, Q. P. Nguyen, C. Cordeiro, M. A. Hassan, N. Yang, R. McKell, E. E. Rosowski, L. Julien, V. Butty, M.-L. Dardé, D. Ajzenberg, K. Fitzgerald, L. H. Young, and J. P. J. Saeij. Transcriptional analysis of murine macrophages infected with different toxoplasma strains identifies novel regulation of host signaling pathways. *PLoS Pathog.*, 9(12):e1003779, Dec. 2013.

M. A. Miller, C. A. Newberry, D. M. Sinnott, F. I. Batac, K. Greenwald, A. Reed, C. Young, M. D. Harris, A. E. Packham, and K. Shapiro. Newly detected, virulent toxoplasma gondii COUG strain causing fatal steatitis and toxoplasmosis in southern sea otters (enhydra lutris nereis). *Front. Mar. Sci.*, 10, Mar. 2023.

R. L. Miller, A. L. Wang, and C. C. Wang. Identification of giardia lamblia isolates susceptible and resistant to infection by the double-stranded rna virus. *Experimental Parasitology*, 66(1):118–123, 1988. ISSN 0014-4894. doi: https://doi.org/10.1016/0014-4894(88)90056-2.

B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37(5):1530–1534, May 2020.

M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. ColabFold: making protein folding accessible to all. *Nat. Methods*, 19(6):679–682, June 2022.

J. G. Montoya and O. Liesenfeld. Toxoplasmosis. *Lancet*, 363(9425):1965–1976, June 2004.

R. K. Narayanasamy, P. Rada, A. Zdrha, M. van Ranst, J. Neyts, and J. Tachezy. Cytidine nucleoside analog is an effective antiviral drug against trichomonasvirus. *J. Microbiol. Immunol. Infect.*, 55(2):191–198, Apr. 2022.

T. Nayeri, S. Sarvi, M. Sharif, and A. Daryani. Toxoplasma gondii: A possible etiologic agent for alzheimer's disease. *Heliyon*, 7(6):e07151, June 2021.

U. Neri, Y. I. Wolf, S. Roux, A. P. Camargo, B. Lee, D. Kazlauskas, I. M. Chen, N. Ivanova, L. Zeigler Allen, D. Paez-Espino, D. A. Bryant, D. Bhaya, RNA Virus Discovery Consortium, M. Krupovic, V. V. Dolja, N. C. Kyrpides, E. V. Koonin, and U. Gophna. Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell*, 185(21):4023–4037.e18, Oct. 2022.

H. M. Ngô, Y. Zhou, H. Lorenzi, K. Wang, T.-K. Kim, Y. Zhou, K. El Bissati, E. Mui, L. Fraczek, S. V. Rajagopala, C. W. Roberts, F. L. Henriquez, A. Montpetit, J. M. Blackwell, S. E. Jamieson, K. Wheeler, I. J. Begeman, C. Naranjo-Galvis, N. Alliey-Rodriguez, R. G. Davis, L. Soroceanu, C. Cobbs, D. A. Steindler, K. Boyer, A. G. Noble, C. N. Swisher, P. T. Heydemann, P. Rabiah, S. Withers, P. Soteropoulos, L. Hood, and R. McLeod. Toxoplasma modulates signature pathways of human epilepsy, neurodegeneration & cancer. *Sci. Rep.*, 7(1):11496, Sept. 2017.

W. Niedelman, D. A. Gold, E. E. Rosowski, J. K. Sprokholt, D. Lim, A. Farid Arenas, M. B. Melo, E. Spooner, M. B. Yaffe, and J. P. J. Saeij. The rhoptry proteins ROP18 and ROP5 mediate toxoplasma gondii evasion of the murine, but not the human, interferon-gamma response. *PLoS Pathog.*, 8(6):e1002784, June 2012.

T. pandas development team. pandas-dev/pandas: Pandas, Feb. 2020. URL `https://doi.org/10.5281/zenodo.3509134`.

G. Pappas, N. Roussos, and M. E. Falagas. Toxoplasmosis snapshots: global status of Toxoplasma gondii seroprevalence and implications for pregnancy and congenital toxoplasmosis. *Int J Parasitol*, 39(12):1385–1394, Oct 2009.

X. Pu, X. Li, L. Cao, K. Yue, P. Zhao, X. Wang, J. Li, X. Zhang, N. Zhang, Z. Zhao, M. Liang, and P. Gong. Giardia duodenalis induces proinflammatory cytokine production in mouse macrophages via TLR9-mediated p38 and ERK signaling pathways. *Front. Cell Dev. Biol.*, 9:694675, July 2021.

M. Raetz, A. Kibardin, C. R. Sturge, R. Pifer, H. Li, E. Burstein, K. Ozato, S. Larin, and F. Yarovinsky. Cooperation of TLR12 and TLR11 in the IRF8-dependent IL-12 response to toxoplasma gondii profilin. *J. Immunol.*, 191(9):4818–4827, Nov. 2013.

J. Rehwinkel and M. U. Gack. RIG-I-like receptors: their regulation and roles in RNA sensing. *Nat. Rev. Immunol.*, 20(9):537–551, Sept. 2020.

A. Rhie, S. A. McCarthy, O. Fedrigo, J. Damas, G. Formenti, S. Koren, M. Uliano-Silva, W. Chow, A. Fungtammasan, J. Kim, C. Lee, B. J. Ko, M. Chaisson, G. L. Gedman, L. J. Cantin, F. Thibaud-Nissen, L. Haggerty, I. Bista, M. Smith, B. Haase, J. Mountcastle, S. Winkler, S. Paez, J. Howard, S. C. Vernes, T. M. Lama, F. Grutzner, W. C. Warren, C. N. Balakrishnan, D. Burt, J. M. George, M. T. Biegler, D. Iorns, A. Digby, D. Eason, B. Robertson, T. Edwards, M. Wilkinson, G. Turner, A. Meyer, A. F. Kautt, P. Franchini, H. W. Detrich, 3rd, H. Svardal, M. Wagner, G. J. P. Naylor, M. Pippel, M. Malinsky, M. Mooney, M. Simbirsky, B. T. Hannigan, T. Pesout, M. Houck, A. Misuraca, S. B. Kingan, R. Hall, Z. Kronenberg, I. Sović, C. Dunn, Z. Ning, A. Hastie, J. Lee, S. Selvaraj, R. E. Green, N. H. Putnam, I. Gut, J. Ghurye, E. Garrison, Y. Sims, J. Collins, S. Pelan, J. Torrance, A. Tracey, J. Wood, R. E. Dagnew, D. Guan, S. E. London, D. F. Clayton, C. V. Mello, S. R. Friedrich, P. V. Lovell, E. Osipova, F. O. Al-Ajli, S. Secomandi, H. Kim, C. Theofanopoulou, M. Hiller, Y. Zhou, R. S. Harris, K. D. Makova, P. Medvedev, J. Hoffman, P. Masterson, K. Clark, F. Martin, K. Howe, P. Flicek, B. P. Walenz, W. Kwak, H. Clawson, M. Diekhans, L. Nassar, B. Paten, R. H. S. Kraus, A. J. Crawford, M. T. P. Gilbert, G. Zhang, B. Venkatesh, R. W. Murphy, K.-P. Koepfli, B. Shapiro, W. E. Johnson, F. Di Palma, T. Marques-Bonet, E. C. Teeling, T. Warnow, J. M. Graves, O. A. Ryder, D. Haussler, S. J. O'Brien, J. Korlach, H. A. Lewin, K. Howe, E. W. Myers, R. Durbin, A. M. Phillippy, and E. D. Jarvis. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856):737–746, Apr. 2021.

J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. Integrative genomics viewer. *Nat. Biotechnol.*, 29(1):24–26, Jan. 2011.

J. R. Rodrigues, S. W. Roy, and R. N. M. Sehgal. Novel RNA viruses associated with avian haemosporidian parasites. *PLoS One*, 17(6):e0269881, June 2022.

S. A. I. Sanford and W. A. McEwan. Type-I interferons in alzheimer's disease and other tauopathies. *Front. Cell. Neurosci.*, 16:949340, July 2022.

C. A. Schneider, D. X. Figueroa Velez, S. B. Orchanian, L. A. Shallberg, D. Agalliu, C. A. Hunter, S. P. Gandhi, and M. B. Lodoen. Toxoplasma gondii dissemination in the brain is facilitated by infiltrating peripheral immune cells. *MBio*, 13(6):e0283822, Dec. 2022.

A. Schoenemeyer, B. J. Barnes, M. E. Mancl, E. Latz, N. Goutagny, P. M. Pitha, K. A. Fitzgerald, and D. T. Golenbock. The interferon regulatory factor, IRF5, is a central mediator of toll-like receptor 7 signaling. *J. Biol. Chem.*, 280(17):17005–17012, Apr. 2005.

Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

W. Shen, S. Le, Y. Li, and F. Hu. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, 11 (10):e0163962, Oct. 2016.

M. Shi, X.-D. Lin, J.-H. Tian, L.-J. Chen, X. Chen, C.-X. Li, X.-C. Qin, J. Li, J.-P. Cao, J.-S. Eden, J. Buchmann, W. Wang, J. Xu, E. C. Holmes, and Y.-Z. Zhang. Redefining the invertebrate RNA virosphere. *Nature*, 540(7634):539–543, Dec. 2016.

M. Shi, X.-D. Lin, X. Chen, J.-H. Tian, L.-J. Chen, K. Li, W. Wang, J.-S. Eden, J.-J. Shen, L. Liu, E. C. Holmes, and Y.-Z. Zhang. The evolutionary history of vertebrate RNA viruses. *Nature*, 556(7700):197–202, Apr. 2018.

S. Silva-Barrios and S. Stäger. Protozoan parasites and type I IFNs. *Front. Immunol.*, 8:14, Jan. 2017.

S. S. Soldan and P. M. Lieberman. Epstein-Barr virus and multiple sclerosis. *Nat. Rev. Microbiol.*, 21(1):51–64, Jan. 2023.

C. Su, D. Evans, R. H. Cole, J. C. Kissinger, J. W. Ajioka, and L. D. Sibley. Recent expansion of Toxoplasma through enhanced oral transmission. *Science*, 299(5605):414–416, Jan 2003.

A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, 102(43):15545–15550, Oct. 2005.

S. Sukla, S. Roy, S. Sundar, and S. Biswas. Leptomonas seymouri narna-like virus 1 and not leishmaniaviruses detected in kala-azar samples from india. *Arch. Virol.*, 162(12):3827–3835, Dec. 2017.

Y. Sun, X. Lin, Q. Zhang, Y. Pang, X. Zhang, X. Zhao, D. Liu, and X. Yang. Genome-wide characterization of lncRNAs and mRNAs in muscles with differential intramuscular fat contents. *Front. Vet. Sci.*, 9:982258, Aug. 2022.

M. van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding, and M. Steinegger. Fast and accurate protein structure search with foldseek. *Nat. Biotechnol.*, May 2023.

Y. Wen, S. Li, G. Bao, J. Wang, X. Liu, J. Hu, F. Zhao, Z. Zhao, B. Shi, and Y. Luo. Comparative transcriptome analysis reveals the mechanism associated with dynamic changes in meat quality of the longissimus thoracis muscle in tibetan sheep at different growth stages. *Front. Vet. Sci.*, 9:926725, July 2022.

R. Werk. How does toxoplasma gondii enter host cells? *Rev. Infect. Dis.*, 7(4):449–457, July 1985.

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

D. Wilkins. *gggenes: Draw Gene Arrow Maps in 'ggplot2'*, 2020. URL `https://CRAN.R-project.org/package=gggenes`. R package version 0.4.1.

Y. I. Wolf, S. Silas, Y. Wang, S. Wu, M. Bocek, D. Kazlauskas, M. Krupovic, A. Fire, V. V. Dolja, and E. V. Koonin. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat. Microbiol.*, 5(10):1262–1270, Oct. 2020.

M. E. J. Woolhouse and K. Adair. The diversity of human RNA viruses. *Future Virol.*, 8(2):159–171, Feb. 2013.

Z. Wu, Y. Han, B. Liu, H. Li, G. Zhu, A. Latinne, J. Dong, L. Sun, H. Su, L. Liu, J. Du, S. Zhou, M. Chen, A. Kritiyakan, S. Jittapalapong, K. Chaisiri, P. Buchy, V. Duong, J. Yang, J. Jiang, X. Xu, H. Zhou, F. Yang, D. M. Irwin, S. Morand, P. Daszak, J. Wang, and Q. Jin. Decoding the RNA viromes in rodent lungs provides new insight into the origin and evolutionary patterns of rodent-borne pathogens in mainland southeast asia. *Microbiome*, 9(1):18, Jan. 2021.

Y. Xie, J. Xiao, X. Zhou, X. Gu, R. He, J. Xu, B. Jing, X. Peng, and G. Yang. Global transcriptome landscape of the rabbit protozoan parasite eimeria stiedae. *Parasit. Vectors*, 14(1):308, June 2021.

H.-Y. Yang, W.-C. Chien, C.-H. Chung, R.-Y. Su, C.-Y. Lai, C.-C. Yang, and N.-S. Tzeng. Risk of dementia in patients with toxoplasmosis: a nationwide, population-based cohort study in taiwan. *Parasit. Vectors*, 14(1):435, Aug. 2021.

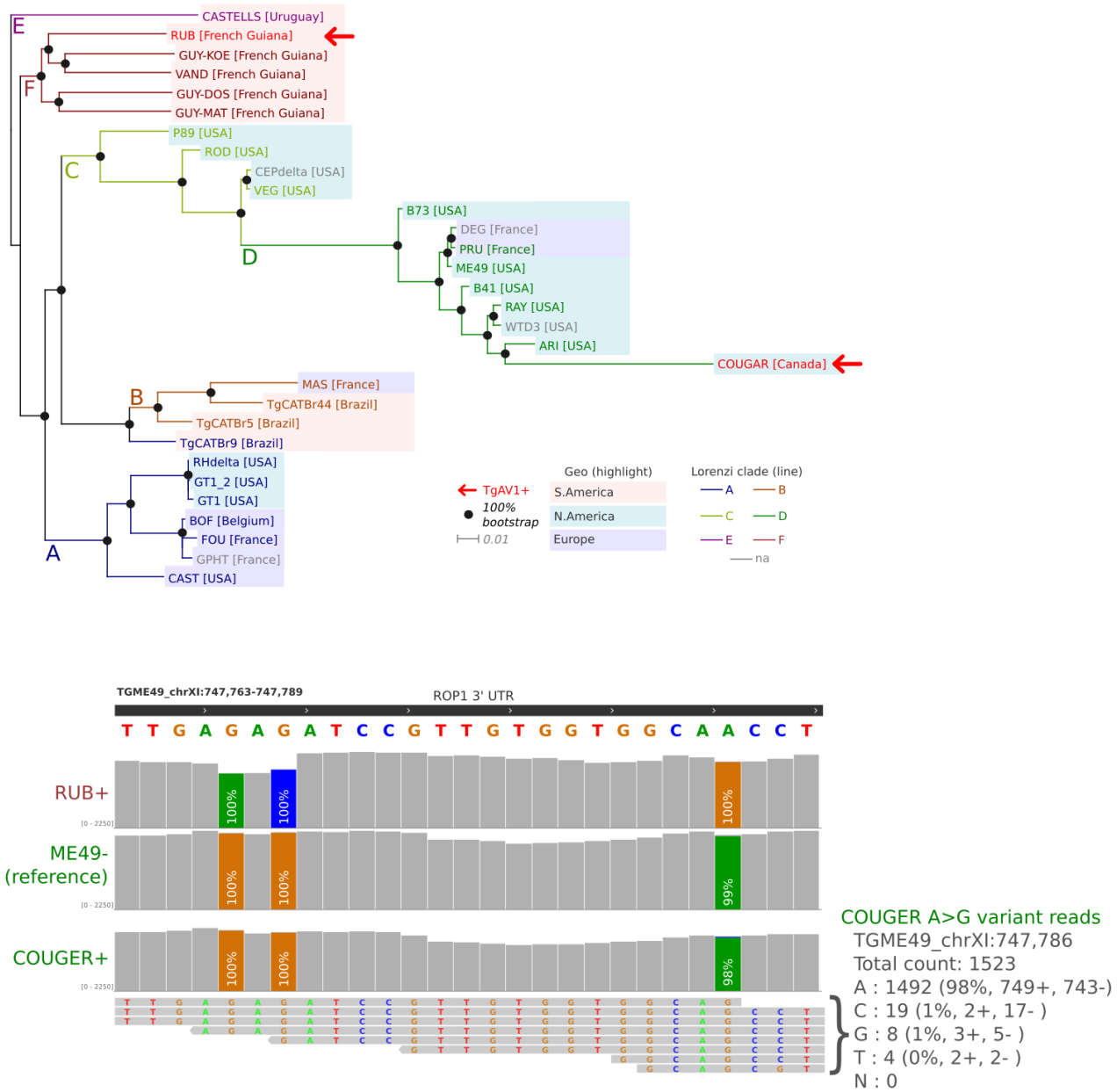
Y. Zhang and J. Skolnick. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33 (7):2302–2309, 01 2005. ISSN 0305-1048. doi: 10.1093/nar/gki524. URL `https://doi.org/10.1093/nar/gki524`.

Z. Zhao, X. Li, N. Zhang, J. Li, N. Zhao, M. Gao, X. Zhang, X. Wang, P. Zhao, L. Li, M. Sun, L. Cao, and P. Gong. Multiple regulations of parasitic protozoan viruses: A double-edged sword for protozoa. *MBio*, 14(1):e0264222, Feb. 2023.

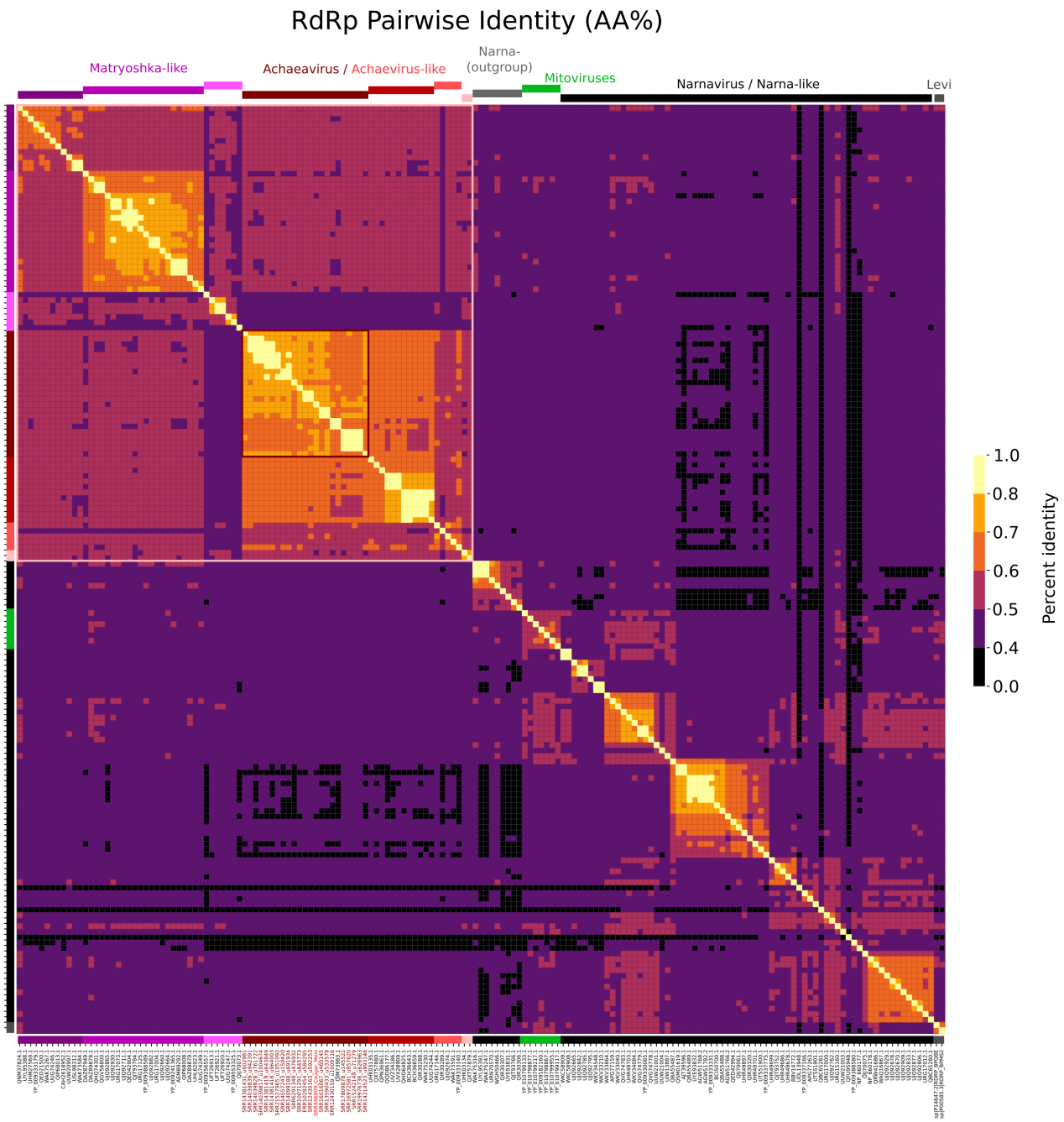
H. Zhou, J. Ji, X. Chen, Y. Bi, J. Li, Q. Wang, T. Hu, H. Song, R. Zhao, Y. Chen, M. Cui, Y. Zhang, A. C. Hughes, E. C. Holmes, and W. Shi. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell*, 184(17):4380–4391.e14, Aug. 2021.

L. Zhou, T. Feng, S. Xu, F. Gao, T. T. Lam, Q. Wang, T. Wu, H. Huang, L. Zhan, L. Li, Y. Guan, Z. Dai, and G. Yu. Ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Brief. Bioinform.*, 23(4), July 2022a.

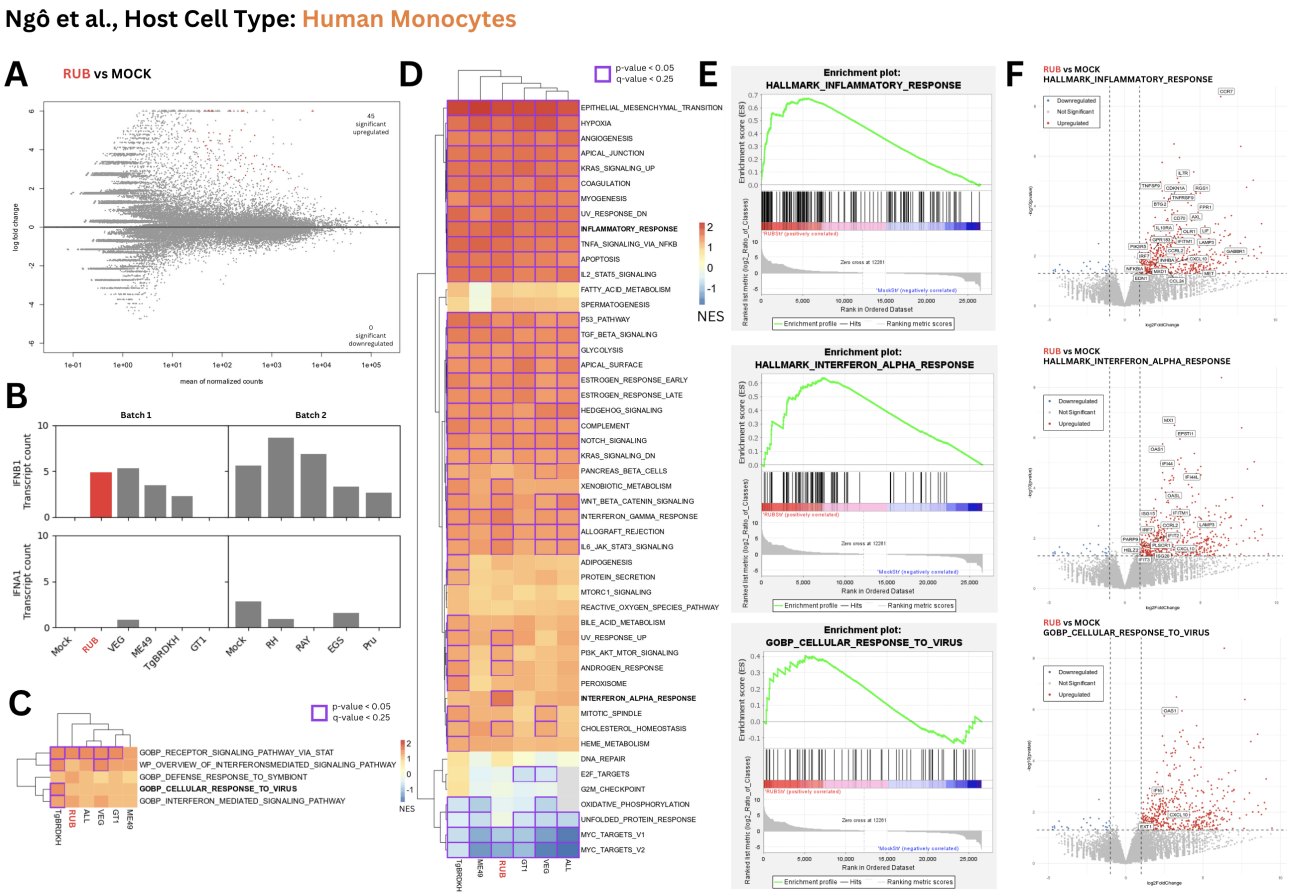
Z. Zhou, K. Li, J. Liu, H. Zhang, Y. Fan, Y. Chen, H. Han, J. Yang, and Y. Liu. Expression profile analysis to identify circular RNA expression signatures in muscle development of wu'an goat longissimus dorsi tissues. *Front. Vet. Sci.*, 9:833946, Apr. 2022b.



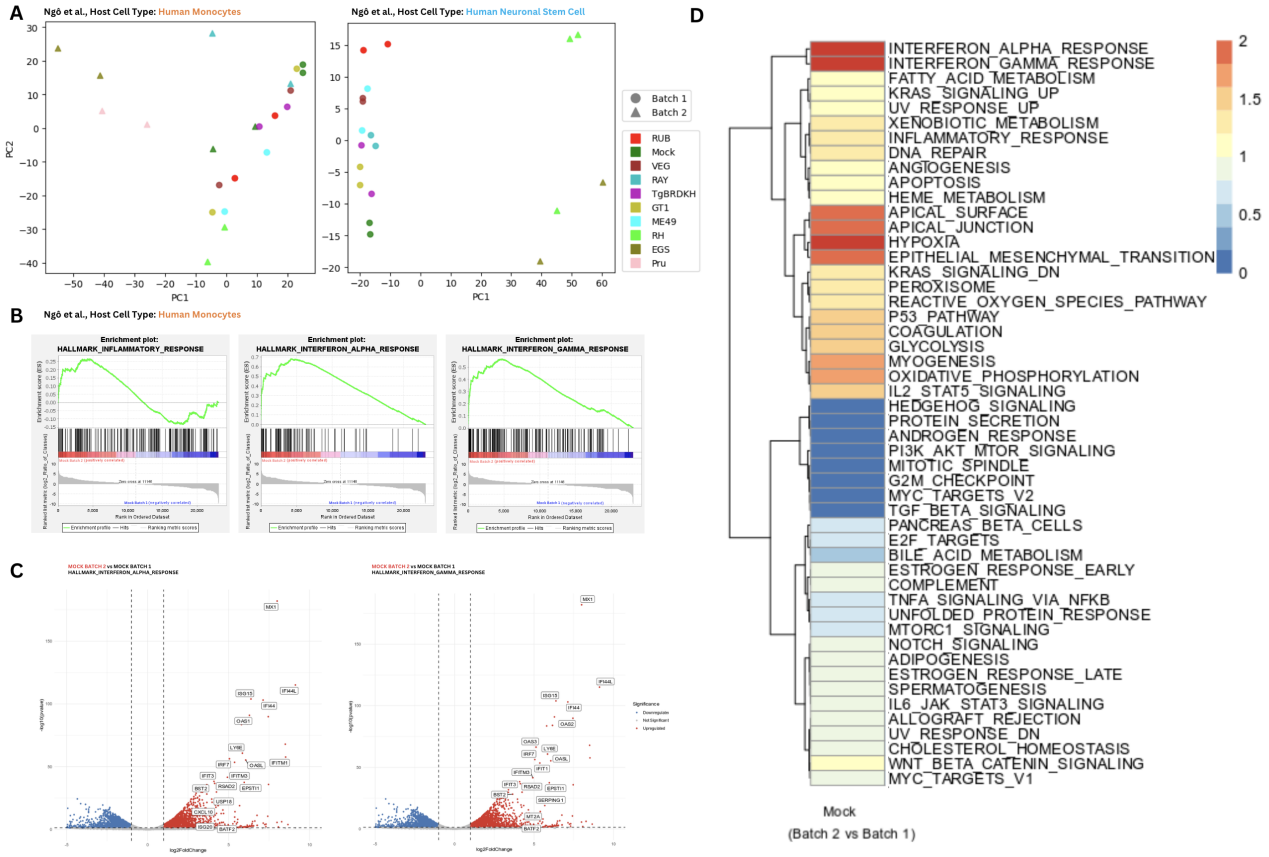
Supplementary Figure S1. Maximum likelihood tree for T. gondii transcriptomes and sample contamination. a. RNA-seq reads were aligned to the T. gondii ME49 reference genome [ToxoDB 64] from each of the 31 T. gondii samples of the Melo study (Melo et al., 2013). Phylogeny was constructed from consensus single nucleotide polymorphisms (SNPs) in expressed annotated exons. 8,201,735/30,124,248 (27.22%) exonic sites were expressed across all samples of which a further 156,773/8,201,735 (1.91%) sites were polymorphic in at least one sample. IQ-TREE (v2.2.2.6; `iqtree2 -s melo.snp-only_nogap_msa.fa -m TEST+ASC -bb 1000 -alrt 1000; model: 'TVM+F+ASC+G4`) was ran. Nodes supported by 100% bootstraps are indicated with a black circle. **b.** Manual inspection of aligned reads in IGV was performed to assess if Ao-positive libraries are cross-contaminated. RUB-specific alleles were not present above the sequencing-error background (1̃ %) in the COUGER sample or vice versa. Where possible, the reads for potential cross-contaminating alleles were inspected at adjacent linked sites.



Supplementary Figure S2. Pairwise percent amino acid identity between Narnavirus RdRp a. Heatmap plotting the pairwise percent identity for each RdRp in the phylogeny. Percent identity was calculated on the basis of pairwise alignments for the RdRp core (motifs F-E) and thumb, gaps were not penalized. The heatmap is symmetrical (i.e. RdRps on the diagonal are identical), and tentative phylogenetic groupings are shown.



Supplementary Figure S3. DGE of human macrophages infected with various *T. gondii* strains. **A.** MA plot of *T. gondii* - RUB vs mock genes (highlighted: Benjamini-Hochberg adjusted p-value < 0.05) **B.** Bar plot of normalized transcription counts of *IFNB1* and *IFNA1* across *T. gondii* strains and mock sequenced in Ngô. et al experiments, seperated by batch. **C.** Heat map of Normalized Enrichment Scores (NES) from Gene Set Enrichment Analysis (GSEA) using gene sets possessing interferon-specific genes, namely *IFNA1* and *IFNB1*, applied to the *T. gondii* strains. **D.** Heat map of NES values from GSEA using the Hallmark gene sets. **E.** GSEA curves comparing RUB vs. Mock strain using inflammatory response, cellular response to virus, and interferon-mediated signalling pathway gene sets. **F.** Volcano plot of differentially regulated genes with genes of notable gene sets being labeled.



Supplementary Figure S4. Inspection of Differential Gene Expression data displaying batch effects. **A.** PCA plot of batch effects visible in Ngô *et al.* datasets for Human Monocytes and Neuronal Stem Cells. **B.** GSEA curves comparing Mock strains between Batches using inflammatory response, interferon alpha response, and interferon gamma response gene sets. **C.** Volcano plot of differentially regulated genes with genes of interferon alpha response, and interferon gamma response gene sets being labeled. **D.** Heat map of NES values from GSEA using the Hallmark gene sets comparing mock batch 1 vs. mock batch 2.

Supplementary Tables

Supplementary Table S1. Ngô sequencing datasets and meta-data. Metadata describing BioProject `PRJNA241125` (Ngô study) including flowcell header metadata extracted from raw (fastq) sequence headers and normalized expression (RPKM) of *Achaeavirus odysseus* per library.

Supplementary Table S2. *Toxoplasma gondii* strains. Strains of *Toxoplasma gondii* analyzed from BioProjects `PRJNA241125` (Ngô study) and `PRJNA114693` (Melo study), and their associated meta-data.

Supplementary Table S3. Achaeaviruses and associated SRA metadata. A. Complete list of exemplar *Achaeaviruses* identified in this study. **B.** Description of each Sequence Read Archive (SRA) library from which a relevant species-like operational taxonomic unit (sOTU) was identified. **C.** The pairwise percentage identity between exemplar *Achaeaviruses*. **D.** Duplicate of Table 1 in spreadsheet format