Tilapia lake virus (TiLV): Genomic epidemiology and its early origin

Yuttapong Thawornwattana[1], Ha Dong[2], Kornsunee Phiwsaiya[1], Pakkakul Sangsuriya[3], Saengchan Senapin[1], and Pakorn Aiewsakun[1]

[1]Mahidol University Faculty of Science
[2]Suan Sunandha Rajabhat University
[3]NSTDA

May 6, 2020

Abstract

Tilapia lake virus (TiLV) is an emerging virus that is rapidly spreading across the world. Over the past 6 years (2014–2020), TiLV outbreaks had been reported in at least 16 countries, spanning three continents, including Asia, Africa, and America. Despite its enormous economic impact, its origin, evolution, and epidemiology are still largely poorly characterised. Here, we report eight TiLV whole genome sequences from Thailand sampled between 2014–2019. Together with publicly available sequences from various regions of the world, we estimated the origin of TiLV to be between 2003–2009, 5–10 years before the first report of the virus in Israel in 2014. Our analyses consistently showed that TiLV started to spread in 2000s, and reached its peak in 2014–2016, matching well with the timing of its first report. From 2016 onwards, the TiLV population declined steadily. This could be a result of herd immunity building up in the fish population, and / or a reflection of a better awareness of the virus coupled with a better and more cautious protocol of Tilapia importation. Despite the fact that we included all publicly available sequences, our analyses revealed long unsampled histories of TiLVs in many countries, especially towards its basal diversification. This result highlights the lack and the need for systematic surveillance of TiLV in fish.

Article Type

Original Article

Title

Tilapia lake virus (TiLV): Genomic epidemiology and its early origin

Running Title

Epidemiology and origin of TiLV

Authors

Yuttapong Thawornwattana[1,2], Ha Thanh Dong[3], Kornsunee Phiwsaiya[4,5], Pakkakul Sangsuriya[5,6], Saengchan Senapin[4,5]*, Pakorn Aiewsakun[1,2]*

Author Affiliations

1. Department of Microbiology, Faculty of Science, Mahidol University, 272, Rama VI Road, Ratchathewi, Bangkok, 10400, Thailand.

2. Center of Microbial Genomics (CENMIG), Faculty of Science, Mahidol University, 272, Rama VI Road, Ratchathewi, Bangkok, 10400, Thailand.

3. Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, 10300, Thailand.

4. Fish Health Platform, Center of Excellence for Shrimp Molecular Biology and Biotechnology (Centex Shrimp), Faculty of Science, Mahidol University, Bangkok, 10400, Thailand.

5. National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, 12120, Thailand

6. Aquatic Molecular Genetics and Biotechnology Laboratory, BIOTEC, NSTDA, Pathum Thani, 12120, Thailand

* Corresponding author contacts

Saengchan Senapin: National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, 12120, Thailand; +66 2 564 8000; saengchan@biotec.or.th; ORCiD: 0000-0002-5599-4343

Pakorn Aiewsakun: Department of Microbiology, Faculty of Science, Mahidol University, 272, Rama VI Road, Ratchathewi, Bangkok, 10400, Thailand; +66 2 201 5676; pakorn.aie@mahidol.ac.th; ORCiD: 0000-0002-5665-4041

Emails

Yuttapong Thawornwattana: yuttapong.thawornwattana@gmail.com

Ha Thanh Dong: hathanh.do@ssru.ac.th

Kornsunee Phiwsaiya: Kornsunee.Jam@biotec.or.th (case sensitive)

Pakkakul Sangsuriya: pakkakul.san@biotec.or.th

Saengchan Senapin: saengchan@biotec.or.th

Pakorn Aiewsakun: pakorn.aie@mahidol.ac.th

Summary

Tilapia lake virus (TiLV) is an emerging virus that is rapidly spreading across the world. Over the past 6 years (2014–2020), TiLV outbreaks had been reported in at least 16 countries, spanning three continents, including Asia, Africa, and America. Despite its enormous economic impact, its origin, evolution, and epidemiology are still largely poorly characterised. Here, we report eight TiLV whole genome sequences from Thailand sampled between 2014–2019. Together with publicly available sequences from various regions of the world, we estimated the origin of TiLV to be between 2003–2009, 5–10 years before the first report of the virus in Israel in 2014. Our analyses consistently showed that TiLV started to spread in 2000s, and reached its peak in 2014–2016, matching well with the timing of its first report. From 2016 onwards, the TiLV population declined steadily. This could be a result of herd immunity building up in the fish population, and / or a reflection of a better awareness of the virus coupled with a better and more cautious protocol of Tilapia importation. Despite the fact that we included all publicly available sequences, our analyses revealed

long unsampled histories of TiLVs in many countries, especially towards its basal diversification. This result highlights the lack and the need for systematic surveillance of TiLV in fish.

Keywords

Tilapia lake virus; whole genome sequencing; reassortment, recombination; Bayesian phylodynamics

Introduction

Tilapia lake virus (TiLV) is an emerging virus that has greatly affected global tilapia farming industry and food security (Food and Agriculture Organization of the United Nations, 2017; Jansen et al., 2019). TiLV was discovered in Israel in 2014 (Eyngor et al., 2014; Bacharach et al., 2016) following the first report of a novel disease in Ecuador, called syncytial hepatitis of tilapia (Ferguson et al., 2014). Since then, TiLV outbreaks have been reported in many other countries, including Peru (Pulido et al., 2019), Ecuador (Subramaniam et al., 2019), Egypt (Fathi et al., 2017; Nicholson et al., 2017), India (Behera et al., 2018), Thailand (Dong, Siriroob et al., 2017; Surachetpong et al., 2017), Malaysia (Abdullah et al., 2018; Amal et al., 2018), Bangladesh (Chaput et al., 2020), Tanzania (Mugimba et al., 2018), Uganda (Mugimba et al., 2018), the USA (Ahasan et al., 2020), Colombia (Tsofack et al., 2017), Indonesia (Koesharyani et al., 2018), Chinese Taipei (World Organisation for Animal Health, 2017a), Mexico (World Organisation for Animal Health, 2019), and Philippines (World Organisation for Animal Health, 2017b). High mortality rates (20–90%), and rapid spread of the virus have threatened the livelihoods of millions of tilapia farmers on the global scale (Food and Agriculture Organization of the United Nations, 2017; Jansen et al., 2019).

TiLV is a negative-sense single-stranded RNA virus with a 10-segment genome (Bacharach et al., 2016). The gene on segment 1 is predicted to code for the PB1 protein (Bacharach et al., 2016), which shows a very low, but detectible, similarity to the polymerase proteins of influenza (Bacharach et al., 2016). Other segments, however, showed no detectible similarities to any known genes in the NCBI database, indicating that it is a new virus that is distantly related to all known viruses. Indeed, TiLV is currently classified as the only member of a new genus *Tilapinevirus* (International Committee on Taxonomy of Viruses, 2019), the sole genus in the family *Amnoonviridae* . *Amnoonviridae* is a recently established family in the order *Articulavirales* , which also contains the influenza virus (the family *Orthomyxoviridae*) (International Committee on Taxonomy of Viruses, 2019).

A number of genome sequences of TiLVs have been reported, and analyses of these sequences have led to better understanding of how the virus evolves and spreads across regions (Bacharach et al., 2016; Surachetpong et al., 2017; Pulido et al., 2019; Ahasan et al., 2020; Chaput et al., 2020). Most of the reported sequences are, however, not whole genomes, and a recent study has shown that reassortment is very common for TiLV (Chaput et al., 2020) – a phenomenon in which multiple strains of viruses with segmented genomes co-infect the same host cell and exchange their genetic materials (McDonald et al., 2016; Lowen, 2018). This means that analysis of individual genomic segments might not give a full picture of how TiLV evolves, and on the other hand, analysis of whole genomes without taking reassortment into account might yield an erroneous history of the virus.

Here, we report eight new whole genomes of TiLVs from Thailand, sampled between 2014–2019. Furthermore, their sampling dates distribute relatively evenly across the entire sampling period, providing us sufficient temporal information for dating the early origin of TiLV, and how it evolves and spreads through time. Together with nine publicly available whole genome sequences of TiLVs, phylogenetic analyses were performed to investigate the emergence of TiLV, and how it spread globally.

Materials and Methods

Amplification of TiLV genomic segments and whole genome sequencing

Total RNA extracted from eight TiLV-infected specimens used for amplification of TiLV genomic sequences in the present study were obtained from our previous works (Dong, Ataguba et al., 2017; Taengphu et al., 2019). Segment 1 of all samples were already sequenced (Taengphu et al., 2019). Thus, the remaining TiLV genome segments 2–10 were individually amplified by RT-PCR, using primers and conditions as previously described (Pulido et al., 2019). After agarose gel electrophoresis, the amplicons were gel purified using FavorPrep GEL/PCR purification kit (Favorgen) before being cloned into pGEM-T easy vector (Promega). Recombinant clones were sequenced using T7 promoter, and SP6 promoter primers (Macrogen, South Korea).

Sequence data

Together with whole genomes generated in this study, nine publicly availably whole genome sequences of TiLVs were retrieved from the NCBI database (accessed on 25 Feb 2020), including those from Israel (Bacharach et al., 2016), Ecuador (Subramaniam et al., 2019), Peru (Pulido et al., 2019), USA (Ahasan et al., 2020), Bangladesh (Chaput et al., 2020), and other sequences from Thailand (Surachetpong et al., 2017; Al-Hussinee et al., 2018). Their accession numbers are given in **Table S1** . Sampling dates of publicly available TiLV genomes were retrieved from original publications when available, and from the NCBI GenBank database otherwise, ranging between 2011 and 2019.

Reassortant detection analysis

Multiple nucleotide sequence alignments were first constructed separately for each of the 10 segments using MAFFT v7.453 (Katoh and Standley, 2013). Each alignment was manually curated to remove non-protein coding regions. They were then concatenated and was checked for potential conflicting evolutionary signals by using seven programs (RDP, GENECONV, Chimera, MaxChi, Bootscan, SiScan, and 3Seq), all implemented in Recombination Detection Program v4.99 (RDP4) (Martin et al., 2015). Only those detected by more than four programs were considered. A few reassortants were inferred, and the concatenated alignment was split into four alignments according to the breakpoints of conflicting singals identified (**Table S2**). Potential reassortment events were then checked again within each alignment, but no additional signal was found.

Temporal signal investigation

Four phylogenetic trees were estimated from the four alignments under the maximum likelihood (ML) framework, implemented in IQ-TREE 2 (Minh et al., 2020). The best-fit nucleotide substitution model for each alignment was determined by ModelFinder (Kalyaanamoorthy et al., 2017) under the Bayesian information criterion (**Table S2**), and used for the phylogenetic reconstruction. Bootstrap method was used to assess the clade uncertainty (N = 1,000).

The temporal signals in the nucleotide datasets were assessed using root-to-tip regression analysis, implemented in the ape R package v5.3 (Paradis and Schliep, 2018). The genetic distances were inferred from the estimated ML trees, and the root placements were determined by minimising the residual mean squared errors. To assess the statistical significance of the observed temporal signals (i.e. the regression slopes), a null distribution of the regression slope (N = 1,000) was estimated for each dataset by randomising the tip-dates (Ramsden et al., 2009). A one-tail p-value was computed as the proportion of the slopes in the null distribution that were greater than or equal to the observed slope. We considered the temporal signal as sufficient if the p-value was below 0.05. The results are summarised in **Table S3** .

Tip-dating analysis

Time-calibrated phylogenies were inferred using a Bayesian approach implemented in BEAST v1.10.4 (Rambaut et al., 2018). The strict clock and the Bayesian Skyride coalescent prior (Minin et al., 2008) were applied for all trees. The root age was assigned a uniform prior over the interval of 7.67–200 years, where 7.67 (i.e., the minimum root height) was the age difference between the youngest and the oldest samples. The rough estimates of the root height of about 10–20 years from the root-to-tip analyses lied within this interval. For each analysis, two independent Markov chain Monte Carlo (MCMC) chains were run for 100,000,000 iterations. The first 10,000,000 were discarded as burnin, and parameter values were logged every 10,000[th] iteration. Effective sample sizes of all parameters were between 1,000–5,000, indicating that all parameters were sufficiently well sampled. Convergence of the MCMC chains was assessed by comparing the two independently estimated posteriors. Maximum clade credibility trees with common ancestor node heights were used to summarise the tree posterior distributions.

Sensitivity analyses

To explore the robustness of our results, we repeated the analyses under different choices of priors and model assumptions. First, we relaxed the strict clock assumption by using the uncorrelated relaxed clock model with a lognormal distribution of the rates among branches (Drummond et al., 2006). To compare the two models, we estimated the Bayes factor by using the generalised stepping stone (GSS) sampling technique to estimate the marginal likelihood for each of the models (Baele et al., 2015) (**Table S4**). For the GSS estimator, we used 100 path steps with 10,000,000 iterations of MCMC at each step. The product of exponential distribution was used as the starting distribution.

Second, the Skyride coalescent model assumed that the population size changes coincide with the internal nodes of the phylogeny. We relaxed this assumption by using the Skygrid coalescent model (Gill et al., 2012), which allows the population size to change at regular grid points. Instead of performing a separate analysis for each alignment, this method also allowed a multilocus analysis. Each region had its own (best-fit) nucleotide substitution process, and a phylogenetic tree structure, but they all shared a mean substitution rate. To specify the grid points in the Skygrid model, the time cutoff value was set to 14 years (beyond which the population size was assumed to be constant). This was based on the root age estimates of around 2005 from the main analyses. The population size was allowed to change every 0.5 years. We also varied the time cutoff parameter while the grid points remained 0.5 years apart. We found that all analyses yielded highly similar estimates (**Table S5**).

Third, there were a number of incomplete TiLV genome sequences deposited in the NCBI database, available as individual genome segments. We retrieved them from the database. Short sequences (i.e. those with length < 90% of the shortest sequence in the main dataset) were excluded to allow for more information to be retained in the alignments. In total, four segments had additional sequences, including segments 1, 3, 5, and 9 (**Table S2** and **Table S6**). For each segment, the same set of analyses, including the ML tree estimation, temporal signal assessment, and Bayesian phylodynamic inference, was performed as aforementioned.

Results and Discussion

TiLV genome dataset

Individual genome segments of TiLVs isolated from eight infected fish samples were RT-PCR amplified, cloned, and sequenced. The samples were collected in Thailand between 2014–2019. The sequences are available from the GenBank database (accession numbers: MN697695–697774). We were able to obtain complete genomes for all eight of them; each comprised 10 segments, termed Segment 1 to 10 according to their sizes (Segment 1 being the largest; **Figure S1**). Together with nine previously published TiLV

complete genomes, we estimated the past transmission history and origin of TiLV. The information of all 17 complete genomes used in this study is summarised in **Table S1** .

Reassortant detection analysis

A recent study showed that TiLVs can and do undergo reassortment process (Chaput et al., 2020). This means that different genome segments may have different evolutionary histories, and analysing multiple genomic regions with conflicting evolutionary signals together can result in an erroneous estimation of the virus evolutionary history.

Potential conflicting evolutionary signals were checked in a concatenated nucleotide alignment of segments 1–10 by using several programs (see **Materials and Methods**). A few reassortants were detected in our dataset (**Figure 1**). The detected signal breakpoints coincided well with the segment boundaries. Our results showed that segments 1–4 shared the same evolutionary history, while segments 5 and 6 had their own distinct histories. Segments 7–10 also shared a common history, but which was distinct from those of others. The concatenated alignment was thus split accordingly (**Table S2**).

Our analysis identified three potential reassortment events: one involving exchanging of segments 1–4 between (the ancestors of) TH-2016-CN isolate and an unknown TiLV, another one involving exchanging of segment 5 between (the ancestors of) TH-2018-K and BD-2017, and the other one involving an exchange of segments 5 and 6 between (the ancestors of) IL-2011 and EC-2012. These results were consistent with the findings previously reported by Chaput et al. (2020). They compared tree topologies estimated from individual genomic segments, and found evidence for reassortment of segments 5 and 6 involving IL-2011 and EC-2012, and that BD-2017 might be involved in a reassortment event. This suggested a role of reassortment in emergence of new combinations of TiLV genomic materials, similar to the one well-established in its distant relatives, influenza viruses (Steel and Lowen, 2014).

Phylogenetic analyses of TiLVs

Chaput et al. (2020) analysed each individual genome segment separately, and yielded phylogenetic trees with high estimation uncertainties (Chaput et al., 2020). This was likely because each segment on its own had relatively low phylogenetic signals. In this study, we therefore grouped the segments based on their shared evolutionary history, and analysed them together. This allowed us to construct relatively longer nucleotide alignments with more phylogenetically informative nucleotide positions (**Table S2**), and in turn gave us more power to resolve the phylogenetic relationships and yield more robust evolutionary estimates (**Figure 2**). Our sequence collection was also "time-stamped", enabling us to perform tip-dating analyses. In this study, we applied a Bayesian phylogenetic framework to simultaneously co-estimate the TiLVs' phylogenetic relationships, divergence dates, the rate of evolution, and their population dynamics.

The temporal signal in each sequence dataset was first quantified using root-to-tip regression analysis (Rambaut et al., 2016), and we found that all datasets had a statistically significant signal ($p < 0.05$; **Table S3**), sufficient for tip-dating analysis. Across all four genomic regions, our analyses consistently estimated the time to most recent common ancestor (tMRCA) of all TiLVs to be between 2003–2009, with the rates of evolution ranging between ˜1.81–3.47×10^{-3} substitutions per site per year (s/n/y) (**Figure 2** and **Table 1**). These rate estimates fell within the expected range RNA viruses' rate (Sanjuán et al., 2010). The analysis of segments 1–4 yielded the most precise date estimate of around mid 2008 (95% highest posterior density (HPD) interval = 2007.5–2009.6). Segment 6 gave the oldest estimate in 2003, but it was also the most uncertain one (95% HPD interval = 1992–2008). This might be indicative of a low, but detectible, temporal signal in the data. Alternatively, this could genuinely reflect its unique early origin and reassortment history. Population dynamics of TiLV estimated from segments 1–4 and 7–10 yielded a potential peak in the population size around 2014–2016, followed by a continuous decline in the population from 2016 onwards. Analyses of segments 5 and 6 gave similar results, although the peaks and the declines were much more modest (**Figure 3**).

6

Similar estimates of the dates, rates, and demographic histories were obtained under the relaxed clock model (**Figures S2** –**S3** , and **Table S4**). Indeed, our analyses did not favour the relaxed clock over the strict clock for all analyses (**Table S4** ; log Bayes factor < 3), suggesting relatively constant evolutionary rates over time. Analyses assuming the same evolutionary rate for all alignments based on the Skygrid coalescent model under the strict clock assumption yielded a rate estimate of ~3.0 (2.2–3.4) $\times 10^{-3}$s/n/y, a tMRCA estimate of 2007–2008, and a population size trajectory that peaked around 2014 (**Figures S4** –**S5** , and**Table S5**), again, all comparable to those from the main analyses. The estimated phylogenetic relationships were also consistent with those obtained from the main analyses. Finally, molecular clock analyses of larger datasets for segments 1, 3, 5, and 9 yielded largely consistent estimates of tMRCA, evolutionary rates, and demographic histories (**Figures S6** –**S9** , **Table S4** , and**Table S7**). Altogether, these analyses showed that our date, and rate estimates, as well as demographic histories were robust to the choice of model assumptions.

The estimated tMRCA of 2003–2009 matched well with the first report of a sharp decline in tilapia populations in Israel from 2005 (Eyngor et al., 2014), and the reported mass die-offs in farmed tilapia in Ecuador and Israel staring in 2009 (Bacharach et al., 2016). This also suggested the presences of TiLV in Israel and Ecuador many years before the first report of the virus outbreak in 2014, which, incidentally, was also around the time when the virus population was at its peak (**Figure 3**). Moreover, the analyses consistently showed that the TiLV population steadily declined from 2016 onwards, 1–2 years following the first report of the virus. This could be a result of natural resistance and herd immunity building up in the fish population against the virus, and / or a reflection of a better awareness of the virus, and a better-prepared and more cautious Tilapia importation protocol following the first virus report.

Examination of the four phylogenies revealed complex histories of TiLVs (**Figure 2**). We found that TiLVs from Thailand, including our new samples, often appeared to cluster together, but not always as a single group. They formed a well-supported cluster in the Segment 1–4 tree (clade support $= 0.99$), and the same was seen in the Segment 7–10 tree, but the grouping was not well supported (clade support < 0.80). In contrast, the virus appeared to split into several well-supported clusters in the Segment 5 and 6 trees; however, due to the low support of deep and long branches, it was therefore perhaps too premature to make any detailed inferences about how TiLVs emerged or were introduced to Thailand. Our analyses, nonetheless, dated the tMRCAs of Thai TiLVs to be between 2008–2011 for the well-supported group in the Segment 1–4 tree. Although TiLV in Thailand was not reported until 2017 (Dong, Siriroob et al., 2017; Surachetpong et al., 2017), retrospective diagnosis of archived fish specimens revealed the TiLV must have been present in Thailand before 2012 (Dong, Ataguba et al., 2017), consistent with and further supporting our results. In addition, analysis of segment 6 robustly inferred TH-2014 to be most closely related to Middle East and South American isolates (clade support $= 0.99$), while it was found to group with other Thai TiLVs in the Segment 1–4 tree (clade support $= 0.99$). This suggested that reassortment could indeed happen, even between viruses that are geographically very far apart.

Regarding TiLVs from other countries, the two USA isolates from a tilapia farm in Idaho were consistently shown to be embedded within the clade of Thai TiLVs, having relatively short branches. In particular, they were found to have a clear grouping structure with two Thai isolates, namely TH-2018-N and TH-2019, forming a small clade in all trees (clade support $= 1$). Their MRCA was dated to be around 2016–2017. This coincided with a recent record that the farm in the USA imported live tilapia from Thailand (Ahasan et al., 2020).

Similarly, the Bangladesh isolate, BD-2017, was inferred to be clustering with Thai isolates in all trees, except in the Segment 5 tree, which showed that it was most closely related to Middle East and South American TiLVs although not well supported. The youngest branching date of the BD-2017 was estimated to be 2012 (segment 6; 95% HPD interval $= 2006$–2015). Assuming that there was a single introduction of the virus to Bangladesh, this result suggested that TiLV was introduced to Bangladesh after the early 2010s. All of these findings were perfectly congruent with an earlier forecast (Dong, Ataguba et al., 2017), and the results previously reported by Chaput et al. (2020).

Regarding the four Middle East and South American TiLVs, they always appeared to form a group in all

trees. Specifically, TiLVs from Ecuador and Peru were inferred to be sisters in all trees, except in the Segment 6 tree, which supported a sister relationship between the Ecuador isolate EC-2012 and the Israeli isolate IL-2011 (Til-4-2011). In contrast, the two isolates from Israel (Til-4-2011 and AD-2016) did not form a clade, and were often found to be the most basal ones in the group. The only exception was that they formed a clade in the Segment 7–10 tree, but then again, the clade was not well-supported, and it was placed relatively deeply towards the root. Together, these findings suggested a single introduction of TiLV into South America perhaps from the Middle East, with an estimated timing around 2010–2013.

Our results were inconclusive about the likely geographical origin of TiLV as a whole. The Segment 1–4 tree showed a well-supported cascading structure of Israeli TiLVs basal to all other isolates, and the most parsimonious interpretation of this was that TiLV originated in Israel or nearby regions. We noted that, in the Segment 5 and 6 trees, the isolates from Israel (and South America) were nested within Thai isolates, suggesting that this might not be the case; however, the support of such pattern was lacking. In fact, our analyses of all segments apart from segments 1–4 consistently showed deep and long branch structures towards the root of the trees. Coupled with the facts that deep branching structures were not well-supported, and that the samples were quite sparse, it was unclear where TiLV actually originated. Additional sequences from diverse locations, and sampling times are thus required to conclusively determine the geographical origin of TiLV.

Conclusion and final remarks

This study reports eight new whole genomes of TiLVs from Thailand, collected between 2014 and 2019. Their regularly spacing sampling dates provided us crucial temporal information for the estimation of TiLV early origin and epidemiological dynamics. Together with publicly available whole genomes of TiLVs, our analyses revealed evidence for reassortment among different virus isolates, and provided new insights into the early origin of TiLV.

Our ability to infer evolutionary history of TiLV was limited by data availability. Given the global presence of the virus across multiple continents, the small samples we analysed herein were unlikely to capture the true scale of the epidemic. Our results supported that TiLV can spread very rapidly across multiple continents, and reassortment is a common evolutionary feature of this virus that can happen even among viruses that are geographically very far apart. This was perhaps facilitated by cross-country importation network of tilapia. Furthermore, our analyses consistently showed long and deep branch structures for many lineages of TiLVs, including those from Thailand, Bangladesh, and Peru. These structures were indicative of long unsampled histories and perhaps hidden genetic diversity of TiLVs in many parts of the world, including Asia and South America, highlighting the lack and the need for systematic surveillance of this virus. Indeed, our results suggested that TiLV likely originated around 2003–2009, 5–10 years prior the first report of the virus. More complete TiLV genomes from these and other geographical regions will help improve the estimation of how the virus evolves and spreads. This information will in turn enable better management, control, and surveillance of the virus, reducing its global social and economic burden.

Supplementary Information

Table S1. Complete TiLV genome data used in this study.

Table S2. Summary of alignment regions used in this study.

Table S3. Root-to-tip regression analyses.

Table S4: tMRCA and evolutionary rate estimates of TiLV (median with 95% HPD interval) under the uncorrelated relaxed molecular clock assumption with lognormal rate distribution.

8

Table S5. tMRCA and evolutionary rate estimates of TiLV (median with 95% HPD interval) under the strict molecular clock and the Skygrid coalescent prior.

Table S6. Additional sequences from individual segments.

Table S7. tMRCA and evolutionary rate estimates of individual segments with additional sequences (median with 95% HPD interval) under the strict molecular clock.

Figure S1. RT-PCR amplified TiLV segments (from segments 1 to 10) from representative infected sample in 2019. M, DNA marker (New England Biolabs).

Figure S2. Time-calibrated phylogenies under the relaxed molecular clock assumption with uncorrelated lognormal rate distribution.

Figure S3. Demographic histories under a relaxed molecular clock with uncorrelated lognormal rate distribution.

Figure S4. Time-calibrated phylogenies under the Skygrid coalescent model and strict molecular clock. The time cutoff value was 14 years.

Figure S5. Demographic history under the Skygrid coalescent model and strict molecular clock. The time cutoff value was 14 years.

Figure S6. Time-calibrated phylogenies of segments 1, 3, 5, and 9 with additional sequence data under the strict molecular clock.

Figure S7. Demographic histories of segments 1, 3, 5, and 9 with additional sequence data under the strict molecular clock.

Figure S8. Time-calibrated phylogenies of segments 1, 3, 5, and 9 with additional sequence data under a relaxed molecular clock with uncorrelated lognormal rate distribution.

Figure S9. Demographic histories of segments 1, 3, 5, and 9 with additional sequence data under a relaxed molecular clock with uncorrelated lognormal rate distribution.

Data Availability

The data that were generated and support the findings of this study are openly available in GenBank at https://www.ncbi.nlm.nih.gov/genbank/, accession numbers: MN697695–697774.

Ethics Statement

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required as no animals were used in this study. Virus sequences were generated from archived samples.

Acknowledgments

Competing Interests

The authors declare no conflict of interest.

Contribution Statement

HD, SS, and PA conceived and designed the experiments; KP and PS generated the raw data; YT, HD, SS, PA analysed the data and wrote the manuscript; all authors reviewed and approved of the manuscript.

References

Abdullah, A., R. Ramly, M.S. Mohammad Ridzwan, F. Sudirwan, A. Abas, K. Ahmad, M. Murni, and B.C. Kua, 2018: First detection of tilapia lake virus (TiLV) in wild river carp (*Barbonymus schwanenfeldii*) at Timah Tasoh Lake, Malaysia. *J. Fish Dis.* **41** , 1459–1462, DOI: 10.1111/jfd.12843.

Ahasan, M.S., W. Keleher, C. Giray, B. Perry, W. Surachetpong, P. Nicholson, L. Al-Hussinee, K. Subramaniam, and T.B. Waltzek, 2020: Genomic characterization of tilapia lake virus isolates recovered from moribund Nile tilapia (*Oreochromis niloticus*) on a farm in the United States. *Microbiol. Resour. Announc.* **9** , e01368-19, DOI: 10.1128/mra.01368-19.

Al-Hussinee, L., K. Subramaniam, M.S. Ahasan, B. Keleher, and T.B. Waltzek, 2018: Complete genome sequence of a tilapia lake virus isolate obtained from Nile tilapia (*Oreochromis niloticus*). *Genome Announc.* **6** , e00580-18, DOI: 10.1128/genomeA.00580-18.

Amal, M.N.A., C.B. Koh, M. Nurliyana, M. Suhaiba, Z. Nor-Amalina, S. Santha, K.P. Diyana-Nadhirah, M.T. Yusof, M.Y. Ina-Salwany, and M. Zamri-Saad, 2018: A case of natural co-infection of Tilapia Lake Virus and *Aeromonas veronii* in a Malaysian red hybrid tilapia (*Oreochromis niloticus* × *O. mossambicus*) farm experiencing high mortality. *Aquaculture* **485** , 12–16, DOI: 10.1016/j.aquaculture.2017.11.019.

Bacharach, E., N. Mishra, T. Briese, M.C. Zody, J.E.K. Tsofack, R. Zamostiano, A. Berkowitz, J. Ng, A. Nitido, A. Corvelo, N.C. Toussaint, S.C. Abel Nielsen, M. Hornig, J. del Pozo, T. Bloom, H. Ferguson, A. Eldar, and W.I. Lipkin, 2016: Characterization of a novel orthomyxo-like virus causing mass die-offs of Tilapia. *MBio* **7** , e00431-16, DOI: 10.1128/mBio.00431-16.

Baele, G., P. Lemey, and M.A. Suchard, 2015: Genealogical working distributions for bayesian model testing with phylogenetic uncertainty.*Syst. Biol.* **65** , 250–264, DOI: 10.1093/sysbio/syv083.

Behera, B.K., P.K. Pradhan, T.R. Swaminathan, N. Sood, P. Paria, A. Das, D.K. Verma, R. Kumar, M.K. Yadav, A.K. Dev, P.K. Parida, B.K. Das, K.K. Lal, and J.K. Jena, 2018: Emergence of Tilapia Lake Virus associated with mortalities of farmed Nile Tilapia *Oreochromis niloticus*(Linnaeus 1758) in India. *Aquaculture* **484** , 168–174, DOI: 10.1016/j.aquaculture.2017.11.025.

Chaput, D.L., D. Bass, M.M. Alam, N. Al Hasan, G.D. Stentiford, R. Van Aerle, K. Moore, J.P. Bignell, M.M. Haque, and C.R. Tyler, 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, DOI: 10.3390/v12030258.

Dong, H.T., G.A. Ataguba, P. Khunrae, T. Rattanarojpong, and S. Senapin, 2017: Evidence of TiLV infection in tilapia hatcheries in Thailand from 2012 to 2017 reveals probable global spread of the disease.*Aquaculture* **479** , 579–583, DOI: 10.1016/j.aquaculture.2017.06.035.

Dong, H.T., S. Siriroob, W. Meemetta, W. Santimanawong, W. Gangnonngiw, N. Pirarat, P. Khunrae, T. Rattanarojpong, R. Vanichviriyakit, and S. Senapin, 2017: Emergence of tilapia lake virus in Thailand and an alternative semi-nested RT-PCR for detection. *Aquaculture***476** , 111–118, DOI: 10.1016/j.aquaculture.2017.04.019.

Drummond, A.J., S.Y.W. Ho, M.J. Phillips, and A. Rambaut, 2006: Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4** , 699–710, DOI: 10.1371/journal.pbio.0040088.

Eyngor, M., R. Zamostiano, J.E.K. Tsofack, A. Berkowitz, H. Bercovier, S. Tinman, M. Lev, A. Hurvitz, M. Galeotti, E. Bacharach, and A. Eldar, 2014: Identification of a novel RNA virus lethal to tilapia. *J. Clin. Microbiol.* **52** , 4137–4146, DOI: 10.1128/JCM.00827-14.

Fathi, M., C. Dickson, M. Dickson, W. Leschen, J. Baily, F. Muir, K. Ulrich, and M. Weidmann, 2017: Identification of Tilapia Lake Virus in Egypt in Nile tilapia affected by 'summer mortality' syndrome. *Aquaculture* **473** , 430–432, DOI: 10.1016/j.aquaculture.2017.03.014.

Ferguson, H.W., R. Kabuusu, S. Beltran, E. Reyes, J.A. Lince, and J. del Pozo, 2014: Syncytial hepatitis of farmed tilapia, *Oreochromis niloticus* (L.): A case report. *J. Fish Dis.* **37** , 583–589, DOI: 10.1111/jfd.12142.

Food and Agriculture Organization of the United Nations, 2017: Outbreaks of Tilapia lake virus (TiLV) threaten the livelihoods and food security of millions of people dependent on tilapia farming. *Glob. Inf. Early Warn. Syst. Food Agric. - Spec. Alert* **338** , a-i7326e.

Gill, M.S., P. Lemey, N.R. Faria, A. Rambaut, B. Shapiro, and M.A. Suchard, 2012: Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30** , 713–724, DOI: 10.1093/molbev/mss265.

International Committee on Taxonomy of Viruses, 2019: Taxonomy, Virus Taxonomy: 2018b Release [Online] Available at https://ictv.global/taxonomy.

Jansen, M.D., H.T. Dong, and C.V. Mohan, 2019: Tilapia lake virus: a threat to the global tilapia industry? *Rev. Aquac.* **11** , 725–739, DOI: 10.1111/raq.12254.

Kalyaanamoorthy, S., B.Q. Minh, T.K.F. Wong, A. Von Haeseler, and L.S. Jermiin, 2017: ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14** , 587–589, DOI: 10.1038/nmeth.4285.

Katoh, K., and D.M. Standley, 2013: MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30** , 772–780, DOI: 10.1093/molbev/mst010.

Koesharyani, I., L. Gardenia, Z. Widowati, K. Khumaira, and D. Rustianti, 2018: Studi kasus infeksi Tilapia lake virus (TiLV) pada ikan nila (*Oreochromis niloticus*). *J. Ris. Akuakultur* **13** , 85–92, DOI: 10.15578/jra.13.1.2018.85-92.

Lowen, A.C., 2018: It's in the mix: Reassortment of segmented viral genomes. *PLoS Pathog.* **14** , e1007200, DOI: 10.1371/journal.ppat.1007200.

Martin, D.P., B. Murrell, M. Golden, A. Khoosal, and B. Muhire, 2015: RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1** , vev003, DOI: 10.1093/ve/vev003.

McDonald, S.M., M.I. Nelson, P.E. Turner, and J.T. Patton, 2016: Reassortment in segmented RNA viruses: Mechanisms and outcomes. *Nat. Rev. Microbiol.* **14** , 448–460, DOI: 10.1038/nrmicro.2016.46.

Minh, B.Q., H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, and R. Lanfear, 2020: IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* DOI: 10.1093/molbev/msaa015.

Minin, V.N., E.W. Bloomquist, and M.A. Suchard, 2008: Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25** , 1459–1471, DOI: 10.1093/molbev/msn090.

Mugimba, K.K., A.A. Chengula, S. Wamala, E.D. Mwega, C.J. Kasanga, D.K. Byarugaba, R.H. Mdegela, S. Tal, B. Bornstein, A. Dishon, S. Mutoloki, L. David, Evensen, and H.M. Munang'andu, 2018: Detection of tilapia lake virus (TiLV) infection by PCR in farmed and wild Nile tilapia (*Oreochromis niloticus*) from Lake Victoria. *J. Fish Dis.* **41** , 1181–1189, DOI: 10.1111/jfd.12790.

Nicholson, P., M.A. Fathi, A. Fischer, C. Mohan, E. Schieck, N. Mishra, A. Heinimann, J. Frey, B. Wieland, and J. Jores, 2017: Detection of Tilapia Lake Virus in Egyptian fish farms experiencing high mortalities in

2015. *J. Fish Dis.* **40** , 1925–1928, DOI: 10.1111/jfd.12650.

Paradis, E., and K. Schliep, 2018: ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics***35** , 526–528, DOI: 10.1093/bioinformatics/bty633.

Pulido, L.L.H., C.M. Mora, A.L. Hung, H.T. Dong, and S. Senapin, 2019: Tilapia lake virus (TiLV) from Peru is genetically close to the Israeli isolates. *Aquaculture* **510** , 61–65, DOI: 10.1016/j.aquaculture.2019.04.058.

Rambaut, A., T.T. Lam, L. Max Carvalho, and O.G. Pybus, 2016: Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2** , vew007, DOI: 10.1093/ve/vew007.

Rambaut, A., P. Lemey, G. Baele, M.A. Suchard, D.L. Ayres, and A.J. Drummond, 2018: Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4** , vey016, DOI: 10.1093/ve/vey016.

Ramsden, C., E.C. Holmes, and M.A. Charleston, 2009: Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol. Biol. Evol.* **26** , 143–153, DOI: 10.1093/molbev/msn234.

Sanjuán, R., M.R. Nebot, N. Chirico, L.M. Mansky, and R. Belshaw, 2010: Viral mutation rates. *J. Virol.* **84** , 9733–9748, DOI: 10.1128/JVI.00694-10.

Steel, J., and A.C. Lowen, 2014: Influenza A virus reassortment.*Curr. Top. Microbiol. Immunol.* **385** , 377–401, DOI: 10.1007/82_2014_395.

Subramaniam, K., H.W. Ferguson, R. Kabuusu, and T.B. Waltzek, 2019: Genome sequence of tilapia lake virus associated with syncytial hepatitis of tilapia in an Ecuadorian aquaculture facility.*Microbiol. Resour. Announc.* **8** , e00084-19, DOI: 10.1128/mra.00084-19.

Surachetpong, W., T. Janetanakit, N. Nonthabenjawan, P. Tattiyapong, K. Sirikanchana, and A. Amonsin, 2017: Outbreaks of tilapia lake virus infection, Thailand, 2015–2016. *Emerg. Infect. Dis.* **23** , 1031–1033, DOI: 10.3201/eid2306.161278.

Taengphu, S., P. Sangsuriya, K. Phiwsaiya, P.P. Debnath, J. Delamare-Deboutteville, C. V. Mohan, H.T. Dong, and S. Senapin, 2019: Genetic diversity of Tilapia lake virus genome segment 1 from 2011 to 2019 and a newly validated semi-nested RT-PCR method. *bioRxiv* DOI: 10.1101/848903.

Tsofack, J.E.K., R. Zamostiano, S. Watted, A. Berkowitz, E. Rosenbluth, N. Mishra, T. Briese, W.I. Lipkin, R.M. Kabuusu, H. Ferguson, J. Del Pozo, A. Eldar, and E. Bacharach, 2017: Detection of Tilapia lake virus in clinical samples by culturing and nested reverse transcription-PCR.*J. Clin. Microbiol.* **55** , 759–767, DOI: 10.1128/JCM.01808-16.

World Organisation for Animal Health, 2017a: Disease Notification Report, Tilapia lake virus disease, Chinese Taipei (China) [Online] Available at https://www.oie.int/wahis_2/public/wahid.php/Reviewreport/Review?page_refer=MapFullEventReport&reportid=24033 (accessed April 23, 2020).

World Organisation for Animal Health, 2017b: Disease Notification Report, Tilapia Lake Virus (TiLV), Philippines [Online] Available at https://www.oie.int/wahis_2/public/wahid.php/Reviewreport/Review?page_refer=MapFullEventReport&reportid=25278 (accessed April 23, 2020).

World Organisation for Animal Health, 2019: Disease Notification Report, Tilapia lake virus, Mexico [Online] Available at https://www.oie.int/wahis_2/public/wahid.php/Reviewreport/Review?page_refer=MapFullEventReport&reportid=27650 (accessed April 23, 2020).

Tables

Table 1. Estimates of the tMRCA and evolutionary rate of TiLV (median with 95% HPD interval) under the strict molecular clock assumption.

Region	tMRCA	Evolutionary rate ($\times 10^{-3}$ s/n/y)
Segment 1–4	2008.64 (2007.51, 2009.56)	3.47 (2.82, 4.07)
Segment 5	2005.08 (1999.90, 2008.97)	2.38 (1.55, 3.29)
Segment 6	2003.10 (1992.78, 2008.64)	2.12 (1.03, 3.23)
Segment 7–10	2006.74 (2000.58, 2010.22)	1.81 (0.90, 3.08)

Figure Legends

Figure 1. Reassortant detection analysis. Three reassortants were found by RDP4 (Martin et al., 2015) among the 17 TiLVs analysed, namely TH-2016-CN, IL-2011 and TH-2018-K. The inferred breakpoints of conflicting evolutionary signals appeared to match well with the segment boundaries (vertical dotted lines). The segment numbers are shown below the diagram. The analysis suggested that segments 1–4 shared the same evolutionary history. Segment 5 and 6 had their own unique histories. Segments 7–10 shared the same history.

Figure 2. Evolutionary dating analyses . Time-calibrated phylogenies for four genomic regions of the 17 TiLVs were inferred using BEAST v1.10.4 (Rambaut et al., 2018). See Methods and Materials for more details. Node bars represent 95% HPD intervals of date estimates. The numbers on nodes are posterior clade probabilities, shown for those with values > 0.80. Tips are coloured according to their countries of origin (black, Thailand; blue, Bangladesh; red, USA; yellow, Israel; green, Peru; and orange, Ecuador).

Figure 3. Demographic history of TiLV. The past population dynamics of the 17 TiLVs was inferred for various genomic regions by using the GMRF Bayesian Skyride coalescent model, implemented in BEAST v1.10.4 (Rambaut et al., 2018). Solid line is the posterior median of the effective population size through time. Dashed lines represent the 95% HPD interval. The y-axis shows the log-scaled effective population size $N_e\tau$ where N_e is the effective population size and τ is the viral generation time in years.

Hosted file

Epidemiology and origin of TiLV.Table 1.200430.docx available at https://authorea.com/users/317733/articles/447824-tilapia-lake-virus-tilv-genomic-epidemiology-and-its-early-origin

Segments 1–4

Segment 5

Segment 6

Segments 7–10

15