A giant virus awakens virophage-like PLVs in the green alga Tetraselmis

Sonia Bouchard[1], Malik Blanc[1], Jean-Pierre Baudouin[2], Julien Andreani[2], Bernard La Scola[2], Guillaume Blanc[1]

[1]Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France

[2]Aix-Marseille Univ, Institut Recherche Développement, Assistance Publique - Hôpitaux de Marseille, Microbe, Evolution, Phylogeny Infection, Marseille, France.

Abstract : Giant double-stranded DNA viruses infecting microalgae often interact with polinton-like viruses (PLVs), yet their dynamics within natural hosts remain poorly understood. Here, we report the isolation of a new *Tetraselmis striata* giant virus (*Allomimiviridae*) from Mediterranean coastal waters, which we designate *Oceanusvirus lionense* (TetV-2). From the same sample, we co-isolated a PLV, *Tsa-S2b*, that replicates concomitantly with TetV-2. Remarkably, TetV-2 infection also reactivates at least two additional PLVs integrated into the host genome. We show that these PLVs exert a virophage-like effect, reducing the fitness of the giant virus. Comparative genomics reveals that these virophagic PLVs are related to the autonomous lytic PLV *TsV-N1*, sharing structural but not replication-associated genes. This newly described pathosystem provides a unique model to investigate the evolution of PLVs and the parasitic mechanisms underlying virophage–giant virus interactions.

Introduction

The green algal genus *Tetraselmis* (*Chlorodendrophyceae*, *Chlorophyta*) comprises more than 30 species widespread in coastal and estuarine ecosystems worldwide, where they contribute to primary production and occasionally form dense blooms [1–4]. Several species are of biotechnological importance as aquaculture feed and for biofuel production [5,6]. Genomic and transcriptomic resources are now available for multiple *Tetraselmis* lineages [7–12], providing opportunities to investigate their interactions with viruses.

Tetraselmis species also serve as hosts for a diversity of double-stranded DNA viruses. The first are giant Oceanusviruses (family *Allomimiviridae*, order *Imitervirales*), represented by *Tetraselmis virus 1* (*Oceanusvirus kanoehense*), which was isolated off the Hawaiian coast using an unidentified local *Tetraselmis* strain as bait [13,14]. TetV-1 possesses a 226–257 nm icosahedral capsid and a 668-kb circular genome encoding 653 predicted proteins and 10 tRNAs, making it one of the largest known algal viruses. Its genome encodes atypical metabolic enzymes, including those involved in mannitol metabolism, saccharide degradation, and fermentation, and its replication cycle has been described in detail [15]. In contrast, *Tetraselmis striata* virus N1 (TsV-N1), isolated from a Norwegian fjord, represents a distinct group of viruses known as polinton-like viruses (PLVs) [16]. PLVs are small dsDNA viruses with genome sizes of 20–31 kb, encoding capsid proteins related to those of virophages and polintons [17–21]. The TsV-N1 genome, although only partially sequenced (31 kb), encodes 33 predicted proteins and forms ~60 nm icosahedral particles. A closely related isolate from the Black Sea, *T. viridis* virus S1 (TvV-S1), shares >99% nucleotide identity with TsV-N1 [22]. Both TsV-N1 and TvV-S1 displays a lytic lifestyle, completely clearing host cultures within 3–5 days.

Comparative genomics revealed that Chlorodendrophyceae genomes are rich in viral insertions, including large integrated fragments related to both Oceanusvirus and PLVs [23]. Recent evidence suggests that some PLVs can adopt a virophage-like lifestyle.

The Gezel-14T virus, isolated from *Phaeocystis globosa*, depends on co-infection with a giant virus for replication, reduces giant virus yield, and integrates into the algal genome [24,25]. These properties mirror those of bona fide virophages (family *Lavidaviridae*), which parasitize the replication factories of giant viruses [21,26–28]. Whether PLVs integrated in *Tetraselmis* genomes can similarly be reactivated by giant virus infection remains unknown.

Here, we report the isolation from Mediterranean coastal waters and complete genome sequencing of a new Oceanusvirus, **TetV-2**, infecting *T. striata*. We propose that TetV-2 defines a new species, *Oceanusvirus lionense*. Importantly, we show that TetV-2 infection triggers the replication of both a co-isolated PLV and a subset of the endogenous PLV elements integrated in the *T. striata* genome, all closely related to TsV-N1. These TsV-like genomes are assembled into virus-like particles, providing evidence that integrated PLVs in *Tetraselmis* can be reactivated by giant virus infection. Moreover, the TsV-like elements exert a negative impact on TetV-2 replication, indicating a virophage-like lifestyle. Together, our findings expand the known repertoire of virophage-like interactions in marine microalgae and underscore the complexity of virus–virus–host dynamics in coastal ecosystems.

Results:

Virus isolation and culture

A culture of *Tetraselmis striata* was inoculated with 0.45 µm–filtered environmental water collected in May 2022 from the Lion lagoon, southern France. Within the first 24 hours post-infection (h.p.i.), cells exhibited a marked loss of motility accompanied by deflagellation. By 72 h.p.i., the culture had largely cleared, with a pronounced decline in viable cell numbers observed microscopically. To confirm the stability of the lytic activity and fulfill Koch’s postulates, several successive rounds of infection were carried out by inoculating fresh *T. striata* cultures with 0.45 µm–filtered lysates.

Sequencing of the TetV-2 genome

Diagnostic short-read sequencing of the initial 0.45 µm–filtered lysate yielded 20 contigs encoding proteins predominantly most similar to homologs from *Oceanusvirus kaneohense* (TetV-1). These contigs spanned a total of 632 kb and displayed a relatively uniform sequencing depth of 792× ± 90. The proteins encoded by these contigs matched distinct, non-overlapping regions of the TetV-1 genome, indicating that they likely represent fragments of a single virus genome closely related to TetV-1. However, attempts to reconstruct a complete genome from the short-read dataset, including multiple contig extension trials, were unsuccessful.

To overcome this limitation, we isolated a single viral clone from the lysate, designated TetV-2, through several rounds of serial dilution to extinction, and sequenced its genome using a long-read approach. Assembly produced a single circular contig of 660,680 bp, encoding 640 predicted proteins and 4 tRNAs. A large fraction of these proteins (n = 389; 61%) shared significant similarity with TetV-1 homologs, of which 348 had their best BLAST match in TetV-1 (**Fig. 1A**). Reciprocal best-hit analysis identified 343 orthologous pairs. The four TetV-2 tRNAs corresponded to a subset of the TetV-1 tRNA repertoire (10 genes) and, as in TetV-1, were clustered within a single genomic locus.

TetV-2 lacked several genes present in TetV-1, including two fermentation-related genes encoding pyruvate formate-lyase (TetV_428) and its activating enzyme

(TetV_456) [13], as well as a photolyase (TetV_298) and a deoxyribodipyrimidine photolyase-related protein (TetV_522), both likely involved in UV-induced DNA repair. Notably, TetV-2 also lacked thymidylate synthase (TetV_240) and thymidylate kinase (TetV_225), which catalyze successive steps in the conversion of dUMP to dTDP. In contrast, 54 TetV-2 proteins had no detectable similarity to TetV-1 but did display significant matches to bacterial, eukaryotic, or viral proteins. Among these were enzymes involved in lipid metabolism, including a very long-chain fatty acid elongase (TetV2_00470) and a lysophospholipid acyltransferase (TetV2_00549). We also identified two genes encoding putative mitochondrial proteins: an "altered inheritance of mitochondria" protein 24 (AIM24; TetV2_00382) and a calcium-binding mitochondrial carrier protein (TetV2_00259). A substantial fraction of TetV-2 proteins (n = 264; 41%) had no detectable homologs in the NR database; these were globally shorter than conserved proteins (average length 200 vs. 377 amino acids, respectively). On average, orthologous proteins shared 49% sequence identity at the amino acid level (**Fig. 1B**).

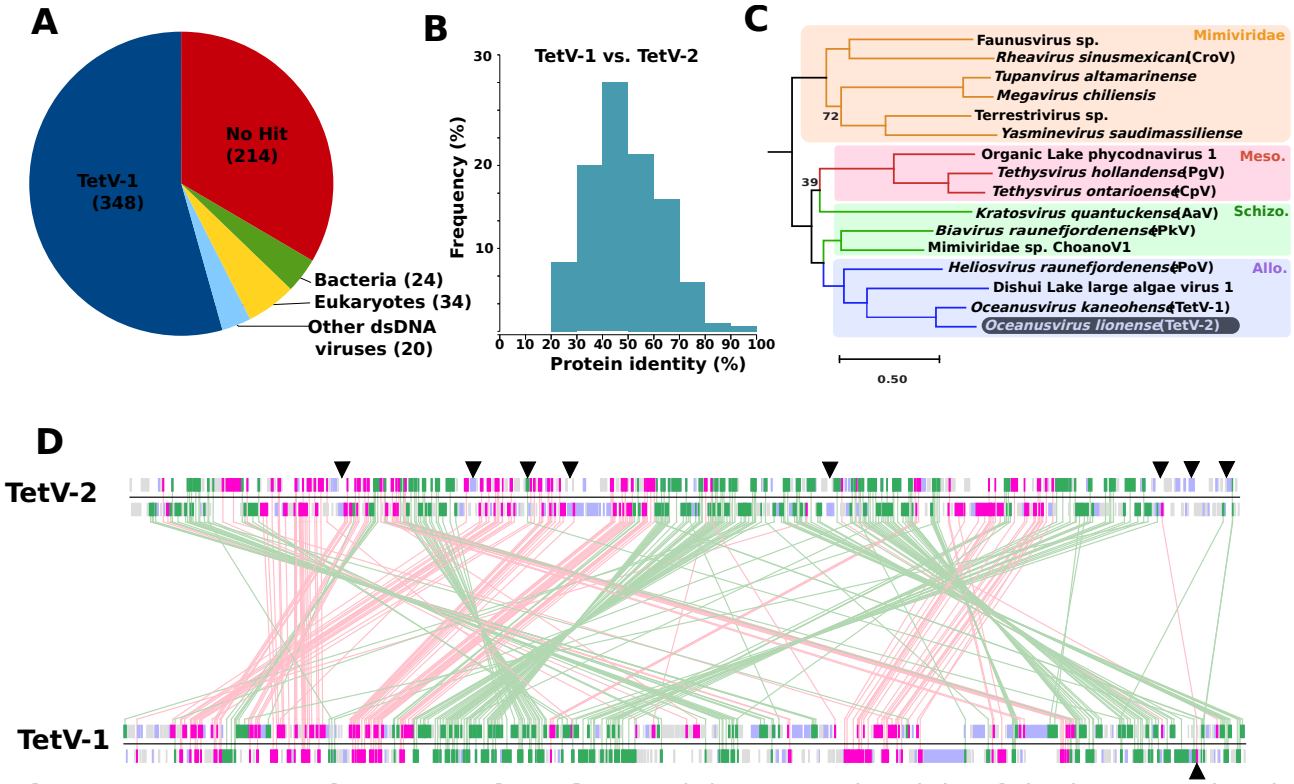

Figure 1 : Comparative genomics of TetV. *(A) Taxonomic origin of the best matches in the GenBank NR database to TetV-2 proteins.(B) Sequence identity distribution among reciprocal best matches between TetV-1 and TetV-2 proteins.(C) Maximum likelihood phylogenetic reconstruction (IQ-TREE) based on a concatenated alignment of the A18 helicase, packaging ATPase, and DNA polymerase. The tree was rooted using chlorovirus and mamiellovirus homologs as outgroup sequences (not shown). Bootstrap support values <90% are indicated.(D) Gene colinearity between TetV-1 and TetV-2 genomes. Genes are represented as colored boxes: pink, conserved genes on the same strand; green, conserved genes on opposite strands; purple, genes with significant NR matches but lacking reciprocal best hits between the two viruses; gray, hypothetical genes without significant matches. Triangles shows gene doublets encoding a bacterial transposase-related protein and an uncharacterized "transposase-associated" protein.*

Phylogenetic placement of TetV-2 within the Allomimiviridae [29] was supported by analyses based on concatenated alignments of three hallmark Megaviricetes proteins: the A18 helicase, packaging ATPase, and DNA polymerase (**Fig. 1C**). Despite infecting distinct *Tetraselmis* hosts, TetV-1 and TetV-2 were more closely related to one another

than to any endogenous TetV-like elements integrated within *Tetraselmis* genomes, including those of *T. striata*, the host of TetV-2 (**Fig. S1**). This indicates that the diversity of *Tetraselmis*-infecting viruses extends beyond that represented by TetV-1 and TetV-2.

Base composition differed substantially between the two viruses: TetV-1 and TetV-2 had GC contents of 41.2% and 57.8%, respectively. Even higher GC levels were observed among endogenous TetV-like elements in *T. striata*, *T. suecica*, and *T. chui* (ranging from 58.8% to 73.9%). These compositional differences are unlikely to reflect adaptive convergence toward host GC content, since host genomes themselves vary only modestly (58% in *T. striata* vs. 53% in *T. suecica* and *T. chui*).

Regions of gene colinearity were apparent between TetV-1 and TetV-2 but were frequently interrupted by large-scale genomic rearrangements (**Fig. 1D**). The TetV-2 genome contained eight nearly identical copies (97.5–100% nucleotide identity) of a gene doublet encoding a bacterial transposase-related protein and an uncharacterized “transposase-associated” protein. These duplications likely contributed to the failure of short-read assembly. By contrast, TetV-1 contained only a single copy of this doublet. Phylogenetic analyses indicated that the duplications occurred after divergence of the TetV-1 and TetV-2 lineages (**Fig. S2**). Several TetV-2 doublets were positioned at the boundaries of colinear regions, suggesting a role in mediating homologous recombination and genome rearrangements.

Given their close relationship yet distinct host specificities, we propose that TetV-2 be assigned to the genus *Oceanusvirus* as the first representative of a new species, *O. lionense*, named after the “salin du Lion” site where it was first isolated.

Reactivation and genomic features of endogenous Tsv-like viruses

Analysis of the initial short-read assembly revealed five contigs (1,775–17,559 bp) matching endogenous Tsv elements (ETEs) previously identified in the T. striata LANL1001 genome, with >99% nucleotide identity across their entire length (**Fig. S3**) [23]. Short-read mapping against the LANL1001 reference uncovered ten additional regions with significant coverage, all but one overlapping known ETEs, including four fully spanned between terminal inverted repeats (TIRs; C2208, C1731, C0566, C2184). Two of these (C2208, C2184) were also supported by long reads from the TetV-2 lysate, indicating that their corresponding viral genomes were present in an independent infection experiment. No other LANL1001 regions were supported by long-read data.

Although the BG host strain used here differs from LANL1001, both belong to the same *T. striata* species group and share near-identical marker genes (100% identity over a 780 bp RBCL fragment; 99.8% identity over a 1,542 bp 18S fragment; **Fig. S4 and Supplementary Data**). Thus, the observed read recruitment cannot be attributed to host DNA release during lysis, which would yield random mapping patterns. Instead, the data indicate that the lysates contained additional Tsv-like viruses closely related to *T. striata* ETEs.

To reconstruct these genomes, we performed ETE-guided local assemblies for the four fully covered elements. Each yielded a single contig flanked by TIRs, corresponding to full-length Tsv-like genomes designated Tsv-S2a, Tsv-S2a.bis, Tsv-S2b, and Tsv-S3b (hereafter S2a, S2a.bis, S2b, and S3b; **Fig. 2**).

PCR assays with subtype-specific primers confirmed the presence of **S2a** and **S3b** sequences both in uninfected *Tetraselmis* BG host DNA and as encapsidated DNA in filtered lysates, consistent with the reactivation of endogenous elements. In contrast, **S2b** DNA was detected exclusively in lysates, suggesting that it was co-isolated with TetV-2 from the environmental sample rather than derived from the host genome. This interpretation was further supported by serial passages of purified TetV-2, which consistently yielded lysates containing S2a and S3b but never S2b, demonstrating that S2b is not generated through endogenous element reactivation. **S2a.bis** could not be assessed, as its sequence lacks unique regions suitable for primer design that would allow discrimination from other TsV variants.

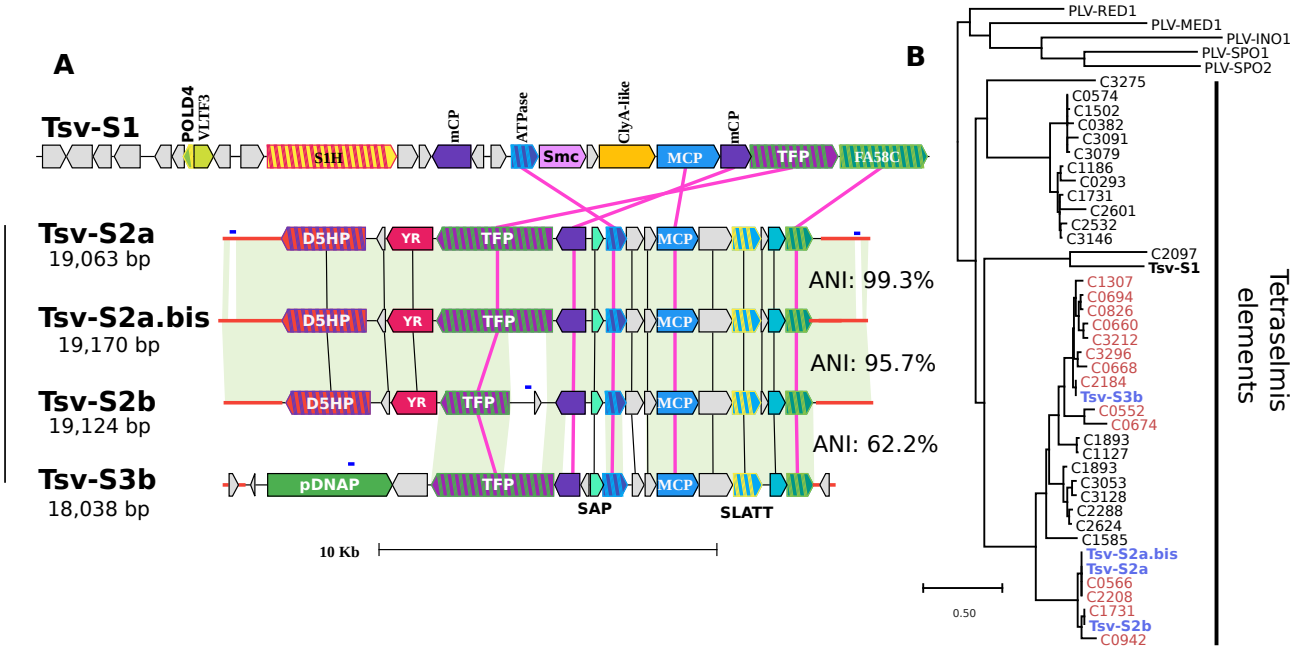

Figure 2. Tetraselmis striata Tsv genomes

(A) Genome maps of Tetraselmis striata viruses. Genes with annotated functions are shown as colored rectangles; hypothetical genes are shown in gray. Genes with significant protein-level similarity (BLASTP) are connected by lines, and regions with significant nucleotide similarity (BLASTN) are highlighted in green. Pink lines indicate genes conserved across all five viruses. Terminal inverted repeats are marked in red at the genome extremities. Average nucleotide identities (ANI) between genomes were calculated as the mean across BLASTN high-scoring pairs (HSPs). (B) Maximum-likelihood phylogenetic tree of the MCPs. Branch support values correspond to approximate likelihood-ratio test (aLRT) local supports (only values <80 are shown). PLV sequences used as outgroups (RED1, MED1, INO1, SPO1, and SPO2) were obtained from Yutin et al. (2015). Tsv-like elements detected by lysate sequencing are highlighted in red. Abbreviations: POLD4, DNA polymerase delta subunit 4; VLTF3, very late transcription factor 3; S1H, superfamily 1 helicase fused to an inactivated pDNAP domain; mCP, minor capsid protein; ATPase, DNA packaging ATPase; Smc, structural maintenance of chromosomes domain; Clya-like, ClyA-like α-pore-forming toxin domain; MCP, major capsid protein; TFP, tail fiber protein; FA58C, coagulation factor 5/8 C-terminal (discoidin) domain; D5HP, helicase–primase D5; SAP, SAP DNA-binding domain protein; AOx, PLN02976 amine oxidase domain; pDNAP, protein-primed family B DNA polymerase; SLATT_4, SMODS and SLOG-associating 2TM effector domain family 4; YR, tyrosine recombinase.

Insufficient sequencing depth prevented assembly of the remaining ETEs, but mapping patterns suggested additional low-abundance, likely truncated, Tsv-like fragments. The four genomes (18,038–19,170 bp; **Table S1**) were substantially smaller than the free-living TsV-S1 genome (26,407 bp [16]). TIRs ranged from 583 bp (S3b) to 1,724 bp

(S2b) and were absent from TsV-S1. The three S2x genomes (i.e., S2a, S2a.bis and S2b) terminated with monotonous “AT” dinucleotide repeats (**Fig. S5**), also present in their ETE guides, whereas S3b and its guide C2184 lacked such repeats. Each contig shared very high nucleotide identity with its ETE guide (ANI = 99.1–100%; **Table S2**). Notably, S2a.bis was slightly more similar to S2a (ANI = 99.3%) than to its guide C0566 (ANI = 99.1%), suggesting it does not derive from reactivation of a C0566 ortholog in BG. The S2b contig also differed from its guide, which carried a 5,441 bp tandem duplication absent from the viral genome (**Fig. S6**). By contrast, S3b was more divergent (ANI = 62.2–66.1% with the S2x group), with similarity limited to the distal two-thirds of the genome. No significant nucleotide similarity was detected between TsV-S1 and any contig by BLASTN, although BLASTP readily identified homologous proteins.

Phylogenetic analysis of the major capsid protein (MCP) confirmed these relationships (**Fig. 2B)**: the four novel Tsv-like viruses clustered more closely with each other than with TsV-S1, forming two clades—S2x and S3b. Short-read-covered LANL1001 ETEs (highlighted in red in **Fig. 2B**) also grouped into these two clades, indicating that ETE reactivation was not random but restricted to subsets of related elements, possibly responding to regulatory cues triggered by TetV-2 replication. We cannot exclude, however, that fewer Tsvs were reactivated and that lower/partial coverage of other ETEs reflects spurious read alignments due to sequencing errors.

Gene content analysis showed that S2a, S2a.bis, and S3b each encoded 16 predicted proteins, whereas S2b encoded 17. All four shared five conserved genes with TsV-S1: the major and minor capsid proteins, a DNA-packaging ATPase, a tail fiber protein (TFB), and a coagulation factor 5/8 C-terminal domain (FA58C)-like protein likely involved in carbohydrate binding [30]. These conserved genes were clustered on one side of the genome, with clade-specific genes located on the opposite side. Unique regions were enriched in replication functions. TsV-S1 encoded a superfamily 1 helicase fused to an inactive protein-primed DNA polymerase domain (S1H) [31]; S3b encoded a protein-primed family B DNA polymerase (pDNAP) [32]; and the S2x group encoded a D5 helicase–primase (D5HP) adjacent to a tyrosine recombinase, potentially mediating proviral integration and excision [33]. Beyond the core capsid and replication modules, the novel TsVs shared five additional genes absent from TsV-S1, including a SAP DNA-binding protein, a SLATT-domain protein related to predicted membrane-perforating toxins [34], and three uncharacterized proteins.

Ultrastructural analysis

Transmission electron microscopy (TEM) provided direct evidence of viral infection within Tetraselmis cells and revealed the presence of two distinct types of virus-like particles (**Fig. S7**). Negative staining of cell lysates showed large icosahedral particles measuring 214–239 nm in diameter, consistent in size and morphology with *Tetraselmis* virus 1 (TetV-1; 226–257 nm) and therefore identified as *TetV*-2 virions. A second population of smaller icosahedral particles, 66–75 nm in diameter, corresponded to the size range previously reported for *Tetraselmis* virus N1 (TsV-N1; 49–73 nm). The coexistence of these two particle classes indicates concurrent production of a giant virus (*TetV*-2) and smaller TsV-like viruses.

To characterize the replication dynamics of *TetV*-2, thin-section TEM was performed on *T. striata* cells at 24, 48, and 72 hours post-inoculation (hpi). At 24 hpi, infected cells displayed extensive cytoplasmic reorganization, with large regions occupied by empty

and filled *TetV*-2 capsids arising from membranous precursors (**Fig. S7 A1-C2**). These features closely resembled those described for *Tetraselmis* sp. infected with *TetV*-1 [15], including putative cytoplasmic viral factories with spatially organized capsid maturation, immature particles located centrally and mature virions peripherally. A minority population of smaller icosahedral particles with electron-dense cores was also observed, likely corresponding to TsV-like viruses replicating in parallel with *TetV*-2.

By 48 hpi, the number of *TetV*-2 particles had markedly declined, while TsV-like particles became predominant, frequently accumulating at the cell periphery near chloroplast membranes rather than in the central cytoplasm where *TetV*-2 factories had formed (**Fig. S7 D1-F2**). This shift in viral abundance suggests a temporal succession in which TsV replication becomes predominant after the peak of TetV-2 particle production.

At 72 h.p.i., TetV-2 particles were rarely detected, and the cytoplasm was almost entirely filled with TsV-like particles (**Fig. S7 G1-I2**). These smaller particles often formed crystalline arrays, a hallmark of advanced viral replication and accumulation. Collectively, these ultrastructural observations demonstrate that the polinton-like virus (PLV) genomes identified by metagenomic sequencing correspond to actively replicating viruses that co-occur with, but are distinct from, the giant *TetV*-2.

To verify encapsidation of the TsV-like viruses, we applied a size-fractionation and DNase-protection assay previously used to demonstrate the virion nature of Gezel-14T [24]. Fresh lysates were sequentially filtered through 0.45-µm and 0.1-µm membranes to remove intact cells and giant viruses, respectively. Aliquots of the raw lysate, clarified (0.45-µm) filtrate, and 0.1-µm filtrate were either left untreated or boiled to release encapsidated nucleic acids, followed by DNase treatment to degrade non-encapsidated DNA. PCR amplification of specific marker genes revealed *TetV*-2 DNA only in the non-boiled raw and clarified fractions, whereas TsV-like markers were detected in all non-boiled fractions. These results indicate that *TetV*-2 and TsV-like viruses are independently encapsidated, forming large (>0.1 µm) and small (<0.1 µm) viral particles, respectively.

TsVs exhibit a virophage-like lifestyle

To evaluate whether TsVs act as virophages, we investigated their impact on the replication dynamics of the giant virus TetV-2. Host cultures were coinfected with a fixed concentration of TetV-2 (3.10^6 ml^{-1}) and varying ratios of TsV to TetV-2. Two distinct TsV cocktails derived from independent lysates were tested: one dominated by S2b (i.e., 99.4% S2b, 0.6% S3b; 0,0 % S2a but still detectable by qPCR [est. 8.2 copies per ml]) and the other dominated by S2a (99.4% S2a, 0.6% S3b; S2b undetectable). Cultures were incubated for 13 days, after which both host cells and viral particles were enumerated.

Control infections with TsVs alone yielded no detectable TsV progeny (**Fig. S8**) and did not increase host mortality (**Fig. 3A**). Instead, TsV abundance declined by more than 98% between inoculation and the end of the experiment, likely reflecting adsorption to host cell surfaces and virion degradation. This confirmed that TsVs cannot replicate autonomously in this host and are not directly harmful to the host. By contrast, coinfections with TetV-2 revealed a strong inhibitory effect of both TsV cocktails on giant virus production (**Fig. 3B**). Increasing the TsV/TetV-2 ratio in the inoculum caused a log-linear, dose-dependent reduction in TetV-2 yields. These findings indicate that

TsVs require TetV-2 for propagation and, in turn, suppress TetV-2 replication, consistent with a virophage-like lifestyle. Nevertheless we did not observe a measurable reduction in host mortality in cultures co-infected with TetV-2 and TsVs; by day 13, all such cultures had visibly clarified.

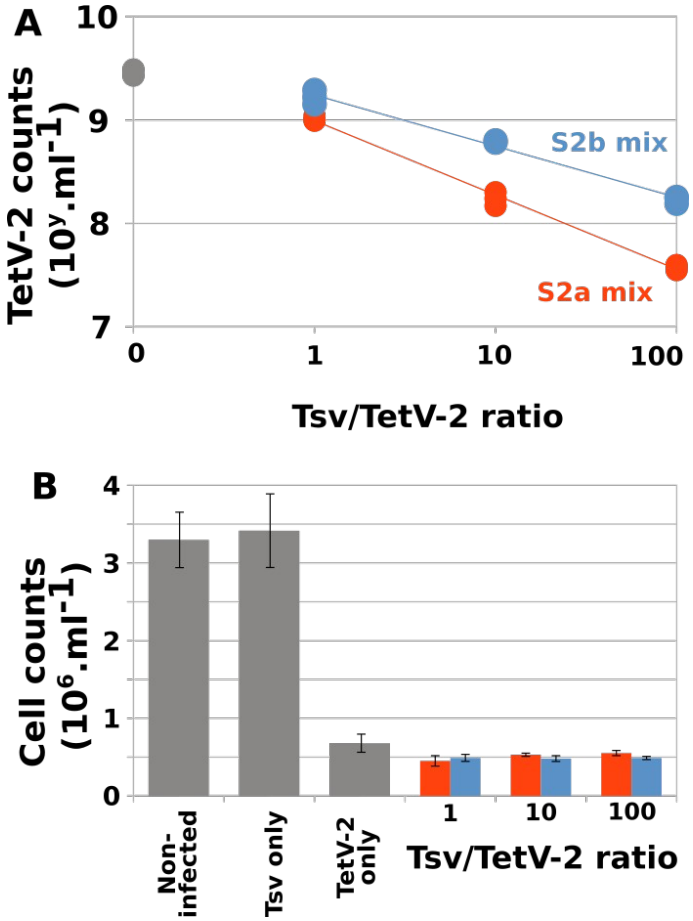

Figure 3. Virophage-associated life traits of TsV. (A) *Abundance of TetV-2 particles at 13 days post infection (dpi) as a function of the initial TsV:TetV-2 inoculation ratio. Cultures were co-infected with TetV-2 (3×10^6 particles mL^{-1}) and a TsV mixture dominated by strain TsV-S2a (red) or TsV-S2b (blue). The TetV-2–only control is shown in gray. Each point represents an independent biological replicate (n = 3), and lines depict the corresponding regression fits.* ***(B)*** *Concentration of viable Tetraselmis cells at 13 dpi. Gray bars indicate control cultures: uninfected, infected with the TsV-S2b–dominant mix only (10×10^6 particles mL^{-1} total TsV), or infected with TetV-2 alone. Red and blue bars correspond to co-infections with TetV-2 and the TsV-S2a– or TsV-S2b–dominant mixes, respectively. Bars represent mean values (n = 3), and error bars denote standard deviations.*

Replication cycle and latency phase

We investigated the replication dynamics of TetV-2 and reactivation of Tsv through a triplicate infection experiment, using cultures inoculated with purified TetV-2 particles. Samples were collected over an 8-day period post-inoculation. Host and viral abundances were monitored by flow cytometry, while extracellular encapsidated Tsv genome copies were quantified by qPCR.

The infection dynamics revealed that the latency phase of TetV-2 last between 21 and 24 hours (**Fig. 4**). After this period, TetV-2 abundance increased rapidly until ~3 days post-infection (d.p.i.), after which it stabilized and remained constant until the end of the experiment. Detection of the two endogenous polinton-like viruses (Tsv-S2a and

Tsv-S3b) was delayed relative to TetV-2: Tsv-S2a was first detected at 24 hours post-infection (h.p.i.), while Tsv-S3b appeared only at 2 d.p.i. Unlike TetV-2, both PLVs continued to accumulate throughout the experiment without reaching a plateau. By the final sampling point, Tsv-S2a abundance exceeded that of TetV-2 by ~9-fold and that of Tsv-S3b by ~452-fold.

These quantitative patterns are consistent with TEM observations. At 24 h.p.i., the cytoplasm of infected cells contained predominantly TetV-2 particles, with only a few Tsv particles visible, reflecting the initial burst of giant virus replication. By 72 h.p.i., however, giant viral particles were largely absent, and cells were densely filled with small Tsv particles. This transition in particle composition mirrors the dynamics observed in abundance profiles: the stabilization of TetV-2 levels after 3 d.p.i. corresponds to the progressive and sustained amplification of the virophages, which continued to accumulate until the end of the experiment.

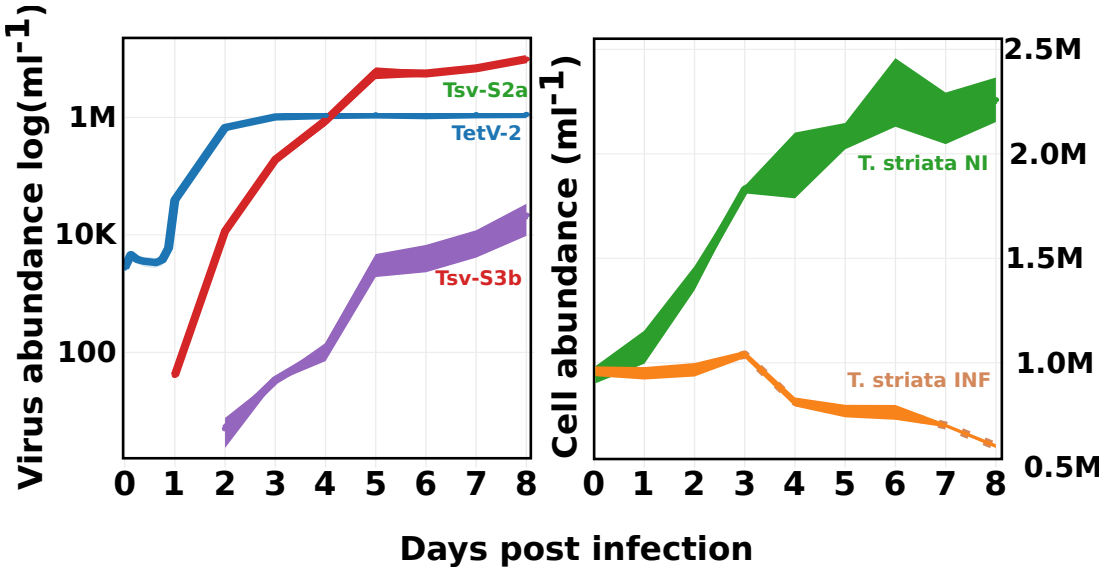

Figure 4. Infection dynamics of Tetraselmis by TetV-2. (A) *Replication dynamics of TetV-2, TsV-S2a, and TsV-S3b following infection at a ratio of 30 TetV-2 particles per host cell. TetV-2 particle abundance was quantified by flow cytometry, while encapsidated DNA copy numbers of TsV-S2a and TsV-S3b were determined by qPCR.* ***(B)*** *Host cell survival over time, assessed by flow cytometry in infected and uninfected control cultures. Each point represents the mean of three biological replicates, and the shaded areas indicate the standard deviation.*

Host range of TetV-2

The host range of TetV-2 was assessed across seven *Tetraselmis* strains, including six *T. striata* isolates and the original *T. striata* BG strain from which TetV-2 was isolated (**Fig. S9**). Cultures were inoculated with purified TetV-2 and monitored after 13 d.p.i.. Cell concentrations and TetV-2 particle abundances were quantified by flow cytometry, while TsV DNA copy numbers were determined by qPCR.

The tested *Tetraselmis* strains displayed variable susceptibility to TetV-2, reflected in host mortality and viral productivity (**Fig. 5**). The highest mortality occurred in *T. striata* BG and RCC126, with cell densities reduced by 80–90% relative to uninfected controls at 13 d.p.i. In contrast, strains TetD, TetF, and LANL1001 displayed moderate

reductions in cell abundance (7–36%), while *T. suecica* exhibited enhanced growth in inoculated cultures (+39%). This latter result suggests that the material inoculated (including submicron cell debris not removed by the 0.45 µm prefiltration) may stimulate growth in this strain.

Quantification of TsV-S2a revealed distinct strain-dependent patterns. TsV-S2a DNA was consistently detected across all *Tetraselmis* strains, with copy numbers broadly correlating with TetV-2 production. Notably, TsV-S2a was detected in infected LANL1001 cultures despite the absence of TetV-2 replication, and low but reproducible TsV-S2a production was also observed in *T. suecica*, which did not support TetV-2 propagation. Control experiments excluded contamination or genomic integration as potential sources: (i) no amplification occurred when *T. suecica* genomic DNA was tested with TsV-S2a-specific primers, (ii) the *T. suecica* genome contained no sequences with significant similarity to primer targets, and (iii) no qPCR signal was detected in Conway medium inoculated with TetV-2 filtrates alone. These results suggest that TsV-S2a, present at initially undetectable levels in the TetV-2 inoculum, was able to replicate in the presence of *T. suecica* and TetV-2, despite the absence of productive TetV-2 infection. TsV-S3b was detected in most *Tetraselmis* cultures, those of except *T. suecica*. In LANL1001, amplification was inconsistent: one replicate yielded no detectable signal, while the other two showed weak amplification close to the detection threshold, suggesting either stochastic presence or very low levels of the TsV-S3b target DNA.

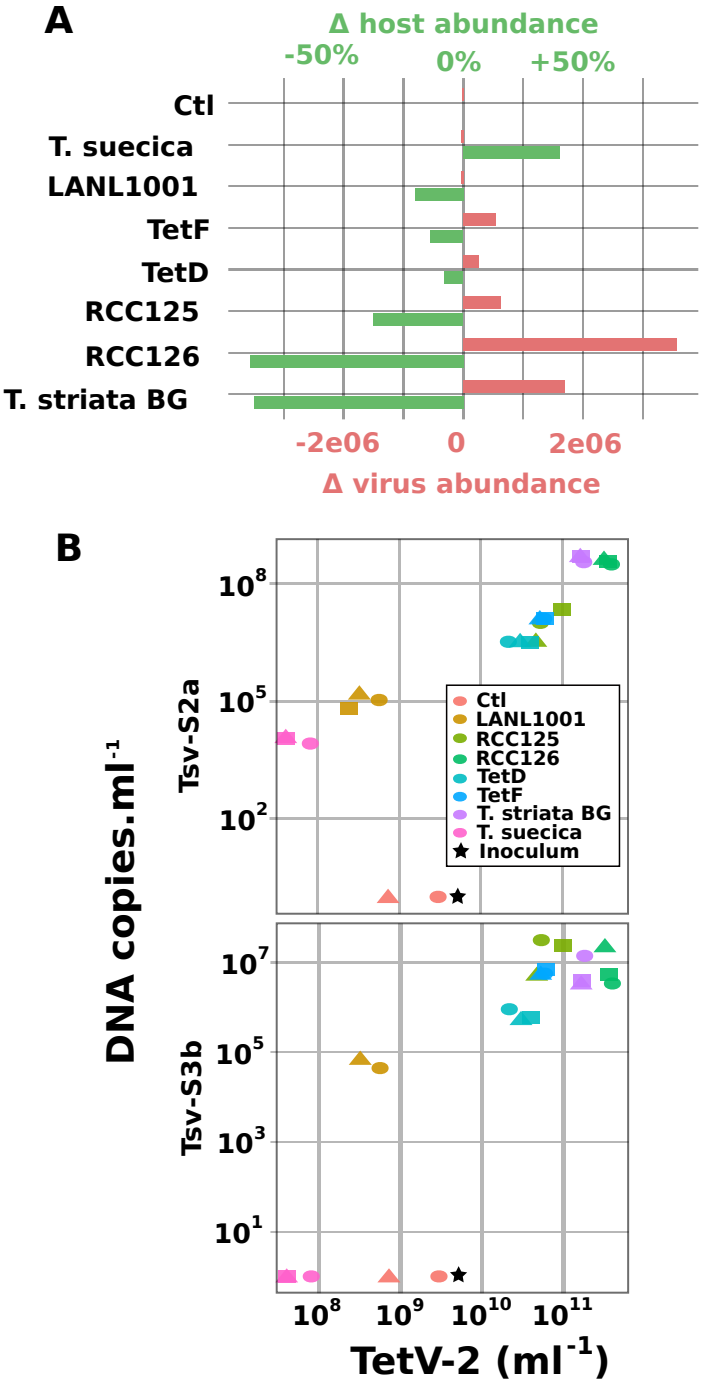

Figure 5:

REFERENCES :

1. Fistarol GO, Coutinho FH, Moreira APB, Venas T, Cánovas A, De Paula SEM, et al. Environmental and Sanitary Conditions of Guanabara Bay, Rio de Janeiro. Front Microbiol. 2015;6. doi:10.3389/fmicb.2015.01232

2. Jones JB, Rhodes LL. Suffocation of pilchards (*Sardinops sagax*) by a green microalgal bloom in Wellington Harbour, New Zealand. New Zealand Journal of Marine and Freshwater Research. 1994;28: 379–383. doi:10.1080/00288330.1994.9516627

3. Pizarro M. An Unusual and Massive Bloom of Tetraselmis Sp. in the Valparaiso Bay, Chile. OFOAJ. 2018;7. doi:10.19080/OFOAJ.2018.07.555717

4. Mata MT, Cameron H, Avalos V, Riquelme C. Identification and Characterization of a Novel Microalgal Strain from the Antofagasta Coast Tetraselmis marina AC16-MESO (Chlorophyta) for Biotechnological Applications. Plants. 2023;12: 3372. doi:10.3390/plants12193372

5. Hemaiswarya S, Raja R, Ravi Kumar R, Ganesan V, Anbazhagan C. Microalgae: a sustainable feed source for aquaculture. World J Microbiol Biotechnol. 2011;27: 1737–1746. doi:10.1007/s11274-010-0632-z

6. Fon-Sing S, Borowitzka MA. Isolation and screening of euryhaline Tetraselmis spp. suitable for large-scale outdoor culture in hypersaline media for biofuels. J Appl Phycol. 2016;28: 1–14. doi:10.1007/s10811-015-0560-2

7. Steadman Tyler CR, Hovde BT, Daligault HE, Zhang XL, Kunde Y, Marrone BL, et al. High-Quality Draft Genome Sequence of the Green Alga *Tetraselmis striata* (Chlorophyta) Generated from PacBio Sequencing. Stajich JE, editor. Microbiol Resour Announc. 2019;8: e00780-19. doi:10.1128/MRA.00780-19

8. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 2014;12: e1001889. doi:10.1371/journal.pbio.1001889

9. Yao C, Ai J, Cao X, Xue S, Zhang W. Enhancing starch production of a marine green microalga Tetraselmis subcordiformis through nutrient limitation. Bioresource Technology. 2012;118: 438–444. doi:10.1016/j.biortech.2012.05.030

10. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 2019;574: 679–685. doi:10.1038/s41586-019-1693-2

11. Lim DKY, Schuhmann H, Thomas-Hall SR, Chan KCK, Wass TJ, Aguilera F, et al. RNA-Seq and metabolic flux analysis of Tetraselmis sp. M8 during nitrogen starvation reveals a two-stage lipid accumulation mechanism. Bioresource Technology. 2017;244: 1281–1293. doi:10.1016/j.biortech.2017.06.003

12. Lauritano C, De Luca D, Amoroso M, Benfatto S, Maestri S, Racioppi C, et al. New molecular insights on the response of the green alga Tetraselmis suecica to nitrogen starvation. Sci Rep. 2019;9: 3336. doi:10.1038/s41598-019-39860-5

13. Schvarcz CR, Steward GF. A giant virus infecting green algae encodes key fermentation genes. Virology. 2018;518: 423–433. doi:10.1016/j.virol.2018.03.010

14. Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, et al. Taxonomic update for giant viruses in the order Imitervirales (phylum Nucleocytoviricota). Arch Virol. 2023;168: 283. doi:10.1007/s00705-023-05906-3

15. Gajigan AP, Schvarcz CR, Conaco C, Edwards KF, Steward GF. Ultrastructural and transcriptional changes during a giant virus infection of a green alga. npj Viruses. 2025;3: 47. doi:10.1038/s44298-025-00128-7

16. Pagarete A, Grébert T, Stepanova O, Sandaa R-A, Bratbak G. Tsv-N1: A Novel DNA Algal Virus that Infects Tetraselmis striata. Viruses. 2015;7: 3937–3953. doi:10.3390/v7072806

17. Roux S, Fischer MG, Hackl T, Katz LA, Schulz F, Yutin N. Updated Virophage Taxonomy and Distinction from Polinton-like Viruses. Biomolecules. 2023;13: 204. doi:10.3390/biom13020204

18. Kapitonov VV, Jurka J. Self-synthesizing DNA transposons in eukaryotes. Proc Natl Acad Sci USA. 2006;103: 4540–4545. doi:10.1073/pnas.0600833103

19. Pritham EJ, Putliwala T, Feschotte C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene. 2007;390: 3–17. doi:10.1016/j.gene.2006.08.008

20. Krupovic M, Bamford DH, Koonin EV. Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. Biol Direct. 2014;9: 6. doi:10.1186/1745-6150-9-6

21. Duponchel S, Fischer MG. Viva lavidaviruses! Five features of virophages that parasitize giant DNA viruses. Condit RC, editor. PLoS Pathog. 2019;15: e1007592. doi:10.1371/journal.ppat.1007592

22. Sizov DV, Polischuk VP. Cultivation, purification and crystallization of virus of green algae Tetraselmis viridis. Biopolym Cell. 2006;22: 243–245. doi:10.7124/bc.000737

23. Chase EE, Desnues C, Blanc G. Integrated viral elements suggest the dual lifestyle of *Tetraselmis* spp. Polinton-like viruses. Virus Evolution. 2022;8: veac068. doi:10.1093/ve/veac068

24. Roitman S, Rozenberg A, Lavy T, Brussaard CPD, Kleifeld O, Béjà O. Isolation and infection cycle of a polinton-like virus virophage in an abundant marine alga. Nat Microbiol. 2023;8: 332–346. doi:10.1038/s41564-022-01305-7

25. Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, et al. Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. PNAS. 2013;110: 10800–10805. doi:10.1073/pnas.1303251110

26. Fischer MG, Hackl T. Host genome integration and giant virus-induced reactivation of the virophage mavirus. Nature. 2016;540: 288–291. doi:10.1038/nature20593

27. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, et al. The virophage as a unique parasite of the giant mimivirus. Nature. 2008;455: 100–104. doi:10.1038/nature07218

28. Blanc G, Gallot-Lavallée L, Maumus F. Provirophages in the Bigelowiella genome bear testimony to past encounters with giant viruses. PNAS. 2015; 201506469. doi:10.1073/pnas.1506469112

29. Aylward FO, Abrahão JS, Brussaard CPD, Fischer MG, Moniruzzaman M, Ogata H, et al. Taxonomic update for giant viruses in the order Imitervirales (phylum Nucleocytoviricota). Arch Virol. 2023;168: 283. doi:10.1007/s00705-023-05906-3

30. Gilbert GE, Baleja JD. Membrane-binding peptide from the C2 domain of factor VIII forms an amphipathic structure as determined by NMR spectroscopy. Biochemistry. 1995;34: 3022–3031. doi:10.1021/bi00009a033

31. Krupovic M, Yutin N, Koonin EV. Fusion of a superfamily 1 helicase and an inactivated DNA polymerase is a signature of common evolutionary history of Polintons, polinton-like viruses, Tlr1 transposons and transpovirons. Virus Evol. 2016;2: vew019. doi:10.1093/ve/vew019

32. Redrejo-Rodríguez M, Ordóñez CD, Berjón-Otero M, Moreno-González J, Aparicio-Maldonado C, Forterre P, et al. Primer-Independent DNA Synthesis by a Family B DNA Polymerase from Self-Replicating Mobile Genetic Elements. Cell Reports. 2017;21: 1574–1587. doi:10.1016/j.celrep.2017.10.039

33. Barth ZK, Hicklin I, Thézé J, Takatsuka J, Nakai M, Herniou EA, et al. Genomic analysis of hyperparasitic viruses associated with entomopoxviruses. Virus Evolution. 2024;10: veae051. doi:10.1093/ve/veae051

34. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res. 2015;43: 10633–10654. doi:10.1093/nar/gkv1267

Supplementary Tables

Table S1 : Features of Tetraselmis viruses

Contig	Length (bp)	TIR length (bp)	Genome conformation	Number of protein genes	Number of tRNA genes
TetV-2	660 680	None	Circular	640	4
Tsv-S2a	19 063	1632	Linear	17	0
Tsv-S2a.bis	19 170	1711	Linear	17	0
Tsv-S2b	19 124	1724	Linear	18	0
Tsv-S3b	18 038	583	Linear	17	0

Table S1 : Average nucleotide identity between Tetraselmis viruses and their ETE guides

Average Nucleotide Identity (ANI)	C2208 (S2a)	C1731 (S2b)	C2184 (S3b)	C0566 (S2a.bis)	Tsv-S2a	Tsv-S2a.bis	Tsv-S2b
C1731 (S2b)	94,58 %						
C2184 (S3b)	65,31 %	62,03 %					
C0566 (S2a.bis)	98,53 %	96,05 %	65,81 %				
Tsv-S2a	99,98 %	94,62 %	65,37 %	98,52 %			
Tsv-S2a.bis	99,13 %	95,58 %	65,56 %	99,05 %	99,26 %		
Tsv-S2b	94,78 %	99,81 %	62,06 %	96,01 %	94,79 %	95,72 %	
Tsv-S3b	65,85 %	62,18 %	99,43 %	65,96 %	65,92 %	66,11 %	62,20 %

Supplementary Figures :

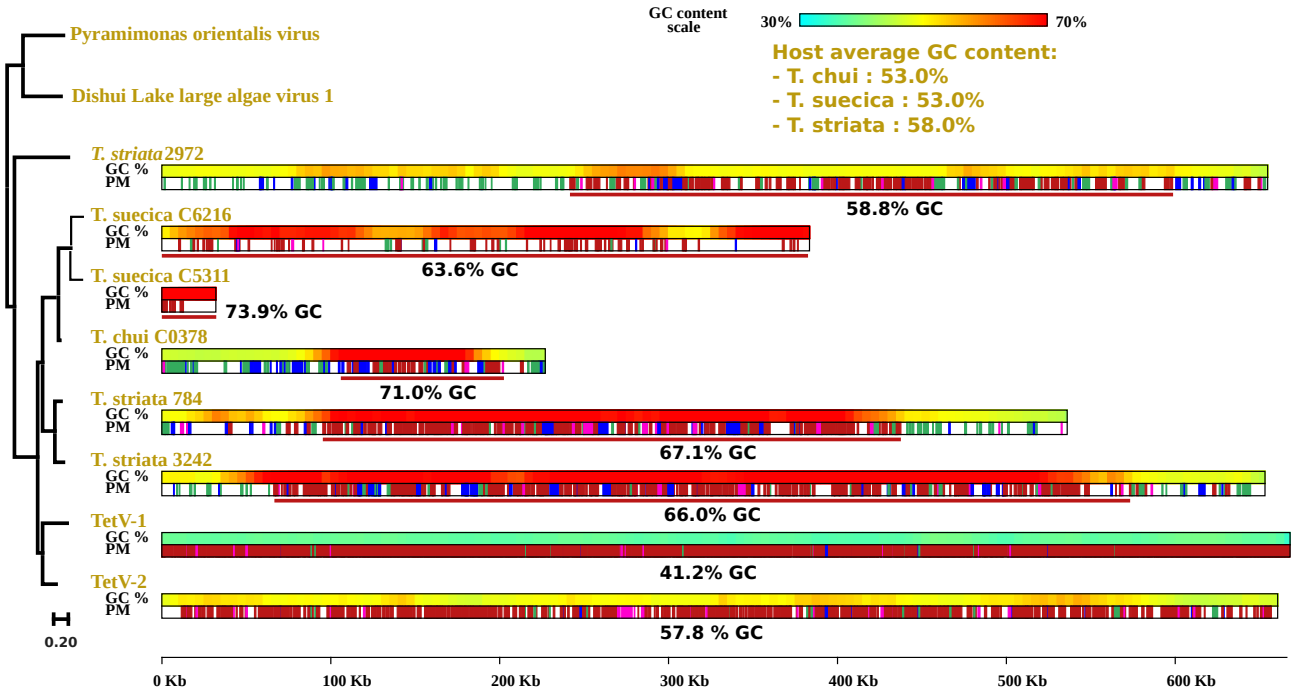

Figure S1 | Endogenous TetV-like elements in Tetraselmis genomes. *Left: Phylogenetic tree of viral DNA polymerases encoded by TetV-1, TetV-2, and TetV-like insertions in Tetraselmis genomes (data from Chase et al., 2018; DOI: 10.1093/ve/veac068). The TetV-1 DNA polymerase was used as a TBLASTN query to identify homologs in algal genomes. Protein sequences were aligned with MAFFT, and phylogenetic reconstruction was performed with IQ-TREE under default parameters. All branches received >80% ultrafast bootstrap support. Right: Genomic contigs encoding viral DNA polymerases. Two color-coded tracks indicate local GC content and the taxonomic affiliation of the best BLASTX protein match in GenBank NR. Colors: red, virus; green, green alga; blue, other eukaryotes; pink, prokaryotes. Viral insertion regions are underlined in red, with the average GC content of each insertion shown below.*

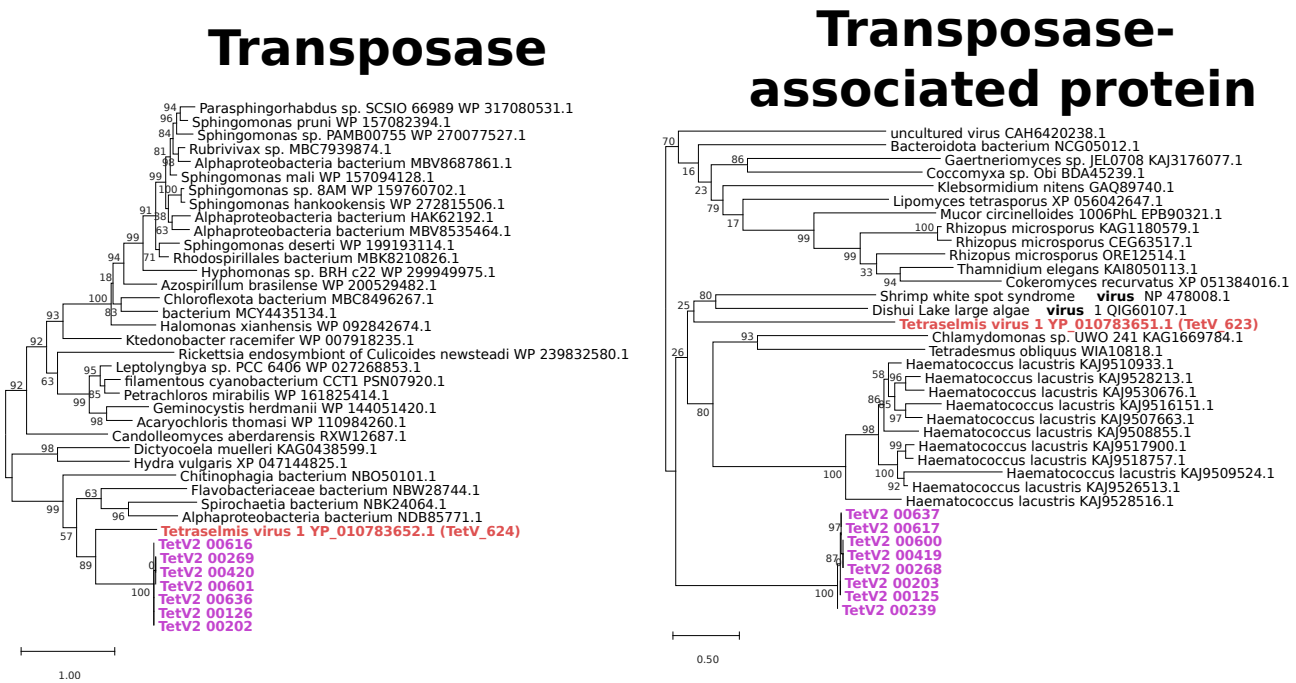

Figure S2: phylogenetic trees of the transposases and transposase-associated proteins encoded by duplicated gene doublets. *Phylogenetic reconstructions were generated with FastTree using default parameters. Branch support values were estimated with the SH-like local support method implemented in FastTree.*

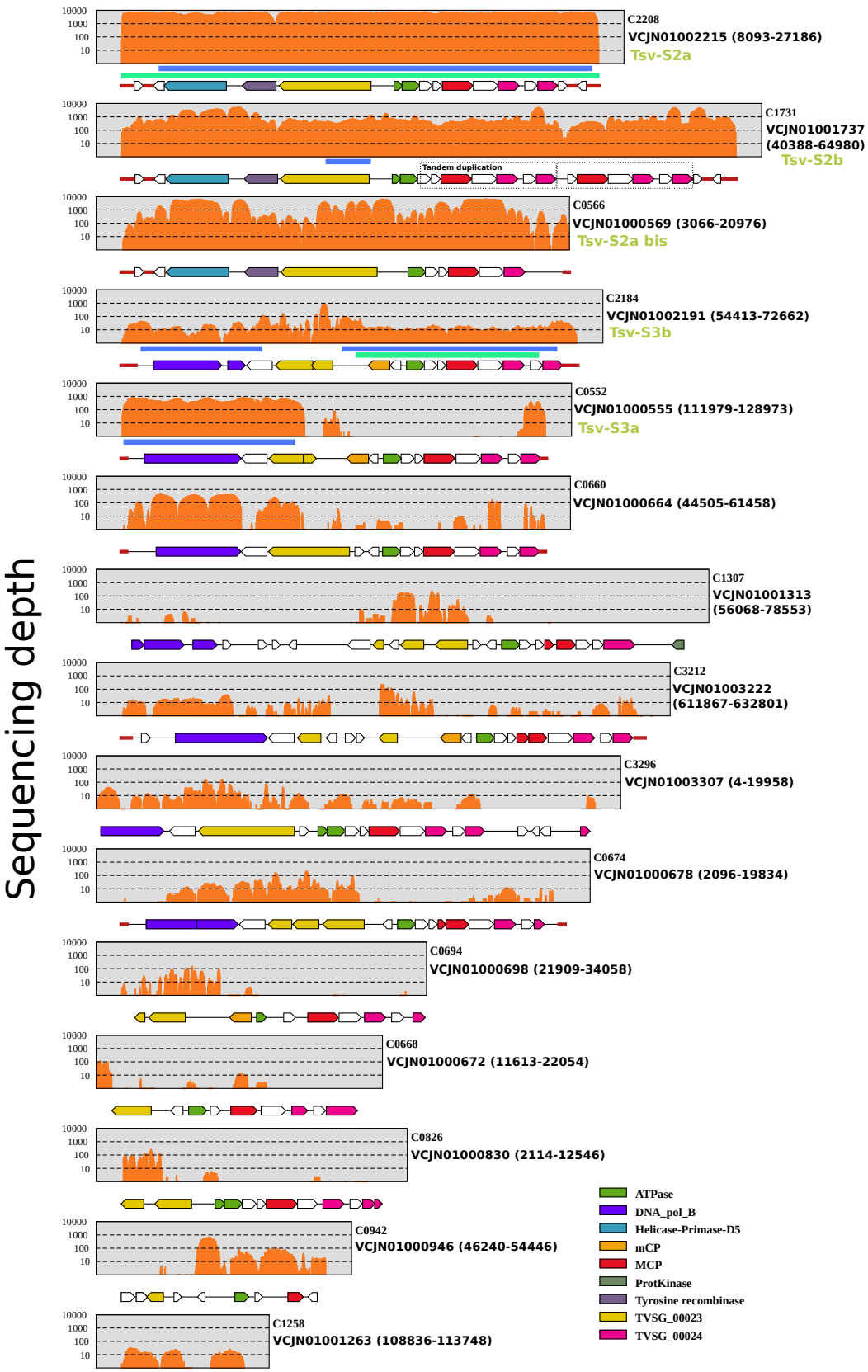

Figure S3 : Regions of the LANL1001 genome covered by Illumina short reads.
Each panel shows position-specific read coverage (orange, y-axis). Below the coverage plots are the corresponding regions assembled from short reads (thick blue lines) and long reads (thick green lines), together with annotations of endogenous TetV-like elements (ETEs), including genes (colored arrows) and terminal inverted repeats (TIRs, thick red lines). Gene color codes are indicated in the figure. For each region, the region identifier, contig GenBank accession number, and genomic coordinates within the contig are provided.

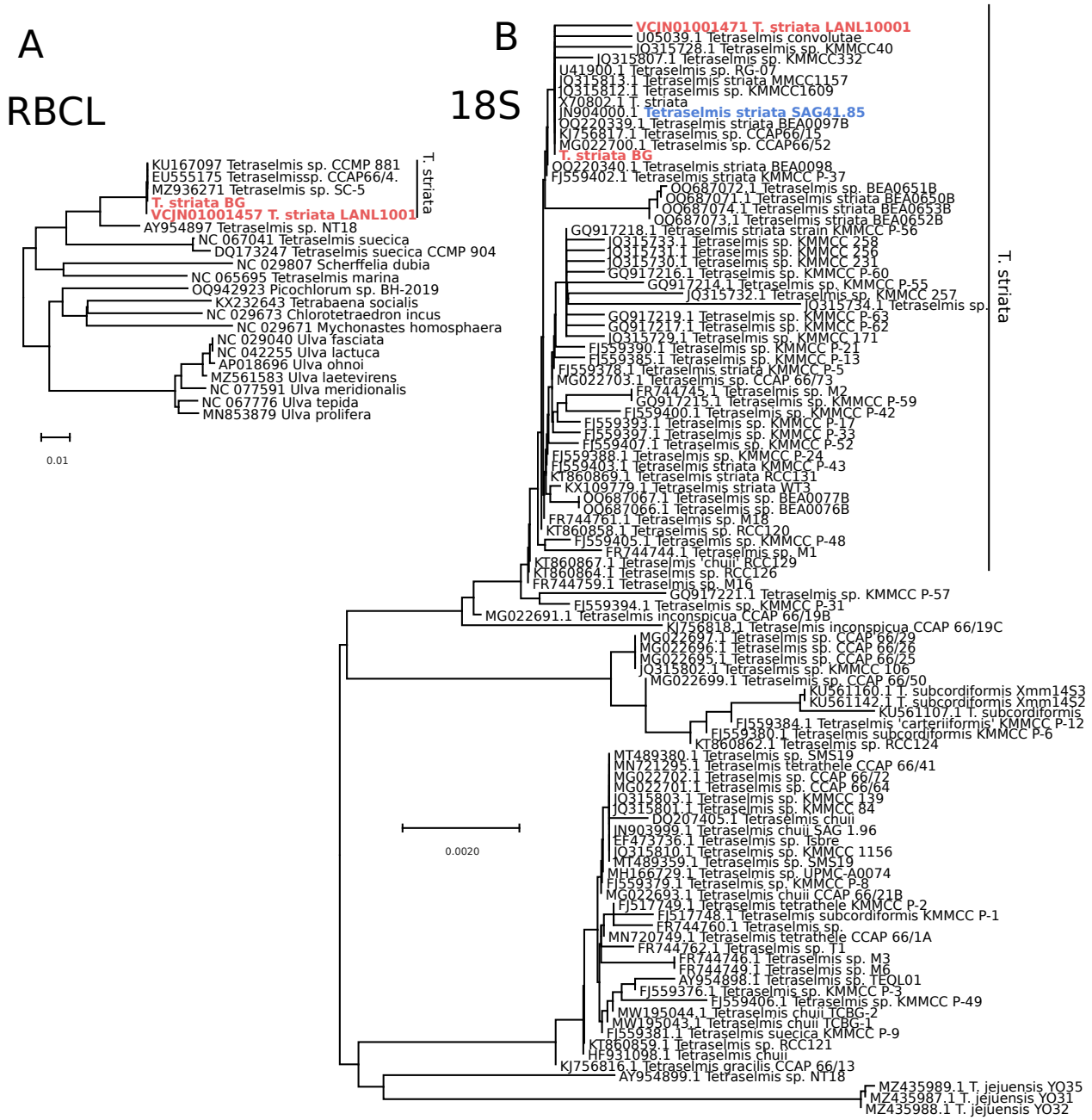

Figure S4: *Minimum evolution phylogenetic trees of RBCL (A) and 18S rDNA (B) sequences generated using MEGA X.*

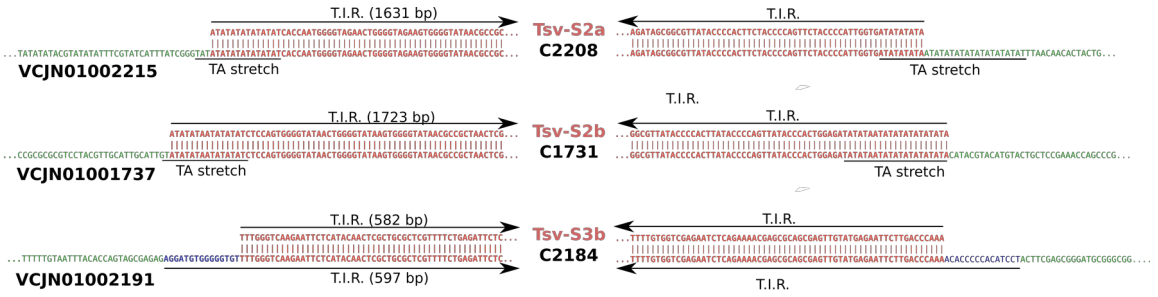

Figure S5 : Sequence alignments at insertion sites between Tsv genomes and ETE guides.

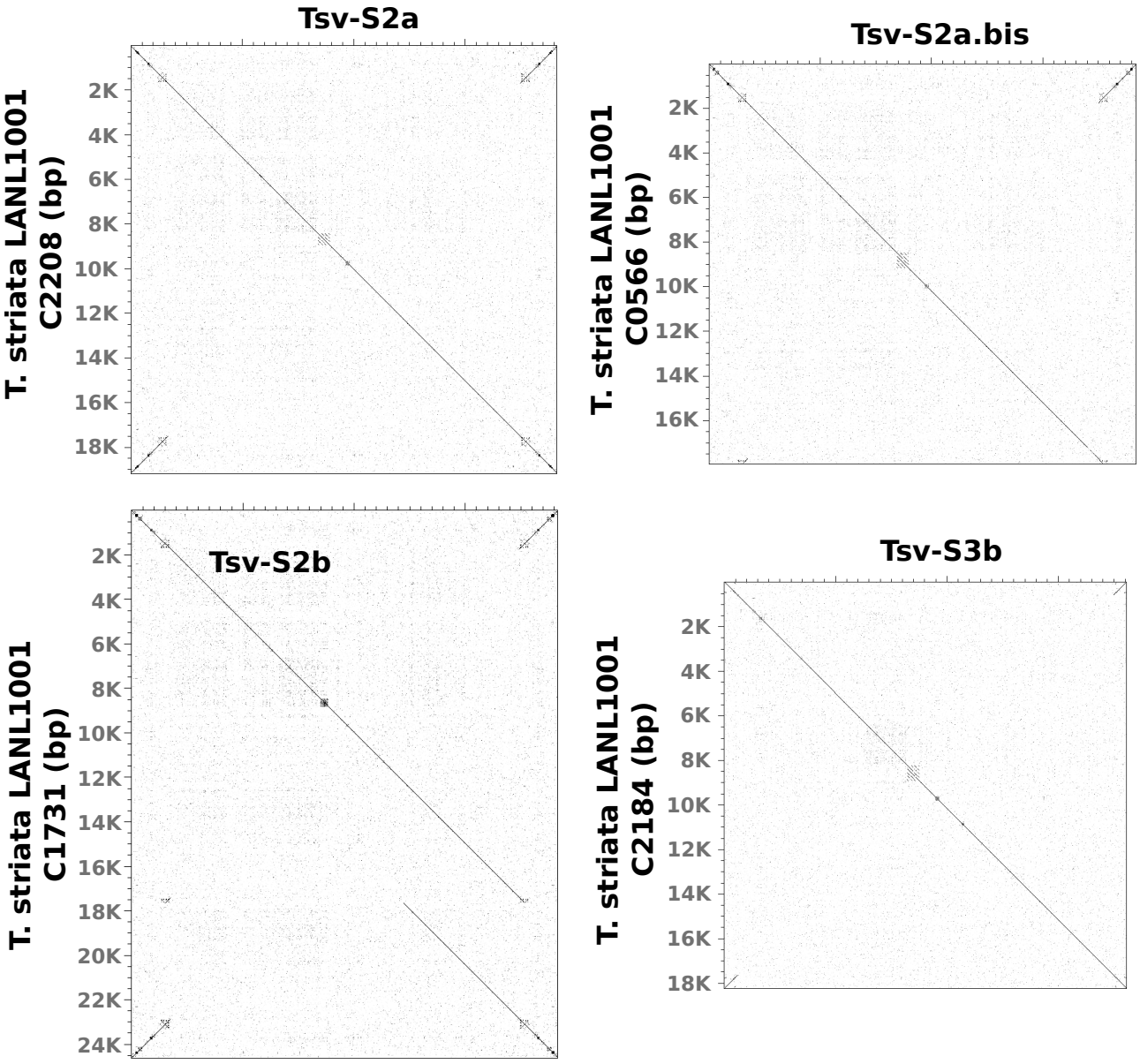

Figure S6 *: Alignment of assembled Tetraselmis virus (Tsv) genomes against endogenous Tsv element (ETE) reference regions used for Illumina short-read recruitment. Pairwise alignments were generated using Dotter from the SeqTools package (Sonnhammer & Durbin, 1995; Gene 167: GC1–10; doi:10.1016/0378-1119(95)00714-8).*

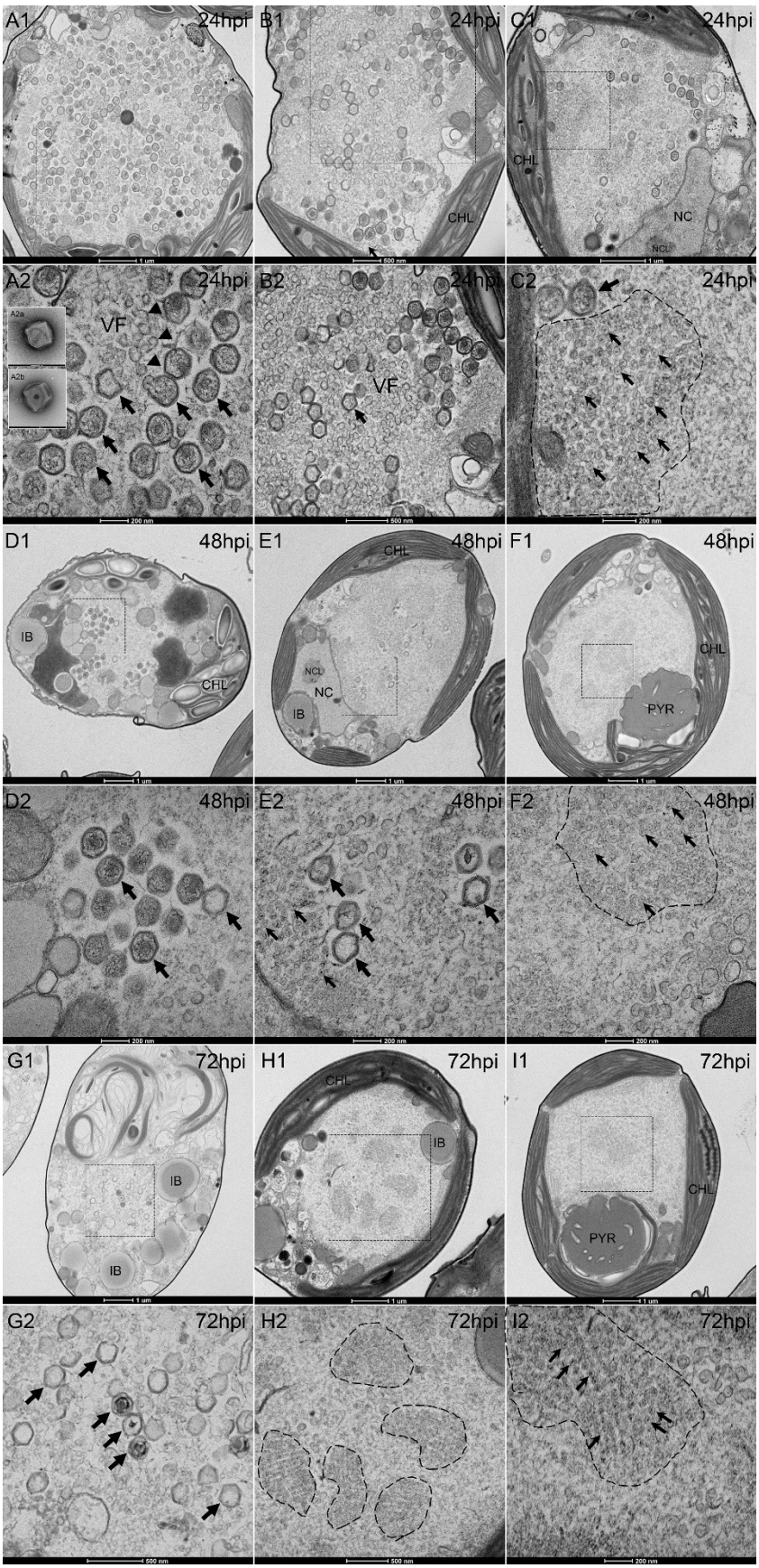

Figure S7. Transmission electron microscopy of Tetraselmis viruses during infection of Tetraselmis striata (TS).

(A1–C2) *Ultra-thin sections of infected cells at 24 hours post-infection (hpi). Panels A1, B1, and C1 show overviews of cellular organization; panels A2, B2, and C2 display higher magnifications of*

the boxed regions. In A2 and B2, high-magnification views reveal assembly of giant virus particles within putative viral factories (VF), with membrane fragments (arrowheads), as well as empty and partially filled particles. In C2, smaller Tetraselmis virus-like particles (Tsv) are concentrated (dashed line) at the cell periphery. Panels A2a and A2b show negatively stained mature giant virus particles at the same scale as A2. ***(D1–F2)*** *Ultra-thin sections of infected TS cells at 48 hpi. D1, E1, and F1 show cellular overviews, while D2, E2, and F2 depict corresponding higher-magnification views. D2 and E2 reveal mature and immature empty giant virus particles, whereas F2 shows Tsv particles concentrated in a central region (dashed line).* ***(G1–I2)*** *Ultra-thin sections of infected TS cells at 72 hpi. G1, H1, and I1 show overviews of cellular ultrastructure, with G2, H2, and I2 providing higher-magnification views of the boxed areas. Most giant virus particles appear empty (G2). Peripheral and central regions rich in Tsv particles (dashed lines in H2 and I2) are evident, with Tsv assembling into pseudo-crystalline arrays. Abbreviations: NC, nucleus; NCL, nucleolus; IB, inclusion body; CHL, chloroplast; PYR, pyrenoid.*

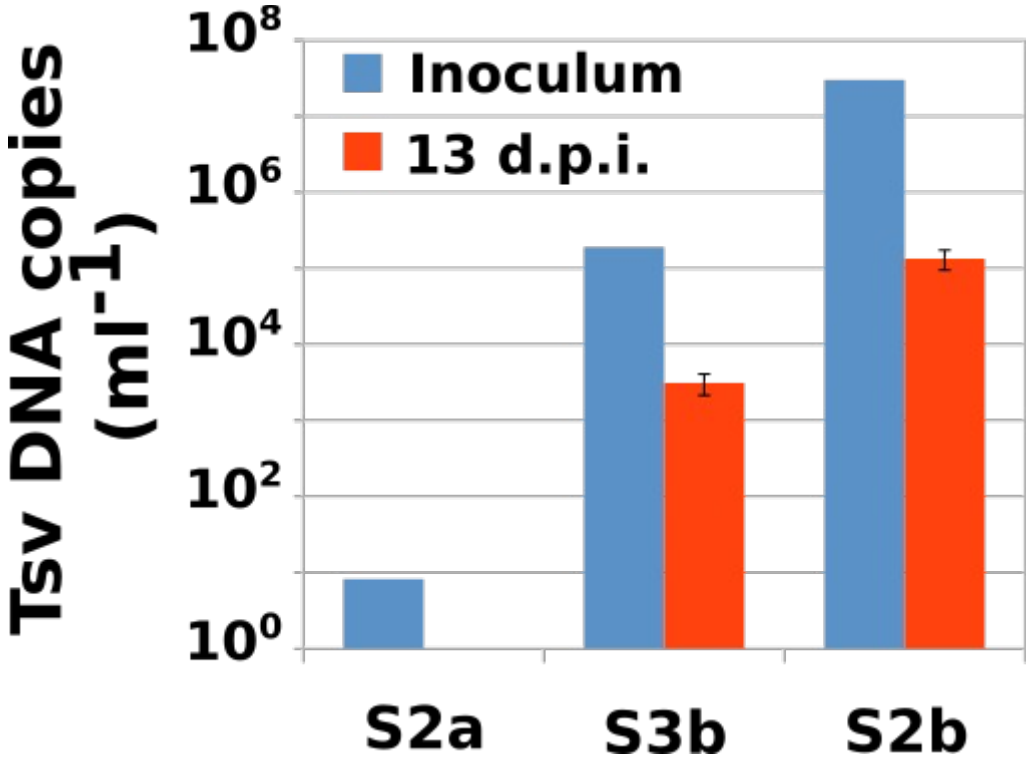

Figure S8. Quantification of TsV DNA copy numbers at 0 and 13 days post infection (dpi).

TsV abundance was determined by quantitative PCR (qPCR) using strain-specific primer sets. The inoculum consisted of a TsV mixture dominated by strain TsV-S2b. Bars represent the mean values of three biological replicates, and error bars indicate the standard deviation.

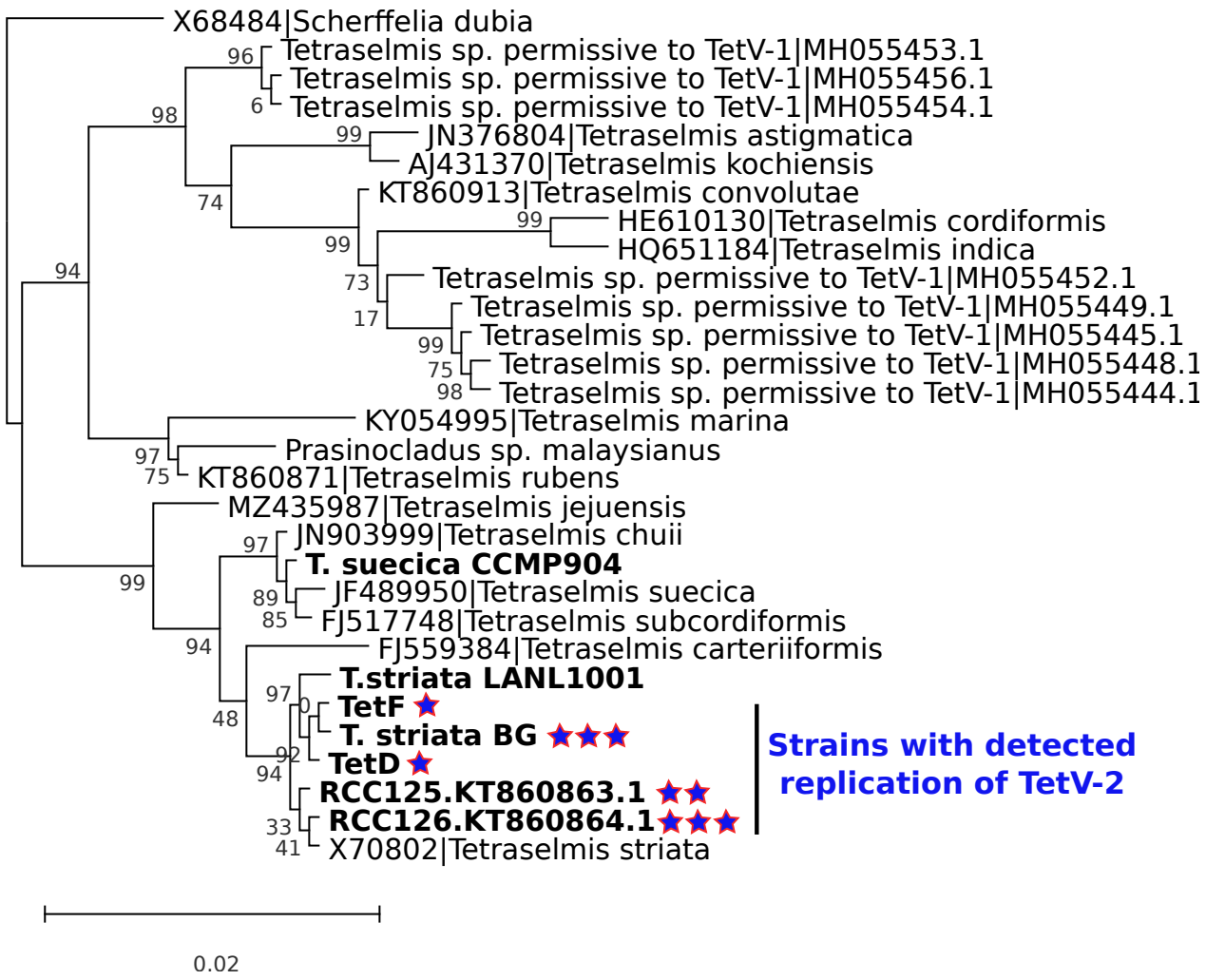

Figure S9. 18S rRNA phylogenetic relationships among Tetraselmis strains used in the host range assay. *Strains tested in this study are shown in bold. TetV-2 replication levels are represented by stars, where an increasing number of stars denotes higher viral replication.*

Supplementary Data :

Sequences of Tetraselmis striata strain GB gene markers :

>RBCL-T.striata.sp.BG

GCGTACCCTTTAGATTYATTTGAAGAAGGTTCAGTTACAAACTTATTTACTTCTATTGTAG
KTAACGTTTTTGGTTTCAAAGCTTTACGTGCTCTTCGTTTAGAAGATTTACGTATTCCAGT
TGCTTACTGTAAAACTTTTACAGGTGCTCCGCACGGTATTCAAGTTGAACGTGATAAATT
AAACAAATATGGACGTGGTTTATTAGGTTGTACTATTAAACCAAAATTAGGTCTTTCTGCT
AAAAACTACGGTCGTGCTTGTTACGAATGTTTACGTGGTGGTTTAGATTTTACAAAAGAT
GATGAAAACGTAAACTCACAAGCATTTATGCGTTGGAGAGATCGTTTCCTATTCGTATCA
GAAGCTATCTACAAATCACAAGCAGAAACTGGTGAAATTAAAGGTCACTACTTAAACGT
AACAGCAGGTACTTGTGAAGAGATGATGAAGCGTGCTGAATGTGCCGCAGGTTTCGGTG
TACCAATTGTTATGCACGATTACTTAACAGGTGGTTTTACAGCTAACACTTCATTAGCTAT
TTACTGTCGTGATAATGGTTTACTATTACACATTCACCGTGCTATGCACGCAGTTATTGAC
CGTCAACGTAATCACGGAATTCACTTCCGTGTTTTAGCAAAAGCTTTACGTATGTCTGGT
GGGGATCACCTTCACTCAGGTACTGTTGTAGGTAAATTAGAAGGTGAACGTGAAGTTAC
TTTAGGTTTCGTAGATTTAATGCGTGATGCTTACGTAGAAAAAGATCGTTCTCGTGGT

>18S-T.striata.sp.BG

aatcatgataacttcacgaatcgcatggcctccgcgccggcgatgtttcattcaaatttctgccctatcaatttgcgatggtaggatagaggcctac
catggtggtaacgggtgacggaGaattagggttcgattccggagagggagcctgagaaacggctaccacatccaaggaaggcagcaggcg
cgcaaattacccaatcctgacacagggaggtagtgacaataaataacaataccgggcttttcaagtctggtaattggaatgagtacaatctaaaca
accttaacgaggatccattggagggcaagtctggtgccagcagccgcggtaattccagctccaatagcgtatatttaagttgctgcagttaaaaa
gctcgtagttggatttcggatgggatttgccggtccgccgtttcggtgtgcactggccagtcccatcttgttgtcggggactagctcctgggcttca
ctgtccgggactaggagctgacgaggttactttgagtaaattagagtgttcaaagcaagcctacgctctgaatacattagcatggaataacatgat
aggactctggcttatcttgttggtctgtgagaccagagtaatgattaagagggacagtcgggggcattcgtatttcattgtcagaggtgaaattcttg
gatttatgaaagacgaacttctgcgaaagcatttgtcaaggatgttttcattaatcaagaacgaaagttgggggctcgaagacgattagataccgtc
ctagtctcaaccataaacgatgccgactagggattggcagacgtttttttgatgactctgccagcaccttatgagaaatcaaagtttttgggttccgg
ggggagtatggtcgcaaggctgaaacttaaaggaattgacggaagggcaccaccaggcgtggagcctgcggcttaatttgactcaacacggg
aaaacttaccaggtccagacatagtgaggattgacagattgagagctctttcttgattctatgggtggtggtgcatggccgttcttagttggtgggtt
gccttgtcaggttgattccggtaacgaacgagacctcagcctgctaaatagttactcctactttggtaggaggtgaacttcttagagggactattgg
cgtttagccaatggaagtgtgaggcaataacaggtctgtgatgcccttagatgttctgggccgcacgcgcgctacactgatgcattcaacgagcc
tagccttgaccgagaggtccgggtaatctttgaaactgcatcgtgatggggctagattattgcaattattaatcttcaacgaggaatgcctagtaag
cgtgattcatcagatcgcgttgattacgtccctgccctttgtacacaccgcccgtcgctcctaccgattgaatgtgttggtgaggagttcggattggc
agtttgtggtggttcgccactgcttacagctgagaagttctccaaaccgccccatttagaggaaggagaagtcgtaacaaggtttccgtaggtgaa
cctgc