Virus Research 308 (2022) 198625

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Reassortment and evolutionary dynamics of tilapia lake virus genomic segments

Dev Kumar Verma [a], Neeraj Sood [a], Anutosh Paria [a], T.R. Swaminathan [b], C.V. Mohan [c], K. V. Rajendran [d], P.K. Pradhan [a,*]

[a] ICAR-National Bureau of Fish Genetic Resources, Canal Ring Road, P.O. Dilkusha, Lucknow, Uttar Pradesh 226002, India
[b] Peninsular and Marine Fish Genetic Resources Centre, ICAR-NBFGR, CMFRI Campus, Kochi, Kerala 682018, India
[c] WorldFish, Penang, Malaysia
[d] ICAR-Central Institute of Fisheries Education, Versova, Andheri (W), Mumbai, Maharashtra 400061, India

ARTICLE INFO

Keywords:
Tilapia lake virus (TiLV)
Genome
Substitution rate
Selection
Mutation

ABSTRACT

The tilapia lake virus (TiLV), a highly infectious negative-sense single-stranded segmented RNA virus, has caused several outbreaks worldwide since its first report from Israel in 2014, and continues to pose a major threat to the global tilapia industry. Despite its economic importance, little is known about the underlying mechanisms in the genomic evolution of this highly infectious viral pathogen. Using phylogenomic approaches to the genome sequences of TiLV isolates from various geographic regions, we report on the pervasive role of reassortment, selection, and mutation in TiLV evolution. Our findings provided the evidence of genome-wide reassortment in this newly discovered RNA virus. The rate of non-synonymous (dN) to synonymous (dS) substitutions was less than one (dN/dS = 0.076 to 0.692), indicating that each genomic segment has been subjected to purifying selection. Concurrently, the rate of nucleotide substitution for each genomic segment was in the order of 1–3 $\times$ 10^{-3} nucleotide substitutions per site per year, which is comparable to the rate of other RNA viruses. Collectively, in line with the results of the previous studies, our results demonstrated that reassortment is the dominant force in the evolution and emergence of this highly infectious segmented RNA virus.

Tilapia lake virus (TiLV), an emerging viral pathogen that affects both wild and farmed tilapia, is considered as a threat to the global tilapia industry and has affected the livelihood of millions of tilapia farmers (FAO, 2017; Jansen et al., 2018). The pathogen has spread across 16 countries representing four continents, including Asia, Africa, South America, and North America, since its first report in Israel in 2014, and continues to cause frequent outbreaks in the respective regions (Dong et al., 2017; Jansen et al., 2018; Surachetpong et al., 2017, 2020).

TiLV is described as a novel enveloped, negative-sense, single-stranded RNA virus with genome size of 10.323 kb and consists of ten segments (Eyngor et al., 2014; Bacharach et al., 2016a). Of the 10 segments, nine segments have no similarity to other sequences in the GenBank nucleotide databases as well as no homology with protein databases of any known virus. However, the segment 1 predicted a protein which had 17 % amino acid identity to the RNA-dependent RNA polymerase (RdRp) subunit of influenza C virus PB1 (Bacharach et al., 2016a). Initially, based on the sequence similarity of the genome segment 1 and the presence of partial complementary sequences at the 5′ and 3′ non-coding termini of all segments, the virus was classified as an orthomyxo-like virus (Bacharach et al., 2016a). Since TiLV was not found to be directly related to any of the currently classified orthomyxoviruses, therefore, the virus has been recently named as *Tilapia tilapinevirus* under the genus *Tilapinevirus* and family *Amnoonviridae* (Bacharach et al., 2016b). Nonetheless, given its rapid spread and significant economic loss to the global aquaculture industry, understanding the evolutionary trajectory of this highly infectious fish RNA virus is critical. The current study seeks to assess the pervasive role of reassortment, selection, and mutations in the evolution and emergence of TiLV by using genome sequences of TiLV isolates representing multiple spatio-temporal regions.

For the analysis, a total of 19 complete TiLV genome sequences from six different countries along with the genome sequence of an Indian TiLV isolate were used (Table S1). For generating Indian TiLV genome sequences, the virus isolated from an outbreak in West Bengal, India was used (Behera et al., 2018). RNA sequencing was carried out from the

* Corresponding author.
E-mail address: pradhanpk1@gmail.com (P.K. Pradhan).

https://doi.org/10.1016/j.virusres.2021.198625
Received 22 July 2021; Received in revised form 2 November 2021; Accepted 6 November 2021
Available online 12 November 2021

liver of infected fish as described previously on Illumina NextSeq500 platform (Illumina Inc., San Diego, CA, USA) (Sood et al., 2021). The raw reads obtained were mapped to the *Oreochromis niloticus* reference genome, and the unmapped reads were aligned to the available TiLV genomes in NCBI GenBank. The reads which mapped with the TiLV genomes, were *de novo* assembled and the total length of genome sequence was 10221 nucleotides.

MEGA ver 10.0.5 was used to align genome sequences using the CLUSTALW algorithm (Kumar et al., 2018). Each alignment was manually curated to remove non-protein coding regions, and appropriate nucleotide substitution models were selected according to the Akaike Information Criterion (Posada and Buckley, 2004) as implemented in jModelTest ver 2 (Darriba et al., 2012). To detect reassortment breakpoints, five different algorithms, namely RDP, GENECONV, MaxChi, Chimaera, and 3Seq implemented in RDP ver 4.97 were used (Martin et al., 2015). To estimate the nucleotide substitution rates and the time to the most recent common ancestor (TMRCA), all available dated complete nucleotide sequences of genomes, along with the corresponding genomic segments, were obtained from NCBI GenBank (Table S1). For each segment, reassortment was checked and sequences having reassortment were not included in the subsequent analyses. A Bayesian Markov Chain Monte Carlos (MCMC) approach implemented in BEAST ver 1.10.4 was employed for estimating TMRCA and nucleotide substitution rate under the strict and relaxed clock model in the Bayesian coalescent prior (Suchard et al., 2018; Drummond et al., 2006). The coefficient of variation (CoV) was used to select the best-fit clock model (Drummond et al., 2006), and the phylogenies were evaluated using a MCMC chain length of 10–30 million states under HKY. Tracer ver 1.7 (available at: http://beast.bio.ed.ac.uk/Tracer) was used to run multiple chains for evaluating tree convergence. The uncertainty of the data was described by the 95% highest posterior density (HPD) intervals. Trees were summarized as maximum clade credibility (MCC) using the Tree Annotator program and visualized using Fig Tree ver 1.4.4 (available at: http://tree.bio.ed.ac.uk/software/figtree/).

The selection pressure on specific amino acid sites was determined by comparing the rates of non-synonymous (dN) and synonymous (dS) substitutions, where the dN/dS (ω) ratio is described by the parameter ω. Values of $\omega > 1$ indicate positive selection, whereas a value of $\omega < 1$ indicates the operation of purifying selection (Yang and Bielawski, 2000). Selection analyses were carried out using random effects likelihood (REL) and mixed-effects model of evolution (MEME) methods implemented in the Datamonkey webserver's HyPhy package (Kosakovsky Pond et al., 2006, 2011; Delport et al., 2010).

The current study used phylogenomic approaches to investigate the underlying forces that led to the evolution of TiLV. Our findings revealed that (1) genetic reassortment is important in the evolution of TiLV (Table 1; Fig. 1), a mechanism that is more pronounced for segmented RNA viruses (Han and Worobey, 2011); (2) this virus appears to have evolved at a rate of $1–3 \times 10^{-3}$ nucleotide substitutions per site per year (Table 2), which is comparable to other RNA viruses (Duffy et al., 2008; Padhi and Verghese, 2012); (3) based on these mutation rates, the TMRCA of TiLV is estimated to be between 1992 and 2010 (Table 2), which is consistent with the previous findings (Thawornwattana et al., 2021); (4) the overall rate of non-synonymous to synonymous substitution in each genomic segment is between 0.076 and 0.62 (Table S2), indicating that each segment has been subjected to purifying selection and (5) finally, with the exclusion of the reassortment sequences, the complete genome phylogeny revealed that the Indian isolate formed a unique cluster with isolates from Israel, Peru, and Ecuador (Fig. 2).

In line with previous studies (Khatchikian et al., 1989; Yoon et al., 2014; Chaput et al., 2020), the present study found evidence of reassortment in TiLV, a phenomenon which is more pronounced in the segmented RNA viruses (Han and Worobey, 2011). Although our results are consistent with the fact that reassortment is the major driving force in the evolution of TiLV, it's unclear as to why the distance-based recombination detection approaches used in the present study couldn't detect the reassortment events that have been previously detected in segments 5 & 6 (Thawornwattana et al., 2021). Therefore, we performed the maximum likelihood-based phylogenetic analyses and found the strong evidence of reassortment events between the segments 5 & 6 (Fig. S1), which is in agreement with the previous studies (Thawornwattana et al., 2021). Nevertheless, based on the results of the present study, it appears that while five TiLV isolates from Thailand (TH-2013, TH-2014, TH-2016, TH-2017, TH-2018-K) were found to have reassortment events, one TiLV isolate (BD-2017-181) from Bangladesh also showed evidence of reassortment (Fig. 1). In contrast, no reassortment was found in TiLV isolates from Ecuador, Peru, Israel, the United States, and India. Although these findings clearly show a geographic specificity in the distribution of reassortant isolates, more samples from other geographic regions are needed to determine the ubiquity of the reassortant isolates in the natural populations of the respective regions. Nonetheless, although it is unclear what drives such unique occurrences of multiple reassortant genome sequences sampled over multiple years, a factor such as within-host reassortment, caused by the frequent introduction of infected hosts across regions via intensive aquaculture practices, could be one of the plausible explanations. Concurrently, previous studies have found that within-host recombination is a major driving force in the genomic diversity of Human Pegivirus (HPgV) in human populations (Wu et al., 2016). However, more research is needed to validate this hypothesis.

The Bayesian coalescent-based analyses revealed that the mean nucleotide substitution rate in the TiLV genome is between $1\text{-}3 \times 10^{-3}$ substitutions per site per year (Table 2), which is comparable to the TiLV mutation rates reported in previous studies (Thawornwattana et al., 2021). Consistent with previous studies, TiLV mutation rates appear to be relatively higher, but comparable to the mutation rates of other RNA viruses (Troyer and Kurath, 2003; Einer-Jensen et al., 2004; Padhi and Verghese, 2012; Shao et al., 2017; Pachetti et al., 2020). TiLV's high nucleotide substitution rates could be attributed to intensive tilapia aquaculture practices, as has been proposed for other fish RNA viruses such as infectious hematopoietic necrosis virus (Troyer and Kurath, 2003), spring viremia of carp virus (Padhi and Verghese, 2012), and viral hemorrhagic septicemia virus (Troyer and Kurath, 2003). Anthropogenic factors associated with intensive tilapia aquaculture

Table 1
Detection of reassortment by five different methods implemented in RDP4 v4.97. P-value under each detection method for the respective reassortment event is mentioned. *P*-value < 0.001 is considered significant.

Reassortment event number	Recombinant genome	Detection methods				
		RDP	GENECONV	Maxchi	Chimaera	3Seq
1	TH-2013	3.22E-25	3.53E-23	1.48E-13	2.21E-13	1.26E-13
2	TH-2013	1.75E-16	2.32E-10	1.41E-12	8.98E-10	1.26E-13
3	TH-2016-CU	1.67E-02	-	1.60E-06	4.62E-02	6.83E-03
4	TH-2013	-	2.94E-10	1.28E-01	8.58E-01	2.30E-05
5	TH-2018-K	-	-	2.92E-04	5.21E-06	6.86E-01
6	TH-2017	-	-	7.30E-04	1.93E-04	1.13E-01
7	BD-2017-181	-	-	-	1.02E-01	2.83E-03
9	TH-2014	3.22E-25	3.53E-23	1.42E-13	2.21E-13	1.26E-13

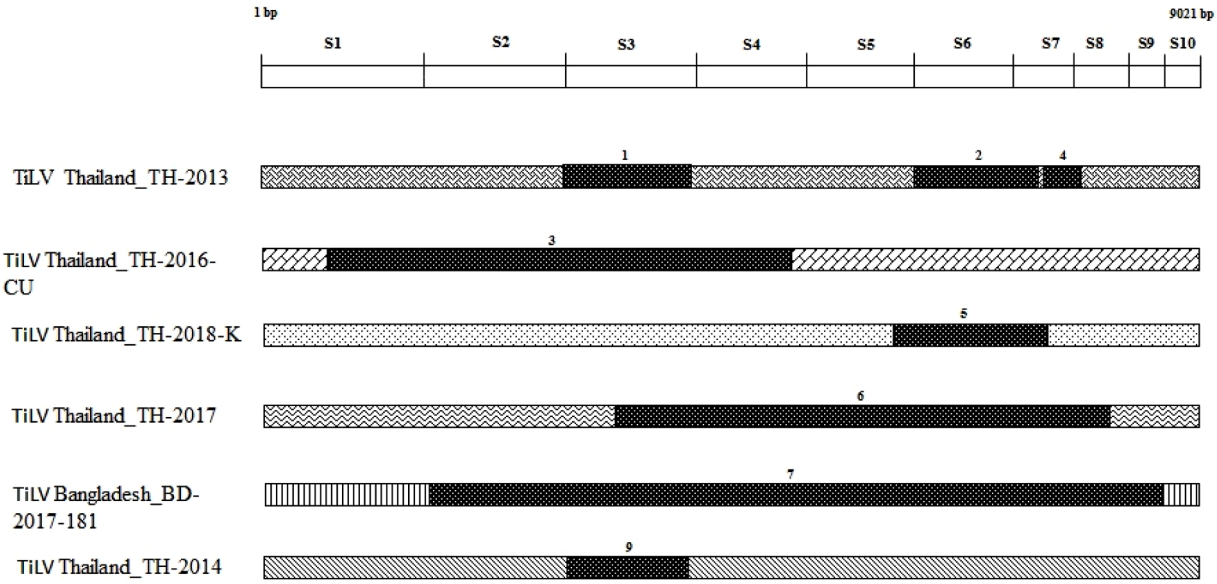

Fig. 1. Reassortment breakpoint map of the isolates detected by different methods implemented in RDP4 v4.97 program (Martin et al., 2015). The nucleotide positions (in alignment) for the coding region of each segment (S1–S10) are shown above. Regions that correspond to different isolates are pattern-coded. The numbers above the breakpoints are corresponding to the reassortment event number as shown in Table 1.

Table 2
Bayesian estimates of evolutionary rates (nucleotide substitutions per site per year) and TMRCAs (in year) inferred from genomic segments of tilapia lake virus. Number of sequences (n) used for each dataset are mentioned

Genome segment	*n*	Date range	Clock model	Mean rate (x 10^{-3}/site/year) (95% HPD)	TMRCA in year (95% HPD)	CoV (95% HPD)
All	14	2011-2020	Strict	1.41 (1.10-1.72)	2000 (1997-2004)	NA
			Relaxed	1.53 (0.83-2.29)	2000 (1992-2007)	0.22 (0.09-0.36)
1	33	2011-2020	Strict	2.37 (1.77-2.96)	2004 (2001-2007)	NA
			Relaxed	2.43 (1.33-3.46)	2003 (1996-2008)	0.44 (0.17-0.74)
2	23	2011-2020	Strict	1.82 (1.19-2.45)	2002 (1997-2006)	NA
			Relaxed	1.98 (1.01-2.94)	2001 (1992-2008)	0.39 (0.09-0.73)
5	23	2011-2020	Strict	1.92 (1.29-2.55)	2001 (1995-2006)	NA
			Relaxed	1.95 (1.21-2.65)	2001 (1993-2007)	0.16 (0.00-0.39)
8	23	2011-2020	Strict	1.70 (0.86-2.59)	2005(2000-2010)	NA
			Relaxed	1.68 (0.85-2.6)	2004(1995-2010)	0.61 (0.00-1.17)

HPD: highest posterior density; TMRCA: time to the most recent common ancestor; NA: not applicable; CoV: coefficient of variation.

practices may eventually raise stress levels, making the host more vulnerable to disease and facilitating increased viral replication (Einer-Jensen et al., 2004). The analyses of selection pressures using the dN/dS ratio (ω = 0.076 to 0.692; Table S2) of different TiLV segments

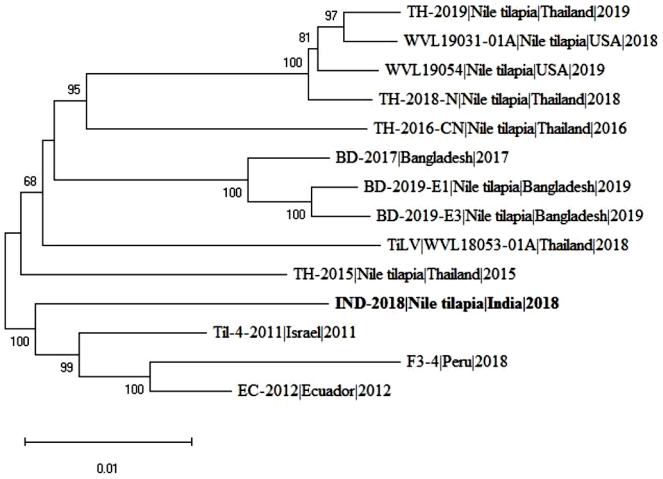

Fig. 2. Neighbor-joining tree inferred from the concatenated coding nucleotide sequences of all 10 segments of TiLV depicting the phylogenetic placement of Indian isolate. Bootstrap supports are mentioned at the base of nodes.

revealed no evidence of positive Darwinian selection, rather each TiLV protein coding genomic segment has been subjected to intense purifying selection, which is most common mechanism in other RNA viruses as well (Jenkins et al., 2002).

The current study found evidence of reassortment in TiLV as well as an elevated rate of nucleotide substitutions/site per year, which is comparable to other RNA viruses, and each genomic segment has been subjected to purifying selection.

Data availability

The assembled genome sequences of the Indian TiLV isolate IND-2018 is available in NCBI GenBank, accession numbers MZ297923 - MZ297932 (segments 1-10).

CRediT authorship contribution statement

Dev Kumar Verma: Data curation, Writing – original draft. **Neeraj**

Sood: Writing – original draft, Conceptualization. **Anutosh Paria:** Writing – review & editing. **T.R. Swaminathan:** Writing – original draft, Conceptualization. **C.V. Mohan:** Writing – review & editing. **K.V. Rajendran:** Conceptualization, Funding acquisition, Supervision. **P.K. Pradhan:** Conceptualization, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful to four anonymous reviewers for the insightful comments which greatly improved the manuscript. The authors are grateful to Director, ICAR-NBFGR and Director, ICAR-CIFE for providing facilities for carrying out this research. Dr. C.V. Mohan was supported by the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish and the ICAR-CGIAR W3 research collaboration between WorldFish and ICAR Fisheries Research Institutions. The support provided by Dr. B. K. Behera is thankfully acknowledged. The funding support from National Agricultural Science Fund, Indian Council of Agricultural Research (Grant Number: NASF/ABA-8006/2019-20/162) is gratefully acknowledged.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2021.198625.

References

Bacharach, E., Mishra, N., Briese, T., Zody, M.C., Kembou Tsofack, J.E., Zamostiano, R., Berkowitz, A., Ng, J., Nitido, A., Corvelo, A., Toussaint, N.C., Abel Nielsen, S.C., Hornig, M., Del Pozo, J., Bloom, T., Ferguson, H., Eldar, A., Lipkin, W.I., 2016a. Characterization of a novel orthomyxo-like virus causing mass die-offs of tilapia. MBio 7 e00431–e00416.

Bacharach, E., Mishra, N., Briese, T., Eldar, A., Lipkin, W.I., Kuhn, J.H., 2016b. ICTV Taxonomic Proposal 2016.016a-dM.A.v2.Tilapinevirus. Create the Unassigned Genus Tilapinevirus. http://www.ictv.global/proposals-16/2016.016a-dM.A.v2.Tilapinevirus.pdf (accessed 12 May 2021).

Behera, B.K., Pradhan, P.K., Swaminathan, T.R., Sood, N., Paria, P., Das, A., Verma, D.K., Kumar, R., Yadav, M.K., Dev, A.K., Parida, P.K., Das, B.K., Lal, K.K., Jena, J.K., 2018. Emergence of tilapia lake virus associated with mortalities of farmed Nile tilapia *Oreochromis niloticus* (Linnaeus 1758) in India. Aquaculture 484, 168–174.

Chaput, D.L., Bass, D., Alam, M.M., Al Hasan, N., Stentiford, G.D., Van Aerle, R., Moore, K., Bignell, J.P., Haque, M.M., Tyler, C.R., 2020. The segment matters: probable reassortment of tilapia lake virus (TiLV) complicates phylogenetic analysis and inference of geographical origin of new isolate from Bangladesh. Viruses 12, 258.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 (8), 772.

Delport, W., Poon, A.F.Y., Frost, S.D.W., Kosakovsky Pond, S.L., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26, 2455–2457.

Dong, H.T., Ataguba, P., Khunrae, T., Rattanarojpong, T., Serapin, S., 2017. Evidence of TiLV infection in tilapia hatcheries from 2012 to 2017 reveals probable global spread of the disease. Aquaculture 479, 579–583.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4 (5), e88.

Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267–276.

Einer-Jensen, K., Ahrens, P., Forsberg, R., Lorenzen, N., 2004. Evolution of the fish rhabdovirus viral haemorrhagic septicaemia virus. J. Gen. Virol. 85, 1167–1179.

Eyngor, M., Zamostiano, R., Tsofack, J.E.K., Berkowitz, A., Bercovier, H., Tinman, S., Lev, M., Hurvitz, A., Galeotti, M., Bacharach, E., Eldar, A., 2014. Identification of novel RNA virus lethal to tilapia. J. Clin. Microbiol. 52, 4137–4146.

Food and Agriculture Organization of the United Nations (FAO), 2017. Outbreaks of Tilapia lake virus (TiLV) threaten the livelihoods and food security of millions of people dependent on tilapia farming. http://www.fao.org/documents/card/en/c/3ce1da5b-1529-4e7c-8b88-7adfef8d138c/. (accessed 30 May 2021).

Han, G.Z., Worobey, M., 2011. Homologous recombination in negative sense RNA viruses. Viruses 3 (8), 1358–1373.

Jansen, M.D., Dong, H.T., Mohan, C.V., 2018. Tilapia lake virus: A threat to the global tilapia industry? Rev. Aquacult. 11 (3), 725–739.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 156–165.

Khatchikian, D., Orlich, M., Rott, R., 1989. Increased viral pathogenicity after insertion of a 28S ribosoma1 RNA sequence into the hemagglutinin gene of an influenza virus. Nature 340, 156–l57.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22 (24), 3096–3098.

Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D., Delport, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. 28 (11), 3033–3043.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6), 1547–1549.

Martin, D.P., Ben, M., Michael, G., Arjun, K., Brejnev, M., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1 (1), vev003.

Pachetti, M., Marini, B., Benedetti, F., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J. Transl. Med. 18, 179.

Padhi, A., Verghese, B., 2012. Molecular evolutionary and epidemiological dynamics of a highly pathogenic fish rhabdovirus, the spring viremia of carp virus (SVCV). Vet. Microbiol. 156 (1-2), 54–63.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst. Biol. 53 (5), 793–808.

Shao, W., Li, X., Goraya, M.U., Wang, S., Chen, J.L., 2017. Evolution of influenza A virus by mutation and re-assortment. Int. J. Mol. Sci. 18 (8), 1650.

Sood, N., Verma, D.K., Paria, A., Yadav, S.C., Yadav, M.K., Bedekar, M.K., Kumar, S., Swaminathan, T.R., Mohan, C.V., Rajendran, K.V., Pradhan, P.K., 2021. Transcriptome analysis of liver elucidates key immune-related pathways in Nile tilapia Oreochromis niloticus following infection with tilapia lake virus. Fish Shellfish Immunol. 111, 208–219.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 84 (1), vey016.

Surachetpong, W., Janetanakit, T., Nonthabenjawan, N., Tattiyapong, P., Sirikanchana, K., Amonsin, A., 2017. Outbreaks of tilapia lake virus infection, Thailand, 2015–2016. Emerg. Infect. Dis. 23, 1031–1033.

Surachetpong, W., Roy, S.R.K., Nicholson, P., 2020. Tilapia lake virus: the story so far. J. Fish Dis. 43 (10), 1115–1132.

Thawornwattana, Y., Dong, H.T., Phiwsaiya, K., Sangsuriya, P., Senapin, S., Aiewsakun, P., 2021. Tilapia lake virus (TiLV): genomic epidemiology and its early origin. Transbound. Emerg. Dis. 68 (2), 435–444.

Troyer, R.M., Kurath, G., 2003. Molecular epidemiology of infectious hematopoietic necrosis virus reveals complex virus traffic and evolution within southern Idaho aquaculture. Dis. Aquat. Organ. 55, 175–185.

Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15 (12), 496–503.

Yoon, S.W., Webby, R.J., Webster, R.G., 2014. Evolution and ecology of influenza A viruses. Microbiol. Rev. 56 (1), 152–179.

Wu, H., Padhi, A., Xu, J., Gong, X., Tien, P., 2016. Evidence for within-host genetic recombination among the human pegiviral strains in HIV infected subjects. PLoS One 11, e0161880.