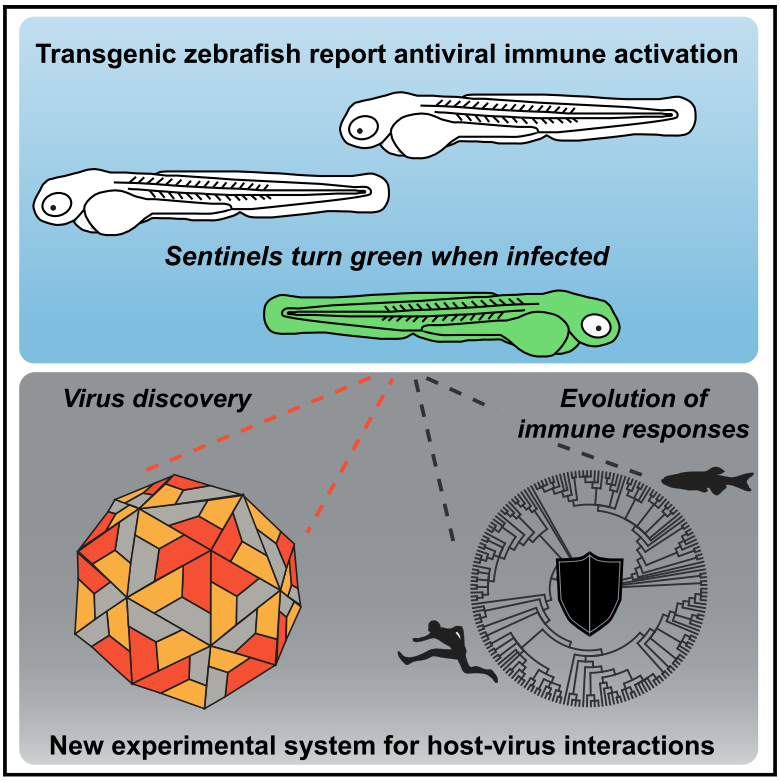
Article

Current Biology

# Linking Virus Discovery to Immune Responses Visualized during Zebrafish Infections

## Graphical Abstract




## Authors

Keir M. Balla, Marlen C. Rice, James A. Gagnon, Nels C. Elde

## Correspondence

nelde@genetics.utah.edu

## In Brief

Many of the host immune responses to viral infection are invisible to outside observation, hindering the experimental surveillance of infections. Here, Balla et al. visualize spontaneous activation of the zebrafish antiviral immune system and link this to the discovery of an endemic picornavirus that infects the gut, brain, and other tissues.

## Highlights

- Transgenic *isg15:GFP* zebrafish enable visualization of antiviral immune activity
- Spontaneous activation of this reporter led to the discovery of a picornavirus
- Zebrafish picornavirus is widespread in RNA-seq data and provokes immune responses
- Infections are transmitted horizontally and amenable to experimental manipulation

Balla et al., 2020, Current Biology *30*, 2092–2103
June 8, 2020 
https://doi.org/10.1016/j.cub.2020.04.031

CellPress

CellPress

Current Biology

Article

# Linking Virus Discovery to Immune Responses Visualized during Zebrafish Infections

Keir M. Balla,[1] Marlen C. Rice,[2] James A. Gagnon,[2] and Nels C. Elde[1,3,*]
[1]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA
[2]School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA
[3]Lead Contact
*Correspondence: nelde@genetics.utah.edu
https://doi.org/10.1016/j.cub.2020.04.031

## SUMMARY

The discovery of new viruses currently outpaces our capacity for experimental examination of infection biology. To better couple virus discovery with immunology, we genetically modified zebrafish to visually report on virus infections. After generating a strain that expresses green fluorescent protein (GFP) under an interferon-stimulated gene promoter, we repeatedly observed transgenic larvae spontaneously expressing GFP days after hatching. RNA sequencing comparisons of co-housed GFP-positive and GFP-negative zebrafish revealed a naturally occurring picornavirus that induced a canonical interferon-mediated response and hundreds of antiviral defense genes not observed following immunostimulatory treatments or experimental infections with other viruses. Among the many genes induced by picornavirus infection was a large set encoding guanosine triphosphatase (GTPase) of immunity-associated proteins (GIMAPs). The GIMAP gene family is massively expanded in fish genomes and may also play a crucial role in antiviral responses in mammals, including humans. We subsequently detected zebrafish picornavirus in publicly available sequencing data from seemingly asymptomatic zebrafish in many research institutes and found that it altered gene expression in a previous study of zebrafish development. Experiments revealed a horizontal mode of virus transmission, highlighting a system for studying the spread of picornavirus infections within and between individuals. Our study describes a naturally occurring picornavirus that elicits strong antiviral responses in zebrafish and provides new strategies for simultaneously discovering viruses and their impact on vertebrate hosts.

## INTRODUCTION

Viruses are estimated to outnumber their cellular hosts in most environments [1, 2], indicating that animals regularly encounter and interact with viruses. The ubiquity of viruses is demonstrated by the widespread success of virus discoveries made through metagenomic sequencing of animal feces and tissues [3–5]. Although surveys of virus prevalence in animals can reveal close associations, most metagenomic virus discovery studies are agnostic as to how, or even whether, the virus and animal interact. Metagenomic investigations of animals with undiagnosed disease have resulted in discoveries of new viruses [6–8], thereby establishing a strong potential link between virus and host. However, the high frequency at which viruses are discovered in asymptomatic hosts suggests that extreme disease phenotypes result from only a small fraction of virus-host interactions.

Most consequential interactions between viruses and animals involve host immune responses. Thus, one sensitive strategy for discovering previously unknown viruses with clear host interactions is to select for engagement of antiviral immune responses instead of overt disease. Although typically invisible, antiviral immune responses can be visualized in some live hosts with the genetic tools that are available for a growing collection of experimentally accessible organisms [9, 10]. Furthermore, carrying out virus discovery in research organisms holds the potential for developing new tractable systems for studying host-virus interactions, as demonstrated through metagenomics approaches in worms, flies, and mice [11–13].

Virus discovery in vertebrates has predominantly been carried out in mammals [14], but recent wide-ranging metagenomic surveys revealed that fish are infected by an extensively overlapping set of diverse virus families [5, 15]. Viruses discovered in fish can therefore illuminate facets of virus and host biology that are shared across a much wider interval of evolutionary history than traditionally considered. Additionally, increasing our knowledge of the viruses that infect fish is an important goal for improving marine and freshwater aquaculture systems that increasingly play a central role in emerging food supply challenges [16].

Zebrafish are versatile research organisms that can be experimentally infected with viruses from heterologous fish and mammals [10, 17–19]. Larvae are mostly transparent, allowing for the visualization of antiviral immune responses in live transgenic animals [10]. Two naturally occurring viruses have been detected in diseased adult zebrafish from a research

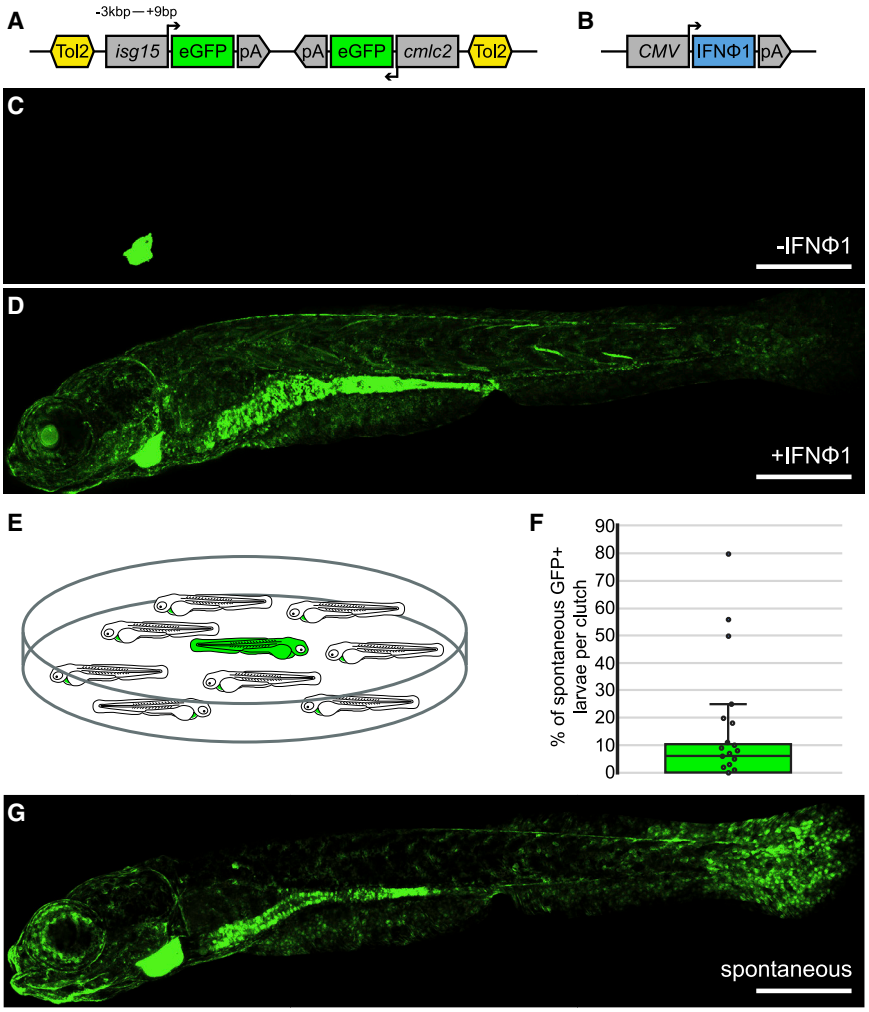



**Figure 1. Induced and Spontaneous GFP Expression in Transgenic Zebrafish that Report on Interferon Signaling**

(A) Transgenic animals were generated with a Tol2 transposon construct bearing the promoter from *isg15* fused to the coding sequence for GFP along with a heart-specific (*cmlc2*) transgenesis marker in the opposite orientation.

(B) Illustration of the construct used to drive type I interferon expression in zebrafish, with interferon ϕ1 (IFNϕ1) driven by the *CMV* promoter.

(C) An untreated *isg15:GFP* zebrafish at 8 dpf.

(D) An IFNϕ1-treated *isg15:GFP* zebrafish at 8 dpf.

(E) Illustration of a dish of untreated *isg15:GFP* zebrafish larvae in which spontaneous GFP expression was observed.

(F) Frequency at which spontaneous GFP was observed across independent clutches of *isg15:GFP* zebrafish. Each dot represents the percent of larvae within a clutch that spontaneously expressed GFP.

(G) An untreated *isg15:GFP* zebrafish that spontaneously expressed GFP at 8 dpf.

All microscope images are maximum intensity projections of confocal Z stacks. Scale bars are 500 μm; dpf, days post fertilization.

See also Video S1 and Table S4.

facility and the pet trade [20, 21], and a third virus was recently discovered from an asymptomatic adult [22]. These observations highlight the possibility that zebrafish regularly interact with viruses even in carefully maintained environments like research facilities, which might manifest in the induction of host antiviral signaling.

To advance the visualization of immune responses as a tool for virus discovery, we generated transgenic zebrafish that report on antiviral interferon signaling. We observed seemingly spontaneous induction of the reporter in zebrafish larvae during normal rearing in our research facility. With RNA sequencing, we discovered a picornavirus in animals exhibiting antiviral immune responses that was distantly related to all other picornaviruses known at the time of discovery. Beyond our research facility, we found virus reads in RNA sequencing datasets derived from zebrafish housed in research institutions around the world. Most notable were virus matches in datasets from a recently acquired wild zebrafish and striking levels of picornavirus from a clonal laboratory line (CG2), pointing to both natural and immunocompromised hosts. Finally, we connect our analysis of virus burden with a previously unrecognized and substantial effect on host gene expression enriched for antiviral immune responses in otherwise apparently asymptomatic hosts.

## RESULTS

### Spontaneous Induction of Interferon Responses in Zebrafish Larvae

We generated transgenic zebrafish that report on interferon signaling to visualize antiviral immune responses in live animals. Interferons are signaling proteins that induce the expression of hundreds of genes collectively termed interferon-stimulated genes, the products of which provide the main antiviral immune defenses in vertebrates [23]. A 3.4 kbp promoter from zebrafish interferon-stimulated gene 15 (*isg15*) was cloned upstream of the coding sequence for GFP in a construct that also drives constitutive GFP expression in the heart, which was inserted into the genome by *Tol2* transposon-mediated transgenesis (Figure 1A). As in other vertebrates, the zebrafish *isg15* promoter contains regulatory transcription factor binding site motifs called interferon-stimulated response elements [24]. To test the interferon responsiveness of *isg15:GFP* animals, we cloned the coding sequence for a type I interferon (*ifnϕ1*) into an expression plasmid (Figure 1B) and injected this construct into transgenic animals. Most untreated animals exhibited GFP expression that was restricted to the heart, which confirmed that they carried the transgenic insertion (Figure 1C). In contrast, more than 80% of animals injected with the interferon plasmid expressed GFP broadly (Figure 1D; Video S1), demonstrating that the *isg15:GFP* transgenic animals provide a conspicuous readout for antiviral immune signaling.

We were surprised to observe that some uninjected *isg15:GFP* transgenic fish spontaneously expressed GFP in multiple tissues

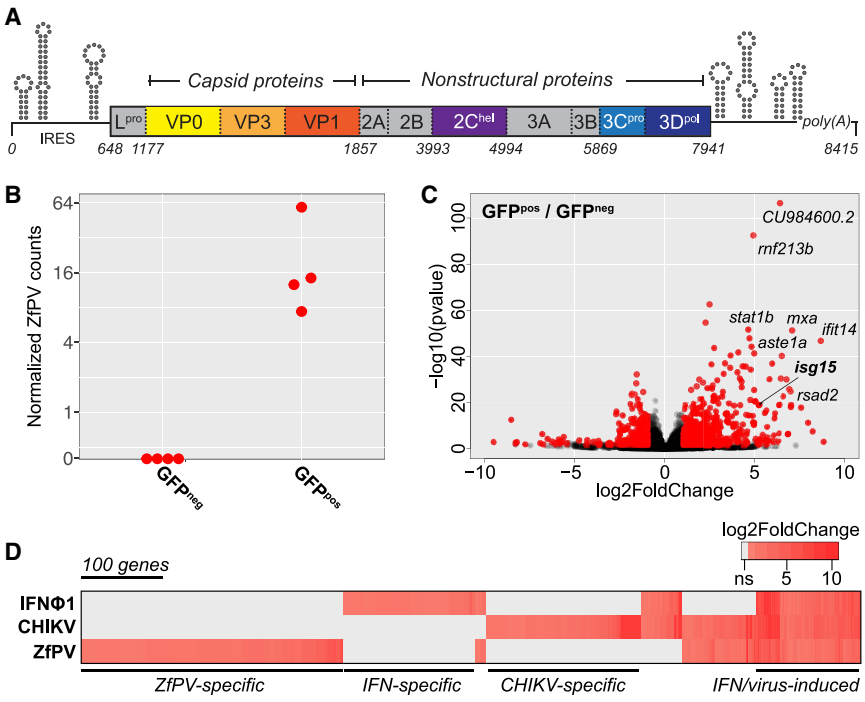



**Figure 2. A Naturally Occurring Picornavirus Sequenced from Zebrafish Larvae Induces Extensive Antiviral Immune Responses**

(A) Genome organization of *Zebrafish picornavirus* (ZfPV). Untranslated regions of the genome are depicted as lines. Schematics above lines represent secondary structures in the UTRs with high base-pairing probabilities based on minimum free energy predictions. The 5′ UTR is predicted to contain an internal ribosome entry site (IRES). Coding regions are depicted as boxes. Dashed lines indicate approximate cleavage positions within the polyprotein. Gray regions lack recognizable protein domain structures and do not share sequence similarity with any other organism. Other colors denote regions that have homology to other picornaviruses. Names in boxes are based on similarity to other picornavirus proteins or are inferred based on the conserved nature of picornavirus genome structures. Numbers below the boxes refer to nucleotide positions.

(B) Normalized virus counts in the 8 dpf RNA-seq samples that were used to identify ZfPV. Samples are defined based on spontaneous expression of the *isg15:GFP* reporter. Only GFP-positive animals contained virus reads.

(C) Differential gene expression in the same *isg15:GFP* animals spontaneously expressing GFP compared to GFP-negative animals. Genes that were differentially expressed greater than 2-fold with adjusted $p < 0.05$ are plotted in red. A subset of induced genes are labeled.

(D) Gene expression in response to experimentally administered interferon or chikungunya virus [27] compared to naturally occurring ZfPV infection. 971 genes were significantly induced in at least one condition with adjusted $p < 0.05$.

See also Figures S1–S4 and Tables S1, S2, S3, and S4.

by 8 days post-fertilization (8 dpf), which typically occurred at low frequency among animals from clutches in shared environments (Figure 1E). The distribution of spontaneous expression across multiple independent clutches was skewed toward low frequencies but varied widely (Figure 1F). Spontaneous GFP expression in *isg15:GFP* animals was most evident in tissues exposed to the environment, particularly the intestine and other epithelial surfaces (Figure 1G).

## *isg15:GFP* Animals Couple Virus Discovery with Extensive Antiviral Host Responses

We hypothesized that the spontaneous GFP expression we observed in *isg15:GFP* animals reflected instances of virus infection. To identify prospective viruses, we isolated RNA from GFP-positive and GFP-negative *isg15:GFP* and wild-type (WT) animals for RNA sequencing (RNA-seq), generating approximately 175 million reads. We assembled reads that did not map to the zebrafish genome into approximately 750,000 larger contigs and assigned a preliminary species identity for each contig by taking the best hit from nucleotide or amino acid sequence similarity searches. The majority of these contigs (92%) were most similar to zebrafish sequences (Figure S1), which evidently failed to map to the zebrafish genome as individual reads during alignment. Contigs with similarity to bacteria sequences were mostly from genera that are known constituents of the zebrafish microbiome [25]. 5% of the contigs were short (150 bp) and highly repetitive, with no significant identity to any known sequences. One contig 1,182 bp in length had 33% amino acid identity (e-value $5 \times 10^{-40}$) to the helicase of rosavirus 2, a picornavirus recovered from humans [26]. No other contigs had significant similarity to known exogenous viruses.

We used the contig sequence with similarity to rosavirus 2 to design rapid amplification of cDNA ends (RACE) experiments that we carried out with cDNA template from spontaneously GFP-positive *isg15:GFP* animals. We used the resulting sequences to assemble a non-segmented positive-sense RNA genome of approximately 8 kbp (GenBank: MN524064) that encodes a polyprotein with features that are characteristic of picornaviruses (Figure 2A). Sequences in the 5′ UTR showed evidence for strong local secondary RNA structures. An internal ribosome entry site (IRES) was predicted within the 5′ UTR, although no sequence similarity between the 5′ UTR and the IRES elements of other picornaviruses was observed. The structural and some nonstructural genes have sequence similarity to other picornaviruses, but portions of the genome do not share significant nucleotide or amino acid sequence similarity with anything and lack recognizable protein domains. These observations indicate that the virus we identified from *isg15:GFP* animals is distantly related to known picornaviruses.

Subsequent to our sequence analyses described above, another study published a picornavirus sequence that was obtained through metagenomic sequencing of enriched viral capsid-protected nucleic acids from the intestine of an asymptomatic adult zebrafish [22]. The authors of the study called this virus strain zebrafish picornavirus (ZfPV), which is 97% identical to the nucleotide sequence of the virus that we assembled from *isg15:GFP* larvae. We describe the picornavirus in the present study as ZfPV for the sake of consistency.

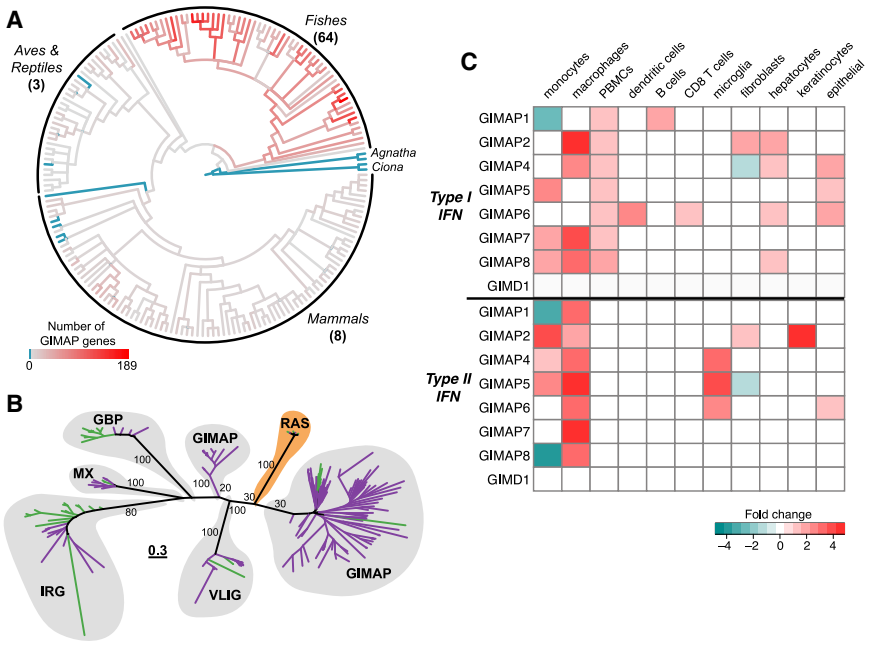



**Figure 3. A Large Family of GIMAP Genes Induced by ZfPV Is a Major Component of the IFN-Inducible Immune GTPase Superfamily**

(A) Number of GIMAP genes in all vertebrate genomes represented in Ensembl. Blue branches indicate lineages that do not contain any genes with AIG1 domains. All other lineages contain at least one GIMAP gene, with numbers scaled by the intensity of red. The median number of GIMAP genes identified among all species within each vertebrate group is shown next to the group name. (B) Maximum likelihood phylogeny of the GTP-binding domains from GIMAP proteins and the interferon-inducible GTPase superfamily proteins from zebrafish (purple branches) and mouse (green branches). Included are domains from all members of the GTPase of immunity-associated protein (GIMAP), very large inducible GTPase (VLIG), immunity-related GTPase (IRG), myxoma resistance (MX), and guanylate-binding protein (GBP) subfamilies. The RAS GTPase family (orange) is included as an outgroup and is not regulated by virus infection or interferon expression. Bootstrap support values are shown at branches leading to each subfamily. Scale is in estimated substitutions per site.

(C) Expression of human GIMAP genes in cells treated with interferons compared to untreated controls compiled from the Interferome database [31]. The median fold change was plotted from all combined time points post-treatment where data from more than one time point was available.

To assess the evolutionary relationship of ZfPV to other members of the *Picornaviridae* family, we performed phylogenetic analyses, including at least one species from all ratified picornavirus genera [28] and all available sequences from picornaviruses that have been isolated from fish (Figure S2). Bayesian and maximum likelihood phylogenies were largely congruent (Figures S3A and S3B). We identified a large clade of viruses that appears to be fish specific, which does not include ZfPV. ZfPV is part of a clade of viruses that infects a mixture of other vertebrate groups. The examples of incongruities between viruses and host groups in this phylogeny are consistent with patterns of virus host switching, which was previously estimated to occur frequently in picornavirus evolution [29]. The Wenling triplecross lizardfish picornavirus is the sister species to ZfPV in this phylogeny, although they only share 36% amino acid identity in a pairwise alignment of 3CD (the precursor of mature 3C protease and 3D polymerase) and are therefore only distant relatives (Figure S4). Overall, these observations demonstrate that ZfPV is highly diverged from the other picornaviruses that are currently known.

The RNA-seq datasets we generated were prepared from co-housed *isg15:GFP* and non-transgenic sibling larvae that were either GFP negative (both transgenic and non-transgenic siblings) or GFP positive (only transgenic) at the time of collection. To gauge the specificity of the *isg15:GFP* reporter in these experiments, we quantified the number of virus reads in both groups. Only GFP-positive datasets contained ZfPV reads (Figure 2B), demonstrating that induction of the *isg15:GFP* reporter tracked with the presence of virus reads. This clear separation provided a unique opportunity to use the same RNA-seq datasets that we used to discover ZfPV to also measure an apparent host response to a naturally occurring virus infecting zebrafish. We identified more than 500 host genes (at an adjusted $p < 0.05$) induced by at least 2-fold in ZfPV-positive *isg15:GFP* animals compared to ZfPV-negative animals (Figure 2C; Table S1). The most significant functional classes that were enriched among genes induced in ZfPV-positive animals were related to interferon signaling and defensive responses against viruses (Figure S4). These results reveal that, beyond the correlation of *isg15:GFP* and virus signal, ZfPV-positive larvae express an extensive set of antiviral genes compared to ZfPV-negative larvae.

We next examined how the host responses associated with naturally occurring ZfPV sequences compared to the transcriptional changes induced by experimentally administered interferon or viruses. A previous RNA-seq study found that zebrafish larvae express many genes in response to injection with interferon protein or the human pathogen chikungunya virus [27]. We found that ZfPV infections induce an overlapping set of 130 genes but also hundreds of genes that are not induced by either of these experimental treatments (Figure 2D; Table S2). These comparisons highlight a common set of antiviral response genes in zebrafish and reveal that naturally occurring viruses are associated with a specific signature of gene expression not previously captured by experimental infections with heterologous viruses or treatment with interferons.

## ZfPV-Induced GIMAPs Are Expanded in Fish and Interferon-Regulated in Humans

The enrichment for interferon signaling and antiviral immunity we observed for genes induced by ZfPV (Figure S4) led us to investigate their functional potential by evaluating the representation of InterPro protein domains. We found the strongest enrichment for members of the large family of immunity-related NACHT-domain and leucine-rich-repeat-containing (NLR) proteins (Table S3). There are more than 300 NLR genes in zebrafish, which

have previously been hypothesized to function in pathogen recognition and inflammasome activity [30].

Intriguingly, we observed a significant enrichment for the GTP-binding avrRpt2-induced gene 1 (AIG1) domain (InterPro: IPR006703) among ZfPV-induced genes and found that 15 of more than 100 genes containing this domain in the zebrafish genome (Ensembl release 98) were significantly upregulated in our analysis of ZfPV-infected fish (Table S1). The AIG1 domain is found in guanosine triphosphatase (GTPase) of immunity-associated proteins (GIMAPs), which is encoded by eight genes in the human genome. We surveyed the repertoires of genes containing AIG1 domains in all vertebrate genomes in Ensembl and found that GIMAP genes were highly expanded in fish species compared to other vertebrate groups (Figure 3A). These patterns suggest that the large expansion of GIMAPs in zebrafish is an evolutionary innovation that might be crucial for antiviral responses in fish and other species.

GIMAPs are known to regulate lymphocyte development and survival in mammals [32–34], but other roles in antiviral immunity have not been previously explored. The induction of GIMAPs by virus infection in zebrafish larvae suggests that GIMAP genes might represent a distinct branch of the interferon-inducible GTPase superfamily of effector genes that act in cell-autonomous immunity [31]. A phylogeny of the GTP-binding domains from the zebrafish and mouse proteins in this superfamily revealed clear separation between subfamilies and also highlighted expansions in the zebrafish myxoma resistance (Mx) and very large inducible GTPase (VLIG) interferon-inducible subfamilies (Figure 3B). Additionally, we queried the Interferome database [35] for human GIMAP expression and identified significant responses to interferon treatment in a wide variety of cell types (Figure 3C). These observations suggest that the large expansion of GIMAP genes is unique to fish but that these genes may encode antiviral effectors that also serve cell-autonomous immune functions in humans.

## ZfPV Is Widespread and Alters Gene Expression in the Sequence Read Archive

Our observations of spontaneous induction of GFP expression in independent clutches of *isg15:GFP* larvae were made separately over the course of several months (Figure 1F), which indicates that ZfPV is endemic to our research facility. We surveyed the prevalence of ZfPV in other research facilities by searching for ZfPV reads in 726 zebrafish RNA-seq datasets in the Sequence Read Archive (SRA) and found that 92 contained ZfPV reads (Table S5). We identified virus reads in sequencing libraries prepared by polyA+ selection, rRNA depletion, or by random selection. Our search revealed widespread prevalence of ZfPV in research facilities around the world (Figure 4A). ZfPV reads were identified in RNA-seq datasets derived from a variety of zebrafish strains that were housed in various locations. We were particularly interested by the identification of ZfPV reads in a dataset (SRR891512) from a zebrafish line called Assam that appears to have been recently caught from its natural habitat in Northeastern India or was recently derived from the wild [36, 37]. The identification of ZfPV reads in this dataset supports the possibility that ZfPV also infects zebrafish in natural habitats.

We identified ZfPV in RNA-seq datasets that were derived from several different zebrafish tissues, which allowed us to evaluate the tissue tropism of the virus. The intestine had the highest number of virus reads (Figure 4B), which is consistent with complementary *in situ* hybridization analyses [22] and indicates that ZfPV may predominate as an enteric virus. However, we also observed ZfPV reads in datasets from several additional tissues, including those exposed to the environment, such as the gills and skin, but also internal tissues, including the brain. In contrast, no virus reads were identified in datasets from eggs or testes.

The widespread prevalence of ZfPV in zebrafish datasets indicates that zebrafish are commonly infected. To assess potential host range beyond zebrafish, we searched for ZfPV in all available datasets in the SRA from fish intestinal tissues, which included 775 RNA-seq datasets from more than 100 species of fish, and found no evidence for ZfPV in these hosts (Figure S5), demonstrating that ZfPV is specific to sequencing datasets from zebrafish among all fish species currently represented in the SRA.

Our identification of ZfPV sequences in various previously published RNA-seq datasets led us to ask whether the presence of virus can alter zebrafish gene expression in these studies. To extend our analysis of ZfPV-induced genes in *isg15:GFP* animals to previously published RNA-seq datasets from fish not previously known to contain virus, we searched for datasets with biological replicates that vary in the presence of ZfPV. We identified datasets from a developmental time course (BioProject PRJEB12982) [38] that included multiple 5 dpf replicates that varied in the presence of ZfPV and identified a small set of genes that were significantly induced in the ZfPV-positive replicates compared to the ZfPV-negative replicates (Figure 4C). Of the 13 genes significantly induced by ZfPV in 5 dpf larvae, 12 were also induced in ZfPV-positive 8 dpf *isg15:GFP* larvae (Figure 4D). Furthermore, none of these genes were significantly induced in 5 dpf larvae that were infected with the pathogenic bacteria *Mycobacterium marinum* (Figure 4D; BioProject PRJNA398644) [39] and as such constitute a specific response to virus infection. These observations reveal a signature of ZfPV infection in RNA-seq datasets and highlight a modest but specific example of how this widespread virus can alter zebrafish gene expression studies.

## The CG2 Zebrafish Strain Harbors Higher Levels of ZfPV Than Tübingen Animals

Two SRA datasets from BioProject PRJNA280983 [40] stood out during our survey for ZfPV, in that they had exceptionally high numbers of virus reads (Figure 5A). These datasets were derived from the clonal CG2 strain of zebrafish [41] and are not plotted in Figure 4B because they were generated from a mixture of adult tissues. The CG2 strain is genetically distinct from Tübingen (Tüb) and other WT lines in that CG2 animals harbor an alternative major histocompatibility complex (MHC) haplotype, which has previously been predicted to confer differences in immunity-related phenotypes [40, 42, 43]. When we aligned reads from the two CG2-derived datasets to the ZfPV genome, we recovered the entire genomic sequence up to nearly 165× coverage (Figure 4A; GenBank: BK011179).

In comparison to the CG2 datasets, the majority of RNA-seq datasets that we queried in the SRA either had very few or no ZfPV reads relative to total (Figure 4B). This observation is

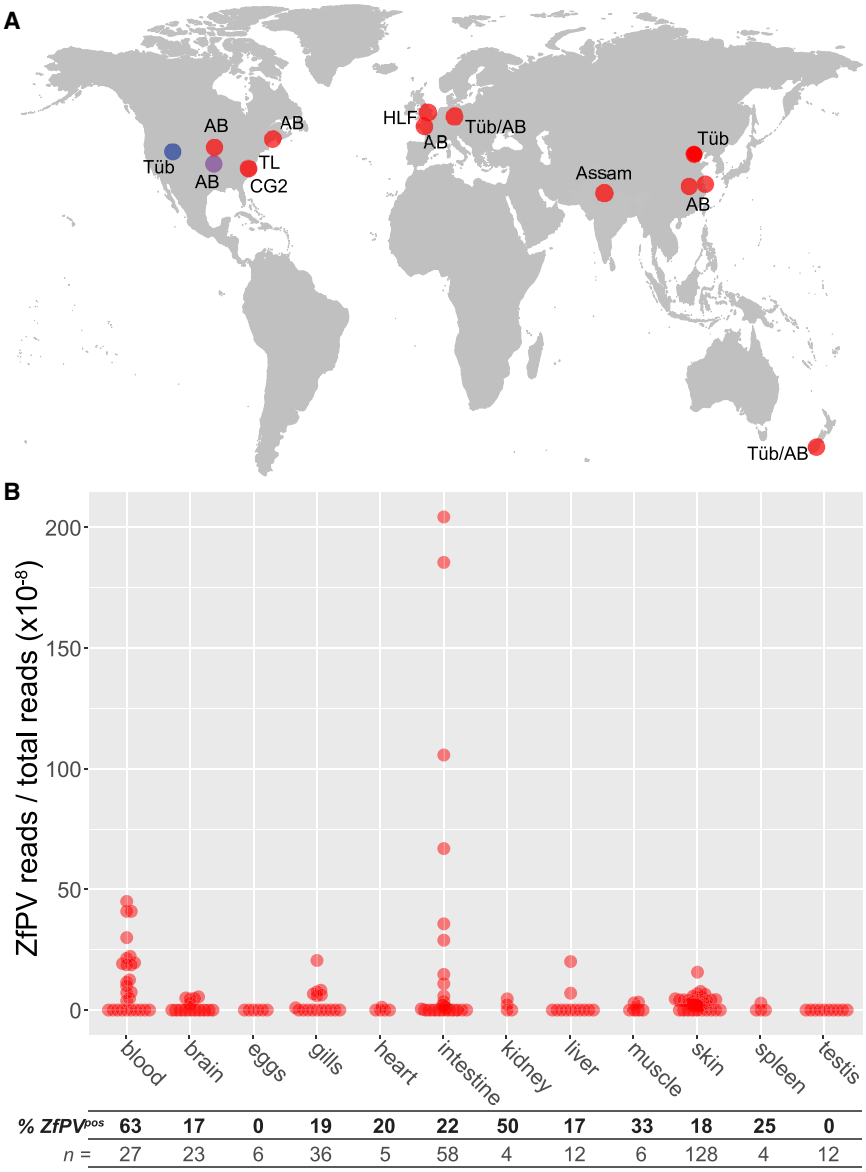



**Figure 4. ZfPV Is Widespread in Publicly Available RNA-Seq Datasets and Alters Zebrafish Gene Expression**

(A) Locations of institutions that have submitted zebrafish RNA-seq projects to the Sequence Read Archive (SRA) that contain reads from ZfPV. Red dots denote locations in which ZfPV reads were identified in the SRA. The blue dot denotes the location of the present study, where ZfPV was initially discovered. The purple dot denotes the location of a recent study where ZfPV was also discovered [22]. Names next to the dots refer to the strain of zebrafish that was sequenced.

(B) Number of ZfPV reads normalized by library size that were identified in the SRA from adult tissues. Each dot represents a single sequencing sample. The percentage of SRA samples that contained ZfPV reads and the total number of samples that were searched for each tissue are shown beneath the plot.

(C) Differential gene expression in 5 dpf replicates that contained ZfPV reads compared to replicates that did not from BioProject PRJEB12982. Genes that were differentially expressed greater than 2-fold with adjusted $p < 0.05$ are plotted in red.

(D) Comparison of gene sets induced by ZfPV infection in larvae collected at 8 dpf (white circle, current study), at 5 dpf (gray circle, SRA reads from [38]), or induced by *Mycobacterium marinum* infection in larvae collected at 5 dpf (blue circle, reads from [39]). Genes were defined as induced if expressed greater than 2-fold over matched control. Groups with adjusted $p < 0.05$ and numbers of genes are plotted.

See also Figure S5 and Table S5.

perhaps not surprising, given that all of these datasets were generated to investigate host gene expression and that virus RNAs typically account for a miniscule fraction of the total RNAs extracted from vertebrates for metagenomic sequencing [5, 15]. To illustrate the relative rarity of virus reads even in the tissue with the highest ZfPV signal, we combined all 12 intestine-derived datasets that contained ZfPV reads from Tüb animals and aligned these to the ZfPV genome. In contrast to the two CG2 datasets, aligning all 12 ZfPV-positive Tüb intestine datasets yielded low coverage across the ZfPV genome, with some regions that lacked coverage altogether (Figure 5A).

The higher degree of ZfPV coverage we observed for reads from CG2 datasets compared to Tüb datasets led us to hypothesize that the CG2 strain might harbor higher levels of virus than Tüb animals. To test this hypothesis, we isolated intestines, gills, and brains from adult CG2 and Tüb animals, extracted RNA, and generated cDNA to measure virus levels by quantitative real-time PCR. All CG2 animals analyzed in these experiments were reared in a facility at Massachusetts General Hospital, although the Tüb animals were age matched to the CG2 animals but reared at two separate facilities at the University of Utah. Despite the disparate nature of sample selection, we detected ZfPV in the intestines and gills of every animal included

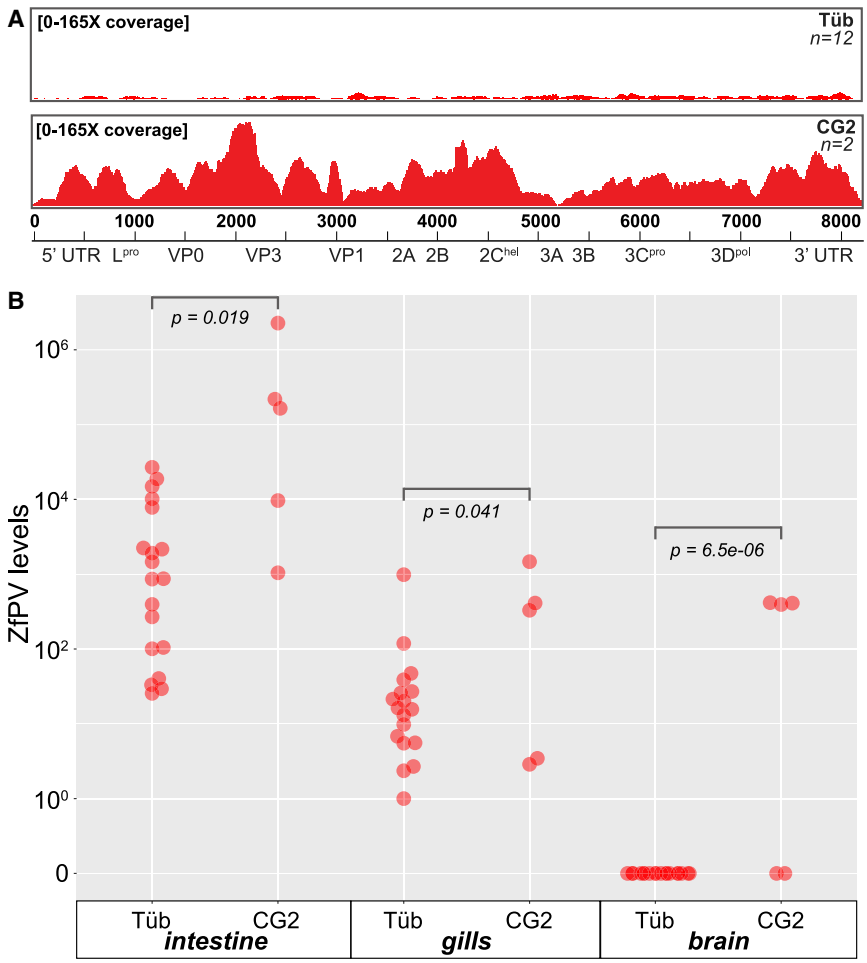



**Figure 5. Virus Levels Are Higher in the CG2 Strain Compared to Tübingen (Tüb) Zebrafish**

(A) Coverage of the ZfPV genome by reads derived from all 12 Tüb intestine SRA datasets that had virus reads from Figure 3 compared to reads in the SRA derived from a single CG2 animal. Nucleotide positions and genomic features of ZfPV are shown below the coverage plots. Scale is in number of reads that cover each position.

(B) Quantitative real-time PCR quantification of ZfPV levels in tissues dissected from adult Tüb and CG2 animals. Each dot represents the relative amount of virus in one individual. Virus levels are scaled by normalizing to the sample with the lowest non-zero amount of ZfPV. p values are reported from unpaired t tests assuming equal variance.

See also Tables S4 and S5.

in these experiments (Figure 5B). In support of our hypothesis, we observed higher levels of ZfPV in CG2 for all three tissues surveyed, including the brain, where virus was only detected in CG2 animals. Although we cannot exclude that environmental or virus differences might influence these comparisons, these data are consistent with the large number of ZfPV reads we observed in RNA-seq datasets from a CG2 animal (reared in a facility at North Carolina State University). Together, these observations reflect a trend for the clonal and potentially immunocompromised strain, highlighting candidate host genetic variation influencing virus levels.

## Experimental Manipulation of ZfPV Infection and Transmission

Every adult animal we examined contained ZfPV transcripts (Figure 5B), which contrasts with variable frequencies of spontaneous GFP expression observed in 8 dpf transgenic larvae (Figure 1F). These discrepancies motivated three parallel experiments investigating how ZfPV establishes and transmits infections (outlined in Figure 6A). First, to test whether zebrafish were naturally infected only after hatching, we bleached eggs at 24 hpf, a common strategy for eliminating pathogenic microbes during rearing [44], and observed a striking reduction in GFP expression at 8 dpf as compared to conventionally reared siblings (Figure S6). Infrequent and low-level GFP expression observed in bleached animals was restricted to a small region of the intestine (GFP-low; Figure S6), and we failed to amplify ZfPV sequences from any bleached GFP-negative or GFP-low animals (Figure 6B), indicating that these patterns do not signal productive infections. In contrast, siblings from the same clutches reared without bleach treatment exhibited high expression of GFP (GFP-positive; Figure S6) that tracked with high levels of *isg15* and ZfPV transcripts (Figure 6B). These results demonstrate that standard bleaching protocols can prevent naturally occurring infections of ZfPV and suggest a horizontal mode of virus transmission.

We next exposed bleach-treated zebrafish to a filtrate from ZfPV-infected animals to test experimental administration of infection. Because ZfPV was present in all adult intestinal tissues sampled (Figure 5B), we homogenized intestines from three adult fish and applied a 0.45 μm filtrate of this homogenate to dishes of 5 dpf transgenic larvae bleached at 24 hpf. In sharp contrast to controls, bleached animals exposed to filtrate were frequently GFP positive and expressed high levels of *isg15* and ZfPV transcripts (Figures 6B and S6). These results suggest that filtrate containing ZfPV particles can be extracted from infected adult animals and administered to zebrafish larvae to establish infections.

To test ZfPV transmission under natural conditions, we next co-housed GFP-positive larvae with bleached GFP-negative larvae and observed high levels of GFP expression and ZfPV transcripts in both donor and recipient animals after 7 days (Figures 6B and S6). In contrast, no GFP expression or ZfPV transcript was detected in animals 7 days after stocking plates in equal numbers with two separately bleached clutches of GFP-negative animals (Figures 6B and S6). We also observed that some co-housed, GFP-positive fish lost GFP signal by the end of the experiment (Figure S6). This result suggests that some animals might clear infections or that the immune response to ZfPV

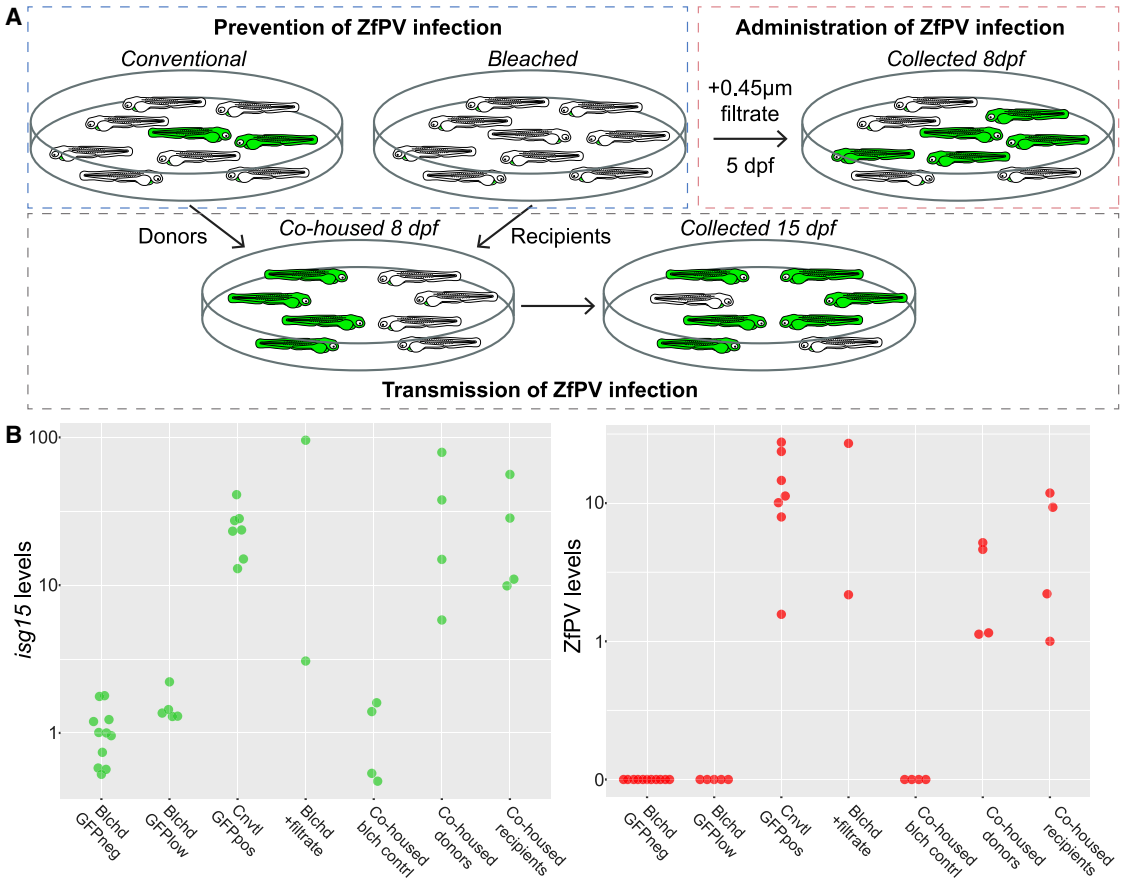



**Figure 6. ZfPV Is Transmitted Horizontally and Can Be Experimentally Controlled**

(A) Three strategies for investigating ZfPV infection and transmission in zebrafish larvae. Half a clutch of transgenic *isg15:GFP* embryos were bleached and compared to conventionally reared siblings at 8 dpf for GFP expression (blue box). Alternatively, bleached embryos were treated with a 0.45 µm filtrate from adult intestinal tissue at 5 dpf and analyzed at 8 dpf (pink box). To assess transmission of ZfPV between larvae, conventionally reared animals exhibiting spontaneous GFP expression were co-housed with bleached animals at 8 dpf. Additionally, two separate clutches of bleached animals were co-housed as a control (not depicted). GFP expression in the donors and recipients was observed at 15 dpf (gray box).

(B) Quantitative real-time PCR quantification of *isg15* and ZfPV levels in zebrafish larvae from the experiments depicted above. Each dot represents the relative amount of *isg15* gene expression or virus in a pool of 25 larvae. A small fraction of bleached animals occasionally exhibited low levels of GFP induction at the time of collection (see Figure S6) and is plotted separately. *isg15* levels are scaled by normalizing to the average level quantified in bleached samples. Virus levels are scaled by normalizing to the sample with the lowest non-zero amount of ZfPV.

See also Figure S6 and Table S4.

might be suppressed as infections progress in some cases. These results demonstrate that GFP-positive larvae can transmit ZfPV infection to uninfected animals in their surroundings, which outlines a new experimental means of investigating natural modes of virus infection and transmission in a tractable vertebrate research organism.

## DISCUSSION

This study reveals naturally induced antiviral responses by combining virus discovery with visualization of host immune gene activation. Although transcriptional responses to experimentally administered viruses have been reported in zebrafish [27, 45], few studies track host gene expression during infections by naturally occurring viruses in any vertebrate system. Broad assessment of the ZfPV-induced gene set assigned a clear signature of interferon- and virus-response-related functions, but surprisingly, more than 300 of the roughly 500 genes induced by these naturally occurring infections were not upregulated by experimental treatment with interferon or a heterologous virus. Many of these uniquely induced genes are members of immune gene families that are highly expanded in zebrafish, such as the previously characterized NLR [30] and tripartite motif (TRIM) [46] families but also the GIMAP gene family, which we find is expanded in several fish species and may serve antiviral roles in vertebrates. Little is known regarding roles of individual genes in these families, and the ZfPV-induced gene set therefore opens new avenues for prioritizing the study of diverse and uncharacterized immune functions.

Zebrafish and humans diverged from a common ancestor with interferon signaling more than 400 mya [23, 47], and studies of zebrafish therefore provide a means for discovering overlapping [27] and distinct [30] gene-encoded functions in vertebrate antiviral immunity. In this study, we catalog a large expansion of the

GIMAP gene family in fish species and highlight an enrichment of GIMAP genes among those induced by ZfPV infection. GIMAPs are critical for lymphocyte apoptosis and maturation in mammals [32–34], but antiviral roles have not been widely considered. GIMAPs may act by a similar mechanism as the interferon-inducible Mx dynamin-like proteins [48], which appear to share both a common GTPase ancestor [49] and a propensity to oligomerize [49]. Our work places the GIMAPs within the interferon-inducible GTPase superfamily and suggests an unexplored involvement of GIMAPs as effectors in vertebrate-virus interactions.

Large expansions in gene families encoding innate immune functions are characteristic of lineages lacking canonical adaptive immune systems, including invertebrates like the purple sea urchin [50]. Zebrafish do not develop an adaptive immune system until several weeks after hatching [51], and we found that they are naturally infected by viruses before this time. We speculate that the expansions of GIMAP, NLR [30], TRIM [46], and other zebrafish immune gene families may reflect pressures applied by pathogens during the long period of developmental time that zebrafish rely exclusively on innate immunity. One strength of the ZfPV-induced gene set we report is that it highlights individual members among massive immune gene families as specific targets for future functional studies during natural infections. The methods we developed for manipulating ZfPV infections also provide new opportunities for assessing the functions of specific host immune responses in future studies.

By carrying out virus discovery in a widely used research organism, we were also able to extend our discovery of ZfPV in zebrafish larvae from our research facility to locations around the world. Our use of publicly available datasets is complementary with recent work that used PCR to detect ZfPV in adult zebrafish samples derived from various research institutions [22]. Furthermore, our observation of virus reads in a zebrafish from India supports the idea that ZfPV circulates in the wild. This provides the opportunity to study natural virus-zebrafish interactions and build on previous *in vivo* studies of zebrafish infections by several viruses isolated from other organisms [19, 45]. Although definitive classification of ZfPV as a virus that infects wild zebrafish awaits direct sampling of animals in natural habitats, the widespread prevalence of ZfPV in research facilities alone offers the possibility of studying host-virus evolution in a variety of controlled environments.

One of the notable findings from our study was the exceptional number of virus reads in datasets from the CG2 strain of zebrafish (Figure 4A). This observation led us to survey ZfPV levels in tissues that we extracted from several CG2 and Tüb adult animals, which revealed that ZfPV levels are higher in the CG2 strain than in Tüb animals (Figure 4B). Tüb and CG2 animals are genetically distinct in at least one important way: Tüb animals carry a heterogeneous mix of distinct MHC core haplotypes and CG2 animals are homozygous for a single core MHC haplotype [42]. Classical MHC molecules are important for presentation of antigens to T cells and are critical for driving adaptive immune responses to infections [52]. Variation at the MHC locus is known to alter susceptibility to infectious diseases in mammals, birds, and fish [52–54], and zebrafish genomes contain extensive MHC gene diversity [43]. Thus, our observation that the CG2 strain tends to harbor higher levels of ZfPV than Tüb animals leads us to speculate that there is host genetic variation in susceptibility to infection and points to specific factors that can be tested in controlled experiments.

ZfPV appears to infect primarily as an enteric pathogen of zebrafish, but the full range of tissue tropism is complex (Figure 3B). This is common for picornaviruses as exemplified by poliovirus, which is thought to primarily cause gastrointestinal infections in humans but can then spread to the blood and occasionally other tissues, including the central nervous system, in a manner that is influenced by the host immune system and other factors [55]. Our observations are consistent with this pattern for ZfPV, which we detected at the highest levels in the intestine but also in several other tissues, including the brain (Figure 3B). Recently published work provided complementary evidence of ZfPV infection in sections of adult intestinal tissue by carrying out *in situ* hybridization analyses [22] but did not report infection of other tissues. Intriguingly, we detected ZfPV in brain tissue that we collected from CG2 fish, but not from Tüb fish (Figure 4C). We interpret this expanded tissue tropism as linking the susceptibility of the CG2 strain to a dynamic infection process that involves both virus and host factors. Therefore, ZfPV infections of CG2 fish may provide a useful new model system for investigating mechanisms of neurotropism common to picornavirus epidemics in humans [56].

The possibility that unrecognized infections by naturally occurring viruses might cause unexplained variation in experiments with zebrafish has been a long-standing concern [57]. We observed gene expression differences within experimental replicates in previously published RNA-seq datasets likely caused by the presence of ZfPV (Figure 3C), and although the number of genes affected may be limited in some cases, these observations support the idea that unrecognized virus infections might contribute to experimental variation in zebrafish studies. The full range of effects that ZfPV has in zebrafish gene expression studies is unclear, as the majority of RNA-seq experiments in the SRA do not include enough replicates with and without virus reads to quantify significant differences. We predict that ZfPV could have a particularly strong effect on single-cell RNA-seq data where specific cell-type expression patterns can be highly sensitive to environmental variables, such as infection [58]. Monitoring the presence of ZfPV in zebrafish gene expression studies to account for host responses to cryptic infections should now be straightforward and informative.

We determined that ZfPV is transmitted horizontally in natural and experimental infections. Several lines of evidence support the idea that infections first occur upon exposure to the environment after hatching. First, we did not observe conspicuous induction of the *isg15:GFP* reporter until approximately 8 dpf. Second, the earliest we identified ZfPV sequences in SRA data was from 5 dpf larvae. Third, surface sterilization of the egg chorion was sufficient to prevent *isg15:GFP* induction and ZfPV infection. Zebrafish larvae typically hatch from their chorions between 2 and 3 dpf, and natural bacterial colonization of the gut begins at 4 dpf [59]. We infer that ZfPV infections and subsequent host responses first occur at this time. These results reveal an early signature of host responses to ZfPV infection and outline a strategy for eliminating virus from zebrafish embryos and larvae that involves a step also used for generating germ-free zebrafish [59], which should be of practical use to

researchers accounting for the microbial constituents that commonly associate with zebrafish.

ZfPV is likely one of many viruses circulating in zebrafish populations, and *isg15:GFP* fish could be useful sentinels for discovering additional viruses that infect zebrafish from a variety of environments. Variations on this strategy could lend further sensitivity and versatility to the discovery process, such as through the combinatorial use of promoters that report on distinct aspects of infection or by generating reporters in immunocompromised host backgrounds. More generally, our study outlines an adaptable strategy that could be deployed in many organisms to illuminate the ubiquitous but typically invisible engagements between viruses and hosts.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Transgenesis
  - Microscopy
  - Induction of type I interferon expression
  - Quantification of spontaneous GFP expression
  - RNA sequencing
  - Virus genome assembly and annotation
  - Phylogenetic analyses
  - Searching the Sequence Read Archive for Zebrafish picornavirus
  - Quantification of virus and isg15 levels by quantitative real-time PCR
  - Differential gene expression analyses
  - Experimental prevention and administration of ZfPV infection
  - Experimental transmission of ZfPV infection
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.04.031.

## ACKNOWLEDGMENTS

We thank Diego Mallo and Tyler Starr for input on phylogenetic analyses; Michelle Culbertson, Kristen Davenport, Diane Downhour, and Zoë Hilbert for critical feedback on the manuscript; Kristen Kwan and members of her lab for sharing Tol2 transgenesis reagents; and David Langenau and Alexandra Veloso for sharing the CG2 fish. We also thank the University of Utah Health Sciences Core Facilities for zebrafish husbandry and imaging support and the Huntsman Cancer Institute High-Throughput Genomics Shared Resource for library preps and sequencing. This work was supported by grants awarded by the National Institutes of Health to K.M.B. (5T32AI055434) and to N.C.E. (R01GM114514 and R35GM134936). Additional support includes funding awarded to N.C.E. from the Burroughs Wellcome Fund (1015462) and the H.A. and Edna Benning Presidential Endowed Chair.

## AUTHOR CONTRIBUTIONS

Conceptualization, K.M.B. and N.C.E.; Data Curation, K.M.B.; Formal Analysis, K.M.B.; Validation, K.M.B.; Investigation, K.M.B., M.C.R., and J.A.G.; Visualization, K.M.B.; Methodology, K.M.B.; Writing – Original Draft, K.M.B.; Writing – Review & Editing, K.M.B., M.C.R., J.A.G., and N.C.E.; Supervision, J.A.G. and N.C.E.; Resources, N.C.E.; Project Administration, N.C.E.; Funding Acquisition, N.C.E.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 2, 2020
Revised: February 28, 2020
Accepted: April 14, 2020
Published: May 14, 2020

## REFERENCES

1. Cobián Güemes, A.G., Youle, M., Cantú, V.A., Felts, B., Nulton, J., and Rohwer, F. (2016). Viruses as winners in the game of life. Annu. Rev. Virol. *3*, 197–214.
2. Suttle, C.A. (2005). Viruses in the sea. Nature *437*, 356–361.
3. Li, L., Victoria, J.G., Wang, C., Jones, M., Fellers, G.M., Kunz, T.H., and Delwart, E. (2010). Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J. Virol. *84*, 6955–6965.
4. Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., et al. (2016). Redefining the invertebrate RNA virosphere. Nature *540*, 539–543.
5. Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S., Shen, J.-J., Liu, L., et al. (2018). The evolutionary history of vertebrate RNA viruses. Nature *556*, 197–202.
6. Finkbeiner, S.R., Allred, A.F., Tarr, P.I., Klein, E.J., Kirkwood, C.D., and Wang, D. (2008). Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog. *4*, e1000011.
7. Greninger, A.L., Runckel, C., Chiu, C.Y., Haggerty, T., Parsonnet, J., Ganem, D., and DeRisi, J.L. (2009). The complete genome of klassevirus - a novel picornavirus in pediatric stool. Virol. J. *6*, 82.
8. Lysholm, F., Wetterbom, A., Lindau, C., Darban, H., Bjerkner, A., Fahlander, K., Lindberg, A.M., Persson, B., Allander, T., and Andersson, B. (2012). Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS ONE *7*, e30875.
9. Lienenklaus, S., Cornitescu, M., Zietara, N., Łyszkiewicz, M., Gekara, N., Jabłónska, J., Edenhofer, F., Rajewsky, K., Bruder, D., Hafner, M., et al. (2009). Novel reporter mouse reveals constitutive and inflammatory expression of IFN-beta in vivo. J. Immunol. *183*, 3229–3236.
10. Palha, N., Guivel-Benhassine, F., Briolat, V., Lutfalla, G., Sourisseau, M., Ellett, F., Wang, C.-H., Lieschke, G.J., Herbomel, P., Schwartz, O., and Levraud, J.P. (2013). Real-time whole-body visualization of Chikungunya Virus infection and host interferon response in zebrafish. PLoS Pathog. *9*, e1003619.
11. Félix, M.-A., Ashe, A., Piffaretti, J., Wu, G., Nuez, I., Bélicard, T., Jiang, Y., Zhao, G., Franz, C.J., Goldstein, L.D., et al. (2011). Natural and experimental infection of Caenorhabditis nematodes by novel viruses related to nodaviruses. PLoS Biol. *9*, e1000586.
12. Roediger, B., Lee, Q., Tikoo, S., Cobbin, J.C.A., Henderson, J.M., Jormakka, M., O'Rourke, M.B., Padula, M.P., Pinello, N., Henry, M., et al. (2018). An atypical parvovirus drives chronic tubulointerstitial nephropathy and kidney fibrosis. Cell *175*, 530–543.e24.
13. Webster, C.L., Waldron, F.M., Robertson, S., Crowson, D., Ferrari, G., Quintana, J.F., Brouqui, J.-M., Bayne, E.H., Longdon, B., Buck, A.H.,

et al. (2015). The discovery, distribution, and evolution of viruses associated with Drosophila melanogaster. PLoS Biol. *13*, e1002210.

14. Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L., and Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. Nature *546*, 646–650.
15. Geoghegan, J.L., Di Giallonardo, F., Cousins, K., Shi, M., Williamson, J.E., and Holmes, E.C. (2018). Hidden diversity and evolution of viruses in market fish. Virus Evol. *4*, vey031.
16. Gentry, R.R., Froehlich, H.E., Grimm, D., Kareiva, P., Parke, M., Rust, M., Gaines, S.D., and Halpern, B.S. (2017). Mapping the global potential for marine aquaculture. Nat. Ecol. Evol. *1*, 1317–1324.
17. Sanders, G.E., Batts, W.N., and Winton, J.R. (2003). Susceptibility of zebrafish (Danio rerio) to a model pathogen, spring viremia of carp virus. Comp. Med. *53*, 514–521.
18. Phelan, P.E., Pressley, M.E., Witten, P.E., Mellon, M.T., Blake, S., and Kim, C.H. (2005). Characterization of snakehead rhabdovirus infection in zebrafish (Danio rerio). J. Virol. *79*, 1842–1852.
19. Van Dycke, J., Ny, A., Conceição-Neto, N., Maes, J., Hosmillo, M., Cuvry, A., Goodfellow, I., Nogueira, T.C., Verbeken, E., Matthijnssens, J., et al. (2019). A robust human norovirus replication model in zebrafish larvae. PLoS Pathog. *15*, e1008009.
20. Bermúdez, R., Losada, A.P., de Azevedo, A.M., Guerra-Varela, J., Pérez-Fernández, D., Sánchez, L., Padrós, F., Nowak, B., and Quiroga, M.I. (2018). First description of a natural infection with spleen and kidney necrosis virus in zebrafish. J. Fish Dis. *41*, 1283–1294.
21. Binesh, C.P. (2013). Mortality due to viral nervous necrosis in zebrafish Danio rerio and goldfish Carassius auratus. Dis. Aquat. Organ. *104*, 257–260.
22. Altan, E., Kubiski, S.V., Boros, Á., Reuter, G., Sadeghi, M., Deng, X., Creighton, E.K., Crim, M.J., and Delwart, E. (2019). A highly divergent picornavirus infecting the gut epithelia of zebrafish (*Danio rerio*) in research institutions worldwide. Zebrafish *16*, 291–299.
23. Secombes, C.J., and Zou, J. (2017). Evolution of interferons and interferon receptors. Front. Immunol. *8*, 209.
24. Seppola, M., Stenvik, J., Steiro, K., Solstad, T., Robertsen, B., and Jensen, I. (2007). Sequence and expression analysis of an interferon stimulated gene (ISG15) from Atlantic cod (Gadus morhua L.). Dev. Comp. Immunol. *31*, 156–171.
25. Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K., and Rawls, J.F. (2011). Evidence for a core gut microbiota in the zebrafish. ISME J. *5*, 1595–1608.
26. Lim, E.S., Cao, S., Holtz, L.R., Antonio, M., Stine, O.C., and Wang, D. (2014). Discovery of rosavirus 2, a novel variant of a rodent-associated picornavirus, in children from The Gambia. Virology *454-455*, 25–33.
27. Levraud, J.P., Jouneau, L., Briolat, V., Laghi, V., and Boudinot, P. (2019). IFN-stimulated genes in zebrafish and humans define an ancient arsenal of antiviral immunity. J. Immunol. *203*, 3361–3373.
28. Zell, R., Delwart, E., Gorbalenya, A.E., Hovi, T., King, A.M.Q., Knowles, N.J., Lindberg, A.M., Pallansch, M.A., Palmenberg, A.C., Reuter, G., et al.; Ictv Report Consortium (2017). ICTV virus taxonomy profile: picornaviridae. J. Gen. Virol. *98*, 2421–2422.
29. Geoghegan, J.L., Duchêne, S., and Holmes, E.C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog. *13*, e1006215.
30. Howe, K., Schiffer, P.H., Zielinski, J., Wiehe, T., Laird, G.K., Marioni, J.C., Soylemez, O., Kondrashov, F., and Leptin, M. (2016). Structure and evolutionary history of a large family of NLR proteins in the zebrafish. Open Biol. *6*, 160009.
31. Kim, B.-H., Shenoy, A.R., Kumar, P., Bradfield, C.J., and MacMicking, J.D. (2012). IFN-inducible GTPases in host cell defense. Cell Host Microbe *12*, 432–444.
32. MacMurray, A.J., Moralejo, D.H., Kwitek, A.E., Rutledge, E.A., Van Yserloo, B., Gohlke, P., Speros, S.J., Snyder, B., Schaefer, J., Bieg, S., et al. (2002). Lymphopenia in the BB rat model of type 1 diabetes is due to a mutation in a novel immune-associated nucleotide (Ian)-related gene. Genome Res. *12*, 1029–1039.
33. Nitta, T., Nasreen, M., Seike, T., Goji, A., Ohigashi, I., Miyazaki, T., Ohta, T., Kanno, M., and Takahama, Y. (2006). IAN family critically regulates survival and development of T lymphocytes. PLoS Biol. *4*, e103.
34. Saunders, A., Webb, L.M.C., Janas, M.L., Hutchings, A., Pascall, J., Carter, C., Pugh, N., Morgan, G., Turner, M., and Butcher, G.W. (2010). Putative GTPase GIMAP1 is critical for the development of mature B and T lymphocytes. Blood *115*, 3249–3257.
35. Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., Chapman, R., and Hertzog, P.J. (2013). Interferome v2.0: an updated database of annotated interferon-regulated genes. Nucleic Acids Res. *41*, D1040–D1046.
36. Patowary, A., Purkanti, R., Singh, M., Chauhan, R., Singh, A.R., Swarnkar, M., Singh, N., Pandey, V., Torroja, C., Clark, M.D., et al. (2013). A sequence-based variation map of zebrafish. Zebrafish *10*, 15–20.
37. Sharma, D., Sehgal, P., Mathew, S., Vellarikkal, S.K., Singh, A.R., Kapoor, S., Jayarajan, R., Scaria, V., and Sivasubbu, S. (2019). A genome-wide map of circular RNAs in adult zebrafish. Sci. Rep. *9*, 3432.
38. White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. eLife *6*, e30860.
39. Hu, W., Yang, S., Shimada, Y., Münch, M., Marín-Juez, R., Meijer, A.H., and Spaink, H.P. (2019). Infection and RNA-seq analysis of a zebrafish tlr2 mutant shows a broad function of this toll-like receptor in transcriptional and metabolic control and defense to Mycobacterium marinum infection. BMC Genomics *20*, 878.
40. Dirscherl, H., and Yoder, J.A. (2015). A nonclassical MHC class I U lineage locus in zebrafish with a null haplotypic variant. Immunogenetics *67*, 501–513.
41. Mizgirev, I.V., and Revskoy, S. (2010). A new zebrafish model for experimental leukemia therapy. Cancer Biol. Ther. *9*, 895–902.
42. McConnell, S.C., Restaino, A.C., and de Jong, J.L.O. (2014). Multiple divergent haplotypes express completely distinct sets of class I MHC genes in zebrafish. Immunogenetics *66*, 199–213.
43. McConnell, S.C., Hernandez, K.M., Wcisel, D.J., Kettleborough, R.N., Stemple, D.L., Yoder, J.A., Andrade, J., and de Jong, J.L.O. (2016). Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. Proc. Natl. Acad. Sci. USA *113*, E5014–E5023.
44. Westerfield, M. (2007). The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish (Danio rerio), Fifth Edition (University of Oregon).
45. Varela, M., Figueras, A., and Novoa, B. (2017). Modelling viral infections using zebrafish: Innate immune response and antiviral research. Antiviral Res. *139*, 59–68.
46. van der Aa, L.M., Levraud, J.-P., Yahmi, M., Lauret, E., Briolat, V., Herbomel, P., Benmansour, A., and Boudinot, P. (2009). A large new subset of TRIM genes highly diversified by duplication and positive selection in teleost fish. BMC Biol. *7*, 7.
47. Hughes, L.C., Ortí, G., Huang, Y., Sun, Y., Baldwin, C.C., Thompson, A.W., Arcila, D., Betancur-R, R., Li, C., Becker, L., et al. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. USA *115*, 6249–6254.
48. Gao, S., von der Malsburg, A., Paeschke, S., Behlke, J., Haller, O., Kochs, G., and Daumke, O. (2010). Structural basis of oligomerization in the stalk region of dynamin-like MxA. Nature *465*, 502–506.
49. Schwefel, D., Fröhlich, C., Eichhorst, J., Wiesner, B., Behlke, J., Aravind, L., and Daumke, O. (2010). Structural basis of oligomerization in septin-like GTPase of immunity-associated protein 2 (GIMAP2). Proc. Natl. Acad. Sci. USA *107*, 20299–20304.
50. Hibino, T., Loza-Coll, M., Messier, C., Majeske, A.J., Cohen, A.H., Terwilliger, D.P., Buckley, K.M., Brockton, V., Nair, S.V., Berney, K.,

et al. (2006). The immune gene repertoire encoded in the purple sea urchin genome. Dev. Biol. *300*, 349–365.

51. Lam, S.H., Chua, H.L., Gong, Z., Lam, T.J., and Sin, Y.M. (2004). Development and maturation of the immune system in zebrafish, Danio rerio: a gene expression profiling, in situ hybridization and immunological study. Dev. Comp. Immunol. *28*, 9–28.
52. Matzaraki, V., Kumar, V., Wijmenga, C., and Zhernakova, A. (2017). The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol. *18*, 76.
53. Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., and Stet, R.J.M. (2003). MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics *55*, 210–219.
54. Hunt, H.D., Jadhao, S., and Swayne, D.E. (2010). Major histocompatibility complex and background genes in chickens influence susceptibility to high pathogenicity avian influenza virus. Avian Dis. *54* (1, Suppl), 572–575.
55. Racaniello, V.R. (2006). One hundred years of poliovirus pathogenesis. Virology *344*, 9–16.
56. Huang, H.-I., and Shih, S.-R. (2015). Neurotropic enterovirus infections in the central nervous system. Viruses *7*, 6051–6066.
57. Crim, M.J., and Riley, L.K. (2012). Viral diseases in zebrafish: what is known and unknown. ILAR J. *53*, 135–143.
58. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-seq. eLife *8*, e43803.
59. Bates, J.M., Mittge, E., Kuhlman, J., Baden, K.N., Cheesman, S.E., and Guillemin, K. (2006). Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. Dev. Biol. *297*, 374–386.
60. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
61. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.
62. Tcherepanov, V., Ehlers, A., and Upton, C. (2006). Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genomics *7*, 150.
63. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. *41*, e121.
64. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36*, W70–W74.
65. Zhao, J., Wu, J., Xu, T., Yang, Q., He, J., and Song, X. (2018). IRESfinder: Identifying RNA internal ribosome entry site in eukaryotic cell using framed k-mer features. J. Genet. Genomics *45*, 403–406.
66. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., and Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. *39*, W13–W17.
67. Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *15*, e1006650.
68. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165.
69. Warren, D.L., Geneva, A.J., and Lanfear, R. (2017). RWTY (R we there yet): an R package for examining convergence of Bayesian phylogenetic analyses. Mol. Biol. Evol. *34*, 1016–1020.
70. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901–904.
71. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
72. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
73. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
74. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.
75. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
76. Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., and Chien, C.-B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev. Dyn. *236*, 3088–3099.
77. Hamming, O.J., Lutfalla, G., Levraud, J.-P., and Hartmann, R. (2011). Crystal structure of Zebrafish interferons I and II reveals conservation of type I interferon structure in vertebrates. J. Virol. *85*, 8181–8187.
78. McCurley, A.T., and Callard, G.V. (2008). Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. BMC Mol. Biol. *9*, 102.
79. Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–1108.
80. White, R.M., Sessa, A., Burke, C., Bowman, T., LeBlanc, J., Ceol, C., Bourque, C., Dovey, M., Goessling, W., Burns, C.E., and Zon, L.I. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell Stem Cell *2*, 183–189.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Bacterial and Virus Strains | | |
| Zebrafish picornavirus strain CZAR/ZfPV/2018/Utah/USA | This manuscript | GenBank MN524064 |
| Chemicals, Peptides, and Recombinant Proteins | | |
| 3-Aminobenzoic acid (Tricaine; MS-222) | Sigma-Aldrich | Cat#127671; CAS#99-05-8 |
| Sodium hypochlorite solution | Millipore Sigma | Cat#425044 |
| Pronase | Sigma-Aldrich | Cat#PRON-RO |
| TRIzol Reagent | ThermoFisher Scientific | Cat#15596018 |
| TRIzol LS Reagent | ThermoFisher Scientific | Cat#10296010 |
| UltraPure Low Melting Point Agarose | ThermoFisher Scientific | Cat#16520050 |
| Critical Commercial Assays | | |
| TruSeq Stranded Total RNA Ribo-Zero H/M/R Gold | Illumina | Cat#RS-122-2301 |
| SMARTer RACE 5′/3′ Kit | Takara Bio Inc. | Cat#634858 |
| SuperScript IV Reverse Transcriptase | ThermoFisher Scientific | Cat#18090010 |
| Power SYBR Green Master Mix | ThermoFisher Scientific | Cat#4367659 |
| Deposited Data | | |
| RNA-sequencing data from zebrafish and other fish species | This manuscript and the NCBI Sequence Read Archive | See Table S5 for all contributed and accessed public datasets |
| Zebrafish picornavirus strain CZAR/ZfPV/2018/Utah/USA, complete genome | This manuscript | GenBank accession MN524064 |
| Zebrafish picornavirus strain NCSZCF/ZfPV/2015/North Carolina/USA, complete genome | This manuscript | GenBank accession TPA: BK011179 |
| Zebrafish picornavirus-1 strain IDEXX/ZfPV-1/2017/USA, complete genome | [22] | GenBank accession MH368041.1 |
| Experimental Models: Organisms/Strains | | |
| Tübingen zebrafish line | Zebrafish International Resource Center, Eugene, OR | Tübingen/Tüb/TU |
| *Tg(isg15:GFP)* zebrafish lines | This manuscript | z207Tg, z208Tg, z209Tg |
| CG2 zebrafish line | David Langenau laboratory, MGH, see [41] | CG2 |
| *roy*$^{a9}$*; mitfa*$^{w2}$ zebrafish line | Zebrafish International Resource Center, Eugene, OR | *casper* |
| Oligonucleotides | | |
| Oligos for cloning the *isg15:GFP* construct, cloning ZfPV, and quantifying ZfPV and gene expression | This manuscript | See Table S4 |
| Recombinant DNA | | |
| Gateway p5E-isg15p entry clone | This manuscript | p5E-isg15 |
| Gateway pME-eGFP entry clone | Tol2kit, Kristen Kwan laboratory | Tol2kit plasmid#383 |
| Gateway p3E-polyA entry clone | Tol2kit, Kristen Kwan laboratory | Tol2kit plasmid#302 |
| Gateway pDestTol2CG2 destination vector | Tol2kit, Kristen Kwan laboratory | Tol2kit plasmid#395 |
| pCS2FA-transposase expression plasmid | Tol2kit, Kristen Kwan laboratory | Tol2kit plasmid#396 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| pCS2-ifnphi1 expression plasmid | This manuscript | pCS2-ifnphi1 |
| Software and Algorithms | | |
| STAR version 2.5.1 | [60] | https://github.com/alexdobin/STAR |
| Trinity version 2.6.6 | [61] | https://github.com/trinityrnaseq/trinityrnaseq |
| BLAST+ | NCBI | ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ |
| TransDecoder | [61] | https://github.com/TransDecoder/ |
| Genome Annotation Transfer Utility | [62] | https://4virology.net/virology-ca-tools/gatu/ |
| HMMER3 | [63] | http://hmmer.org/download.html |
| RNAfold | [64] | http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi |
| IRESfinder | [65] | https://github.com/xiaofengsong/IRESfinder |
| T-Coffee | [66] | http://tcoffee.crg.cat/ |
| BEAST version 2.5.2 | [67] | https://www.beast2.org/ |
| ProtTest version 3.4.2 | [68] | https://github.com/ddarriba/prottest3 |
| RWTY | [69] | https://github.com/danlwarren/RWTY |
| Tracer | [70] | https://github.com/beast-dev/tracer/releases/tag/v1.7.1 |
| FigTree version 1.4.4 | [70] | https://github.com/rambaut/figtree |
| SRA toolkit version 2.5.2 | NCBI | https://github.com/ncbi/sra-tools |
| Bowtie2 | [71] | https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.3.5.1/ |
| SAMtools | [72] | http://www.htslib.org/download/ |
| Integrative Genomics Viewer | [73] | https://software.broadinstitute.org/software/igv/ |
| Salmon version 0.14.0 | [74] | https://github.com/COMBINE-lab/salmon |
| DESeq2 version 1.24.0 | [75] | https://bioconductor.org/packages/release/bioc/html/DESeq2.html |
| R | CRAN | https://www.r-project.org/ |
| ggplot2 | Tidyverse | https://ggplot2.tidyverse.org/ |
| Adobe Illustrator | Adobe Systems | https://www.adobe.com/products/illustrator.html |

## LEAD CONTACT AND MATERIALS AVAILABILITY

All datasets generated in this study are publicly available. Further information and requests for other resources and reagents should be directed to and will be fulfilled by the Lead Contact, Nels Elde (nelde@genetics.utah.edu).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

Zebrafish (*Danio rerio*) were maintained in accordance with approved institutional protocols under the supervision of the Institutional Animal Care and Use Committee (IACUC) of the University of Utah, which is fully accredited by the AAALAC. Wild-type zebrafish were from the Tübingen (Tüb) strain. Adult zebrafish were maintained under standard conditions [44] and kept on a light-dark cycle of 14 hours in light and 10 hours in dark at 27°C. Embryos were collected following natural spawning and reared at 28.5°C in E3 embryo medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$), which was prepared from an autoclaved stock solution. Embryos and larvae were reared in sterile plastic disposable 100mm x 20mm Petri dishes (Fisherbrand) and all larvae hatched naturally except where otherwise stated. Zebrafish bearing the *roy*$^{a9}$ and *mitfa*$^{w2}$ mutations (*casper* fish) were obtained from Kristen Kwan at the University of Utah. CG2 zebrafish [41] were a gift from David Langenau and were euthanized upon arrival at the University of Utah

from the Massachusetts General Hospital. Sex was indeterminate at the stages of development that were used in all experiments with larvae, and only males were used in the experiments with adult animals presented in Figure 5B.

## METHOD DETAILS

### Transgenesis

*Tg(isg15:GFP)* zebrafish were generated with the Tol2kit [76]. A fragment containing 3.4kbp upstream of the start codon of *isg15* along with the first 9bp of coding sequence was amplified from Tüb genomic DNA using primers *isg15_pro_F1 isg15_pro_R1* (Table S1). This fragment was cloned into a Gateway construct (p5E-isg15p) and fused in frame with eGFP with the Tol2kit. 30pg of the *isg15:GFP* DNA construct was injected in Tüb embryos at the one-cell stage along with 25pg of transposase RNA. Three stable independent lines were generated (z207Tg, z208Tg, and z209Tg) and are collectively referred to as *isg15:GFP* animals in the text. These lines were backcrossed to the Tüb line three times before any of the data in this manuscript was collected. All three lines were used to produce the data presented in Figure 1F, while all other data in the text was generated in experiments using only z207Tg.

### Microscopy

Transgenic animals were identified and maintained by screening for GFP expression in the heart using an Olympus SZX12 stereo microscope outfitted with a 100W mercury bulb fluorescent light source (Olympus). The same microscope was used to observe induced and spontaneous GFP expression in *isg15:GFP* animals. For imaging, live 8dpf *isg15:GFP* animals were anaesthetized using buffered 3-aminobenzoic acid (Tricaine, MS-222, Sigma-Aldrich) at a final concentration of 200 μg/mL and mounted in 1% low melt agarose. Animals were imaged using a Nikon A1 laser scan confocal microscope (Nikon Instruments) with an NA 0.45/10x objective. A 20μm step series to a depth of 160μm was acquired using a 48μm pinhole size at a resolution of 0.4 pixels per μm. All images within a single figure were collected on the same day using the same laser power and gain, and exposure was set to the same exact values.

### Induction of type I interferon expression

The coding sequence of *ifnϕ1* (also called *ifnphi1*, NM_207640.1), a type I interferon [77], was synthesized (GENEWIZ) with BamHI and XbaI restriction enzyme sites added on to the 5′ and 3′ ends of the sequence, respectively. This fragment was directionally cloned into the pCS2FA-transposase vector backbone (see Key Resources Table) by cutting out the transposase sequence with BamHI and XbaI and inserting the *ifnϕ1* fragment (pCS2-ifnphi1). This plasmid allows for constitutive expression of *ifnϕ1* in zebrafish from the simian cytomegalovirus (CMV) IE94 promoter. 50pg of plasmid was injected in *isg15:GFP* embryos at the single cell stage to induce interferon signaling. Animals were screened for GFP expression at the end of the first day and then daily up to 9dpf. Transgenic animals from the same clutches that were not injected were screened over the same period of time.

### Quantification of spontaneous GFP expression

Transgenic *isg15:GFP* embryos were collected following natural spawning and maintained at 28.5°C in E3 embryo medium. Embryos were not bleached and hatched naturally. They were screened daily with a fluorescent stereo microscope for GFP expression outside of the heart (which expresses GFP constitutively) up to 9dpf. Animals that expressed GFP spontaneously were first observed at 8dpf. To quantify the frequency of spontaneous GFP expression animals were anaesthetized with tricaine on 8dpf or 9dpf. The number of GFP-positive and GFP-negative animals were counted for each clutch by manually inspecting each individual. There were approximately 200 individuals per clutch, and embryos were split across Petri-dishes at the time of collection to contain 50 individuals per dish. Animals were euthanized after quantification of GFP expression.

### RNA sequencing

Four independent untreated clutches of *isg15:GFP* animals were screened for GFP expression and separated as GFP-negative and GFP-positive at 8dpf. 30 individuals were pooled per group. The GFP-negative samples include two libraries generated from transgenic animals and two libraries generated from non-transgenic siblings (both in the Tübingen background). Non-transgenic sibling samples were initially included (as a separate group from GFP-negative transgenic siblings) to estimate the level of enrichment for virus sequences we might have found by using transgenic larvae compared to non-transgenic. Prior to RNA extraction, embryos were euthanized using an overdose of 300 mg/L Tricaine (Sigma-Aldrich). RNA was extracted by mechanical lysis with TRIzol (ThermoFisher) and DNase-treated. RNA quality was assessed by RNA ScreenTape assay (Agilent Technologies) yielding $RIN^e$ scores of $\geq$ 9.6 for all samples. Libraries were prepared by the High Throughput Genomics Shared Resource at the University of Utah with the Illumina TruSeq Stranded Total RNA Ribo-Zero Gold kit (which generates cDNA from RNA fragments using random hexamer priming), and were sequenced on an Illumina HiSeq 2500 by 125 cycle paired-end sequencing v4. Raw sequencing reads are available in the National Center for Biotechnology Informatics (NCBI) Sequence Read Archive (SRA) under BioProject accession number PRJNA575566.

### Virus genome assembly and annotation

FASTQ reads from all samples were mapped to the GRCz11 zebrafish genome assembly (Ensembl release 94) with STAR version 2.5.1 [60]. Unmapped reads were assembled into larger contigs with Trinity [61] with minimum contig length set to 150. All contigs were searched for nucleotide similarity to all sequences in the NCBI nucleotide (nt) database with BLASTn. For contigs that lacked

similarity to known sequences at the nucleotide level, the longest ORFs and protein sequences were predicted for each contig with TransDecoder (https://github.com/TransDecoder/). Predicted proteins were then searched for similarity to sequences in the UniProt UniRef90 database with BLASTp. One contig (1.2 kbp in length) had significant but distant similarity only to a picornavirus protein. A full-length genome for this prospective virus was assembled by carrying out rapid amplification of cDNA ends (RACE) using cDNA prepared with the SMARTer 5′/3′ RACE kit (Takara Bio USA, Inc.) from the same RNA that was used for sequencing. The full-length genome was annotated with Genome Annotation Transfer Utility [62] and manually searched for conserved sequences and protein domains with BLAST and HMMER3 [63]. Minimum free energy structures in the UTRs were modeled with RNAfold in the Vienna RNA Websuite [64] under default parameters (output files available upon request). The internal ribosome entry site in the 5′ UTR was predicted with IRESfinder [65]. The full genome sequence for the virus we assembled from these sequencing data is available under GenBank accession number MN524064.

### Phylogenetic analyses

Phylogenies were estimated with the concatenated amino acid sequences from the picornavirus 3C protease and 3D RNA-dependent RNA polymerase. Sequences from at least one species of all known genera of picornaviruses were included, as well as all available sequences from picornaviruses that infect fish. All sequences were aligned with T-Coffee in PSI-Coffee homology extension mode [66], resulting in an alignment of 803 amino acid sites across 65 taxa. Evolutionary relationships were inferred using Bayesian phylogenetic analyses carried out in BEAST version 2.5.2 [67]. The Blosum62 substitution model was selected as the best-fit model for the alignment under the Akaike information criterion as determined by ProtTest version 3.4.2 [68]. The rate of evolution was set to 1.0 under a strict clock rate. A Yule model was implemented for the tree prior. The MCMC chain length was set to 10 million generations, storing every 5 thousand. Two independent runs were carried out from unique starting points. Convergence and mixing of phylogenetic tree topologies in these runs were assessed with the RWTY package [69] in R v3.6.0, and reached approximate effective sample sizes (ESS) of greater than 800. Convergence and mixing of all other parameters for these runs were determined with Tracer [70], and all reached approximate ESS values greater than 2000. A maximum clade credibility tree was generated with TreeAnnotator in the BEAST 2 package. Summarized and annotated trees were displayed with FigTree v1.4.4 (https://github.com/rambaut/).

### Searching the Sequence Read Archive for Zebrafish picornavirus

Publicly available RNA-seq datasets were searched with the NCBI SRA toolkit version 2.5.2 using sra-blast with the ZfPV genome as the query sequence. More than 700 SRA datasets were queried in total, which were selected from the archive based on Boolean searches for the terms "zebrafish" and "intestine," "gut," "skin," "gills," "blood," or "brain." SRA datasets from other tissues were included if they were part of the same BioProject as those identified through the searches above. Non-zebrafish datasets were identified based on searching for "fish" and "intestine" or "gut." 775 datasets from non-zebrafish fish species were investigated. Hits were considered significant if the e-value was $\geq$ 1e-10. The location of the institutions in which the zebrafish were housed in and the strains that were used for sequencing was determined by inspecting metadata associated with the SRA submissions and with descriptions in corresponding publications if available. To confirm that SRA dataset reads matched ZfPV, FASTQ files from SRA datasets with significant hits were downloaded and aligned to the ZfPV genome with Bowtie 2 [71]. SAM files were converted to BAM files and sorted with SAMtools [72] to inspect manually with the Integrative Genomics Viewer [73]. We assembled a full zebrafish picornavirus genome sequence from BioProject PRJNA280983 raw reads, which is available under GenBank accession number TPA: BK011179.

### Quantification of virus and isg15 levels by quantitative real-time PCR

Intestines, gills, and brains were dissected from Tübingen and CG2 adult animals (gift from David Langenau, selected at random from a facility in the Massachusetts General Hospital) of approximately one year of age. The Tübingen adults were selected at random (age-matched with the CG2 samples) from two separate facilities at the University of Utah. All animals were males raised in normal rearing conditions and did not exhibit any overt signs of disease. Tissues were homogenized in 250μl of PBS with a THb Tissue Homogenizer (Omni International) before adding 750μl of Trizol LS Reagent (ThermoFisher) to extract RNA. cDNA was prepared with the SuperScript IV kit (ThermoFisher) using 1μg of RNA for each sample. Quantitative real-time PCR was performed by amplifying a ZfPV target (a 161bp fragment in the 2C helicase) with primers ZfPV_F2 and ZfPV_R2 (Table S4), *isg15* with primers isg15_F1 and isg15_R1 (Table S4), and a host reference gene target (*elfa*) with primers elfa_F and elfa_R [78] using Power SYBR Green Mastermix (ThermoFisher) on a QuantStudio 3 Real-Time PCR System (ThermoFisher). Primer efficiencies were measured by standard curve and were found to be 94%, 92%, and 101% for *elfa*, *isg15*, and ZfPV, respectively. The comparative C(T) method [79] was used to determine relative levels of ZfPV and *isg15* per sample, which were normalized by dividing the values for all samples by the lowest non-zero value.

### Differential gene expression analyses

Transcript abundances were estimated from FASTQ files for all samples with Salmon version 0.14.0 [74] in mapping-based mode with the–validateMappings,–seqBias, and–gcBias options flagged. Reads were mapped to a transcript index built from the zebrafish transcripts annotated in Ensembl release 96 and modified to include the full genomic sequence of ZfPV. Gene-level differential expression analysis was carried out by aggregating transcript-level quantification using DESeq2 version 1.24.0 [75]. Four *isg15:GFP* samples that spontaneously expressed GFP were compared to two *isg15:GFP* samples that did not express GFP and two non-transgenic sibling samples. Non-transgenic sibling samples were initially included (as a separate group from GFP-negative transgenic siblings) to estimate the level of enrichment for virus sequences we might have found by using transgenic larvae compared to non-transgenic. Only the GFP-

positive samples had reads that mapped to ZfPV, leading us to conclude that selecting for GFP-positive transgenic animals afforded us high enrichment for virus sequences. Furthermore, gene expression levels were highly similar between GFP-negative transgenic larvae and their non-transgenic siblings in these experiments, so we combined both groups to gain statistical power in comparison with gene expression levels in their GFP-positive transgenic siblings. 5dpf RNA-seq datasets containing ZfPV reads were identified in BioProject PRJEB12982 [38] with sra-blast. FASTQ files for all 20 5dpf datasets from this BioProject were downloaded and differential gene expression analysis was carried out by comparing all replicates that contained ZfPV reads to those that did not.

### Experimental prevention and administration of ZfPV infection

For prevention of infection, clutches of *isg15:GFP* embryos were divided at 24hpf to be conventionally reared (no exogenous treatment, natural hatching) or bleached and dechorionated. Embryos were bleached at 24hpf in 30ppm Sodium hypochlorite solutions freshly diluted from 10%–15% concentrates (Millipore Sigma). Bleaching was performed by immersing embryos in two bleach baths for 5 minutes each interspersed with two deionized water baths for 5 minutes each, followed with two E3 baths for 5 minutes each. Bleached embryos were transferred to new Petri dishes with 20mL of E3 and dechorionated by adding 3μL of 30mg/mL Pronase (Sigma-Aldrich) and shaking until animals were free of chorions. Dechorionated embryos were washed three times with E3 media and reared in 60mL of fresh media. GFP expression was assessed at 8dpf in bleached and conventionally reared siblings (see microscopy methods for imaging details) before euthanizing with an overdose of 300 mg/L Tricaine (Sigma-Aldrich) and collecting in TRIzol (ThermoFisher) for RNA extraction. RNA was prepared from a pool of approximately 25 larvae per sample to ensure that sufficient yields of RNA were acquired. For generating virus preparations to administer to larvae, 3 randomly selected adult Tübingen animals (2 years old) were euthanized by immersion in ice water and intestines were extracted and homogenized together in 2mL of E3 media. This solution was centrifuged at 3000 g for 5 minutes and the supernatant was filtered with a syringe through a sterile 0.45μm cellulose acetate membrane (VWR international). 50μL of fresh 0.45μm filtrate was added to a Petri dish containing 40mL of E3 media and 50 5dpf *isg15:GFP* larvae that had been bleached and dechorionated at 24hpf. GFP expression and virus infection was assessed at 8dpf (3 days after adding filtrate). Quantification of *isg15* and ZfPV expression was made with RNA from pools of approximately 25 larvae per sample of GFP-negative or GFP-low animals from bleached groups, or GFP-positive animals from conventionally reared groups.

### Experimental transmission of ZfPV infection

For testing the transmission of virus infection from GFP-positive to GFP-negative transgenic animals, the *isg15:GFP* transgene (z207Tg) was crossed into the *casper* zebrafish line [80]. The *casper* and Tübingen lines bear dramatic differences in pigmentation as larvae, which provided a means of distinguishing donor from recipient animals during co-housing. Transgenic clutches of embryos (in the Tübingen or *casper* background, labeled *isg15:GFP* or *isg15:GFP-casper*, respectively) were split in half to be conventionally reared or bleached at 24hpf as described above. All animals were fed twice a day with rotifers and algae starting at 5dpf. Animals were lightly anaesthetized with Tricaine (Sigma-Aldrich) at 8dpf and divided into co-housed groups of 30 donor animals and 30 recipient animals in 60mL of fresh E3 media per Petri dish. Three separate conditions of co-housing were carried out per experiment. One dish was populated with 30 bleached *isg15:GFP* animals co-housed with 30 bleached *isg15:GFP-casper* animals. The second dish was populated with 30 bleached *isg15:GFP* animals co-housed with 30 conventionally reared GFP-positive *isg15:GFP-casper* animals. The third dish was populated with 30 bleached *isg15:GFP-casper* animals co-housed with 30 conventionally reared GFP-positive *isg15:GFP* animals. Animals were fed twice daily with rotifers and algae as before co-housing, and were additionally fed once daily with GEMMA Micro 75 powder (Skretting) from 10dpf to 15dpf. Media was removed and fresh E3 was added daily from 10dpf to 15dpf. One week after co-housing (15dpf), larvae were euthanized with an overdose of 300 mg/L Tricaine, assessed for GFP expression, and collected in TRIzol (ThermoFisher) for RNA extraction. RNA was extracted from donor and recipient animals separately.

## QUANTIFICATION AND STATISTICAL ANALYSIS

All specific details regarding quantification methods and statistical analyses are described alongside experimental descriptions in the text, figure legends, and methods. Statistical tests were carried out using the software described in the STAR Methods and Key Resources Table and in the R software environment for statistical computing and graphics. For the experiments we carried out with larval animals each sample that involved RNA extraction was made up of a pool of approximately 25 larvae (always siblings collected from the same Petri dish) to ensure that enough RNA was extracted for downstream applications. The details of how this relates to n for all samples plotted in the figures is provided in the figure legends and methods. Thresholds used for defining significance are stated in the text and figure legends.

## DATA AND CODE AVAILABILITY

The RNA-sequencing datasets we generated for this manuscript are publicly available at the NCBI Sequence Read Archive, BioProject accession number PRJNA575566. The virus genome sequences we assembled are available at GenBank, accession numbers MN524064 and TPA: BK011179. The publicly available datasets generated by others that we used in this manuscript are listed in Table S5.

**Current Biology, Volume *30***

## Supplemental Information

# Linking Virus Discovery to Immune Responses Visualized during Zebrafish Infections

**Keir M. Balla, Marlen C. Rice, James A. Gagnon, and Nels C. Elde**

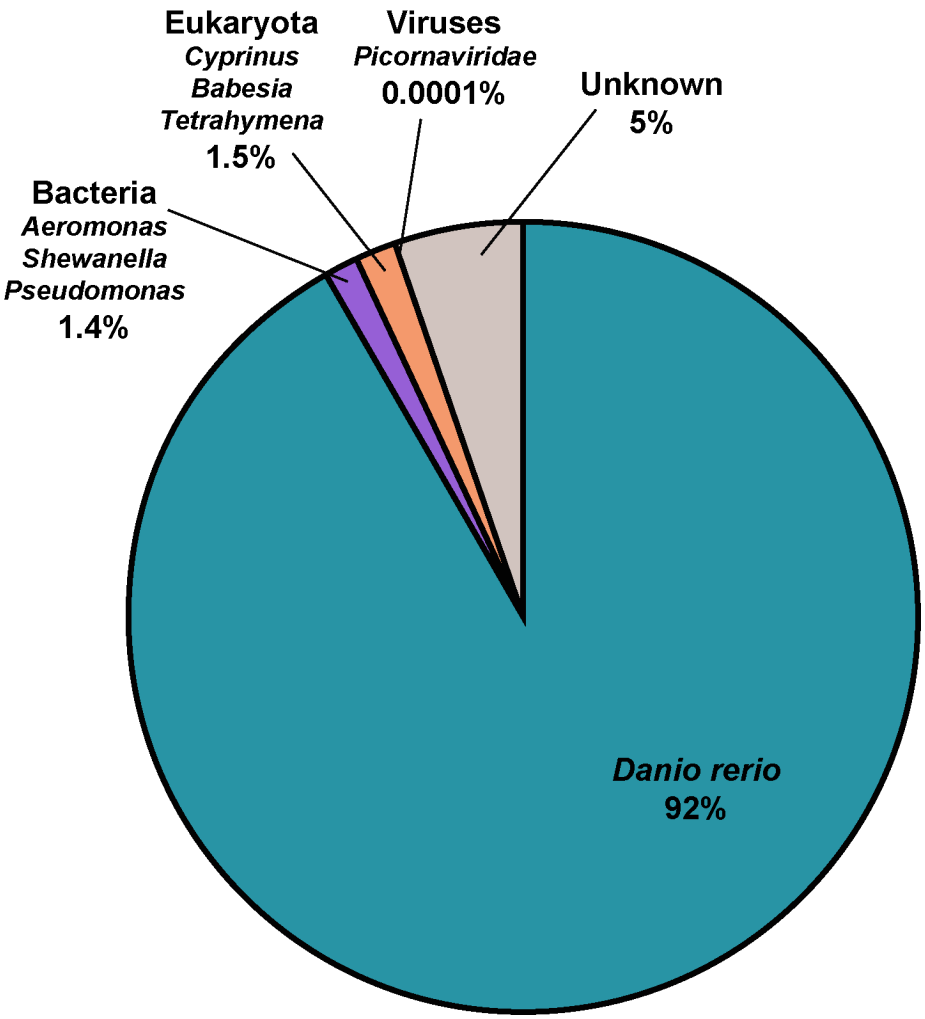



**Figure S1. Species identities of all 737,608 contigs assembled from reads that did not map to the zebrafish genome during alignment. Related to Figure 2.**
Most were determined to be from zebrafish and had failed to map as individual reads. The top 3 most abundant genera from other taxa are listed for Bacteria and Eukaryota. One contig was most similar to viruses. 5% of contigs had no similarity to any known nucleotide or amino acid sequence by BLAST. These were predominantly short (~150bp) and repetitive sequences.

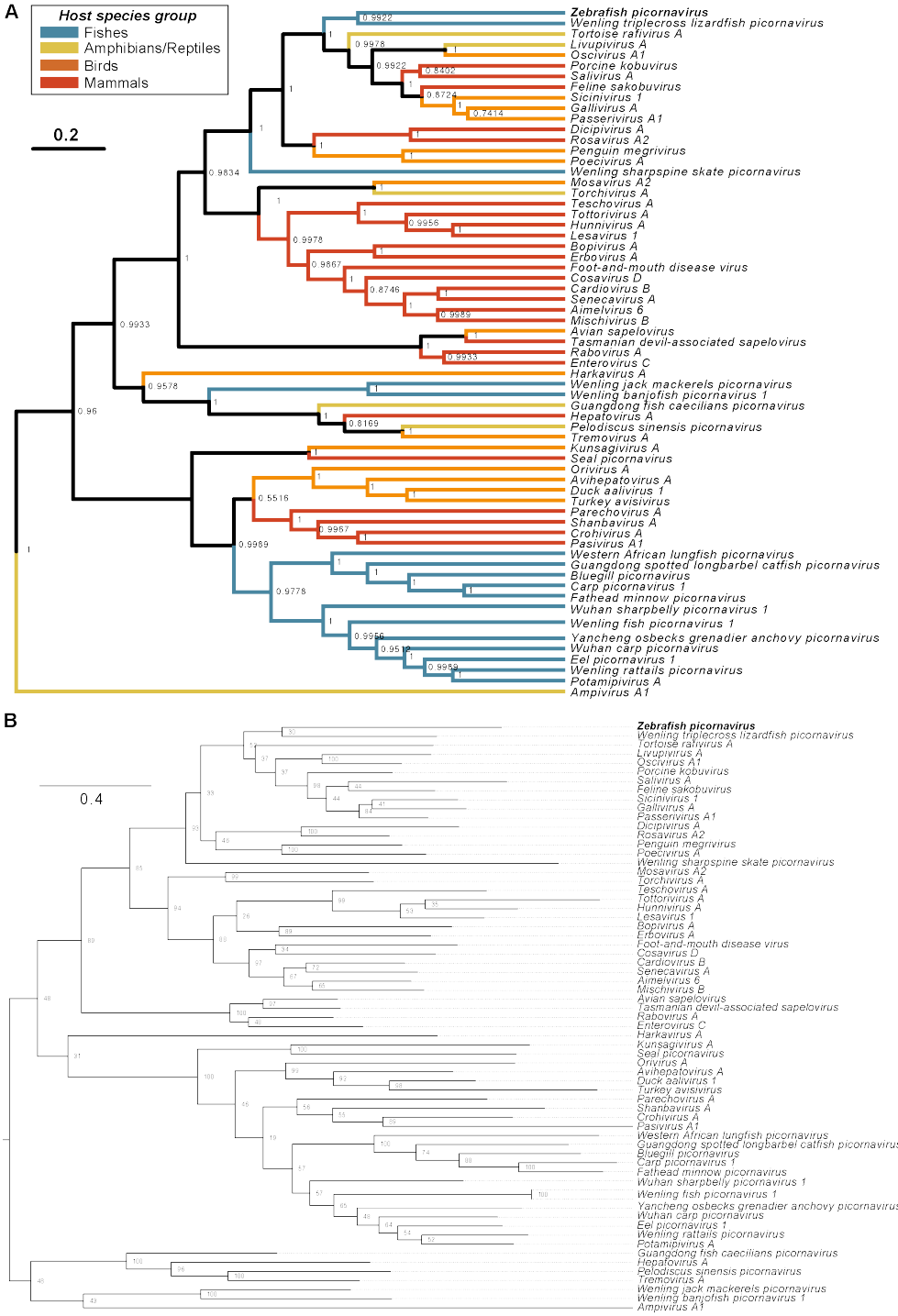



**Figure S2. Bayesian and maximum likelihood phylogenies of picornaviruses. Related to Figure 2.** (A) Maximum clade credibility tree of picornavirus 3C and 3D amino acid sequences. Branches are colored based on the group of vertebrates that serve as hosts with estimated support at each node. Scale is in expected number of substitutions per site. (B) Maximum likelihood phylogenetic tree of the same picornavirus 3C and 3D amino acid sequence alignment used for Bayesian inference. Phylogenetic relationships were estimated using the maximum likelihood method (with 100 bootstrap replicates) implemented in PhyML 3.0 [S1]. Numbers at nodes are bootstrap values. The tree is midpoint rooted. Scale is in substitutions per site.

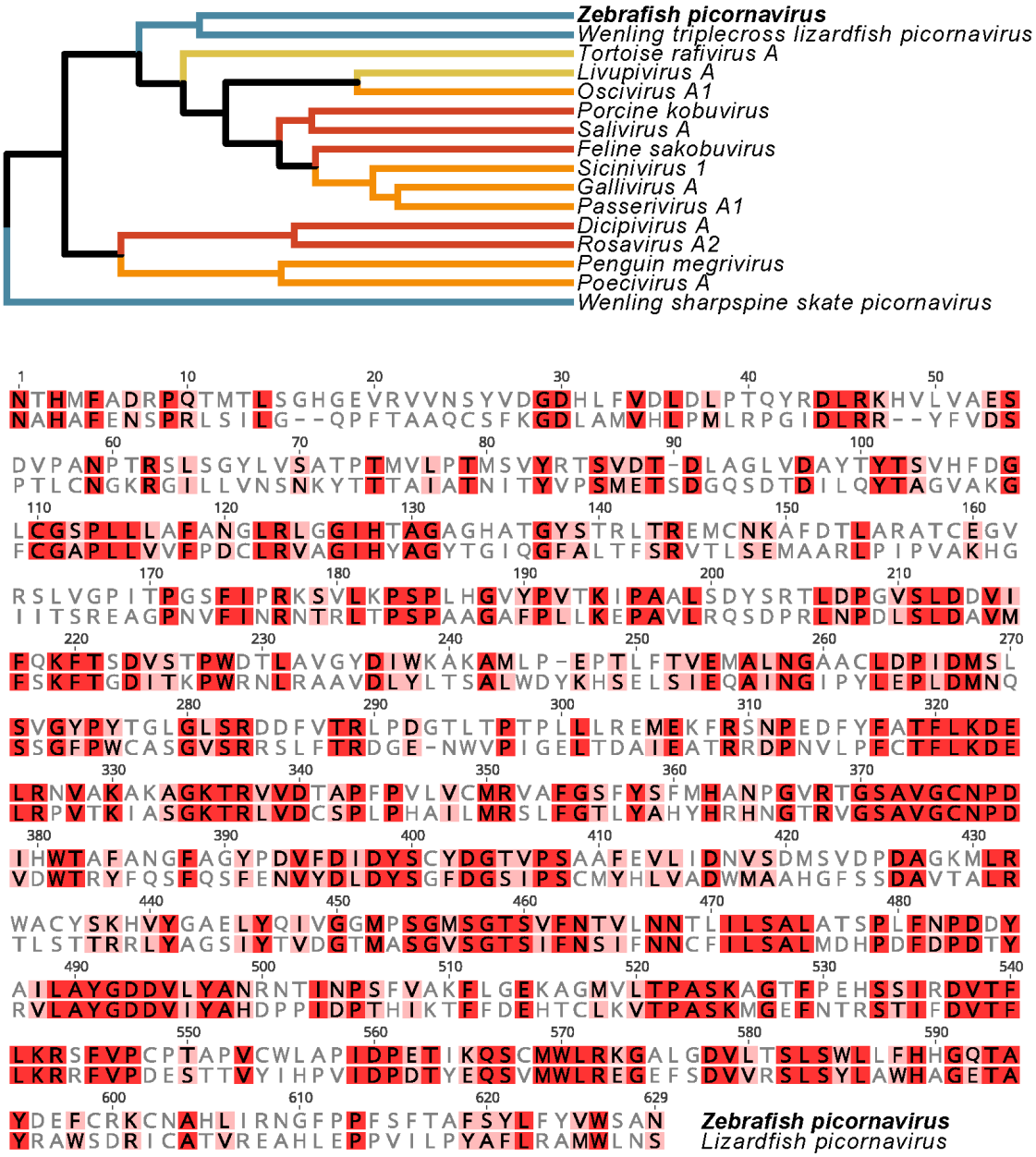



**Figure S3. Pairwise alignment of the 3CD amino acid sequences from *Zebrafish picornavirus* and its closest relative, the *Wenling triplecross lizardfish picornavirus*. Related to Figure 2.**

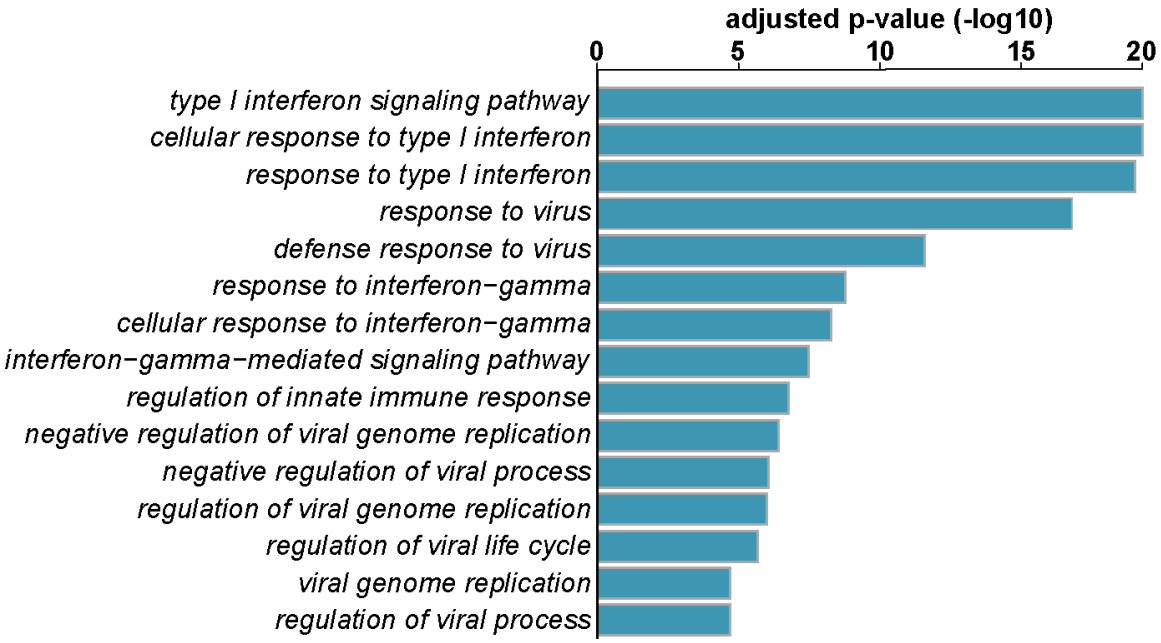



**Figure S4. Gene set analysis of differentially expressed genes in *isg15:GFP* animals spontaneously expressing GFP compared to GFP-negative animals 8dpf. Related to Figure 2.**
The top 15 significantly enriched Gene Ontology (GO) terms identified in the genes induced by ZfPV are plotted. The differentially expressed zebrafish genes were first compared to their homologs in human. GO terms were assigned based on functional classes having significant enrichment for both species. These analyses were carried out with the XGSA statistical framework for cross-species comparisons of gene sets [S2].

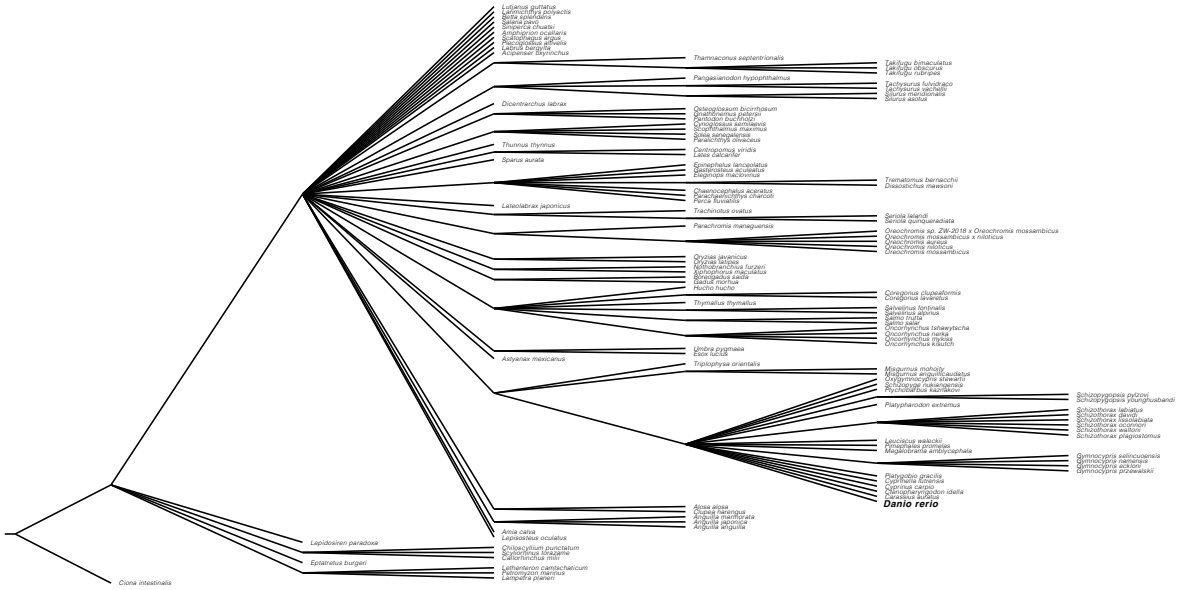



**Figure S5. Common tree from the NCBI Taxonomy browser of all the fish species that were investigated for the presence of ZfPV. Related to Figure 4.**
775 non-zebrafish RNA-seq datasets were searched, but only zebrafish datasets contained ZfPV reads. The tree is rooted on *Ciona intestinalis* (which was not searched for ZfPV).

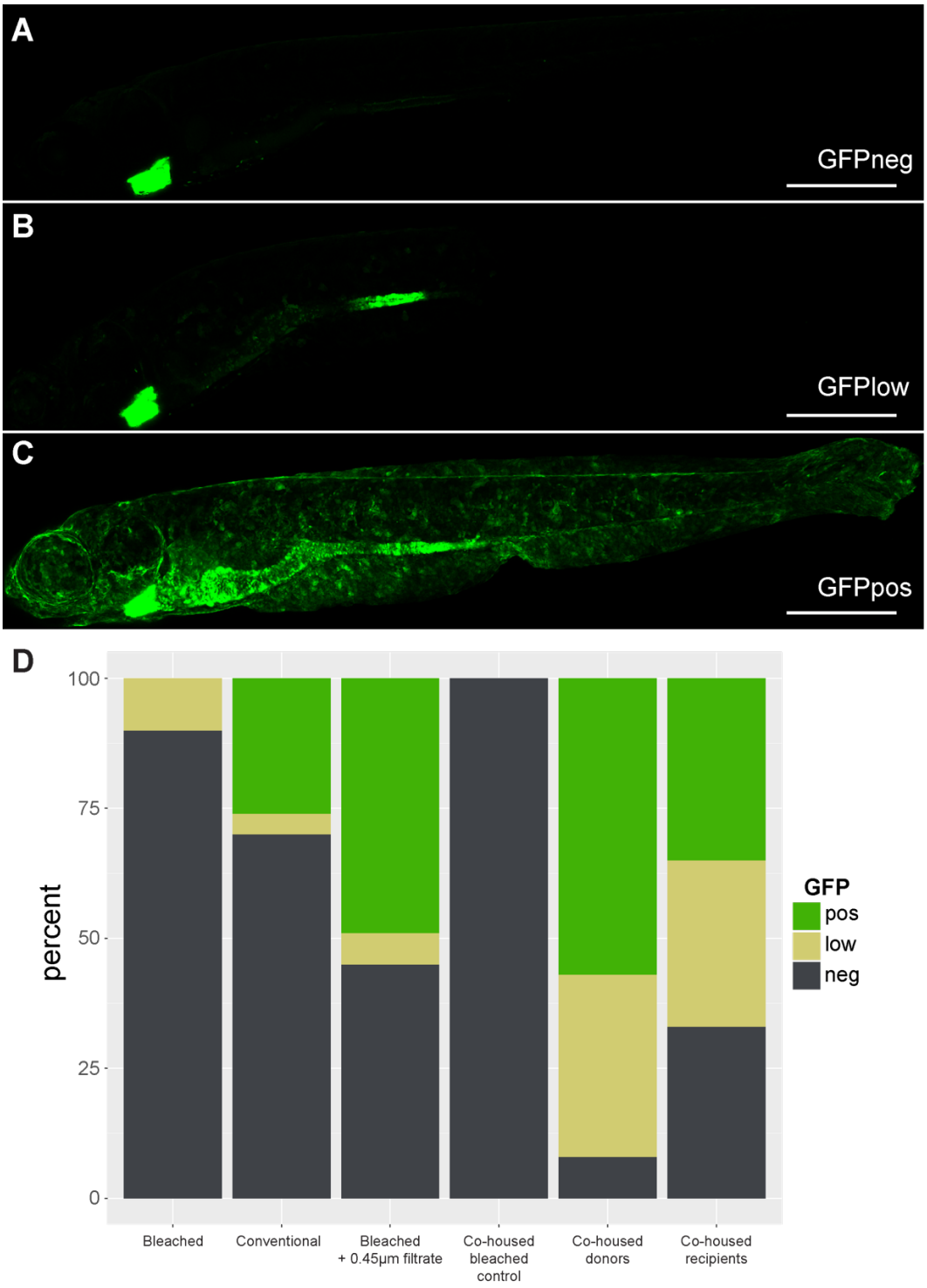



**Figure S6. GFP expression in bleached and conventionally reared transgenic animals that are uninfected or infected with zebrafish picornavirus. Related to Figure 6.**
Confocal images of animals at 8 dpf categorized as (A) GFP-negative, (B) GFP-low, and (C) GFP-positive. All images were acquired using the same settings on the same day. (D) Percentage of animals assigned to the three GFP categories at the time of collection from experiments described in Figure 6. Percentages are averages from 7 independent clutches split in to bleached and conventional groups, 2 independent clutches split in to bleached or filtrate groups, and 4 independent clutches across 2 experiments for the co-housed groups.

| Term | P-Value | Benjamini |
|---|---|---|
| SPRY-associated | 2.10E-29 | 1.10E-26 |
| B30.2/SPRY domain | 7.20E-27 | 1.80E-24 |
| Butyrophylin-like | 4.70E-23 | 8.00E-21 |
| SPla/RYanodine receptor SPRY | 1.50E-22 | 1.90E-20 |
| Concanavalin A-like lectin/glucanase, subgroup | 2.20E-22 | 2.20E-20 |
| Zinc finger, B-box | 1.80E-21 | 1.50E-19 |
| Zinc finger, RING-type, conserved site | 1.40E-18 | 1.00E-16 |
| Zinc finger, RING-type | 2.90E-13 | 1.80E-11 |
| NACHT nucleoside triphosphatase | 5.00E-11 | 2.80E-09 |
| P-loop containing nucleoside triphosphate hydrolase | 8.80E-11 | 4.40E-09 |
| Zinc finger, RING/FYVE/PHD-type | 9.60E-11 | 4.40E-09 |
| Interferon-induced 6-16 | 4.50E-08 | 1.90E-06 |
| Dynamin | 9.70E-08 | 3.80E-06 |
| Poly(ADP-ribose) polymerase, catalytic domain | 1.60E-07 | 5.80E-06 |
| Leucine-rich repeat | 2.80E-06 | 9.30E-05 |
| Heat shock protein 70 family | 2.10E-05 | 6.50E-04 |
| WWE domain | 3.50E-05 | 1.00E-03 |
| Interferon regulatory factor DNA-binding domain | 7.40E-05 | 2.10E-03 |
| Dynamin GTPase effector | 1.40E-04 | 3.70E-03 |
| GTPase effector domain, GED | 1.40E-04 | 3.70E-03 |
| Dynamin central domain | 1.40E-04 | 3.70E-03 |
| Dynamin, GTPase domain | 1.80E-04 | 4.60E-03 |
| Heat shock protein 70, conserved site | 1.80E-04 | 4.60E-03 |
| STAT transcription factor, core | 3.40E-04 | 8.10E-03 |
| STAT transcription factor, all-alpha | 3.40E-04 | 8.10E-03 |
| STAT transcription factor, protein interaction | 3.40E-04 | 8.10E-03 |
| STAT transcription factor, DNA-binding, subdomain | 3.40E-04 | 8.10E-03 |
| STAT transcription factor, DNA-binding | 3.40E-04 | 8.10E-03 |
| STAT transcription factor, coiled coil | 3.40E-04 | 8.10E-03 |
| Nidogen, extracellular domain | 6.70E-04 | 1.50E-02 |
| Interferon regulatory factor-3 | 9.50E-04 | 2.10E-02 |
| Interferon regulatory factor, conserved site | 9.50E-04 | 2.10E-02 |
| Putative helicase MOV-10/Armitage | 1.00E-03 | 2.10E-02 |
| RHD3/Sey1 | 1.00E-03 | 2.10E-02 |
| Chemokine interleukin-8-like domain | 1.10E-03 | 2.20E-02 |
| Uncharacterised protein family Cys-rich | 1.20E-03 | 2.40E-02 |
| SAND domain | 1.20E-03 | 2.40E-02 |
| Tetratricopeptide-like helical | 1.30E-03 | 2.40E-02 |
| Dynamin, GTPase region, conserved site | 2.00E-03 | 3.60E-02 |
| AIG1 | 2.80E-03 | 4.70E-02 |

**Table S3. InterPro protein domains that are enriched among the genes induced by ZfPV (GFPpos/GFPneg 8dpf isg15:GFP animals). Related to Figure 2 and is based on the genes in Table S1.**

| Primer name | Sequence (5'-3') | Application |
|---|---|---|
| **isg15_pro_F1** | GGGGACAACTTTGTATAGAAAAGTTGCTGGAGTCTGGCAAGAAGTTGC | Amplifying *isg15* promoter for Gateway cloning. Blue nucleotides are for Gateway attB4 site. |
| **isg15_pro_R1** | GGGGACTGCTTTTTTGTACAAACTTGACAGCTGCATCGTCACCTGTCAG | Amplifying *isg15* promoter for Gateway cloning. Blue nucleotides are for gateway attB1r site. Pink nucleotide is added to be in frame with middle element (GFP). |
| **ZfPV_3prime_RACE_F1** | GATTACGCCAAGCTTCCGCGGGACGCAAAACTTGGTGCAG | 3' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_3prime_RACE_F3** | GATTACGCCAAGCTTAAATGCTGCCCGCTGCCTCAACGC | 3' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_5primeRACE_R** | GATTACGCCAAGCTTCCTATAGCGGCAGCGACAAGCTGCACCAAG | 5' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_5primeRACE_R1** | GATTACGCCAAGCTTGGCGGGGGAGTTAGTCCCACCAACACGC | 5' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_5primeRACE_R3** | GATTACGCCAAGCTTCACTGTGGCACTCGTGGGAGCACCGC | 5' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_5primeRACE_R4** | gattacgccaagcttCAGGCGTCGACTTCGCGAACGGGGCA | 5' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_5primeRACE_R5** | gattacgccaagcttGCTGCGCACGGTCCCCAAGAACAGTGGAGG | 5' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_5primeRACE_R6** | gattacgccaagcttTCGGTGGCGGGGTCGGCTAGAGACTTCTTC | 5' RACE to clone full sequence of ZfPV. Blue sequence is part of the SMARTer RACE adapter. |
| **ZfPV_F2** | CCGCCCTCATCTACCGAAAA | qRT-PCR amplification of ZfPV. |
| **ZfPV_R2** | ACGCAGAGGGGTCAAATTGT | qRT-PCR amplification of ZfPV. |
| **isg15_F1** | TGTGTTTTATTCTGACAGGTGACG | qRT-PCR amplification of *isg15*. |
| **isg15_R1** | TGGCCCTTCTCGTTCTTGAC | qRT-PCR amplification of *isg15*. |

**Table S4. Oligos used in this study. Related to Figures 1, 2, 5, and 6.**

## Supplemental References

S1. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

S2. Djordjevic, D., Kusumi, K., and Ho, J.W.K. (2016). XGSA: A statistical method for cross-species gene set analysis. Bioinforma. Oxf. Engl. *32*, i620–i628.