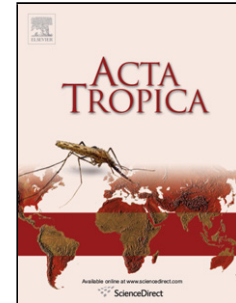


## Accepted Manuscript

Title: First characterization of viruses from freshwater snails of the genus *Biomphalaria*, the intermediate host of the parasite *Schistosoma mansoni*.

Author: Richard GALINIER Guillaume TETREAU Anaïs PORTET Silvain PINAUD David DUVAL Benjamin GOURBAL



PII: S0001-706X(16)30715-X  
DOI: <http://dx.doi.org/doi:10.1016/j.actatropica.2016.12.021>  
Reference: ACTROP 4154

To appear in: *Acta Tropica*

Received date: 16-9-2016  
Revised date: 16-11-2016  
Accepted date: 18-12-2016

Please cite this article as: GALINIER, Richard, TETREAU, Guillaume, PORTET, Anaïs, PINAUD, Silvain, DUVAL, David, GOURBAL, Benjamin, First characterization of viruses from freshwater snails of the genus *Biomphalaria*, the intermediate host of the parasite *Schistosoma mansoni*. *Acta Tropica* <http://dx.doi.org/10.1016/j.actatropica.2016.12.021>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

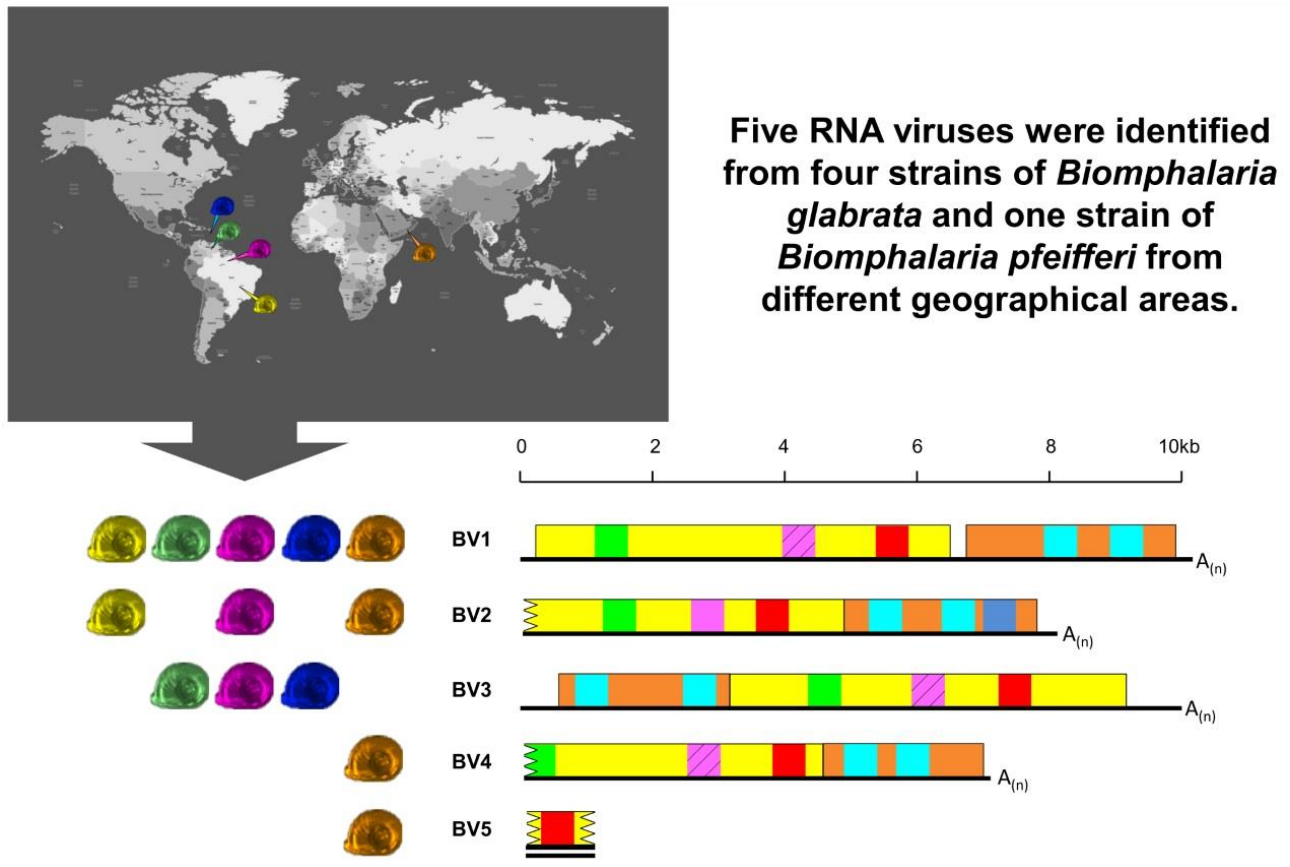
**First characterization of viruses from freshwater snails of the genus *Biomphalaria*, the intermediate host of the parasite *Schistosoma mansoni*.**

Richard GALINIER<sup>1§</sup>, Guillaume TETREAU<sup>1</sup>, Anaïs PORTET<sup>1</sup>, Silvain PINAUD<sup>1</sup>, David DUVAL<sup>1</sup>, Benjamin GOURBAL<sup>1</sup>

<sup>1</sup> CNRS, IHPE UMR 5244, Univ. Perpignan Via Domitia, IFREMER, Univ. Montpellier, F-66860 Perpignan, France

§ Corresponding author

## Graphical abstract



## Highlights

- Five RNA viruses were isolated from five *Biomphalaria* strains
- Four on five viruses may belong to Picorna-like virus superfamily
- One virus presents similarity with *Totiviridae* family
- Viruses are differentially represented between snail strains

## Abstract:

We report the genome sequence and organization of five viruses infecting snails of both *Biomphalaria glabrata* and *Biomphalaria pfeifferi*, which are vectors of the intestinal schistosomiasis. Four viruses presented a polyadenylated positive single strand RNA genome encoding one or two large open reading frames (ORFs) flanked by untranslated region. Conserved protein motifs typical of the picorna-like virus superfamily were identified in these viruses but they all presented different genome organization. Phylogenetic analysis confirmed their assignment to this superfamily. The partially characterized fifth virus presented sequence similarity for *Totiviridae*, a family of non-polyadenylated double-strand RNA viruses. Virus distribution and relative abundance between the five strains of *Biomphalaria* originating from different geographical areas was determined. Our results provide valuable information of new viruses from *Biomphalaria* and pave the way for future studies dedicated to their impact on snail fitness and *Biomphalaria/Schistosoma* interactions.

Keywords: Schistosomiasis vector, *Biomphalaria*, viral microbiota, picorna-like virus, *Totiviridae*,

## 1. Introduction

Schistosomiasis is the second most widespread human parasitic disease after malaria. It is reported to be endemic in more than 78 countries in the tropical and subtropical areas, in Africa, Asia and Americas. Around 200 million people are estimated to be infected all over the world, with 200 000 death per year (Chitsulo et al., 2000). Infection is caused by a trematode parasite of the genus *Schistosoma*, which comprises 22 different species. Three of them, *Schistosoma mansoni*, *S. haematobium* and *S. japonicum* are mainly responsible for human infections. *S. mansoni* is the causative agent of intestinal schistosomiasis and it is vectored by a freshwater snail from the genus *Biomphalaria* (Basch, 1976). The compatibility between the pathogen and its host is determined by two main factors, the susceptibility of the host and the infectivity of the parasite. In *Biomphalaria*, this host/pathogen compatibility is determined by genetic factors but can also be modulated by non-genetic factors, such as temperature (Knight et al., 2015), pollution (Ibrahim, 2006) or UV exposure (Ruelas et al., 2009), and could also be influenced by its microbiota composition.

The microbiota of an organism is constituted of all microbial organisms, such as bacteria, fungi and viruses that live within the host in a mutualistic interaction. The microbiota is known to impact host evolution and adaptation by affecting several fitness parameters (Bosch and McFall-Ngai, 2011; Zilber-Rosenberg and Rosenberg, 2008). Notably, it has been shown to play a key role in the interaction between a host and its pathogens. In the mosquito *Anopheles gambiae*, the microbial flora can limit the infection with *Plasmodium falciparum* by inhibiting the development of the parasite within the host (Dong et al., 2009). More specifically, although viruses have often been characterized by their pathogenic and harmful effects, they are also an integral part of the microbiota and they can modulate host-parasite interaction in invertebrates (Martinez et al., 2015; Pradeu, 2016). For example, PolyDNA viruses are able to manipulate the immune system of parasitoid wasps (Burke and Strand, 2012), while SSRNA DcPv can modify the behavior of a ladybird parasitized by a parasitoid wasp to protect the parasitoid larvae (Dheilly et al., 2015b). More generally, it is widely accepted that viruses can control the immune system of their host for

their own development (Christiaansen et al., 2015), but also for the defense of their host against other pathogens (Bettarel et al., 2015).

In this context, RNAseq data were used to perform the first identification of sequences of resident viruses from *Planorbidae* of the genus *Biomphalaria*. Next-generation sequencing (NGS) has already been successfully used as a powerful tools for the discovery of new viruses (Liu et al., 2011; Marston et al., 2013), even when they are present at a low dose in a latency phase or at higher dose but without any associated symptoms. In this work, we report the identification and characterization of 5 viruses from four strains of *B. glabrata* and one strain of *B. pfeifferi*. The taxonomic position of these *Biomphalaria* viruses was assessed by a phylogenetic analysis using the most conserved domains from protein sequences. Finally, based on virus abundance analysis, we evidenced a high level of virus specificity toward the different snail strains investigated herein.

## 2. Material and methods

### 2.1 Biological material

The experiments were conducted using four strains of *Biomphalaria glabrata* (Bg) and one of *Biomphalaria pfeifferi* (Bp). The four South American *B. glabrata* strains were named as follow: BgBAR (origin Brazil, Belo Horizonte); BgVEN (origin Venezuela, Guaraca); BgBRE (origin Brazil, Recife) and BgGUA (origin Guadeloupe, le Lamentin). The *B. pfeifferi* strain came from Arabica peninsula and was named BpOMAN (origin Oman, Dhofar). These five *Biomphalaria* strains exhibit differential susceptibility to each of the five *Schistosoma* strains, from a generally high compatibility with their sympatric strains (*i.e.*, originating from the same location), to a generally moderate/low compatibility with other allopatric strains (*i.e.*, originating from different locations) (Theron et al., 2014).

## 2.2 RNAseq data

RNA extraction, cDNA library construction and Illumina SOLEXA sequencing were performed as previously described (Dheilly et al., 2015a). Each paired-end cDNA library (72 bp) was constructed from a pool of 30 snails, ranging from juveniles to old adults. *De novo* transcriptome assemblies were performed as previously described (Dheilly et al., 2015a). Finally, five high quality reference transcriptomes were obtained. They comprised 70 533, 79 664, 117 269, 82 500 and 99 507 transcripts for the strains BgBAR, BgVEN, BgBRE, BgGUA and BpOMAN, respectively. Automatic annotation of transcripts was performed using Blast2GO version 2.4.2. Putative virus sequences were identified by focusing on key viral features that are RNA-dependent RNA polymerase, capsid protein and polyprotein domains. Sequence analysis was performed using either the GeneCodes software package (Sequencher version 4.5, Ann Arbor MI) or the BioEdit Sequence Alignment Editor version 7.1.3.0. (Hall, 1999).

## 2.3 PCR and RACE-PCR

RACE-PCR was performed to expand the virus genome sequences at the 5' and/or 3' ends when they were missing. First strand cDNA synthesis and RACE-PCR were performed using the SMARTer RACE cDNA Amplification Kit (Clontech) from the RNA extracted in section 2.2 by following manufacturer's instructions. RACE-PCR products were analyzed by agarose gel electrophoresis and cloned into the pCR4-TOPO vector according to the manufacturer's instructions (ThermoFisher Scientific). Clones were then sequenced using GenoScreen facilities (GenoScreen, France).

PCR were performed to validate the sequences obtained by RNAseq, as well as to complete virus sequence gaps. One microliter of a ten times diluted cDNA (see section 2.2) was used as template to amplify desired fragments using the GoTaq G2 Hot Start polymerase (Promega) using the manufacturer's instructions. PCR products were cloned and sequenced following the same procedure as for RACE-PCR products.

#### 2.4 Sequence analysis and phylogenetic trees

Homologous sequences were searched using BLAST searches against GenBank non redundant database on the National Center for Biotechnology Information web server (Bethesda, USA). A search for the presence of conserved motifs in the sequences obtained by RACE-PCR and PCR was done using the Conserved Domain Search Service (CD-Search) from the NCBI web server (Marchler-Bauer et al., 2015).

For phylogenetic analyses, multiple protein sequence alignments were performed with ClustalW using the BLOSUM62 substitution matrix model. The neighbor-joining method (Poisson substitution model; uniform substitution rate; gaps/missing data treatment: pairwise deletion) was used to generate the phylogenetic trees. A bootstrap analysis of 2000 replications was carried out on the trees inferred from the neighbor-joining method to assess the robustness of the tree branches. Two phylogenetic trees were generated: one using the RdRP protein sequences, the other one using the helicase protein sequences. In addition to the 5 RdRP and 4 helicase sequences from *Biomphalaria* viruses, 76 RdRP and 59 helicase sequences from 9 different families of *Picornavirales* were obtained from NCBI database (Supplementary Table 1). These different phylogenetic groups of viruses were chosen based on the closest domain organization of the virus genome to the *Biomphalaria* viruses and on the best BLAST matches against the NCBI database. All phylogenetic analyses were performed using MEGA 7.0.14 software (Kumar et al., 2016).

#### 2.5 Analysis of viral abundance

A viral transcriptome containing the nucleotide sequences of the five *Biomphalaria* viruses previously characterized was created in order to analyze viral abundance in the five *Biomphalaria* strains. For each RNAseq data of each *Biomphalaria* strain, high quality reads (phred score >26) were aligned to viral transcriptome using Bowtie2 (V2.2.6.2), which was run locally on the Galaxy project server (<http://bioinfo.univ-perp.fr>) (Giardine et al., 2005). The mapping was realized with paired library and sensitive end-to-end mode. Differential abundance of each identified virus was



estimated and normalized by the upper quartile (UQ) within each snail strain (Dillies et al., 2012).

### 3. Results and discussion

#### 3.1 Viral sequence analysis and genome organization

##### 3.1.1 Discovery of five new viral sequences from *Biomphalaria* RNAseq data

The presence of virus-derived transcripts was investigated in four *de novo* assembled transcriptomes from *B. glabrata* coming from different geographical areas and one from *B. pfeifferi*. A total of five viruses were identified from *B. glabrata* and *B. pfeifferi* species, and named BV1 to BV5 for *Biomphalaria* virus 1 to 5, respectively (Figure 1). Sequence of BV1 virus was almost completely obtained from RNAseq data while only partial contig coverage was observed for the four other viral sequences. Joining PCR and RACE-PCR followed by Sanger sequencing allowed fulfilling some gaps and extending viral sequences to generate respectively full or near-full length sequences for BV1 and BV3 while only partial sequences were obtained for BV2, BV4 and BV5 (Figure 1). The obtained viral sequences were deposited in GenBank under the following accession numbers: KY024317-KY024321 for BV1, KY024322 for BV2, KY024323 for BV3, KY024324 for BV4 and KY024325 for BV5. Rearing facilities (water, food, and material used) were screened by PCR for the presence of these viruses in order to avoid misinterpretation. None of the *Biomphalaria* viruses was detected.

##### 3.1.2 *Biomphalaria glabrata* Virus 1 (BV1)

The BV1 genome was found to be 10 116 nucleotides long, excluding the poly(A) tail present at the 3' end (Figure 1). No sequence match was found in GenBank based on the complete nucleotide sequence. The genome was enriched in Adenine/Uracile (27% A, 37% U, 19% C and 17% G), which is typical of Picorna-like viruses. Two large ORFs located on the same positive RNA strand but on different reading frames account for 88% of the total genome length. The other 12% of the genome consisted in a 253 nucleotides long 5' UTR, a 160 nt intergenic region (IGR), and a longer 756 nt 3' UTR ended by a poly(A) tail (Figure 1). No large ORF was found in reverse

orientation, suggesting that BV1 is a positive strand RNA virus.

The 5' proximal ORF (ORF1) started with an AUG codon at position 254 and ended at a UAA stop codon at nucleotide 6403. The predicted product was 2049 amino acid long with a theoretic molecular mass of 227 kDa. It encoded a picornavirus-like H-P-Rep cassette, which included in this order a RNA helicase (pfam00910), a 3C cystein protease (pfam00548) and RNA-dependent RNA polymerase (RdRP) (cd1699) as described for *Picornaviridae* (Koonin and Dolja, 1993). A search for conserved domains in NCBI database revealed the presence of both helicase and polymerase domains but not of protease domain. However, manual analysis of the sequence allowed the identification of a protease domain spanning from amino acid position 1227 to 1427, with the conserved cysteine protease motif  $^{1399}\text{GxCG}^{1402}$  and the putative substrate binding residues  $^{1418}\text{GxHxxG}^{1423}$  (Gorbalenya et al., 1989). In addition, we also found the amino acids that could constitute the protease catalytic triad previously described by (Ryan and Flint, 1997) at position  $\text{H}^{1279}$ ,  $\text{D}^{1331}$  and  $\text{C}^{1401}$ , confirming the presence of the protease domain. The best hit after protein blast analysis was for a *Nora virus* replication polyprotein (GenBank accession no. KP970079), with 26% identity over 653 residues (e-value:  $8^{-42}$ ). *Nora virus* is a recently characterized positive single strand RNA virus, which is still unclassified among the picorna-like viruses (Habayeb et al., 2006).

The 3' proximal ORF (ORF2) started at nucleotide position 6564 with an AUG codon and ended at position 9362 with a UAA stop codon. It encoded a predicted product of 932 amino acids with a theoretical molecular mass of 103 kDa. BLAST analysis against NCBI database didn't retrieve any homologous sequences. Nevertheless, CD search analysis revealed the presence of two picornavirus capsid protein domain-like (cd 00205), typical of picorna-like non-structural virus proteins.

The genome organization of BV1 suggests that it could be a new species in the picorna-like virus superfamily, considering that it possesses all the typical domain of structural and non-structural proteins of these viruses. More precisely, its di-cistronic genome organization suggests that it might be a candidate species for the *Dicistroviridae* family.

The BV1 genome has been sequenced in all *Biomphalaria* strains and compared at the nucleic acid level (Table 1). BV1 nucleotide sequences were highly similar between most snail strains, with 99 to 100% identity for BV1 sequences from BgVEN, BgGUA and BpOMAN, and more than 97% identity with the one from BgBRE. Interestingly, the BV1 sequence from BgBAR strain was the most different, as it only shared 87% identity with the BV1 genome sequences from the four other *Biomphalaria* strains. In addition, BV1 sequence from BgBAR strain presented a longer 5'UTR (261 nt instead of 167 nt in the other strains), a single codon insertion in the ORF1 sequence, and an insertion of 8 codons in the ORF2, which thus encoded a 940 residues product. The polyprotein from ORF1 and ORF2 presented respectively 89% and 93% identity and 95% and 96% similarity with the ones of BV1 from the four other *Biomphalaria* strains, respectively (Table 1). Additional studies based on morphological and biochemical data, as well as host range and tropism specificities will be necessary to determine if BV1 from BgBAR is a distinct species or simply a variant of BV1 identified in the four other *Biomphalaria* species.

### 3.1.3 *Biomphalaria glabrata* Virus 2 (BV2)

BV2 virus was identified thanks to a long contig located in the middle of the genome sequence of the virus (Figure 1). RACE-PCRs allowed to fully sequence the 3' end but only partially the 5' part of the virus genome. The partial genome sequence obtained for BV2 virus was 7 958 nt long. Only one ORF was predicted by *in silico* nucleotide sequence translation with a partial length of 7 749 nt lacking the initial start codon. The 3' UTR was 209 nt long and ended by a poly(A) tail.

The partial ORF encoded a 2 582 amino acid product and included several conserved motifs; at the 5' end of the ORF, typical domains from picorna-like viruses were identified including non-structural proteins, such as the RNA helicase domain (pfam00910), two mingled protease domains, the 3C cystein protease (pfam00548) and the peptidase C3G (pfam12381) domains, and a RdRP (cd1699) domain. At the 3' end, the unique ORF of BV2 encoded two *Picornavirus* capsid protein-

like domains (cd00205) followed by a CRPV capsid protein-like domain (pfam08762), which are all typical domains from picorna-like non-structural virus proteins. It is noteworthy that such genome organization comprising one ORF coding for non-structural proteins at the 5' end and structural proteins at the 3' end was only characterized for the *Marnaviridae* family inside the picorna-like virus supergroup, and it comprises only one species so far. Nevertheless, protein blast analysis provided the highest similarity for the non-structural protein of *Solenopsis invicta virus-1*, a *Dicistrovirus*, with 26% identity over 797 residues (e-value:  $7^{-56}$ ).

#### 3.1.4 *Biomphalaria glabrata* Virus 3 (BV3)

The RNAseq contig coverage of BV3 genome sequence was composed of multiple small overlapping fragments from the 5' side to the two-thirds of the sequence, separated by a gap from two overlapping fragments at the 3' side (Figure 1). The sequence of the gap was completed by PCR and the two ends were obtained by RACE-PCR. The near full-length viral genome sequence generated was constituted of 9 982 nt comprising one ORF of 8 556 nt, a 5' UTR and a 3' UTR of 619 and 807 nt, respectively. A poly(A) tail was also recovered at the 3' end of the genome.

The unique BV3 ORF encoded a 2 852 amino acids product. Two picornavirus capsid protein domain-like (cd00205) were found at the 5' side of BV3 and non-structural protein domains for RNA helicase (pfam00910) and RdRP (cd1699) at the 3' side, which is the reverse organization as compared to the one of BV2. Cysteine protease domain was not detected as it probably differed from already known sequences; however, the cysteine protease motif  $^{1910}\text{GxCG}^{1913}$  was retrieved as well as the potential catalytic triad constituted by H<sup>1804</sup>, E<sup>1846</sup> and C<sup>1912</sup>. Inside the *Picornavirales* order, only *Secoviridae*, *Picornaviridae* and *Iflaviridae* families have the replication block in the C-terminal region of the polyprotein, while the structural proteins are found in the N-terminal region. Protein BLAST analysis provided the best hit for mouse *Encephalomyocarditis virus*, a member of the mammalian *Picornaviridae* family, with 28% identity over 297 residues (e-value:  $1\text{e}^{-17}$ ).

### 3.1.5 *Biomphalaria pfeifferi* Virus 4 (BV4)

The contig coverage of BV4 was the weakest of the five viruses identified in *Biomphalaria* species (Figure 1). The many gaps within the sequence were successfully filled by PCR but RACE-PCR failed to amplify both genome ends. The partial genome sequence of BV4 that we obtained was 7 264 nt long, which is constituted of a unique ORF, incomplete at its 5' side, and of a fragment of the 3'UTR (103 nt).

Partial sequence of BV4 ORF coded for a 2 386 residues product. CD search analysis revealed the same domain organization as BV2, with non-structural proteins at the 5' end and structural proteins at the 3' end (Figure 1). The 5' end of the ORF started inside the RNA helicase domain (pfam00910) and was followed by the RdRP domain (cd1699). No protease domain was detected and its manual identification was less obvious than for other viruses. The cysteine protease motif <sup>1003</sup>GxCG<sup>1006</sup> was clearly identifiable but the putative substrate binding site <sup>1022</sup>GxMxxG<sup>1027</sup> was divergent from the conserved one (GxHxxG). In addition, the first H<sup>898</sup> and last C<sup>1005</sup> residues of the catalytic triad were found, but several candidate residues were found for the second one (E<sup>930</sup>, D<sup>941</sup>, E<sup>949</sup>, E<sup>951</sup> and D<sup>954</sup>). In the structural protein part at the 3' side of the ORF, two picornavirus capsid protein-like domains (cd00205) were identified. Protein BLAST analysis provided the best hit for the polyprotein of *Fisavirus 1* (GenBank accession NC\_025836), with 32% identity over 1 350 residues (e-value: 3e<sup>-179</sup>). *Fisavirus 1* has been recently described as an unclassified *Picornavirales* isolated from fresh water carp feces (Reuter et al., 2015), with a 8 712 nt long genome bearing the same genome organization as BV4.

### 3.1.6 *Biomphalaria pfeifferi* Virus 5 (BV5)

The last virus BV5 was identified by the presence of a few non-overlapping contigs corresponding to the RdRP gene. Sequence gaps were fulfilled by PCR, but only a partial 1013 nt long sequence of the unique ORF of its genome was obtained (Figure 1).

The ORF encoded a 337 residues product. This comprised the RdRP domain (cd1699) and a RdRP

4 domain (pfam02123), which is a RdRP protein found in *Luteovirus*, *Totivirus* and *Rotavirus*. The best BLAST hit was for the RdRP of *Golden shiner totivirus*, with 54% identity over 319 amino acids (e-value:  $3e^{-111}$ ). This virus was recently characterized from a fresh water fish from a North American lake and it presented a double-stranded RNA genome (Mor and Phelps, 2016). Moreover, all the other BLAST hits also corresponded to RdRP from other *Totivirus*, suggesting that BV5 could be a new candidate species for this family.

### 3.2 Taxonomic position of viruses

Both phylogenetic analyses performed using the RdRP domain (Figure 2) and helicase domain sequences (Supp. Fig. Helicase) clustered BV1 with a group of unclassified *Picornaviridae* supported with high bootstrap values for RdRP sequences.

This group of unclassified *Picornaviridae* comprised an isolate from *Spodoptera exigua* and several *Nora* viruses from different *Drosophila* species. In addition to structural and non-structural proteins, each encoded by a specific ORF, *Nora* viruses possess 2 additional ORFs; one is coding for a RNAi suppressor protein, and the other one encodes a protein of unknown function. The *S. exigua virus* possesses 5 ORFs, one that encodes non-structural proteins and the four others that encode proteins without any conserved motif sequence. Therefore, the clustering of BV1 on the basis of RdRP and helicase protein sequence similarity is not consistent with the genome organization of viruses from this cluster. BV1 presents the typical dicistrovirus organization with its dicistronic genome structure comprising the 2 ORFs coding from 5' to 3' end for non-structural and structural proteins, respectively. This phylogenetic relationship of BV1 with *Nora* viruses is surprising as it did not have the same genomic structure. As *Nora* viruses have been previously described to be enteric viruses transmitted by feces (Habayeb et al., 2009); it could be interesting to investigate in the future the tropism of BV1 for snail tissues.

Analysis of RdRP sequences led to the clustering of BV2 with the *Secoviridae*, which is a family of

viruses known to infect plants. But on one hand, this taxonomic position is weakly supported by the bootstrap analysis, and on the other hand, its genomic organization is not that of *Secoviridae*, but rather the one of *Marnaviridae*. A complementary phylogenetic analysis based on the helicase domain (Figure S1) positioned BV2 between BV1 and the *Marnaviridae*, but again supported by a low bootstrap value. To be noted, the helicase sequence used for this analysis was truncated, which could have biased the phylogenetic analysis due to the treatment of the missing amino-acids in the neighbor-joining model used (gap treatment: pairwise deletion). It is obvious that BV2 is a picorna-like virus as it possesses all their typical domains, but to date we were not able to clearly classify it among the already known families by phylogenetic analyses.

The two phylogenetic analyses positioned BV3 with high bootstrap values inside the *Picornaviridae* family, which groups the mammalian picornaviruses. Moreover, the genome organization of BV3 is in accordance with that of the *Picornaviridae* family. This family comprises several genus and BV3 is clustered together with several unclassified *Picornaviridae* identified from different bird species. The discovery of other viruses phylogenetically closer to BV3 may help to give more insights into its exact position within this family.

BV4 clustered with a group of unclassified *Picornavirales* in analyses performed with RdRP and helicase sequences. It also shared a similar domain organization with viruses from this group. All viruses from this group were isolated from feces and it comprises viruses identified from farmed animal's feces like swine and fish, but also from human and fly feces. It would be interesting in the future to study the tissue tropism of BV4 to determine if it also has an intestinal origin. It seems clear that a new group of picorna-like viruses, which includes BV4, is emerging and that it has to be assigned to a new family or to a new group inside an existing family.

The phylogenetic analysis of the RdRP protein sequence of BV5 clustered it with the *Totiviridae* family. It particularly positioned it with *Totiviruses* from other invertebrates, including shrimp, mosquito, and fly. *Totiviruses* are non-polyadenylated double-stranded RNA viruses and they exhibit the same dicistronic genome organization with a large first ORF encoding structural

proteins, and non-structural proteins encoded by a shorter second ORF. BV5 presents a very strong bootstrap support in spite of a very short protein fragment.

### 3.3 Virus distribution and relative abundance in different *Biomphalaria* strains and species

The five viruses identified in this study were not all present in all snail strains studied (Figure 3). This confirms that our snail rearing process and facilities properly isolate the different snail strains with their own virus community and do not allow inadvertent transmissions between all different snail strains. Previous studies have already shown a correlation between viral reads number and viral load (Valenzuela-Miranda et al., 2014; Wilk et al., 2015). Even if the exact number of viral particles cannot be estimated by this way, it gives a relative viral abundance between different biological samples. Therefore, we used our RNA-seq data to estimate the relative abundances of each virus within the five *Biomphalaria* strains. Total viral load was similar in the *Biomphalaria* strains BgVEN, BgBRE, BgGUA and BpOMAN (between 5.2 and 6.1 normalized hit counts) but it was 2.4 to 2.9 times lower in BgBAR. Proportion of the five viruses varied between the different *Biomphalaria* strains studied (Figure 3).

BV1 was present in all tested snail strains, including both *B. glabrata* and *B. pfeifferi* species. It was the most abundant virus within three *B. glabrata* strains (81.0, 81.4 and 79.0% of total viral load in BgVEN, BgBAR and BgGUA, respectively) whereas it represented only 46.2% of total viral load in BgBAR strain. Interestingly, genome sequence of BV1 from BgBAR strain was slightly different from BV1 from other snail strains. In *B. pfeifferi*, BV1 was also present but its abundance was very low (0.9% of total viral load), which could explain why it was more difficult to amplify by PCR from BpOMAN and why we only managed to get 8 732 nt over 10 116 at the 3' side. Such variation in the relative abundance of this virus between the two snail species suggests that it might be differently constrained by the presence of different other viruses and potentially different other members of the microbiota but also by differences in the metabolism, immunity and physiology of



the different species.

In *B. glabrata*, BV2 and/or BV3 were also recovered. Their abundance varied according to the snail strain. It is noteworthy that Bowtie2 analysis revealed the presence of BV2 reads in BgVEN transcriptome, whereas no viral transcript nor PCR amplicon corresponding to BV3 was recovered from BgVEN. We could hypothesize that viral load might be extremely low in this snail strain that could partly explain why PCR-based approaches failed. Another possibility is that mapped reads might belong to another uncharacterized virus from BgVEN having sequence homology with BV3. In BgBRE, BV2 was much more abundant than BV3, while in BgVEN it was the opposite (Figure 3). Interestingly, BV3 abundance within the different *B. glabrata* strains seems to decrease from populations sampled in Caribbean Island and Venezuela to Brazil (from BgGUA to BgVEN, BgBRE and BgBAR) while BV2 load is increasing. This potential trade-off in the geographical distribution of BV2 and BV3 needs to be further investigated. In *B. pfeifferi*, BV2 was as abundant as in BgBRE and BgBAR. Consequently, this virus was able to infect both *Biomphalaria* species, like BV1. The most abundant virus in *B. pfeifferi* was BV4 (78.4% of total viral load), which was only recovered in this strain. Similarly, BV5 is specific to *B. pfeifferi* but it exhibited a 48-fold lower abundance than BV4. Again, this particular viral distribution could be due to a specific tropism of BV4 and BV5 to *B. pfeifferi* species and/or could be linked to the geographical origin of the snail strain. Viral specificity should be investigated by performing viral transmission experiments.

Nevertheless, we have to keep in mind that RNAseq data were obtained from pools of 30 animals. Therefore, it provides a global overview of virus distribution and abundances within each strain of snails but it fails in reaching the inter-individual plasticity of virus abundance and diversity. Future studies are thus also needed to solve this question.

### 3.4 *Biomphalaria* resident viruses as potential new schistosomiasis control agents?

The control of host snail population is one of the fighting strategies proposed by the WHO for schistosomiasis eradication (Rollinson et al., 2013). Limiting snail populations in endemic areas could constrain parasite transmission. In this context, several approaches have been developed or proposed. Molluscicides have been developed (McCullough et al., 1980), but their toxicity to the abiotic and biotic environments is a limit to their massive use in the field. The management of snail habitat together with the introduction of predators or snail competitors for habitat have also been used (Giboda et al., 1997; Guimaraes et al., 2001; Hofkin et al., 1992; Pointier and Jourdane, 2000; Sokolow et al., 2014). Unfortunately, snail population dynamics are highly variable depending on environmental conditions and thus the efficacy of such strategies is therefore difficult to predict.

The use of natural snail pathogens could also be considered for controlling snail populations. However, only few studies have been conducted using natural snail microbes. Two snail pathogens have been reported to become potential control agents. *Brevibacillus lacterosporus* was shown to induce mortality in some but not all *B. glabrata* populations, and only on juvenile snails (3-12 days-old post-eclosion) (de Oliveira et al., 2004). More recently, a new bacterial strain of *Paenibacillus*, *P. glabratella*, was reported to be lethal for *B. glabrata* snails, with a high rate of mortality at all developmental stages of the snail (Duval et al., 2015). Nevertheless, its pathogenicity against other schistosome intermediate host species as well as its safety toward other local endemic species is still to be determined. Our results open a new field of research using *Biomphalaria* viruses. Nowadays, the role of holobiont in the modulation of host immune system is well known in other models (Gilbert et al., 2012). Thus, it will be important to discover the role (positive or negative) played by these viruses on the snail homeostasis, especially during *S. mansoni* parasite infection. Without evident pathogenicity, virus could affect snail life history traits potentially depending on the viral tropism. Future work will investigate viral prevalence at individual level and potential pathogenicity in reciprocal host/virus crosses to investigate if these viruses could be used as a new strategy to control vector snail populations in the field.

#### 4. Conclusion

In this study, we report the identification and characterization of viruses from five *Biomphalaria* snail strains from RNAseq data. We discovered the presence of at least 5 new RNA viruses in both *B. glabrata* and *B. pfeifferi* species. The analysis of their sequence and genome organization revealed that four viruses were associated with the picorna-like virus superfamily, while the fifth one is close to the *Totiviridae* family. These results are supported by a phylogenetic analysis based on the RdRP and helicase protein sequences. The distribution analysis showed that BV1 was present in all the 5 snail strains while the four others were not equally distributed. Some of them, such as BV2 and BV3, seem to have a geographic-specific origin while others, such as BV4 and BV5, might be species-specific as they were only recovered from *B. pfeifferi*. BV1 is clearly the most abundant virus in *B. glabrata*, while it is BV4 in *B. pfeifferi*.

This work is a first step towards the identification of natural biological agents able to infect several populations of *Biomphalaria*, which is the intermediate host of the parasite *S. mansoni*. Future work will investigate viral prevalence at individual level, but also in field populations. The viral tropism, as well as the mode of transmission, has to be studied to better understand the different steps of the contamination process. The role played by these viruses on snail homeostasis also has to be investigated and more particularly during *S. mansoni* parasite infection to know if the presence of these viruses would be adaptive for the host.

#### 5. Acknowledgements

This work was supported by ANR JCJC INVIMORY (number ANR-13-JSV7-0009) funded to BG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 6. References

- Basch, P.F., 1976. Intermediate host specificity in *Schistosoma mansoni*. *Experimental parasitology* 39, 150-169.
- Bettarel, Y., Bouvier, T., Nguyen, H.K., Thu, P.T., 2015. The versatile nature of coral-associated viruses. *Environmental microbiology* 17, 3433-3439.
- Bosch, T.C., McFall-Ngai, M.J., 2011. Metaorganisms as the new frontier. *Zoology (Jena, Germany)* 114, 185-190.
- Burke, G.R., Strand, M.R., 2012. Polydnviruses of Parasitic Wasps: Domestication of Viruses To Act as Gene Delivery Vectors. *Insects* 3, 91-119.
- Chitsulo, L., Engels, D., Montresor, A., Savioli, L., 2000. The global status of schistosomiasis and its control. *Acta tropica* 77, 41-51.
- Christiaansen, A., Varga, S.M., Spencer, J.V., 2015. Viral manipulation of the host immune response. *Current opinion in immunology* 36, 54-60.
- de Oliveira, E.J., Rabinovitch, L., Monnerat, R.G., Passos, L.K., Zahner, V., 2004. Molecular characterization of *Brevibacillus laterosporus* and its potential use in biological control. *Applied and environmental microbiology* 70, 6657-6664.
- Dheilly, N.M., Duval, D., Mouahid, G., Emans, R., Allienne, J.F., Galinier, R., Genthon, C., Dubois, E., Du Pasquier, L., Adema, C.M., Grunau, C., Mitta, G., Gourbal, B., 2015a. A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. *Developmental and comparative immunology* 48, 234-243.
- Dheilly, N.M., Maure, F., Ravallec, M., Galinier, R., Doyon, J., Duval, D., Leger, L., Volkoff, A.N., Misse, D., Nidelet, S., Demolombe, V., Brodeur, J., Gourbal, B., Thomas, F., Mitta, G., 2015b. Who is the puppet master? Replication of a parasitic wasp-associated virus correlates with host behaviour manipulation. *Proceedings* 282, 20142773.
- Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloe, D., Le Gall, C., Schaeffer, B., Le Crom, S., Guedj, M., Jaffrezic, F., 2012. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics* 14, 671-683.
- Dong, Y., Manfredini, F., Dimopoulos, G., 2009. Implication of the mosquito midgut microbiota in the defense against malaria parasites. *PLoS pathogens* 5, e1000423.
- Duval, D., Galinier, R., Mouahid, G., Toulza, E., Allienne, J.F., Portela, J., Calvayrac, C., Rognon, A., Arancibia, N., Mitta, G., Theron, A., Gourbal, B., 2015. A novel bacterial pathogen of *Biomphalaria glabrata*: a potential weapon for schistosomiasis control? *PLoS neglected tropical diseases* 9, e0003489.
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A., 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome research* 15, 1451-1455.
- Giboda, M., Malek, E.A., Correa, R., 1997. Human schistosomiasis in Puerto Rico: reduced prevalence rate and absence of *Biomphalaria glabrata*. *The American journal of tropical medicine and hygiene* 57, 564-568.
- Gilbert, S.F., Sapp, J., Tauber, A.I., 2012. A symbiotic view of life: we have never been individuals. *The Quarterly review of biology* 87, 325-341.
- Gorbalenya, A.E., Donchenko, A.P., Blinov, V.M., Koonin, E.V., 1989. Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS letters* 243, 103-114.
- Guimaraes, C.T., Souza, C.P., Soares, D., 2001. Possible competitive displacement of planorbids by *Melanoides tuberculata* in Minas Gerais, Brazil. *Memorias do Instituto Oswaldo Cruz* 96 Suppl, 173-176.
- Habayeb, M.S., Cantera, R., Casanova, G., Ekstrom, J.O., Albright, S., Hultmark, D., 2009. The *Drosophila* Nora virus is an enteric virus, transmitted via feces. *Journal of invertebrate pathology*

101, 29-33.

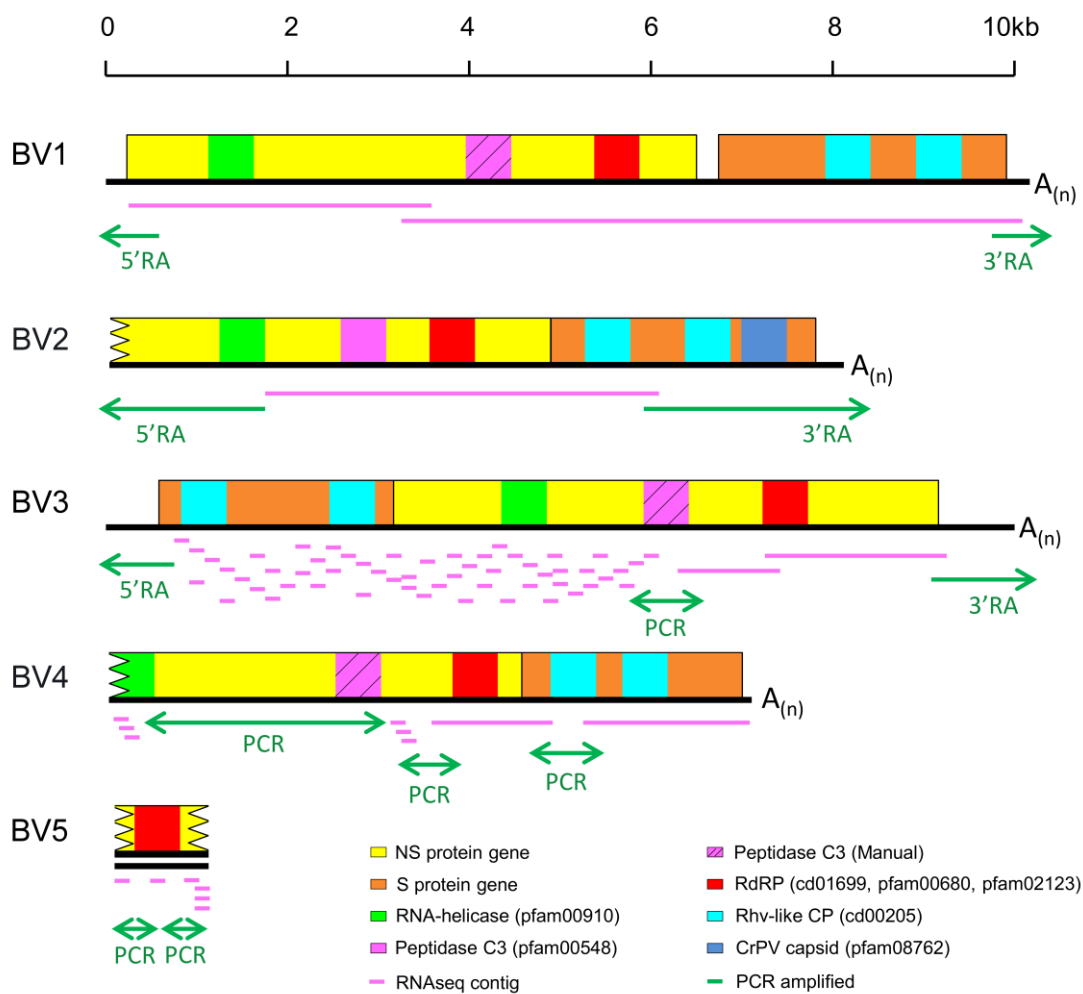
- Habayeb, M.S., Ekengren, S.K., Hultmark, D., 2006. Nora virus, a persistent virus in *Drosophila*, defines a new picorna-like virus family. *The Journal of general virology* 87, 3045-3051.
- Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.
- Hofkin, B.V., Hofinger, D.M., Koech, D.K., Loker, E.S., 1992. Predation of *Biomphalaria* and non-target molluscs by the crayfish *Procambarus clarkii*: implications for the biological control of schistosomiasis. *Annals of tropical medicine and parasitology* 86, 663-670.
- Ibrahim, M.M., 2006. Energy allocation patterns in *Biomphalaria alexandrina* snails in response to cadmium exposure and *Schistosoma mansoni* infection. *Experimental parasitology* 112, 31-36.
- Knight, M., Elhelu, O., Smith, M., Haugen, B., Miller, A., Raghavan, N., Wellman, C., Cousin, C., Dixon, F., Mann, V., Rinaldi, G., Ittiprasert, W., Brindley, P.J., 2015. Susceptibility of Snails to Infection with Schistosomes is influenced by Temperature and Expression of Heat Shock Proteins. *Epidemiology (Sunnyvale, Calif)* 5.
- Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical reviews in biochemistry and molecular biology* 28, 375-430.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* 33, 1870-1874.
- Liu, S., Vijayendran, D., Bonning, B.C., 2011. Next generation sequencing technologies for insect virus discovery. *Viruses* 3, 1849-1869.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. *Nucleic acids research* 43, D222-226.
- Marston, D.A., McElhinney, L.M., Ellis, R.J., Horton, D.L., Wise, E.L., Leech, S.L., David, D., de Lamballerie, X., Fooks, A.R., 2013. Next generation sequencing of viral RNA genomes. *BMC genomics* 14, 444.
- Martinez, J., Fleury, F., Varaldi, J., 2015. Competitive outcome of multiple infections in a behavior-manipulating virus/wasp interaction. *Ecology and evolution* 5, 5934-5945.
- McCullough, F.S., Gayral, P., Duncan, J., Christie, J.D., 1980. Molluscicides in schistosomiasis control. *Bulletin of the World Health Organization* 58, 681-689.
- Mor, S.K., Phelps, N.B., 2016. Molecular detection of a novel totivirus from golden shiner (*Notemigonus crysoleucas*) baitfish in the USA. *Archives of virology* 161, 2227-2234.
- Pointier, J.P., Jourdane, J., 2000. Biological control of the snail hosts of schistosomiasis in areas of low transmission: the example of the Caribbean area. *Acta tropica* 77, 53-60.
- Pradeu, T., 2016. Mutualistic viruses and the heteronomy of life. *Studies in history and philosophy of biological and biomedical sciences*.
- Reuter, G., Pankovics, P., Delwart, E., Boros, A., 2015. A novel posavirus-related single-stranded RNA virus from fish (*Cyprinus carpio*). *Archives of virology* 160, 565-568.
- Rollinson, D., Knopp, S., Levitz, S., Stothard, J.R., Tchuem Tchuente, L.A., Garba, A., Mohammed, K.A., Schur, N., Person, B., Colley, D.G., Utzinger, J., 2013. Time to set the agenda for schistosomiasis elimination. *Acta tropica* 128, 423-440.
- Ruelas, D.S., Karentz, D., Sullivan, J.T., 2009. Effects of UVB on interactions between *Schistosoma mansoni* and *Biomphalaria glabrata*. *Journal of invertebrate pathology* 101, 140-142.
- Ryan, M.D., Flint, M., 1997. Virus-encoded proteinases of the picornavirus super-group. *The Journal of general virology* 78 ( Pt 4), 699-723.
- Sokolow, S.H., Lafferty, K.D., Kuris, A.M., 2014. Regulation of laboratory populations of snails (*Biomphalaria* and *Bulinus* spp.) by river prawns, *Macrobrachium* spp. (Decapoda, Palaemonidae): implications for control of schistosomiasis. *Acta tropica* 132, 64-74.
- Theron, A., Rognon, A., Gourbal, B., Mitta, G., 2014. Multi-parasite host susceptibility and multi-host parasite infectivity: a new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni*

compatibility polymorphism. *Infect Genet Evol* 26, 80-88.

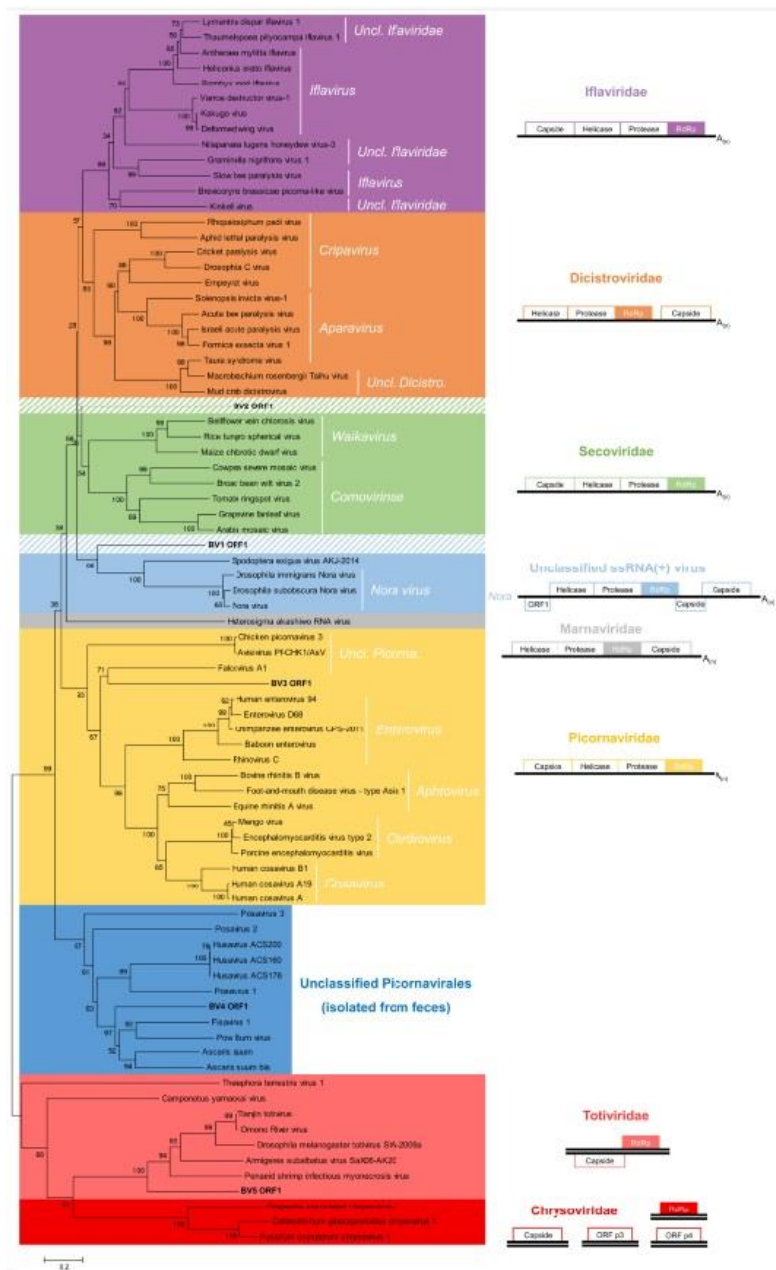
Valenzuela-Miranda, D., Cabrejos, M.E., Yanez, J.M., Gallardo-Escarate, C., 2014. From the viral perspective: infectious salmon anemia virus (ISAV) transcriptome during the infective process in Atlantic salmon (*Salmo salar*). *Marine genomics* 20, 39-43.

Wilk, E., Pandey, A.K., Leist, S.R., Hatesuer, B., Preusse, M., Pommerenke, C., Wang, J., Schughart, K., 2015. RNAseq expression analysis of resistant and susceptible mice after influenza A virus infection identifies novel genes associated with virus replication and important for host resistance to infection. *BMC genomics* 16, 655.

Zilber-Rosenberg, I., Rosenberg, E., 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS microbiology reviews* 32, 723-735.

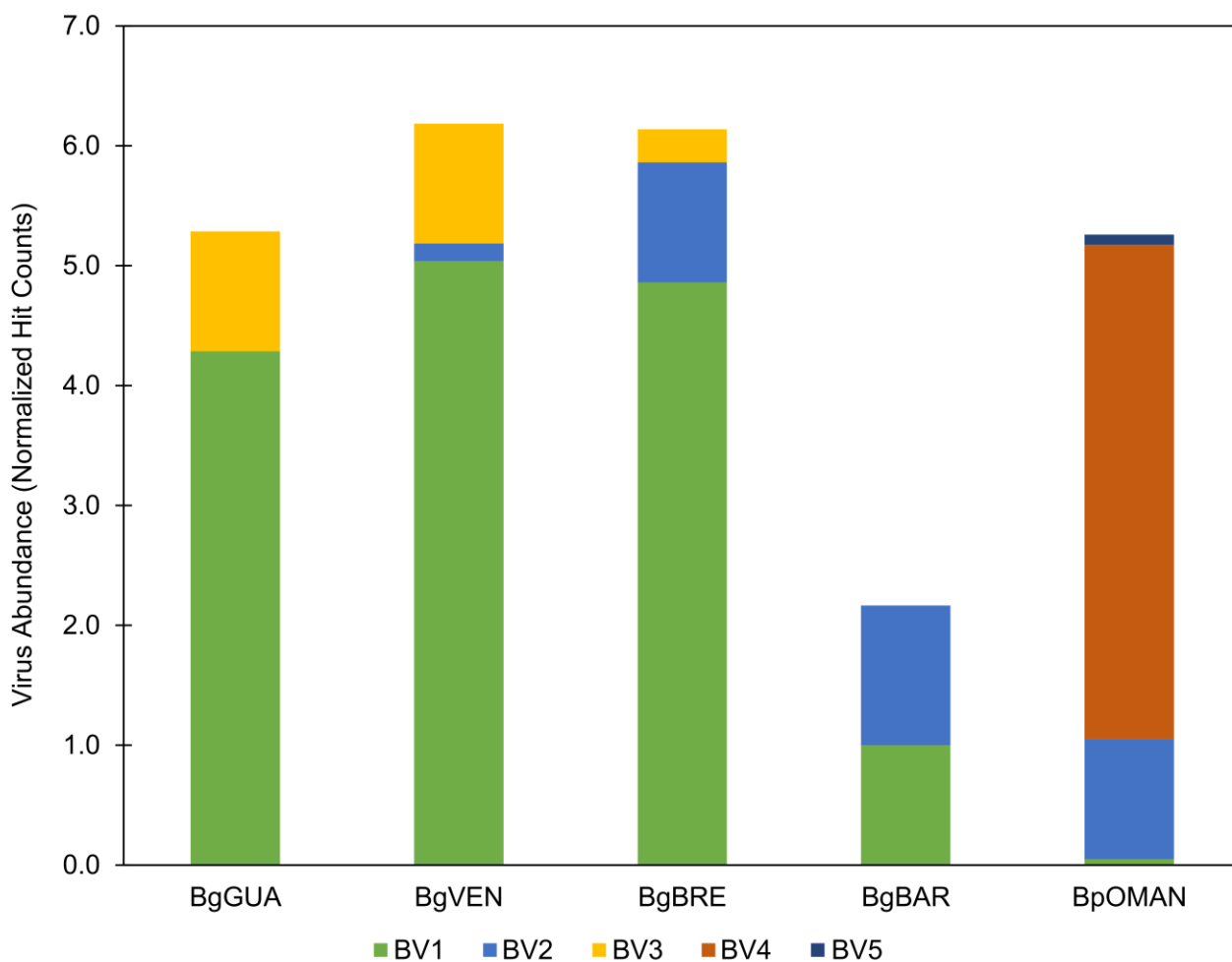


**Figure 1.** Genome structures of *Biomphalaria* RNA-viruses. Black lanes represent the RNA strands (one for BV1, BV2, BV3 and BV4 or two for BV5). Yellow and orange boxes correspond to the region coding for non-structural proteins and structural proteins, respectively. Colored squares in these boxes relate the presence of conserved protein domains identified either by CD-search analysis or manually (hatched). Transcripts from RNAseq data are materialized by pink lines, while gaps and cDNA ends obtained only by PCR and RACE-PCR are illustrated by green arrows.



**Figure 2.** Phylogenetic analysis of the RdRP protein sequence domains from 81 different viruses belonging to the *Iflaviridae* (13; purple), *Dicistroviridae* (12; orange), *Secoviridae* (8; green), Unclassified ssRNA(+) virus (4; light blue), *Marnaviridae* (1; grey), *Picornaviridae* (17; yellow), Unclassified *Picornavirales* isolated from feces (10; dark blue), *Totiviridae* (8; light red) and *Chrysoviridae* (3; dark red) families in addition to the five viruses isolated from *Biomphalaria*. Viruses from *Biomphalaria* (BV1 to BV5) are indicated in bold and are highlighted in the color of the family if they were clearly included in or they are hatched with the color of the closest family. A bootstrap analysis of 2000 replications was carried out on the trees inferred from the neighbor joining method. Bootstrap values are shown at each branch of the trees. The accession number of each protein sequence used is listed in Supplementary Table 1.





**Figure 3:** Relative abundance of virus into *Biomphalaria* strains. Cumulative expression (Hit count normalized by the upper-quartile) of the *Biomphalaria* virus transcripts, in the different *Biomphalaria* strains, determined by Bowtie2 analysis. Each histogram corresponds to a *Biomphalaria* strain (BgGUA, BgVEN, BgBRE, BgBAR and BpOMAN) and each color corresponds to a virus (green for BV1, blue for BV2, yellow for BV3, orange for BV4 and purple for BV5).

**Table 1.** Identity percentages between BV1 nucleic sequences.

	<b>GUA</b>	<b>OMAN</b>	<b>VEN</b>	<b>BRE</b>	<b>BAR</b>
<b>GUA</b>	100				
<b>OMAN</b>	99.79	100			
<b>VEN</b>	99.25	99.39	100		
<b>BRE</b>	97.12	97.32	97.32	100	
<b>BAR</b>	86.7	86.75	86.75	86.56	100

*Pairwise sequence alignment between nucleic sequences of BV1 from the four B. glabrata strain and the B. pfeifferi one was performed using EMBOSS Needle tools from the EBI server (<https://www.ebi.ac.uk/>).*