ACADEMIA JOURNAL OF BIOLOGY 2021, 43(2): 55–71
DOI: 10.15625/2615-9023/16018

GENOMIC ANALYSIS OF *Aeromonas hydrophila* BACTERIOPHAGES ISOLATED IN STRIPED CATFISH FARMS IN THE MEKONG DELTA, VIETNAM

Tu Quang Vinh[1,2], Nguyen Le Hoai Nhan[1,2], Tran Thi Thanh Xuan[1,2], Le Phi Nga[1,2], Hoang Anh Hoang[1,2,*]

[1]Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam
[2]Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam

Received 16 December 2020, accepted 2 June 2021

ABSTRACT

Striped catfish or Vietnamese catfish mainly contributes to national aquaculture exports. However, bacterial diseases result in the decrease of striped catfish production efficiency and the most popular disease is hemorrhagic septicemia caused by *Aeromonas hydrophila*. Bacteriophages or bacterial viruses have been investigated as alternative biological agents to control pathogenic bacterial infections in aquaculture for many decades. A few genomes of striped catfish *A. hydrophila* bacteriophages have been analyzed, although many genomes of *A. hydrophila* bacteriophages in other fish have been investigated. In this study, the whole genome sequences of three new *A. hydrophila* bacteriophages such as PVN03, PVN04 and PVN05 were described. The morphological analysis by transmission electron microscopy indicated that these phages belonged to Myoviridae family. The genome sizes of the phages PVN03, PVN04 and PVN05 were 50,725 bp, 51,721 bp and 51,884 bp, respectively. PVN03 had 64 open reading frames (ORFs), while PVN04 and PVN05 had 65 ORFs and 66 ORFs, respectively. No tRNAs, rRNAs or sRNAs, antibiotic resistance genes, virulence factors, toxin genes and integrase genes were detected in three phage genomes. ANI analysis tool showed that *Aeromonas hydrophila* phages isolated in Vietnam formed a distinct group having a very low similarity with other *Aeromonas hydrophila* phages available on the database. In addition, phylogenetic tree analysis indicated that these phages formed a new genus in the Myrovidae family.

Keywords: *Aeromonas hydrophila*, bacteriophage, striped catfish, comparative genomic analysis.

Citation: Tu Quang Vinh, Nguyen Le Hoai Nhan, Tran Thi Thanh Xuan, Le Phi Nga, Hoang Anh Hoang, 2021. Genomic analysis of *Aeromonas hydrophila* bacteriophages isolated in striped catfish farms in the Mekong delta, Vietnam. *Academia Journal of Biology*, 43(2): 55–71. https://doi.org/10.15625/16018

**Corresponding author email*: hoang.a.hoang@hcmut.edu.vn

INTRODUCTION

The Mekong Delta is the main region for striped catfish *Pangasianodon hypophthalmus* farming in Vietnam. The fish is an important component of national aquaculture exports and also contributes to 90% of striped catfish produced globally (De-Silva & Phuong, 2011). According to Vietnam's Office of Statistics (2019), aquaculture exports generated 8.6 billion dollars for the nation's economy, in which striped catfish contributed 23.2% of such exports (VASEP, 2019). However, bacterial pathogen infections appear more popular and become the main factor affecting the sustainable development of the striped catfish industry. One of the most important bacterial pathogens is *Aeromonas hydrophila* that causes hemorrhagic septicemia in striped catfish. Antibiotics have been commonly used in the prevention and treatment of the disease. However, antibiotic resistance by *A. hydrophila* has become more prevalent. Quach et al. (2014) reported that antibiotic resistance ratio of *A. hydrophila* isolates in ill striped catfish in the Mekong Delta, Vietnam was 100% to ampicillin, amoxicillin, cefalexin, trimethoprim/sunfamethoxazol and was 93% to tetracyclin. Moreover, the development of antibiotic resistance of *A. hydrophila* has resulted in an economic loss for the country. Many consignments from leading Vietnamese producers have been currently rejected due to higher-than-approved-limit of antibiotic residuals by importing markets. Because of these consequences of *A. hydrophila* infection, there is an urgent necessity for an alternative solution to antibiotic usage.

Bacteriophages or bacterial viruses have been investigated as alternative biological agents to control pathogenic bacterial infections for many decades. The method (phage therapy) has become more significant in the aquaculture industry in the last forty years with the widespread development of antibiotic resistant bacteria (Wu et al., 1981; Kowalska et al., 2020). The efficacy of phage therapy for the prevention and treatment of bacterial diseases in fish and shellfish has been reviewed (Doss et al., 2017; Culot et al., 2019). Some studies about phages to control *A. hydrophila* and other bacterial pathogens in striped catfish have been published since 2018 (Le et al., 2018; Hoang et al., 2018; Hoang et al., 2019; Dang et al., 2021).

Phages are considered as the most abundant organisms on the biosphere with an estimated number of 10^{31} (Clokie et al., 2010). However, the diversity also causes difficulty to select suitable candidates used in phage therapy. One of the important criteria for the selection of a phage is its lifestyle, lytic or temperate phage (Philipson et al., 2018). Thus, before the application of phage therapy against *A. hydrophila* infections in striped catfish, it is essential to understand clearly the characterization of the phage genome to be used (Philipson et al., 2018). In addition, other benefits can be derived from phage genome information such as the development of recombinant-based methods for detection of the pathogens (Tanji et al., 2004; Ripp et al., 2008; Hoang et al., 2014), understanding about evolution of the host pathogens and phages (Hatfull & Hendrix, 2011; Scanlan, 2017). Tu et al. (2020) reported the first complete genome sequence of a novel lytic phage infecting *Aeromonas hydrophila* in striped catfish. In this study, genomic analysis of some other *Aeromonas hydrophila* phages isolated in striped catfish farms in the MKDVN is shown and compared to existing phage genomic databases.

MATERIALS AND METHODS

Phage isolation

The bacterial strains *A. hydrophila* 4.3T and *A. hydrophila* 4.4T given by Dang Thi Hoang Oanh (College of Aquaculture and Fisheries, Can Tho University) were used as host bacteria in phage isolation. They were identified at the species level by verifying *aerolysin* gene (Panangala et al., 2007; Le et al., 2010). Water samples collected from striped catfish ponds in Can Tho City, Vietnam were used for bacteriophage isolation. Protocols for phage isolation and purification were described previously (Hoang et al., 2019).

Transmission electron microscope examination

A highly concentrated suspension of phage (approximately 10^{10} PFU mL^{-1}) was prepared as previously described (Ackermann, 2009a). The phage sample was then negatively stained with 5% uranyl acetate and observed by transmission electron microscope (JEOL JEM-1010) operating at a voltage of 80 kV and an instrumental magnification of 25,000−40,000 at the Vietnam National Institute of Hygiene and Epidemiology.

Genome sequencing

Phage nucleic acids were extracted from 1 mL phage stock (approximately 10^9 PFU.mL^{-1}) by using Phage DNA Isolation Kit (Norgen Biotek Inc, Canada) following the manufacturer's instructions. The purified nucleic acids were treated with DNase I, RNase A (Thermo Fisher Scientific, USA), and Mung bean nuclease (NEB, USA) followed by electrophoresis on agarose gel 1% to determine the genome type of phages. The purified nucleic acids were amplified using whole genome amplification techniques by EquiPhi29™ DNA Polymerase (Thermo Fisher Scientific, USA) under the manufacturer's protocol and purified by ethanol precipitation method. The samples were then sent to Microbial Genome Sequencing Center (Pittsburgh, PA, USA) for library preparation using Nextera XT and sequencing using Illumina NextSeq550 (150 bp paired end).

De novo assembly and annotation

Reads from sequencing were trimmed with Trimmomatic v0.39 (Bolger et al., 2014) using default parameters. Assembly the reads was carried out with Unicycler v0.4.8 (Wick et al., 2017) using the "Illumina-only assembly" option and any assembly errors were corrected with Pilon v1.23 (Walker et al., 2014) using default parameters. The genome was initially annotated using Prokka v1.14.6 (Seemann et al., 2014) with default parameters. Further manual annotation was carried out with BLASTp (e-value cutoff, 0.001) (Johnson et al., 2008), using the non-redundant protein database and NCBI Conserved Domain Database. After complete annotation of phages genome, phages safety determination was performed by analysis the genome sequences using ResFinder 4.0 (Bortolaia et al., 2020) and web service of Virulence Factors of Pathogenic Bacteria Database (VFPB) (Liu et al., 2019).

Comparative genome and Pan-genome analysis

Genomes of the phages were pairwise comparative using BLASTN web service (Johnson et al., 2008) on NCBI, tBLASTx through Easy Fig v2.2.2 (Sullivan et al., 2011) and compared with published genome of PVN02 that was also isolated previously in Can Tho, Vietnam (minimum identity cutoff setting was 50%) (Tu et al., 2020). Completed genomes of other *Aeromonas hydrophila* phages on Genbank were downloaded and combined with the genome of phages of our study for pan-genome analysis using GET_HOMOLOGUES v3.3.3 (Contreras-Moreira et al., 2013) with the default setting.

Phylogenetic tree and taxonomy

The terL amino acid sequences of the phages and the most similar phage were aligned with MAFTT (Katoh et al., 2013) and trees were constructed by IQ-TREE (Nguyen et al., 2015) using Maximum Likelihood method with 1000 bootstrap replicates. Further taxonomy analysis with vContact2 (Bin-Jang et al., 2019) software with RefSeq database (December, 2020) from Millard Lab (Leicester University, UK) to confirm the phylogenetic relationship.

RESULTS AND DISCUSSION

Isolation and morphology of phages

Three newly isolated phages named PVN03, PVN04, PVN05 showed their plaques of 2–3 mm diameters (Fig. 1). These phages were purified and their stocks were obtained with concentration of 10^9–10^{10} PFU mL^{-1}.

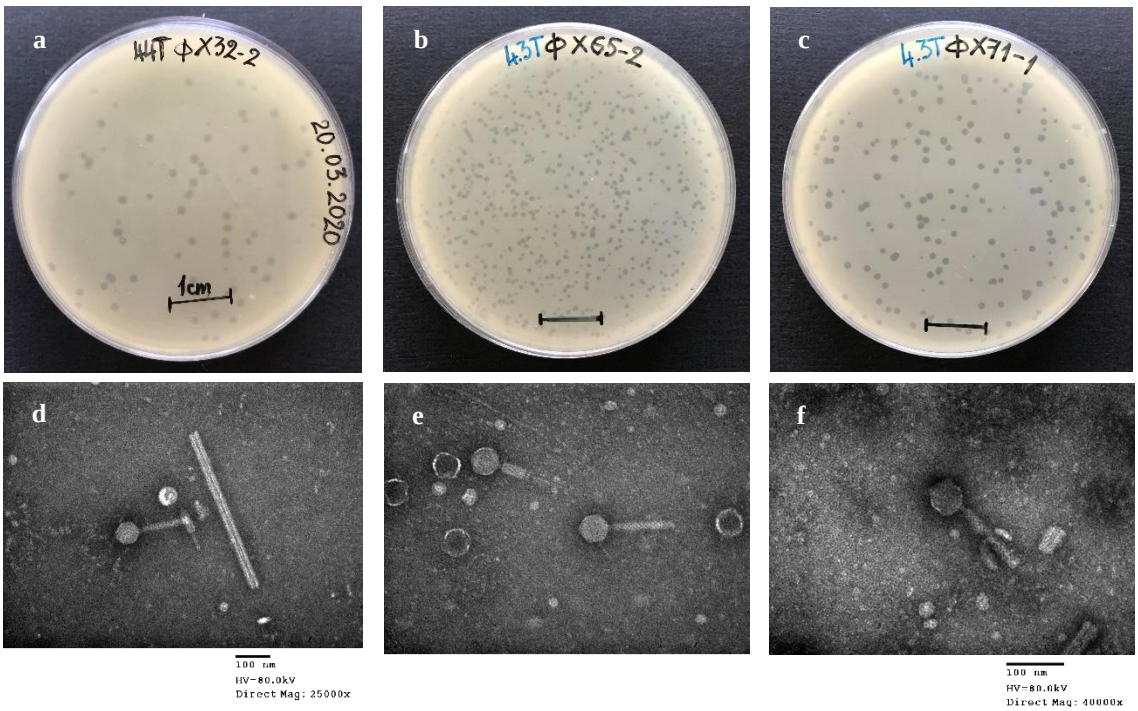

Figure 1. Phage plaques (a, b, c) and electron micrograph (d, e, f) of PVN03 (ϕX32-2) (a and d), PVN04 (ϕX65-2) (b and e) and PVN05 (ϕX71-1) (c and f)

Based on the morphological analysis by transmission electron microscopy (Fig. 1), PVN03, PVN04, PVN05 phages were placed in the Myoviridae family (Ackermann, 2009b). The PVN03 has an icosahedral head that is 65.6 nm in diameter. Its tail is 113 nm in length and 25.1 nm in width. The PVN04 has an icosahedral head that is 68.1 nm in diameter. Its tail is 135 nm in length and 12.8 nm in width. The PVN05 has an icosahedral head that is 56.7 nm in diameter. Its tail is 109 nm in length and 15.6 nm in width.

Genomic characterization

The genome of PVN03, PVN04 and PVN05 after being assembled had lengths of 50,725 bp, 51,721 bp and 51,884 bp, respectively. The percentage of GC content was 52.35% for PVN03, 52.44% for PVN04, and 52.43% for PVN05. No tRNAs, rRNAs, or sRNAs were detected in three genomes.

PVN03 had 64 open reading frames (ORFs) with 12 ORFs of predicted function as putative tail protein, peptidase m15a, hydrolase_2 domain-containing protein, RNA polymerase, amino acid adenylation domain-containing protein, DNA polymerase, 5'-3' exonuclease, DNA ligase, anaerobic NTP reductase large subunit, terminate large subunit and major capsid protein. PVN04 and PVN05 had 65 ORFs and 66 ORFs, respectively, in which 11 ORFs had predicted functions similar to PVN03. Detailed information was shown in table S1 (Supplementary material). NCBI GenBank accession numbers for PVN03, PVN04 and PVN05 were MW380983, MW380984 and MW380985, respectively.

One of the important criteria for the selection of a phage is its lifestyle, lytic or temperate phage (Philipson et al., 2018). The lytic phage infects the host cell and new phages are produced and released during bacterium lysis. In contrast, the temperate phage can operate as a vector that transfers virulent genes or resistant genes into the host genome via transduction. This process can result in a virulent host from a non-virulent

host. In aquaculture, Munro et al. (2003) demonstrated that the presence of the bacteriophage *V. harveyi* myovirus-like could confer virulence to *V. harveyi* strains explaining the large variation in pathogenicity among strains of *V. harveyi*, the causative agent of luminous vibriosis in larval prawns systems. Especially, the Early Mortality Syndrome disease has caused serious loss for shrimp farms in Vietnam since 2010 (Oanh et al., 2018). Tran et al. (2013) indicated that integration of *Vibrio parahaemolyticus* and a bacteriophage causing serious damage in the digestion system of shrimp is the main reason for the disease. Phages used in the phage therapy in aquaculture must be lytic phages, avoiding temperate phages. These three phages in the current study did not have antibiotic resistance genes, virulence factors, toxin genes and integrase genes. Therefore, they are considered safe for usage in phage therapy (Philipson et al., 2018).

Comparative genome and Pan-genome analytic

The genome of the three phages PVN03, PVN04, PVN05 and the previously published phage genome PVN02 (Tu et al., 2020) were compared to each other. The results showed that they have a high nucleotide sequence similarity as described in table 1.

Table 1. Similarity of nucleotide genome sequences of the phages

Phage	Similarity of the phages (%)			
	PVN02	PVN03	PVN04	PVN05
PVN02	100	-	-	-
PVN03	99.47	100	-	-
PVN04	99.60	99.57	100	-
PVN05	99.34	99.77	99.65	100

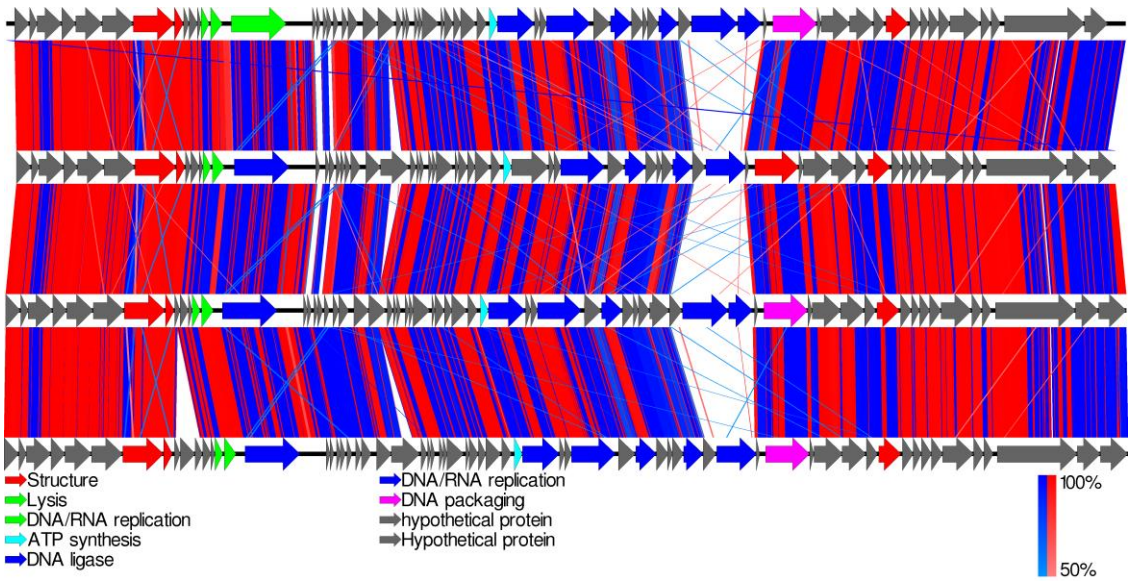

Figure 2. Genome map comparing the structure and distribution of ORFs in the four phage genomes; from the top to the down: PVN02, PVN03, PVN04 and PVN05

The tBLASTx based genome comparison map confirmed the high similarity among these four phages (Fig. 2). The phages had similar genome structure, the difference between genomes lies in ORFs in the replication region with two groups, PVN02

and PVN04 have two ORFs involved in the replication activity, namely ORF 47 and 48, PVN03 and PVN05 with only one ORF are ORF 47 with PVN03. and ORF 49 with PVN05. For the remainder of the ORFs, the differences are mostly SNPs.

Together with PVN02 phage (Tu et al., 2020), the genome of these phages is considered as the first complete genome sequences of phages infecting *Aeromonas hydrophila* in striped catfish.

Analysis results of the pan-genome of 47 *Aeromonas hydrophila* phages on the NCBI Genbank database (access on 24-December-2020) showed that there was no core gene among the phages. Usage of ANI analysis tool to get homologues showed that *Aeromonas hydrophila* phages in Vietnam had a very low similarity with other *Aeromonas hydrophila* phages available on the database. Pan-genome analysis of the four phages in Vietnam showed that there were 56 clusters with a difference in hypothetical genes.

Phylogenetic tree and taxonomy

Phage analysis results based on Terminase large subunit (terL) sequence showed that PVN03, PVN04, PVN05 are in the same branch as PVN02 (Tu et al., 2020). They form a separate new genus in the Myrovidae family (Fig. 3).

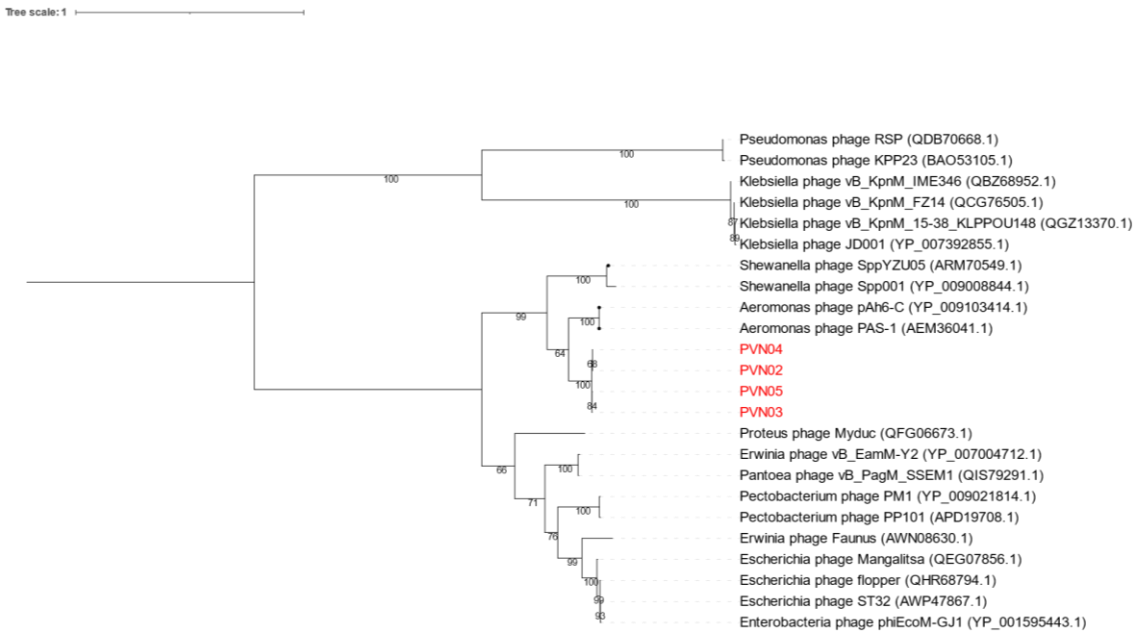

Figure 3. Phylogenetic analysis of three phages and selected phages based on the amino acid sequence of Terminase large subunit. Sequences were ordered using MAFTT software and phylogenetic tree constructed using IQ-TREE software using LG + G4 evolution model with 1000 bootstrap (show bootstrap values > 50)

The analysis using vContact2 software also showed forming a relationship with the Shewanella phage SppYZU05, *Aeromonas* phage pAh6-C and *Shewanella* phage Spp001 (Fig. 4). Besides, phages PVN03, PVN04 and PVN05 are also in a cluster with other phages having similar nucleotide sequences such as *Aeromonas hydrophila* PVN02 phage (LR813619), *Proteus* phage Myduc (MN098326), *Escherichia* phage flopper (MN850594), *Erwinia* phage vB_EamM-Y2 (HQ728264). This result is similar to the phylogenetic tree based on the amino acid terL sequence.

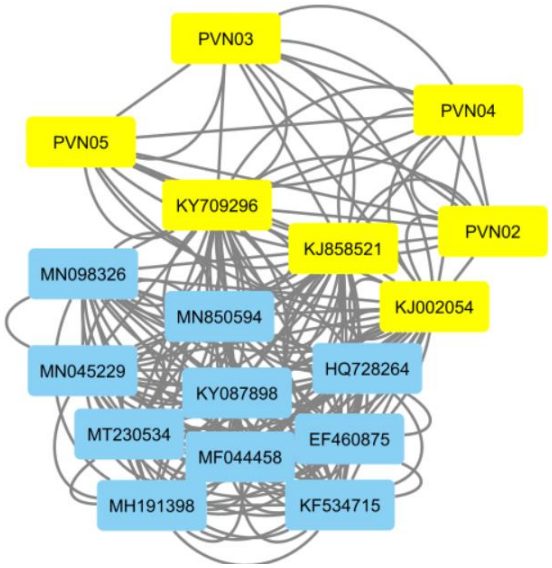

Figure 4. Gene-sharing network inferred by vContact2 and visualized with Cytoscape 3.8.9. The gray lines describe the interaction of genomes. The yellow boxes show a direct interaction among genomes. The blue boxes show an indirect interaction among genomes

CONCLUSION

Next generation sequencing technologies provide unique capacities for the comprehensive assessment of the functions and diversity of bacteriophages. In this study, the whole genome sequences and annotations for three new isolates of striped catfish *A. hydrophila* bacteriophages such as PVN03, PVN04 and PVN05 were shown. Results of genomic and electroscopic morphology analyses both indicated that three phages belonged to Myoviridae family. No tRNAs, rRNAs or sRNAs, antibiotic resistance genes, virulence factors, toxin genes and integrase genes were found in these phages indicating that they are considerable as safe candidates for phage therapy in aquaculture. It also showed that *Aeromonas hydrophila* phages in Vietnam had very low similarity to other *Aeromonas hydrophila* phages available on the database. These are considered as the first complete genome sequences of phages infecting *Aeromonas hydrophila* in striped catfish. In addition, phylogenetic tree analysis showed that these phages form a separate new group in the Myrovidae family. The lytic nature of the phage suggested it might serve as a potential agent to control *Aeromonas hydrophila* in striped catfish in the Mekong Delta, Vietnam.

Acknowledgements: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 10/2019/TN.

REFERENCES

Ackermann H. W., 2009[a]. Basic phage electron microscopy. In Bacteriophages, *Humana Press*, pp. 126.

Ackermann H. W., 2009[b]. Phage classification and characterization. In Bacteriophages, *Humana Press*, pp. 140.

Bin-Jang H., Bolduc B., Zablocki O., Kuhn J. H., Roux S., Adriaenssens E. M., Brister J. R., Kropinski A. M., Krupovic M., Lavigne R., Turner D., Sullivan M. B., 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol*., 37(6): 632–639.

Bolger A. M., Lohse M., Usadel B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114–2120.

Bortolaia V., Kaas R. S., Ruppe E., Roberts M. C., Schwarz S., Cattoir V., ..., Fagelhauer L., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother*., 75(12): 3491–3500.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L., 2009. BLAST+: Architecture and applications. *BMC Bioinformatics*, 10: 1–9.

Clokie M. R., Millard A. D., Letarov A. V., Heaphy S., 2011. Phages in nature. *Bacteriophage*., 1(1): 31–45.

Contreras-Moreira B., Vinuesa P., 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol*., 79(24): 7696–7701.

Culot A., Grosset N., Gautier M., 2019. Overcoming the challenges of phage therapy for industrial aquaculture: A review. *Aquaculture*, 513: 734423.

Dang T. H. O., Xuan T. T., Duyen L. T., Le N. P., Hoang H. A., 2021. Protective efficacy of phage PVN02 against *haemorrhagic septicaemia* in striped catfish *Pangasianodon hypophthalmus* via oral administration. *J. Fish Dis.* https://doi.org/10.1111/jfd.13387

De-Silva S. S., Phuong N. T., 2011. Striped catfish farming in the Mekong Delta, Vietnam: a tumultuous path to a global success. *Rev. Aquaculture.*, 3: 45–73.

Doss J., Culbertson K., Hahn D., Camacho J., Barekzi N., 2017. A Review of Phage Therapy against Bacterial Pathogens of Aquatic and Terrestrial Organisms. *Viruses* 9(3), pp. 50.

Hatfull G. F., Hendrix, R. W., 2011. Bacteriophages and their genomes. *Curr. Opin. Virol.*, 1(4): 298–303.

Hoang H. A., Tran T. T. X., Le P. N., Dang T. H. O., 2019. Selection of phages to control *Aeromonas hydrophila* - an infectious agent in striped catfish, *Biocontrol. Sci*., 24: 23–28.

Hoang H. A., Yen M. H., Ngoan V. T., Nga L. P., Oanh D. T. H., 2018. Virulent bacteriophage of *Edwardsiella ictaluri* isolated from kidney and liver of striped catfish *Pangasianodon hypophthalmus* in Vietnam. *Dis. Aquat. Organ.*, 132(1): 49–56.

Hoang A. H., Abe M., Nakasaki, K., 2014. A novel colorimetric method for the detection of *Escherichia coli* using cytochrome c peroxidase-encoding bacteriophage. *FEMS. Microbiol. Lett.* 352: 97–103.

Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden T. L., 2008. NCBI BLAST: a better web interface. *Nucleic. Acids. Res.*, 36: W5–W9.

Liu B., Zheng D., Jin Q., Chen L., Yang J., 2019. VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic. Acids. Res.*, 47(D1): D687–D692.

Katoh K., Standley D. M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.*, 30(4): 772–780.

Kowalska J. D., Kazimierczak J., Sowińska P. M., Wójcik E. A., Siwicki A. K., Dastych J., 2020. Growing trend of fighting infections in aquaculture environment opportunities and challenges of phage therapy. *Antibiotics* 9(6), pp. 301.

Le T. S., Nguyen T. H., Vo H. P., Doan V. C., Nguyen H. L., Tran M. T., Tran T. T., Southgate P. C., Kurtböke D.İ., 2018. Protective Effects of Bacteriophages against *Aeromonas hydrophila* Causing Motile *Aeromonas Septicemia* (MAS) in Striped Catfish. *Antibiotics*. 7(1): 15−16.

Le H. T., Truong Q. N., Nguyen H. G., Dang T. H. O., 2010. Study on application of multiplex PCR protocol for simultaneous detection of *Edwardsiella ictaluri* and *Aeromonas hydrophila* infection in kidney of striped catfish (*Pangasianodon hypophthalmus*). *J. Sci.* - Can Tho University, 16a: 129–135 (in Vietnamese with English summary).

Munro J., Oakey J., Bromage E., Owens L., 2003. Experimental bacteriophage-mediated virulence in strains of Vibrio harveyi. *Dis. Aquat. Organ.*, 54(3): 187–194.

Nguyen L. T., Schmidt H. A., Von Haeseler A., Minh B. Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol*., 32(1): 268–274.

Oanh D. T. H., Nghia N. T., Tien T. V., Bondad-Reantaso M. G., 2018. Identification and characterization of vibrio bacteria isolated from shrimp infected with early mortality syndrome/acute hepatopancreatic necrosis syndrome (EMS/AHPNS) in Vietnam. *Asian Fish. Sci*, 31: 283–292.

Panangala V. S., Shoemaker C. A., Van S. V. L., Dybvig K., Klesius P. H., 2007. Multiplex-PCR for simultaneous detection of 3 bacterial fish pathogens, *Flavobacterium columnare*, *Edwardsiella ictaluri*, and *Aeromonas hydrophila*. *Dis. Aquat. Organ.*, 74(3): 199–208.

Philipson C. W., Voegtly L. J., Lueder M. R., Long K. A., Rice G. K., Frey K. G., Biswajit B., Regina Z. C., Theron H., Bishop-Lilly K.A., 2018. Characterizing Phage Genomes for Therapeutic Applications. *Viruses*., 10(4): 187−188.

Quach V. C. T., Tu T. D., Dang P. H. H., 2014. The current status antimicrobial resistance in *Edwardsiella ictaluri* and *Aeromonas hydrophila* cause disease on the striped catfish farmed in the Mekong Delta. *J. Sci.* - Can Tho University, Special Issue: Aquaculture 2: 7–14 (in Vietnamese with English summary).

Ripp S., Jegier P., Johnson C. M., Brigati J. R., Sayler G. S., 2008. Bacteriophage-amplified bioluminescent sensing of Escherichia coli O157: H7. *Anal. Bioanal. Chem.*, 391(2): 507–514.

Scanlan P. D., 2017. Bacteria–bacteriophage coevolution in the human gut: implications for microbial diversity and functionality. *Trends. Microbiol.*, *25*(8): 614–623.

Seemann T., 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14): 2068–2069.

Sullivan M. J., Petty N. K., Beatson S. A., 2011. Easyfig: a genome comparison visualizer. Bioinformatics (Oxford, England), 27(7): 1009–1010.

Tanji Y., Furukawa C., Na S. H., Hijikata T., Miyanaga, K., Unno H., 2004. *Escherichia coli* detection by GFP-labeled lysozyme-inactivated T4 bacteriophage. *J. Biotechnol.*, 114(1–2): 11–20.

Tran L., Nunan L., Redman R. M., Mohney L. L., Pantoja C. R., Fitzsimmons K., Lightner D. V., 2013. Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. *Dis. Aquat. Organ.*, 105(1): 45–55.

Tu V. Q., Nguyen T. T., Tran X. T., Millard A. D., Phan H. T., Le N. P., Hoang H. A., 2020. Complete genome sequence of a novel lytic phage infecting *Aeromonas hydrophila*, an infectious agent in striped catfish (*Pangasianodon hypophthalmus*). *Arch. Virol.* 165: 2973–2977.

VASEP, 2019. http://vasep.com.vn/Tin-Tuc/1200_58730/Xuat-khau-thuy-san-nam-2019-can-dich-voi-86-ty-USD.htm (access on December 18, 2020)

Walker B. J., Abeel T., Shea T., Priest M., Abouelliel A., Sakthikumar S., Cuomo C. A., Zeng Q., Wortman J., Young S. K., Earl A. M., 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 9(11). https://doi.org/10.1371/journal.pone.0112963

Wick R. R., Judd L. M., Gorrie C. L., Holt K. E., 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.*, 13(6): 1–22.

Wu J. L., Lin H. M., Jan L., Hsu Y. L., Chang L. H., 1981. Biological control of fish acterial pathogen, Aeromonas hydrophila, by bacteriophage AH 1. *Fish Pathol*., 15: 271–276.

ELECTRONIC SUPPLEMENTARY MATERIAL

Table S1. Predictive ORFs using BLASTP

❖ **PVN03**

CDS	START (bp)	END (bp)	LENGTH (aa)	BLASTP			FINAL PREDICTED
				% Match	E-value	% Query Cover	
orf1	1	675	104	95.19	2E-57	100%	hypothetical protein
*orf*2	687	1037	80	97.5	2E-51	98%	hypothetical protein
*orf*3	1034	2179	565	100	0.0	100%	DNA primase/helicase
*orf*4	2176	2820	122	97.54	2E-83	100%	amino acid adenylation domain-containing protein
*orf*5	2822	4045	146	100	1E-99	100%	hypothetical protein
*orf*6	4042	5469	232	99.57	8E-174	100%	hypothetical protein
*orf*7	5479	7362	114	96.49	2E-77	100%	hypothetical protein
*orf*8	7372	7776	137	99.27	2E-97	100%	hypothetical protein
*orf*9	7803	8036	46	97.83	1E-26	100%	hypothetical protein
*orf*10	8048	8434	258	98.06	0.0	100%	hypothetical protein
*orf*11	8431	8601	57	98.25	2E-34	100%	hypothetical protein
*orf*12	8608	8976	60	100	2E-35	100%	hypothetical protein
*orf*13	9044	9583	59	100	1E-35	100%	hypothetical protein
*orf*14	10056	12590	48	100	7E-29	100%	hypothetical protein
*orf*15	13805	14002	86	98.84	3E-58	100%	hypothetical protein
*orf*16	14257	14442	434	94.74	3E-105	100%	hypothetical protein
*orf*17	14432	14677	231	99.13	6E-171	100%	hypothetical protein
*orf*18	14751	14978	18	99.33	1E-106	100%	hypothetical protein
*orf*19	14959	15144	84	100	5E-57	99%	hypothetical protein
*orf*20	15155	15409	61	-	-	-	No significant similarity found

*orf*21	15427	15876	75	98.67	8E-48	100%	hypothetical protein
*orf*22	16117	16812	81	-	-	-	No significant similarity found
*orf*23	16809	18113	61	96.72	1E-35	100%	hypothetical protein
*orf*24	18180	18440	65	-	-	-	No significant similarity found
*orf*25	18474	18620	844	99.88	0.0	100%	RNA polymerase
*orf*26	18625	18804	179	99.13	6E-171	100%	RNA polymerase
*orf*27	19014	19196	122	100	6E-89	100%	peptidase m15a
*orf*28	19196	19369	56	100	1E-32	100%	hypothetical protein
*orf*29	19371	20147	128	99.22	5E-88	98%	hypothetical protein
*orf*30	20243	20383	77	100	4E-50	100%	hypothetical protein
*orf*31	20395	20808	134	98.51	2E-92	100%	putative tail protein
*orf*32	20823	21167	627	98.88	0.0	100%	putative tail protein
*orf*33	21160	21858	475	100	0.0	100%	hypothetical protein
*orf*34	21905	22345	407	99.02	0.0	100%	hypothetical protein
*orf*35	22482	22850	214	100	2E-156	100%	hypothetical protein
*orf*36	22860	24557	381	99.48	0.0	100%	hypothetical protein
*orf*37	24569	24811	116	99.14	4E-80	100%	hypothetical protein
*orf*38	24808	25122	224	99.55	4E-168	100%	hypothetical protein
*orf*39	25122	27149	397	55.08	2E-153	100%	hypothetical protein
*orf*40	27284	28066	358	100	0.0	100%	hypothetical protein
*orf*41	28082	29038	1229	99.92	0.0	100%	hypothetical protein
*orf*42	29043	29588	142	100	1E-103	100%	hypothetical protein
*orf*43	29545	29781	151	100	7E-108	100%	hypothetical protein
*orf*44	29781	30287	472	99.79	0.0	100%	hypothetical protein
*orf*45	30296	31201	175	98.86	6E-126	100%	hypothetical protein
*orf*46	31201	31791	154	99.35	5E-111	99%	hypothetical protein
*orf*47	31818	33665	123	100	1E-88	98%	hypothetical protein

*orf*48	33643	33804	163	98.77	1E-117	100%	hypothetical protein
*orf*49	34083	36110	344	99.71	0.0	100%	major capsid protein
*orf*50	36121	36303	169	98.22	6E-118	100%	hypothetical protein
*orf*51	36313	37638	366	99.73	0.0	100%	hypothetical protein
*orf*52	37616	38716	441	100	0.0	100%	hypothetical protein
*orf*53	38728	39237	60	100	2E-36	100%	hypothetical protein
*orf*54	39301	40335	675	99.56	0.0	100%	terminase large subunit
*orf*55	40395	40886	53	98.11	3E-31	100%	hypothetical protein
*orf*56	40895	41266	615	63.36	0.0	98%	anaerobic NTP reductase large subunit
*orf*57	41274	41738	196	50.52	6E-56	100%	hypothetical protein
*orf*58	41728	42255	301	93.02	0.0	100%	DNA ligase
*orf*59	42255	43673	168	94.05	8E-119	100%	hypothetical protein
*orf*60	43683	44138	78	97.44	3E-53	100%	hypothetical protein
*orf*61	44158	44586	181	99.45	4E-133	100%	hypothetical protein
*orf*62	44773	48462	318	99.69	0.0	100%	5'-3' exonuclease
*orf*63	48462	49538	260	97.69	0.0	100%	hypothetical protein
*orf*64	49531	50724	662	99.85	0.0	100%	DNA polymerase

❖ **PVN04**

CDS	START (bp)	END (bp)	LENGTH (aa)	BLASTP			FINAL PREDICTED
				% Match	E-value	% Query Cover	
orf1	1	675	173	100	0.0	100%	hypothetical protein
*orf*2	687	1037	257	100	0.0	100%	hypothetical protein
*orf*3	1034	2179	179	99.44	4E-130	100%	hydrolase_2 domain-containing protein
*orf*4	2176	2820	122	100	6E-89	98%	peptidase m15a
*orf*5	2822	4045	56	100	1E-32	100%	hypothetical protein

*orf*6	4042	5469	128	100	3E-89	100%	hypothetical protein
*orf*7	5479	7362	77	100	4E-50	100%	hypothetical protein
*orf*8	7372	7776	134	100	7E-95	100%	putative tail protein
*orf*9	7803	8036	627	99.36	0.0	100%	putative tail protein
*orf*10	8048	8434	475	100	0.0	100%	hypothetical protein
*orf*11	8431	8601	407	99.75	0.0	100%	hypothetical protein
*orf*12	608	8976	214	100	2E-156	100%	hypothetical protein
*orf*13	9044	9583	381	99.74	0.0	100%	hypothetical protein
*orf*14	10000	12534	116	100	8E-81	100%	hypothetical protein
*orf*15	13748	13924	224	99.55	8E-168	100%	hypothetical protein
*orf*16	13927	14106	424	51.91	8E-149	100%	hypothetical protein
*orf*17	14228	14413	358	100	0.0	100%	hypothetical protein
*orf*18	14403	14630	1229	99.67	0.0	100%	hypothetical protein
*orf*19	14704	14931	142	100	1E-103	100%	hypothetical protein
*orf*20	15106	15360	151	100	7E-108	100%	hypothetical protein
*orf*21	15378	15827	472	100	0.0	100%	hypothetical protein
*orf*22	16068	16763	175	98.86	6E-126	100%	hypothetical protein
*orf*23	16760	17539	154	100	7E-112	100%	hypothetical protein
*orf*24	17606	17866	123	100	1E-88	100%	hypothetical protein
*orf*25	17900	18046	163	98.77	1E-117	100%	hypothetical protein
*orf*26	18051	18230	344	99.71	0.0	100%	major capsid protein
*orf*27	18440	18622	169	99.41	8E-120	100%	hypothetical protein
*orf*28	18622	18795	366	99.45	0.0	100%	hypothetical protein
*orf*29	18797	19573	441	100	0.0	100%	hypothetical protein
*orf*30	19668	19808	60	100	2E-36	100%	hypothetical protein
*orf*31	19820	20233	675	99.85	0.0	100%	terminase large subunit
*orf*32	20248	20592	53	98.11	3E-31	100%	hypothetical protein

*orf*33	20585	21283	367	99.73	0.0	100%	ribonucleoside triphosphate reductase beta chain
*orf*34	21330	21770	706	100	0.0	100%	ribonucleotide diphosphate reductase subunit alpha
*orf*35	21907	22275	194	100	7E-143	100%	hypothetical protein
*orf*36	22285	23982	301	100	0.0	100%	DNA ligase
*orf*37	23994	24236	168	99.4	5E-124	100%	hypothetical protein
*orf*38	24233	24547	78	98.72	7E-54	100%	hypothetical protein
*orf*39	24547	26574	181	100	1E-133	100%	hypothetical protein
*orf*40	26709	27491	319	100	0.0	100%	5'-3' exonuclease
*orf*41	27507	28466	260	99.23	0.0	100%	hypothetical protein
*orf*42	28471	29016	675	100	0.0	100%	DNA polymerase
*orf*43	28973	29209	104	95.19	2E-57	100%	hypothetical protein
*orf*44	29209	29715	80	97.5	2E-51	100%	hypothetical protein
*orf*45	29724	30629	565	100	0.0	100%	DNA primase/helicase
*orf*46	30640	31224	122	100	4E-85	100%	amino acid adenylation domain-containing protein
*orf*47	31241	33361	146	100	1E-99	100%	hypothetical protein
*orf*48	33371	34474	232	99.57	8E-174	100%	hypothetical protein
*orf*49	34558	34719	114	99.12	2E-79	100%	hypothetical protein
*orf*50	34998	37025	137	100	3E-98	100%	hypothetical protein
*orf*51	37036	37218	46	100	2E-27	100%	hypothetical protein
*orf*52	37228	38553	258	99.22	0.0	100%	hypothetical protein
*orf*53	38531	39631	57	100	2E-35	100%	hypothetical protein
*orf*54	39643	40152	60	100	2E-35	100%	hypothetical protein
*orf*55	40216	41250	59	100	1E-35	100%	hypothetical protein
*orf*56	41310	41801	48	100	7E-29	100%	hypothetical protein
*orf*57	41810	42181	86	100	4E-59	100%	hypothetical protein
*orf*58	42189	42653	259	100	0.0	98%	hypothetical protein
*orf*59	42643	43170	231	100	3E-172	100%	hypothetical protein

*orf*60	43170	44588	149	100	2E-107	100%	hypothetical protein
*orf*61	44598	45053	84	98.81	6E-56	100%	hypothetical protein
*orf*62	45073	45501	75	100	1E-48	100%	hypothetical protein
*orf*63	45688	49377	75	100	1E-48	100%	hypothetical protein
*orf*64	49377	50453	61	100	1E-37	100%	hypothetical protein
*orf*65	50446	51720	59	100	1E-36	100%	hypothetical protein

❖ **PVN05**

CDS	START (bp)	END (bp)	LENGTH (aa)	BLASTP			FINAL PREDICTED
				% Match	E-value	%Query Cover	
orf1	1	675	337	100	0.0	100%	RNA polymerase
*orf*2	687	1037	58	100	1E-34	100%	hypothetical protein
*orf*3	1034	2179	59	98.31	6E-36	100%	hypothetical protein
*orf*4	2176	2820	61	100	1E-37	100%	hypothetical protein
*orf*5	2822	4045	75	100	1E-48	100%	hypothetical protein
*orf*6	4042	5469	75	100	1E-48	100%	hypothetical protein
*orf*7	5479	7362	84	98.81	6E-56	100%	hypothetical protein
*orf*8	7372	7776	149	99.33	1E-105	100%	hypothetical protein
*orf*9	7874	8131	231	98.7	2E-168	100%	hypothetical protein
orf10	8141	8821	434	94.74	3E-103	65%	hypothetical protein
*orf*11	8830	9162	86	98.84	3E-56	100%	hypothetical protein
*orf*12	9174	9560	48	100	6E-27	100%	hypothetical protein
*orf*13	9557	9727	59	100	1E-33	100%	hypothetical protein
*orf*14	9734	10102	60	100	2E-33	100%	hypothetical protein
*orf*15	10170	10709	47	98.25	2E-32	100%	hypothetical protein
*orf*16	11126	13660	258	98.06	0.0	100%	hypothetical protein

*orf*17	14875	15051	46	97.83	9E-25	100%	hypothetical protein
*orf*18	15054	15233	137	99.27	2E-95	100%	hypothetical protein
*orf*19	15355	15540	114	96.49	2E-75	100%	hypothetical protein
*orf*20	15530	15757	232	99.57	7E-172	100%	hypothetical protein
*orf*21	15831	16058	146	100	1E-97	100%	hypothetical protein
*orf*22	16233	16487	122	97.54	2E-81	100%	amino acid adenylation domain-containing protein
*orf*23	16505	16954	565	100	0.0	100%	DNA primase/helicase
*orf*24	17195	17890	80	97.5	2E-49	100%	hypothetical protein
*orf*25	17887	19191	104	95.19	2E-55	100%	hypothetical protein
*orf*26	19258	19518	675	99.85	0.0	100%	DNA polymerase
*orf*27	19552	19698	260	97.69	0.0	100%	hypothetical protein
*orf*28	19703	19882	318	99.69	0.0	100%	5'-3' exonuclease
*orf*29	20092	20274	181	99.45	4E-131	100%	hypothetical protein
*orf*30	20274	20447	78	97.44	3E-51	100%	hypothetical protein
*orf*31	20449	21225	168	94.05	7E-117	100%	hypothetical protein
*orf*32	21321	21461	301	93.02	0.0	100%	DNA ligase
*orf*33	21473	21886	196	50.52	5E-54	98%	hypothetical protein
*orf*34	21901	22245	615	70.33	0.0	97%	anaerobic ribonucleoside-triphosphate reductase
*orf*35	22238	22936	53	98.11	3E-29	100%	hypothetical protein
*orf*36	22983	23423	675	99.41	0.0	100%	terminase large subunit
*orf*37	23560	23928	60	100	2E-34	100%	hypothetical protein
*orf*38	23938	25635	441	100	0.0	100%	hypothetical protein
*orf*39	25647	25889	366	99.73	0.0	100%	hypothetical protein
*orf*40	25886	26200	169	98.22	5E-116	100%	hypothetical protein
*orf*41	26200	28227	344	99.71	0.0	100%	major capsid protein
*orf*42	28362	29144	163	98.77	1E-115	100%	hypothetical protein
*orf*43	29160	30116	123	100	1E-86	100%	hypothetical protein

*orf*44	30121	30666	154	99.35	5E-109	100%	hypothetical protein
*orf*45	30623	30859	175	98.29	2E-123	100%	hypothetical protein
*orf*46	30859	31365	472	99.15	0.0	100%	hypothetical protein
*orf*47	31374	32279	151	100	6E-106	100%	hypothetical protein
*orf*48	32279	32869	142	100	1E-101	100%	hypothetical protein
*orf*49	32896	34743	1229	99.76	0.0	100%	hypothetical protein
*orf*50	34721	34882	358	100	0.0	100%	hypothetical protein
*orf*51	35161	37188	424	51.91	8E-147	98%	hypothetical protein
*orf*52	37199	37381	224	100	2E-154	100%	hypothetical protein
*orf*53	37391	38716	116	99.14	4E-78	100%	hypothetical protein
*orf*54	38694	39794	381	99.48	0.0	100%	hypothetical protein
*orf*55	39806	40315	214	100	2E-154	100%	hypothetical protein
*orf*56	40379	41413	407	99.51	0.0	100%	hypothetical protein
*orf*57	41473	41964	475	99.79	0.0	100%	hypothetical protein
*orf*58	41973	42344	627	99.2	0.0	100%	putative tail protein
*orf*59	42352	42816	134	97.76	7E-90	100%	putative tail protein
*orf*60	42806	43333	85	42.86	1E-15	98%	hypothetical protein
*orf*61	43333	44751	226	46.41	1E-41	67%	hypothetical protein
*orf*62	44761	45216	110	97.37	4E-17	34%	hypothetical protein
*orf*63	45236	45664	128	99.22	5E-88	100%	hypothetical protein
*orf*64	45851	49540	56	100	1E-32	100%	hypothetical protein
*orf*65	49540	50616	122	100	6E-89	100%	peptidase m15a
*orf*66	50609	51883	179	99.44	2E-130	100%	hydrolase_2 domain-containing protein