Herpes-like viral elements and universal subtelomeric ribosomal RNA genes in a chromosome-scale thraustochytrid genome assembly

Jackie L. Collier*^[1], Joshua S. Rest*[2], Lucie Gallot-Lavallée[3], Erik Lavington[2], Alan Kuo[4], Jerry Jenkins[4,5], Chris Plott[4,5], Jasmyn Pangilinan[4], Chris Daum[4], Igor V. Grigoriev[4,6], Gina V. Filloramo[3], Anna M. G. Novák Vanclová[7], John M. Archibald[3]

*these authors contributed equally
^corresponding author

[1] jackie.collier@stonybrook.edu, School of Marine and Atmospheric Sciences, Stony Brook University
[2] joshua.rest@stonybrook.edu, enlaving@gmail.com, Department of Ecology and Evolution, Stony Brook University
[3] Lucie.Gallot-Lavallee@dal.ca, Gina.Filloramo@dal.ca, jmarchib@dal.ca, Department of Biochemistry & Molecular Biology, Dalhousie University
[4] akuo@lbl.gov, JLPangilinan@lbl.gov, cgdaum@lbl.gov, ivgrigoriev@lbl.gov, U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory
[5] jjenkins@hudsonalpha.org, cplott@hudsonalpha.org, HudsonAlpha Institute for Biotechnology
[6] Department of Plant and Microbial Biology, University of California Berkeley
[7] novak@biologie.ens.fr, Faculty of Science, Charles University, BIOCEV, Vestec, Czechia.

ABSTRACT

We used long-read sequencing to produce a telomere-to-telomere genome assembly for the heterotrophic stramenopile protist *Aurantiochytrium limacinum* MYA-1381. Its ~62 Mbp nuclear genome comprises 26 linear chromosomes with a novel configuration: subtelomeric rDNAs are interspersed with long repeated sequence elements denoted as LOng REpeated - TElomere And Rdna Spacers (LORE-TEARS). These repeats may play a role in chromosome end maintenance. A ~300 Kbp circular herpesvirus-like genomic element is present at a high copy number. A 269 Kbp related virus-like element was found to reside between two complete sets of rRNA and LORE-TEAR sequences on one end of chromosome 15, indicating recent recombination between the viral and nuclear genome. Our data reveal new types of giant endogenous viral elements originating from herpes-like viruses and existing as either 'stand-alone' or integrated elements.

KEYWORDS

subtelomere structure, endogenous viral element, mirusvirus, long tandem repeat

MAIN

Labyrinthulomycetes are ecologically important osmoheterotrophic marine stramenopiles which possess interesting biochemical and cell biological novelties (Fossier Marchan et al. 2018). For example, the possession of a polyketide synthase-like fatty acid biosynthesis pathway has made some strains workhorses for production of the essential polyunsaturated fatty acid docosahexaenoic acid, and they may also be suitable for large-scale production of isoprenoid compounds like squalene and carotenoids (Rius et al. 2023; X. Xu et al. 2020). Labyrinthulomycetes were the source of pioneering discoveries in molecular biology, with notable examples published in *Science*. In the 1960s, Moens and Perkins (1969) demonstrated that *Labyrinthula* has a sexual cycle, haploid number of 9 chromosomes, and that the chromosome ends are associated with the nuclear membrane by tracing the synaptonemal complexes of prophase nuclei in serial sections by electron microscopy. In the 1970s, the first herpes-like virus detected outside of vertebrates was reported by Kazama and Schornstein (1972). Notably, neither of these important observations were followed up in the phylum in which they were discovered. The advent of long-read DNA sequencing (Rhoads and Au 2015; Branton et al. 2008) can rapidly advance the connection of old observations like these to the omic era by enabling the assembly of highly contiguous genomes. Chromosome-scale genomic sequences can reveal the abundance and distribution of repeat elements and ribosomal RNA (rRNA) gene clusters, the number and location of centromeres, the presence of endogenized viral genes and genomes, and the structure of telomeres and subtelomeric regions (Filloramo et al. 2021; Guérin et al. 2021; Fang et al. 2020; Fletcher et al. 2019; Moniruzzaman et al. 2020). Genome-scale leaps in biological insight can be particularly valuable for non-model organisms like the labyrinthulomycetes (Collier and Rest 2019), analysis of which can expand perspectives on eukaryotic biology. Here we present genome-wide discoveries in the labyrinthulomycete protist *Aurantiochytrium limacinum* ATCC MYA-1381 (a thraustochytrid) that corroborate and extend early reports about the genome structure and viral associates of labyrinthulomycetes.

RESULTS AND DISCUSSION

Chromosome Architecture and Subtelomere-Associated Elements in the *A. limacinum* Genome

We performed short-read 454 (454 Life Sciences) and long-read Nanopore (Oxford Nanopore Technologies (Jain et al. 2016)) sequencing of the *A. limacinum* nuclear genome, which independently yielded assemblies of 60.93 Mbp (Newbler) and 63.71 Mbp (Canu; (Koren et al. 2017)), respectively (Additional File 1: **Text S1, Fig. S1;** Additional File 2: **Table S1, Table S2**, **Additional File 3**). Neither assembly was rich in repetitive sequence, with ~4% of the assemblies containing repetitive sequence (mostly simple repeats (Chen 2004) (Additional File 1: **Table S3**). Multiple metrics suggested that both assemblies were highly complete: 99.66% of RNA-seq reads mapped to the 454 assembly and 96% to the Nanopore assembly (Additional File 1: **Text S1**), and we detected 91.4% and 87.9% of Eukaryota BUSCO genes in the 454 and Nanopore assemblies, respectively (Simão et al. 2015) (8.6% and 12.1% missing BUSCOs, respectively; Additional File 1: **Table S4**).

We found that the 26 largest Nanopore contigs likely represent complete or nearly complete *A. limacinum* physical chromosomes. These contigs range in size from ~1.02 Mbp to ~4 Mbp (**Fig. 1**) and total 61.41 Mbp (96.4% of the complete Nanopore assembly); they align with 37 of the longest 454 scaffolds (referred to here as the primary 454 assembly) containing 59.93 Mbp (98.4%) of the total 454 assembly; Additional File 1: **Fig. S1B, Fig. S2;** Additional File 2: **Table S1, S2**). The Nanopore contig sizes are consistent with our examination of the genome by pulsed-field gel electrophoresis, which detected chromosome-sized bands ranging from ~1.05 Mbp to >3 Mbp (Additional File 1: **Fig. S3**). This genome structure is similar to other stramenopiles for which both chromosome number and genome size are known: three diatoms and a eustigmatophyte have smaller genomes (31-36.5 Mbp) with similar numbers of chromosomes (24 to 33), while

the oomycete *Phytophthora sojae* has an 82.6 Mbp genome with ~12-14 chromosomes (Filloramo et al. 2021; Diner et al. 2017; Guérin et al. 2021).

Most *A. limacinum* chromosomes were sequenced telomere-to-telomere. Among the 52 predicted chromosome ends, 39 terminate with telomeric repeats of sequence TTAGG ~500 bp in length (mean 480 bp, median 499 bp) (**Fig. 2**). The telomeric repeats identified in the Nanopore assembly are slightly shorter than the TTAGGG repeats of vertebrates and other eukaryotes, including some fungi, plants, and protists such as the photosynthetic stramenopile *Pelagomonas calceolata* (Guérin et al. 2021), but identical to TTAGG repeats reported from diverse insects and a few other eukaryotes (http://telomerase.asu.edu/sequences_telomere.html). Telomeric repeats are missing from 13 of the chromosome ends; this likely reflects assembly issues (Additional File 1: **Text S1**).

Almost all of the sub-telomeric regions of the *A. limacinum* chromosomes unexpectedly contain 18S, 5.8S, and 28S rRNA gene clusters interspersed with long repeated sequences (**Fig. 1, 2**). These elements are evident as extensive sequence matches between the ends of all contigs (**Fig. 1**, inset; in contrast to the 454 assembly; Additional File 1: **Fig. S4**). Specifically, in 37 of the 39 Nanopore contig ends with telomeres, an 18S-28S rRNA gene cluster (small subunit or 18S rRNA, ITS1, 5.8S rRNA, ITS2, large subunit or 28S rRNA; average length = 5551bp) transcribed toward the telomere is found ~9.4 Kbp (median) from the telomeric repeat (**Fig. 2**). In 30 of these 37 contig ends, a 5S rRNA gene transcribed away from the telomeric repeat is found ~10 Kbp (median) further from the telomere. In 17 of these 30 cases, no other rRNA genes were identified, and the 454 assembly scaffold mapped to the approximate location of the 5S gene. The remaining contig ends vary from this pattern. In some cases both ends of a contig have the same organization (Chr17, Chr15, Chr6), but more commonly the two ends are different. Only one rRNA gene (a 5S on Chr3) was found in the opposite orientation, and only three 5S genes (on Chr21, Chr22, Chr15) and one 18S-28S rRNA gene cluster (on Chr15) were identified in the more central regions of Nanopore assembly contigs (**Fig. 1)**.

Between each of these telomeric and subtelomeric elements are characteristic long repeated sequences (**Fig. 2;** Additional File 1: **Table S5;** identified with Tandem Repeats Finder (Benson 1999)). We call these LOng REpeated - TElomere And RDna Spacers (LORE-TEARS), which are built from 366 - 529 bp units repeated between 1.8 and 22.6 times and lacking similarity to sequences in GenBank. Several distinct LORE-TEARS families occur in regular positions with respect to the chromosome ends. Just downstream of the 28S rRNA genes, there is usually one ‘Group 1’ element containing ~4 repeated units, each ~406 bp long. Closer to the telomere, there is usually at least one ‘Group 2’ element with ~6 repeated units, each ~366 bp long. Usually upstream from the 18S gene nearest the telomere and between it and the nearest 5S gene is a ‘Group 0’ element, containing ~9 repeated units of ~385 bp. Where two consecutive 5S rRNA genes are detected, there is often a ‘Group 5’ element between them (~5 repeats of a ~421 bp unit). Among the 15 putative chromosomes with telomeric repeats assembled at both ends, seven have a ‘Group 3’ element (~3 repeats of a ~529 bp unit) between the ‘Group 1’ and ‘Group 2’ elements at only one end, while one has a ‘Group 3’ element between the ‘Group 1’ and ‘Group 2’ elements at both ends and seven have no ‘Group 3’ elements. We detected G-quadruplexes associated with rRNA genes and some LORE-TEARS, which is consistent with a regulatory function for these elements at the chromosome ends (Juranek and Paeschke 2012; Paeschke et al. 2008; Biffi, Tannahill, and Balasubramanian 2012; Wang et al. 2012) (**Fig. 2**, Additional File 2: **Table S6**).

The organization of rRNA genes in the subtelomeres of the *A. limacinum* chromosomes suggests a consistent and specific relationship with telomeric processes. This arrangement is highly unusual, both in the nature of the repeats and their location. Eukaryotic rRNA genes are most commonly organized in a few large tandem arrays (e.g., one in yeast, five in humans) often associated with sites of genomic fragility, and 5S genes and 18S-28S gene clusters are typically not associated with one another (Kobayashi 2011; Torres-Machorro et al. 2009). Subtelomeric rRNA tandem repeats have been found in plants (Dvořáčková, Fojtová, and Fajkus 2015; Roa and Guerra 2012) and in some metazoans (in aphids at one telomere of the X chromosome (Criniti et al. 2009)), and in the protist parasite *Giardia* (Tůmová et al. 2015; F. Xu, Jex, and

Svärd 2020). The multicellular stramenopile *Saccharina japonica* (the kelp kombu) has a typical tandem 45S array in the middle of one chromosome, and a tandem 5S array at the subtelomere of another (L. Liu et al. 2017). In all these cases, rRNA gene arrays reside on the ends of only one or a few chromosomes. Unlinked 18S-28S rRNA gene clusters (i.e., not in tandem arrays) are found in the red alga *Cyanidioschyzon merolae* (Maruyama et al. 2004; Matsuzaki et al. 2004) and in several apicomplexan parasites (Torres-Machorro et al. 2010) including *Plasmodium falciparum* (Gardner et al. 2002). The 5S and 18S-28S coding regions in the Nanopore assembly of *A. limacinum* are more closely spaced than is usual for organisms where this linkage occurs, but not as tightly linked as in the brown alga (stramenopile) *Scytosiphon lomentaria* and some other protists, where the 5S is just downstream of the 18S-28S (Kawai et al. 1997, 1995).

The most similar sub-telomeric chromosome architecture to that of *A. limacinum* is found in the microsporidian parasites *Encephalitozoon cuniculi* and *E. intestinalis*, which have one subtelomeric, divergently transcribed 18S-28S rRNA gene cluster near the end of each of their 11 chromosomes separated from the telomeric repeats by two types of telomere-associated repeat elements (TAREs) with ~30 to 70 bp repeat units (Mascarenhas Dos Santos et al. 2023). However, the 5S rRNA genes are not subtelomeric in *Encephalitozoon* spp. Subtelomeric 18S-28S rRNA gene clusters are also a chromosomal feature of the endosymbiotically-derived 'nucleomorph' genomes of cryptomonads (Douglas et al. 2001; Kim et al. 2022) and chlorarachniophytes (Suzuki et al. 2015). Brugère et al. (2000) suggested that the subtelomeric location of rDNA might be related to selective pressure associated with genome reduction, but the ~62 Mbp genome of *A. limacinum* is not notably small among free-living stramenopiles, suggesting that genomic streamlining is not a factor here. The ends of chromosomes tend to be different from internal portions in exhibiting a higher frequency of recombination (Jensen-Seaman et al. 2004; McKim, Howell, and Rose 1988), lower level of gene expression (T. Liu et al. 2011), and higher rate of sequence evolution (Perry and Ashworth 1999). The selective forces and molecular mechanisms (e.g., convergent evolution by frequent inter- and intra-chromosomal homogenization) acting to maintain the consistent structure and homogeneous rDNA and LORE-TEARS sequences at the chromosome ends of *A. limacinum* offer novel avenues for future research, particularly if similar arrangements are found broadly in labyrinthulomycetes or other unexplored corners of protist diversity. The rRNA genes, LORE-TEARS, and/or associated subtelomeric sequences in *A. limacinum* may be involved in chromosome end maintenance and replication, including the maintenance of rDNA stability and/or nucleolar structure (Torres-Machorro et al. 2009), comparable to the repetitive subtelomeric sequences that are functionally important in other species (Tashiro et al. 2017; Scherf, Figueiredo, and Freitas-Junior 2001).

Discovery of herpes-like viral elements in *A. limacinum*: Characterization of CE1 and LE-Chr15

We also detected a 298 Kbp chromosome of probable viral ancestry in *A. limacinum*, dubbed CE1 (circular element 1) (**Fig. 1**), with several remarkable features. This 27th genomic element is present in both genome assemblies (Additional File 2: **Table S1, S2**) and is consistent with a ~0.35 Mbp band in the pulsed-field gel electrophoresis (Additional File 1: **Fig. S3)**. CE1 is predicted to be circular (Additional File 1: **Fig. S5**), and has read coverage (reads/bp) ~9X higher than the other chromosomes, suggesting that it is present at a much higher copy number (Additional File 2: **Table S1**). CE1 lacks the predicted rRNA genes, LORE-TEARS, and telomeric repeats found on the other chromosomes (**Fig. 1**)**.** GC content and mapped transcript abundance are similar to other chromosomes (Additional File 2: **Table S1**), but a smaller proportion of predicted genes have functional annotations, orthologous gene assignments, and predicted introns, and CE1 contains no BUSCO proteins (Additional File 1: **Fig. S6**). Of the 177 predicted genes on CE1, 128 are ORFans (i.e., do not hit any known proteins), 21 have best BLAST hits to bacteria, 22 to eukaryotes, four to archaea and one to viruses (when excluding the thraustochytrid *Hondaea fermentalgiana*; see below) (Additional File 2: **Table S7**).

VirSorter2 (Guo et al. 2021) and ViralRecall (Aylward and Moniruzzaman 2021) were initially used to identify both CE1 and the left end of Chr15 (LE-Chr15) as possible nucleocytoplasmic large DNA viruses

(NCLDV) of the *Nucleocytoviricota* (Additional File 1: **Text S2, Table S8, Table S9**). Subsequent detailed sequence similarity searches using specific virion proteins of various groups of viruses as queries (via blastp and HMMsearch) against the *A. limacinum* genome revealed more genes on CE1 and LE-Chr15 related to key genes of herpes-like viruses recently identified as '*Mirusviricota*' (Gaïa et al. 2023) than to core genes of *Nucleocytoviricota* (**Fig. 3A**; Additional File 2: **Table S7**; Additional File 1: **Text S2**). For example, we detected *Mirusviricota*-like major capsid protein (MCP) coding regions on CE1 and LE-Chr15 but no *Nucleocytoviricota*-like MCPs. A terminase-like homolog was also found to be shared between mirusviruses, CE1, and LE-Chr15 (the terminase protein packs the freshly synthesized genome into newly formed capsids). Divergent mirusvirus-like homologs were detected on both CE1 and LE-Chr15 for the remaining virion proteins as well (i.e., capsid maturation protease, portal protein, triplex 1 and triplex 2), as were other core mirusvirus genes, including a heliorhodopsin (on CE1), a histone H3 (CE1), TATA-binding protein (CE1 and LE-Chr15), subunits alpha and beta of ribonucleotide reductase (LE-Chr15), and additional proteins of unknown function (**Fig 3A;** Additional File 2: **Table S7;** Additional File 1: **Text S2**). We also detected core informational viral genes such as PolB (identified previously by Gallot-Lavallée and Blanc (2017), RNAPolB large subunit (RNAPL), and superfamily II helicase proteins (**Fig. 3A;** Additional File 1: **Table S7**). We found no sequence similarity between CE1 or LE-Chr15 and the lytic large DNA virus previously reported to infect the thraustochytrid *Sicyoidochytrium minutum* (SmDNAV) (Takao et al. 2007; Murakoshi et al. 2021) (Additional File 1: **Text S2**).

Some virus-like genes on CE1 and LE-Chr15 are found only in mirusviruses or herpesviruses (MCP, **Fig. 3D**, terminases, **Fig. S7A**). For genes with broader distribution, phylogenetic analyses also support relationships of several CE1 and LE-Chr15 viral genes to mirusviruses, as well as to nucleocytoviruses, which share several informational genes with miruviruses. The resolvase, helicase, and DNAP trees show *A. limacinum* viral sequences branching specifically with mirusviruses (**Fig. 3B**; Additional File 1: **Fig. S7B and Fig. S7C**), while the topoisomerase and nuclease trees show relatedness of our *A. limacinum* viral sequences to nucleocytoviruses (**Fig. 3C; Fig. S7D**). In contrast, the *A. limacinum* viral TATA-binding proteins group with archaeal sequences, rather than mirusviruses (**Fig. S7E).** Homologs of thraustochytrid viral and cellular arylsulfatase genes are detected only in various bacteria (**Fig. S7F**). The RNAPL genes of CE1 and LE-Chr15 are particularly unusual. As seen in some other viruses (e.g., *P. sibericum* (YP_009001268.1 and YP_009001052.1) and other pithoviruses, and cells (e.g., many archaea, (Langer et al. 1995)), the RNAPL coding region is split: the N- and C-terminal domains are encoded by separate ORFs located far apart from one another on both CE1 and LE-Chr15, and in *H. fermentalgiana*. The CE1, LE-Chr15, and *H. fermentalgiana* homologs branch robustly together in independent phylogenies of both RNAPL domains (**Fig. S7G and S7H**) but the precise evolutionary origin(s) of RNAPL in *A. limacinum* is unclear from the data in hand. On balance, these data suggest that CE1 and the viral-like element of Chr15 are most closely related to mirusviruses described in (Gaïa et al. 2023), but with genes derived from other sources as well .

The putative viral region at the left end of Chr15 (LE-Chr15) provides a nexus between the striking subtelomere structure and viral content of the *A. limacinum* genome. LE-Chr15 has rRNAs, LORE-TEARS, and telomeric repeats on one end and rRNAs and LORE-TEARs on the other (**Fig. 1, Fig. 2**); this is the only place in the assembly with internal (non-telomeric) arrays of rRNA genes and LORE-TEARs. The GC content of the putative viral integrant is 41.6%, slightly lower than the rest of the chromosome (45.0%), consistent with a foreign origin. The putative virus-like elements detected on CE1 and LE-Chr15 are related but distinct from one another (Additional File 1: **Fig. S8**). Comparing CE1's 177 predicted proteins to the 152 proteins encoded by the virus-like region of Chr15, only 48 are each other's reciprocal best BLAST hits (Additional File 2: **Table S7**), but many of their shared homologs branch together in phylogenetic trees (**Fig. 3B and D**; Additional File 1: **Fig. S7; Text S2**).

Mirusvirus virion particles have yet to be isolated. Our data show that *A. limacinum* is a probable natural host. CE1 could be an active viral genome capable of yielding viral particles: CE1 is circular and has an apparently complete virion module and full-length DNA polymerase, and viral particles consistent with the

presence of an endogenous herpes-like virus have previously been identified in thraustochytrids (Kazama and Schornstein 1972, 1973). It is noteworthy that CE1 encodes proteins with ParA (Aurli_135839) and Fic (Aurli_13050) domains, which have been associated with plasmid segregation (Łobocka and Gągała 2020): this may speak to how CE1 is maintained as an episomal element in *A. limacinum*. Endogenization of LE-Chr15 appears to have occurred via sub-telomeric recombination. The pace of discovery of endogenous viral elements is accelerating thanks to growth of genome-scale resources in diverse organisms (Feschotte and Gilbert 2012; Schulz, Abergel, and Woyke 2022); to our knowledge, the differences between CE1 and LE-Chr15 make this the first example of related 'stand-alone' and integrated virus-like elements recovered through routine eukaryotic genome sequencing. Interestingly, we found several close homologs of the *A. limacinum* virus-like sequences in the fragmented genome assembly of another thraustochytrid, *Hondaea fermentalgiania* (Dellero et al. 2018) (**Fig. 3,** Additional File 1: **Fig. S7**). This suggests that mirusvirus-like viruses have been associated with this protist lineage for some time.

Conclusions

Long-read sequencing has revealed that the genome of *A. limacinum* is dynamic and structurally innovative, an epic advance in light of their role as subjects first detailing chromosome counts in protists (Moens and Perkins 1969). We observed arrays of sub-telomeric rRNA and position-specific classes of long repeated elements (LORE-TEARS) on all chromosome ends, suggesting their role in chromosome maintenance. We also caught the 'superposition' of two complete sets of these elements surrounding an endogenized viral genome-like element at the end of Chr15. Our comparative genomic investigation reveals that CE1 and LE-Chr15 are specifically related to the recently discovered mirusviruses that are predicted to be "…among the most abundant and active eukaryotic viruses characterized in the sunlit oceans" (Gaïa et al. 2023). Mirusviruses have not, however, been linked to specific microbial eukaryotic hosts: Labyrinthulomycetes such as *A. limacinum* appear to be one such host. The functionality of both this integrated mirusvirus-like entity and of the circular, high-copy mirusvirus-like genome we identified are unclear, but both appear to correspond to (or be derived from) novel giant endogenous viral genomes, and one appears to maintain itself independently as a plasmid-like entity. It is noteworthy that Kazama and Schornstein described their thraustochytrid culture as 'virogenic' because they were never able to isolate strains free of the virus - even when establishing cultures from single zoospores - and because viral particles were only produced under permissive growth conditions (Kazama and Schornstein 1972, 1973). The combination of the union of highly conserved cellular elements (rRNAs), novel classes of repetitive elements (LORE-TEARs), and viral integration events at chromosome ends suggests new opportunities for future investigation of mechanisms of chromosome maintenance and nucleolus formation. There is still much eukaryotic diversity to be surveyed with long-read technology, and we will soon learn whether these features are unique to *A. limacinum* or a general feature of labyrinthulomycetes or broadly distributed among eukaryotic diversity.

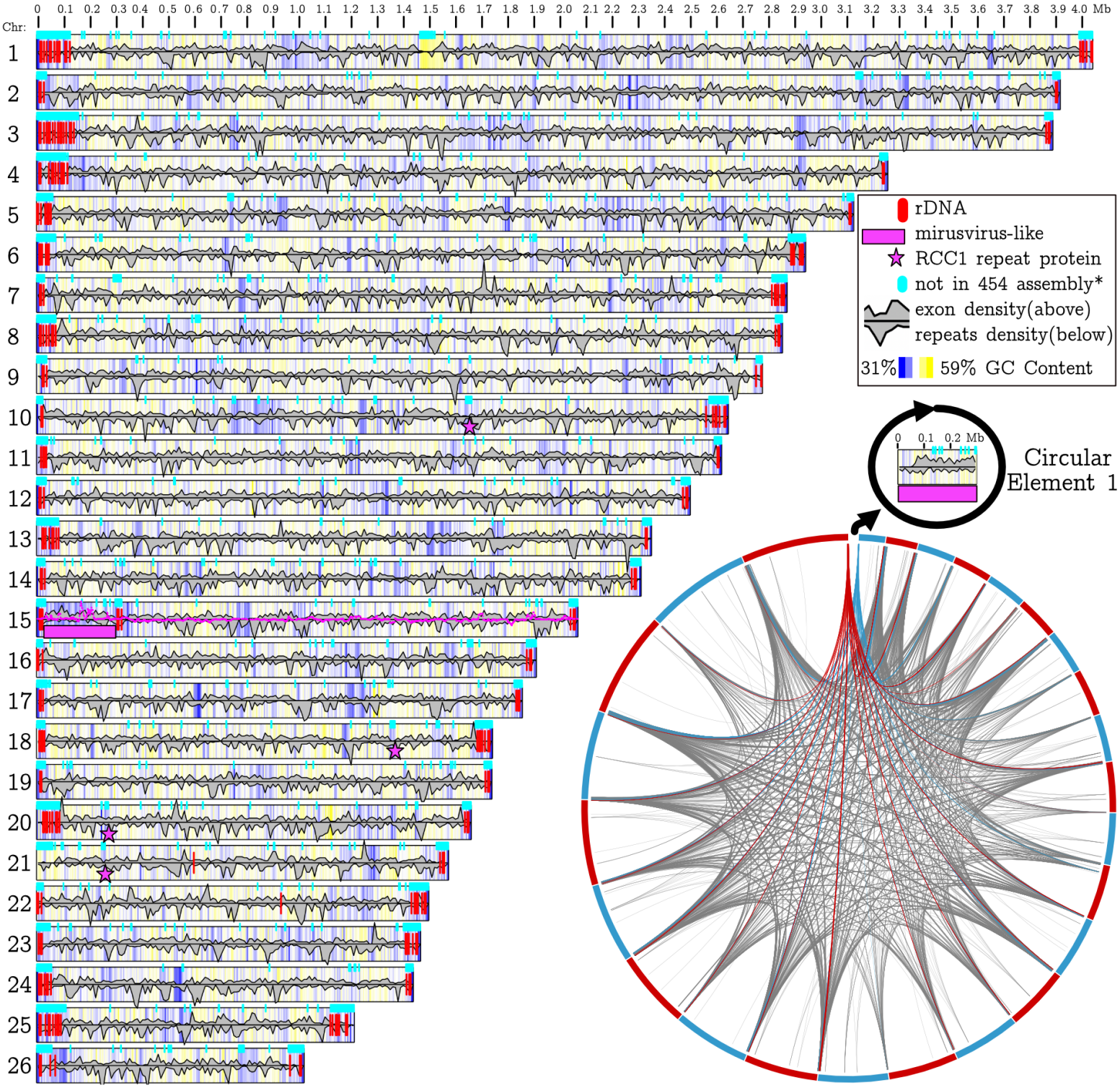


Fig. 1. Size and select features of the 26 putative linear chromosomes and a Circular Element in *Aurantiochytrium limacinum* ATCC MYA-1381. Circular Element 1 is predicted to be circular, but is displayed as linear. A scale in megabases is provided along the top of the plot (Gel and Serra 2017). Vertical red lines represent locations of predicted (Lagesen et al. 2007) rRNA gene regions, and are found almost exclusively at the ends of the linear chromosomes. Cyan boxes represent regions that did not align (Darling et al. 2004) with the primary 454 assembly (*some regions may be present in short 454 scaffolds; see Supplemental Fig. S2). GC content (5 Kbp windows) is indicated by background color, with darker shades of blue indicating regions of lower GC content and darker shades of yellow indicating regions with higher GC content; low GC content at the linear chromosome ends reflect telomere content. The gray density plot above the midline of each

chromosome indicates the relative density of exons, based on mapping of predicted exons from the JGI assembly to each Nanopore chromosome. The gray density plot below the midline (reflected so that higher values form valleys) indicates the relative density of repetitive sequences identified by RepeatMasker (Chen 2004). The magenta line (shown only for Chromosome 15) is a plot of the VirusRecall rolling score of NCLDV content (original range -17.6 to 3.9; negative and positive scores rescaled linearly above and below the axis, respectively). **Inset**: Chord plot showing matching sequence regions of at least 1 Kbp between contigs (Delehelle et al. 2018). An arbitrary set of chords are colored (arbitrarily red and blue) to highlight the directional nature of the repeats at the scaffold ends. A blank space in the chord plot represents the shortest scaffold, which has no matching sequence regions on other scaffolds.

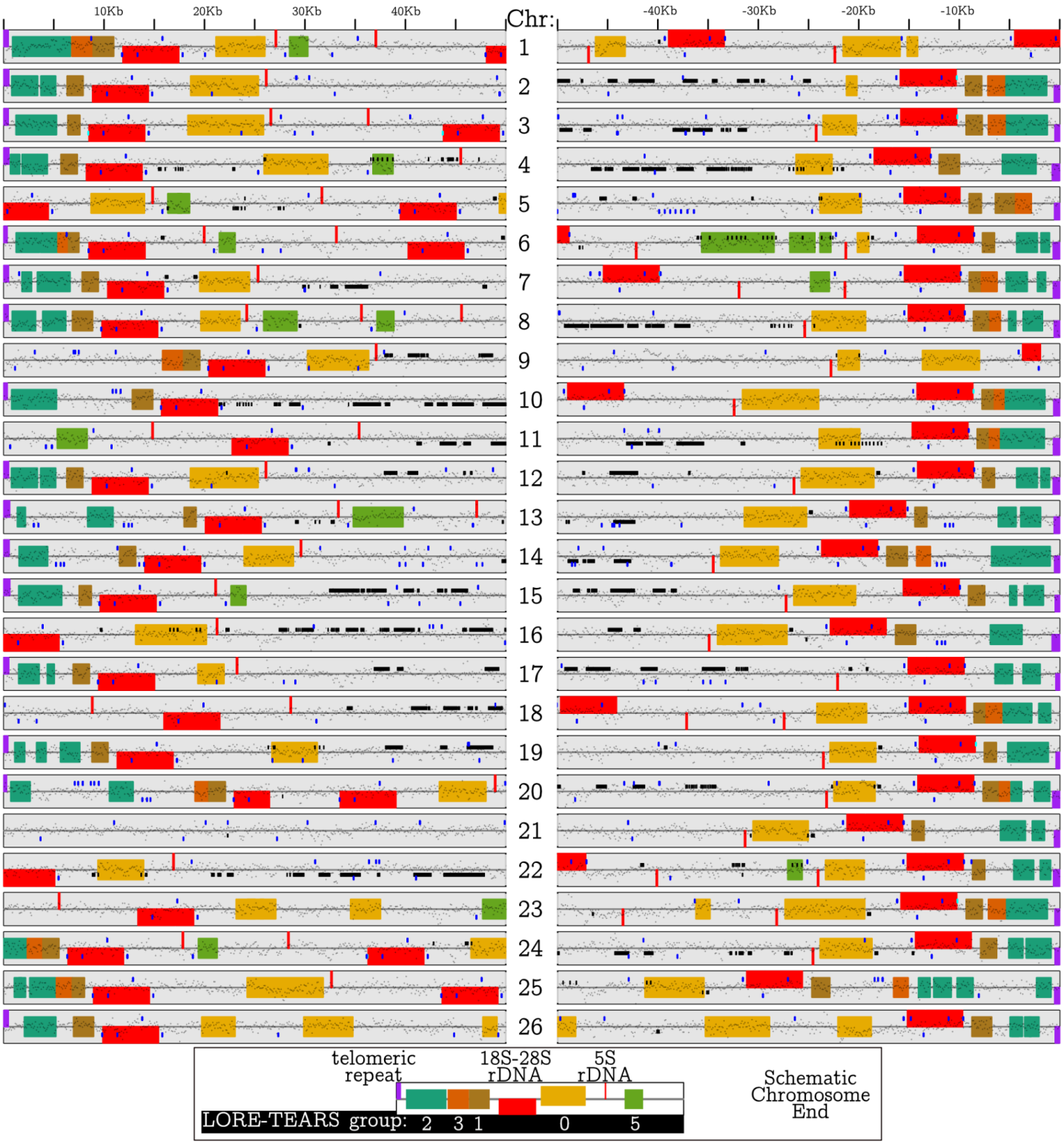
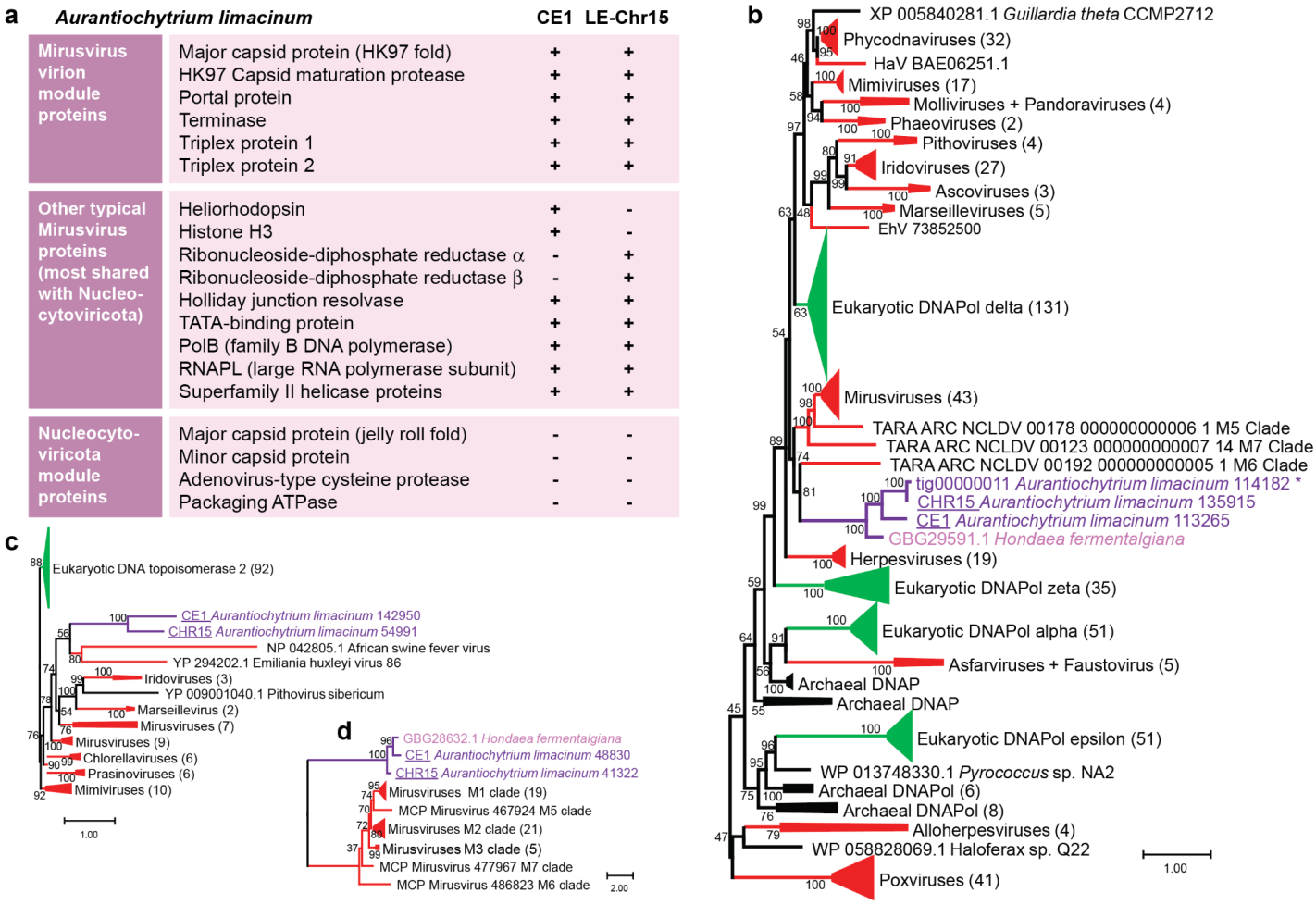


Fig. 2. Focused view of the 50 Kbp of both ends of the *Aurantiochytrium limacinum* chromosomes (Chr) from Figure 1. Locations of predicted rRNA genes (red) as in Fig. 1, with a large 18S-5.8S-28S rDNA cluster transcribed toward the telomere, and a small 5S rDNA transcribed away from the telomere. Purple boxes on ends of contigs indicate the presence of telomeric repeats in our assembly. Five classes of LORE-TEARS (0,1,2,3,5) are shown (see inset for key). G-quadruplexes are plotted as blue lines; note the regular positioning of G-quadruplexes within and around the rRNAs. Black lines are exons mapped using BLAST. GC content is plotted along the centerline. **Bottom Inset:** Schematic view of a typical arrangement of elements at the ends of a chromosome. LORE-TEARS elements are colored and labeled by their sequence similarity group.

286

a

	Aurantiochytrium limacinum	CE1	LE-Chr15
Mirusvirus virion module proteins	Major capsid protein (HK97 fold)	+	+
	HK97 Capsid maturation protease	+	+
	Portal protein	+	+
	Terminase	+	+
	Triplex protein 1	+	+
	Triplex protein 2	+	+
Other typical Mirusvirus proteins (most shared with Nucleo-cytoviricota)	Heliorhodopsin	+	-
	Histone H3	+	-
	Ribonucleoside-diphosphate reductase α	-	+
	Ribonucleoside-diphosphate reductase β	-	+
	Holliday junction resolvase	+	+
	TATA-binding protein	+	+
	PolB (family B DNA polymerase)	+	+
	RNAPL (large RNA polymerase subunit)	+	+
	Superfamily II helicase proteins	+	+
Nucleocyto-viricota module proteins	Major capsid protein (jelly roll fold)	-	-
	Minor capsid protein	-	-
	Adenovirus-type cysteine protease	-	-
	Packaging ATPase	-	-



287

288 **Fig. 3:** Viral content detected in *Aurantiochytrium limacinum* predicted proteins. **A**). Presence (+) and absence
289 (-) of select viral proteins on CE1 and LE-Chr15 relative to key *Mirusviricota* and *Nucleocytoviricota* (NCLDV)
290 proteins. Note that the RNAPL coding region is split into two discrete ORFs, as in pithoviruses and many
291 archaea. **B**) Maximum likelihood phylogenetic tree of virus-like family B DNA polymerases (DNAPol) proteins
292 encoded on CE1 and LE-Chr15 of *A. limacinum* (purple) and homologs in *Hondaea fermentalgiana* (pink). The
293 sequences were aligned with MAFFT, and sites with less than 20% gaps were retained for phylogenetic
294 reconstruction. Note that in addition to the homologs found in CE1 and LE-Chr15, a viral-like DNAPol is also
295 found in *A. limacinum* tig00000011; it shows signs of pseudogenization. **C)** Phylogeny of viral DNA
296 topoisomerases rooted with eukaryotic homologs. Sequences were aligned with MAFFT-linsi prior to
297 phylogenetic reconstruction. **D)** Phylogenetic tree of Mirusvirus major capsid proteins and their homologs in *A.*
298 *limacinum* and *H. fermentalgiana*. The sequences were aligned with MAFFT-linsi and sites with less than 30%
299 gaps were retained for phylogenetic reconstruction. Viral sequences are in red, eukaryotic homologs are in
300 green, and bacterial/archaeal sequences are in black. Scale bars indicate inferred number of amino acid
301 substitutions per site.

302

303

ONLINE METHODS

Strain cultivation and nucleic acid preparation

Aurantiochytrium (formerly *Schizochytrium*) *limacinum* Honda et Yokochi ATCC MYA-1381 (also designated NIBH SR21 or IFO 32693, GenBank Accession AB022107; (Honda et al. 1998; Yokoyama and Honda 2007) was isolated from seawater in a mangrove area of the Yap Islands, Micronesia. For sequencing at JGI, *Aurantiochytrium* ATCC MYA-1381 cultures were grown in 2 liters of ATCC 790 By+ medium (5 g glucose, 1 g yeast extract, 1 g peptone, 30 g Instant ocean per liter) distributed in four large tissue culture flasks (500 ml each) at room temperature without shaking. Cultures were harvested after 7 days, producing 2.4 g wet weight. Genomic DNA was extracted from 0.507 g wet biomass and RNA from 1.057 g wet biomass following the protocols of (Lippmeier et al. 2009) and subject to JGI QA/QC protocols. For Nanopore sequencing, *Aurantiochytrium* ATCC MYA-1381 was cultured for three days in 50 ml ATCC 790 By+ medium. Genomic DNA was extracted based on a previously published protocol (https://dx.doi.org/10.17504/protocols.io.n83dhyn). The precipitated DNA was left to dissolve in water by spontaneous diffusion for 48+ hours at room temperature to avoid shearing and subsequently purified using QIAGEN Genomic-tip 20/G. Agarose gel electrophoresis (1%) was used to visually assess and confirm the integrity of high molecular weight (20+ Kb) DNA. DNA quality was evaluated using a NanoPhotometer P360 (Implen) to measure A260/280 (~1.8) and A260/230 (2.0-2.2) ratios. The quantity of DNA was calculated using a Qubit 2.0 Fluorometer (ThermoFisher Scientific) with the dsDNA broad range assay kit.

Sequencing and assembly

Short-read sequencing was performed by the Joint Genome Institute on the 454 sequencing platform, and assembly was accomplished with Newbler followed by annotation with the JGI Annotation Pipeline; details are provided in Additional File 1: **Text S1**. Draft genome sequence for *Aurantiochytrium limacinum* ATCC MYA-1381 is available via the PhycoCosm Genome Portal, (https://mycocosm.jgi.doe.gov/Aurli1/Aurli1.info.html).

Long-read sequencing was performed using the Oxford Nanopore Technology (ONT) MinION sequencing platform and assembly was accomplished with Canu; details are provided in Additional File 1: **Text S1**. The raw fast5 MinION data has been deposited in the NCBI SRA database BioProject PRJNA680238 (WT accession: SRR13108467; KO32 accession: SRR13108466; KO33 accession: SRR13108465).

rRNA, G-quadruplex, and repetitive element predictions

rRNA gene locations were predicted using RNAmmer 1.2 (Lagesen et al. 2007). Tandem Repeat Finder 4.09 (Benson 1999) was used to identify tandem repeats (TR) with maximum repeat unit set as large as possible (2000 bp). The vast majority of the 29640 identified TRs were a few bp in length, and the number of TRs declined with TR length until ~350 bp, where a peak appeared. The sequences of the 673 TR elements longer than 299 bp were dereplicated by removing overlapping TRs, keeping the shorter repeat unit and clustering with cd-hit. Telomeric repeats in the Nanopore assembly were identified in this output as repeats with unit 5, 10, or 15 matching the motif TTAGG (or CCTAA). More than 98% of TR elements were shorter than 200bp. 490 Tandem repeat (TR) units longer than 200bp were dereplicated (retaining one TR to represent each locus) and 398 subjected to clustering by cd-hit-est (Huang et al. 2010) with default parameters except sequence identity cutoff 0.8 and -r yes. Manual examination and alignment of the resulting 25 clusters (which excluded 34 singletons) revealed 4 types of TRs grouped by location relative to rRNA genes and telomeric repeats. Additional repetitive content was identified with RepeatMasker (Chen 2004). G-quadruplexes were

predicted with G4-iM Grinder (Belmonte-Reche and Morales 2020) using Methods 2 and 3 and filtering for scores greater than 20. Genomic features were visualized using karyoploteR (Gel and Serra 2017).

Viral gene predictions and phylogenetics

Virion proteins from various groups of dsDNA viruses, including mirusviruses and nucleocytoviruses, were used to screen the *A. limacinum* assembly using blastp and HMMsearches. HMMs were generated from alignments in Gaïa *et al.* (2023) (Additional File 3). Following the detection of mirusvirus-like structural proteins in CE1 and LE-Chr15, all mirusvirus ORFs (>=90 amino acids; predicted from the mirusvirus contigs from (Gaïa et al. 2023)) were used as blastp queries to detect additional mirusvirus homologs. In parallel, ViralRecall (Aylward and Moniruzzaman 2021) and VirSorter2 (Guo et al. 2021) were used to evaluate viral gene content in both the 454 and Nanopore *A. limacinum* assemblies. The results of similarity searches against nr were then used to characterize additional CE1 and LE-Chr15 proteins.

Unless specified otherwise, *A. limacinum* virus-like proteins were aligned with homologs in diverse viruses, prokaryotes and eukaryotes using MAFFT, with BMGE (default parameters) used to perform site selection. Maximum likelihood phylogenetic trees were constructed with IQTree (Nguyen et al. 2015) v1.6.3 model C60+G4 with 1000 ultrafast bootstraps replicates.

Availability of data and materials

The short-read (454) data, assembly, and annotation are publicly available via the JGI genome portal in PhycoCosm (https://phycocosm.jgi.doe.gov/). The raw fast5 MinION data has been deposited in the NCBI SRA database BioProject PRJNA680238 (WT accession: SRR13108467; KO32 accession: SRR13108466; KO33 accession: SRR13108465).

Additional File 1 contains supplemental text, tables, and figures.
Additional File 2 contains supplemental tables S1, S2, S6 and S7.
Additional File 3: The Canu nanopore assembly, annotations, HMM files, alignments, and trees, are at doi:10.5061/dryad.2fqz612t6: https://datadryad.org/stash/share/QYfnoRY6ZF9b-xA1lCXvm7xfAjmeqY-A5b-nSv15YNI

Competing interests

The authors declare that they have no competing interests.

Funding

The work (proposal:10.46936/10.25585/60007256) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. The construction and analysis of the *crtIBY* mutants in the Collier and Rest labs was supported by grant from the Gordon and Betty Moore Foundation (GBMF4982). Research in the Archibald Lab, including Oxford Nanopore long-read sequencing, was supported by a grant from the Gordon and Betty Moore Foundation (GBMF5782) and a Discovery Grant (RGPIN-2019-05058) from the Natural Sciences and Engineering Research Council of Canada.

Authors' contributions

JC and JR conceived the study, performed the analyses and drafted the manuscript. EL built the combined Canu assembly and performed initial comparative analyses. CD, JJ, JP, CP, AK, IVG performed the short-read sequencing, assembly, and annotation. GF and AMGNV carried out Nanopore long-read sequencing and performed the PFGE. JC, JR and LG-L performed detailed comparative genomic analyses of virus-like sequences, and LG-L and JMA performed and interpreted phylogenetic reconstructions. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Bruce A. Curtis for technical support with Nanopore sequencing and feedback on the final manuscript.

References

Aylward, Frank O., and Mohammad Moniruzzaman. 2021. “ViralRecall-A Flexible Command-Line Tool for the Detection of Giant Virus Signatures in ’Omic Data.” *Viruses* 13 (2): 150.

Belmonte-Reche, Efres, and Juan Carlos Morales. 2020. “G4-iM Grinder: When Size and Frequency Matter. G-Quadruplex, I-Motif and Higher Order Structure Search and Analysis Tool.” *NAR Genomics and Bioinformatics* 2 (1): lqz005.

Benson, G. 1999. “Tandem Repeats Finder: A Program to Analyze DNA Sequences.” *Nucleic Acids Research* 27 (2): 573–80.

Biffi, Giulia, David Tannahill, and Shankar Balasubramanian. 2012. “An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2.” *Journal of the American Chemical Society* 134 (29): 11974–76.

Branton, Daniel, David W. Deamer, Andre Marziali, Hagan Bayley, Steven A. Benner, Thomas Butler, Massimiliano Di Ventra, et al. 2008. “The Potential and Challenges of Nanopore Sequencing.” *Nature Biotechnology* 26 (10): 1146–53.

Brugère, J. F., E. Cornillot, G. Méténier, A. Bensimon, and C. P. Vivarès. 2000. “Encephalitozoon Cuniculi (Microspora) Genome: Physical Map and Evidence for Telomere-Associated rDNA Units on All Chromosomes.” *Nucleic Acids Research* 28 (10): 2026–33.

Chen, Nansheng. 2004. “Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences.” *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* Chapter 4 (1): Unit 4.10.

Collier, Jackie L., and Joshua S. Rest. 2019. “Swimming, Gliding, and Rolling toward the Mainstream: Cell Biology of Marine Protists.” *Molecular Biology of the Cell* 30 (11): 1245–48.

Criniti, Angela, Gabriele Simonazzi, Stefano Cassanelli, Mario Ferrari, Davide Bizzaro, and Gian Carlo Manicardi. 2009. “Distribution of Heterochromatin and rDNA on the Holocentric Chromosomes of the Aphids Dysaphis Plantaginea and Melanaphis Pyraria (Hemiptera: Aphididae).” *European Journal of Entomology* 106 (2): 153–57.

Darling, Aaron C. E., Bob Mau, Frederick R. Blattner, and Nicole T. Perna. 2004. “Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements.” *Genome Research* 14 (7): 1394–1403.

Delehelle, Franklin, Sylvain Cussat-Blanc, Jean-Marc Alliot, Hervé Luga, and Patricia Balaresque. 2018. “ASGART: Fast and Parallel Genome Scale Segmental Duplications Mapping.” *Bioinformatics* 34 (16): 2708–14.

Dellero, Younès, Olivier Cagnac, Suzanne Rose, Khawla Seddiki, Mathilde Cussac, Christian Morabito, Josselin Lupette, et al. 2018. “Proposal of a New Thraustochytrid Genus Hondaea Gen. Nov. and Comparison of Its Lipid Dynamics with the Closely Related Pseudo-Cryptic Genus Aurantiochytrium.” *Algal Research* 35 (November): 125–41.

Diner, Rachel E., Chari M. Noddings, Nathan C. Lian, Anthony K. Kang, Jeffrey B. McQuaid, Jelena Jablanovic, Josh L. Espinoza, et al. 2017. “Diatom Centromeres Suggest a Mechanism for Nuclear DNA

Acquisition." *Proceedings of the National Academy of Sciences of the United States of America* 114 (29): E6015–24.

Douglas, S., S. Zauner, M. Fraunholz, M. Beaton, S. Penny, L. T. Deng, X. Wu, M. Reith, T. Cavalier-Smith, and U. G. Maier. 2001. "The Highly Reduced Genome of an Enslaved Algal Nucleus." *Nature* 410 (6832): 1091–96.

Dvořáčková, Martina, Miloslava Fojtová, and Jiří Fajkus. 2015. "Chromatin Dynamics of Plant Telomeres and Ribosomal Genes." *The Plant Journal: For Cell and Molecular Biology* 83 (1): 18–37.

Fang, Yufeng, Marco A. Coelho, Haidong Shu, Klaas Schotanus, Bhagya C. Thimmappa, Vikas Yadav, Han Chen, et al. 2020. "Long Transposon-Rich Centromeres in an Oomycete Reveal Divergence of Centromere Features in Stramenopila-Alveolata-Rhizaria Lineages." *PLoS Genetics* 16 (3): e1008646.

Feschotte, Cédric, and Clément Gilbert. 2012. "Endogenous Viruses: Insights into Viral Evolution and Impact on Host Biology." *Nature Reviews. Genetics* 13 (4): 283–96.

Filloramo, Gina V., Bruce A. Curtis, Emma Blanche, and John M. Archibald. 2021. "Re-Examination of Two Diatom Reference Genomes Using Long-Read Sequencing." *BMC Genomics* 22 (1): 379.

Fletcher, Kyle, Juliana Gil, Lien D. Bertier, Aubrey Kenefick, Kelsey J. Wood, Lin Zhang, Sebastian Reyes-Chin-Wo, et al. 2019. "Genomic Signatures of Heterokaryosis in the Oomycete Pathogen Bremia Lactucae." *Nature Communications* 10 (1): 2645.

Fossier Marchan, Loris, Kim J. Lee Chang, Peter D. Nichols, Wilfrid J. Mitchell, Jane L. Polglase, and Tony Gutierrez. 2018. "Taxonomy, Ecology and Biotechnological Applications of Thraustochytrids: A Review." *Biotechnology Advances* 36 (1): 26–46.

Gaïa, Morgan, Lingjie Meng, Eric Pelletier, Patrick Forterre, Chiara Vanni, Antonio Fernandez-Guerra, Olivier Jaillon, et al. 2023. "Mirusviruses Link Herpesviruses to Giant Viruses." *Nature*, April. https://doi.org/10.1038/s41586-023-05962-4.

Gallot-Lavallée, Lucie, and Guillaume Blanc. 2017. "A Glimpse of Nucleo-Cytoplasmic Large DNA Virus Biodiversity through the Eukaryotic Genomics Window." *Viruses* 9 (1). https://doi.org/10.3390/v9010017.

Gardner, Malcolm J., Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W. Hyman, Jane M. Carlton, et al. 2002. "Genome Sequence of the Human Malaria Parasite Plasmodium Falciparum." *Nature* 419 (6906): 498–511.

Gel, Bernat, and Eduard Serra. 2017. "karyoploteR: An R/Bioconductor Package to Plot Customizable Genomes Displaying Arbitrary Data." *Bioinformatics* 33 (19): 3088–90.

Guérin, Nina, Marta Ciccarella, Elisa Flamant, Paul Frémont, Sophie Mangenot, Benjamin Istace, Benjamin Noel, et al. 2021. "Genomic Adaptation of the Picoeukaryote Pelagomonas Calceolata to Iron-Poor Oceans Revealed by a Chromosome-Scale Genome Sequence." *bioRxiv*. https://doi.org/10.1101/2021.10.25.465678.

Guo, Jiarong, Ben Bolduc, Ahmed A. Zayed, Arvind Varsani, Guillermo Dominguez-Huerta, Tom O. Delmont, Akbar Adjie Pratama, et al. 2021. "VirSorter2: A Multi-Classifier, Expert-Guided Approach to Detect Diverse DNA and RNA Viruses." *Microbiome* 9 (1): 37.

Honda, Daiske, Toshihiro Yokochi, Toro Nakahara, Mayumi Erata, and Takanori Higashihara. 1998. "Schizochytrium Limacinum Sp. Nov., a New Thraustochytrid from a Mangrove Area in the West Pacific Ocean." *Mycological Research* 102 (4): 439–48.

Huang, Ying, Beifang Niu, Ying Gao, Limin Fu, and Weizhong Li. 2010. "CD-HIT Suite: A Web Server for Clustering and Comparing Biological Sequences." *Bioinformatics* 26 (5): 680–82.

Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. "Erratum to: The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community." *Genome Biology* 17 (1): 256.

Jensen-Seaman, Michael I., Terrence S. Furey, Bret A. Payseur, Yontao Lu, Krishna M. Roskin, Chin-Fu Chen, Michael A. Thomas, David Haussler, and Howard J. Jacob. 2004. "Comparative Recombination Rates in the Rat, Mouse, and Human Genomes." *Genome Research* 14 (4): 528–38.

Juranek, Stefan A., and Katrin Paeschke. 2012. "Cell Cycle Regulation of G-Quadruplex DNA Structures at Telomeres." *Current Pharmaceutical Design* 18 (14): 1867–72.

Kawai, Hiroshi, Hideki Muto, Tetsuya Fujii, and Atsushi Kato. 1995. "A LINKED 5S rRNA GENE IN SCYTOSIPHON LOMENTARIA (SCYTOSIPHONALES, PHAEOPHYCEAE)1." *Journal of Phycology* 31 (2): 306–11.

Kawai, Hiroshi, Takeshi Nakayama, Isao Inouye, and Atsushi Kato. 1997. "LINKAGE OF 5S RIBOSOMAL DNA TO OTHER rDNAS IN THE CHROMOPHYTIC ALGAE AND RELATED TAXA1." *Journal of Phycology* 33 (3): 505–11.

Kazama, F. Y., and K. L. Schornstein. 1972. “Herpes-Type Virus Particles Associated with a Fungus.” *Science* 177 (4050): 696–97.

———. 1973. “Ultrastructure of a Fungus Herpes-Type Virus.” *Virology* 52 (2): 478–87.

Kim, Jong Im, Goro Tanifuji, Minseok Jeong, Woongghi Shin, and John M. Archibald. 2022. “Gene Loss, Pseudogenization, and Independent Genome Reduction in Non-Photosynthetic Species of Cryptomonas (Cryptophyceae) Revealed by Comparative Nucleomorph Genomics.” *BMC Biology* 20 (1): 227.

Kobayashi, Takehiko. 2011. “Regulation of Ribosomal RNA Gene Copy Number and Its Role in Modulating Genome Integrity and Evolutionary Adaptability in Yeast.” *Cellular and Molecular Life Sciences: CMLS* 68 (8): 1395–1403.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36.

Lagesen, Karin, Peter Hallin, Einar Andreas Rødland, Hans-Henrik Staerfeldt, Torbjørn Rognes, and David W. Ussery. 2007. “RNAmmer: Consistent and Rapid Annotation of Ribosomal RNA Genes.” *Nucleic Acids Research* 35 (9): 3100–3108.

Langer, D., J. Hain, P. Thuriaux, and W. Zillig. 1995. “Transcription in Archaea: Similarity to That in Eucarya.” *Proceedings of the National Academy of Sciences of the United States of America* 92 (13): 5768–72.

Lippmeier, J. Casey, J. Casey Lippmeier, Kristine S. Crawford, Carole B. Owen, Angie A. Rivas, James G. Metz, and Kirk E. Apt. 2009. “Characterization of Both Polyunsaturated Fatty Acid Biosynthetic Pathways in *Schizochytrium* Sp.” *Lipids*. https://doi.org/10.1007/s11745-009-3311-9.

Liu, Li, Qi-Fan Yang, Wu-Shan Dong, Yan-Hui Bi, and Zhi-Gang Zhou. 2017. “Characterization and Physical Mapping of Nuclear Ribosomal RNA (rRNA) Genes in the Haploid Gametophytes of Saccharina Japonica (Phaeophyta).” *Journal of Applied Phycology* 29 (5): 2695–2706.

Liu, Tao, Andreas Rechtsteiner, Thea A. Egelhofer, Anne Vielle, Isabel Latorre, Ming-Sin Cheung, Sevinc Ercan, et al. 2011. “Broad Chromosomal Domains of Histone Modification Patterns in *C. Elegans*.” *Genome Research*. https://doi.org/10.1101/gr.115519.110.

Łobocka, Małgorzata, and Urszula Gągała. 2020. “Prophage P1: An Example of a Prophage That Is Maintained as a Plasmid.” In *Bacteriophages: Biology, Technology, Therapy*, edited by David R. Harper, Stephen T. Abedon, Benjamin H. Burrowes, and Malcolm L. McConville, 1–13. Cham: Springer International Publishing.

Maruyama, Shinichiro, Osami Misumi, Yasuyuki Ishii, Shuichi Asakawa, Atsushi Shimizu, Takashi Sasaki, Motomichi Matsuzaki, et al. 2004. “The Minimal Eukaryotic Ribosomal DNA Units in the Primitive Red Alga Cyanidioschyzon Merolae.” *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 11 (2): 83–91.

Mascarenhas Dos Santos, Anne Caroline, Alexander Thomas Julian, Pingdong Liang, Oscar Juárez, and Jean-François Pombert. 2023. “Telomere-to-Telomere Genome Assemblies of Human-Infecting Encephalitozoon Species.” *BMC Genomics* 24 (1): 237.

Matsuzaki, Motomichi, Osami Misumi, Tadasu Shin-I, Shinichiro Maruyama, Manabu Takahara, Shin-Ya Miyagishima, Toshiyuki Mori, et al. 2004. “Genome Sequence of the Ultrasmall Unicellular Red Alga Cyanidioschyzon Merolae 10D.” *Nature* 428 (6983): 653–57.

McKim, K. S., A. M. Howell, and A. M. Rose. 1988. “The Effects of Translocations on Recombination Frequency in Caenorhabditis Elegans.” *Genetics* 120 (4): 987–1001.

Moens, P. B., and F. O. Perkins. 1969. “Chromosome Number of a Small Protist: Accurate Determination.” *Science* 166 (3910): 1289–91.

Moniruzzaman, Mohammad, Alaina R. Weinheimer, Carolina A. Martinez-Gutierrez, and Frank O. Aylward. 2020. “Widespread Endogenization of Giant Viruses Shapes Genomes of Green Algae.” *Nature* 588 (7836): 141–45.

Paeschke, Katrin, Stefan Juranek, Tomas Simonsson, Anne Hempel, Daniela Rhodes, and Hans Joachim Lipps. 2008. “Telomerase Recruitment by the Telomere End Binding Protein-β Facilitates G-Quadruplex DNA Unfolding in Ciliates.” *Nature Structural & Molecular Biology* 15 (6): 598–604.

Perry, J., and A. Ashworth. 1999. “Evolutionary Rate of a Gene Affected by Chromosomal Position.” *Current Biology: CB* 9 (17): 987–89.

Rhoads, Anthony, and Kin Fai Au. 2015. “PacBio Sequencing and Its Applications.” *Genomics, Proteomics & Bioinformatics* 13 (5): 278–89.

Rius, Mariana, Joshua S. Rest, Gina V. Filloramo, Anna M. G. Novák Vanclová, John M. Archibald, and Jackie

L. Collier. 2023. “Horizontal Gene Transfer and Fusion Spread Carotenogenesis among Diverse Heterotrophic Protists.” *Genome Biology and Evolution*, February. https://doi.org/10.1093/gbe/evad029.

Roa, Fernando, and Marcelo Guerra. 2012. “Distribution of 45S rDNA Sites in Chromosomes of Plants: Structural and Evolutionary Implications.” *BMC Evolutionary Biology* 12 (November): 225.

Scherf, A., L. M. Figueiredo, and L. H. Freitas-Junior. 2001. “Plasmodium Telomeres: A Pathogen’s Perspective.” *Current Opinion in Microbiology* 4 (4): 409–14.

Schulz, Frederik, Chantal Abergel, and Tanja Woyke. 2022. “Giant Virus Biology and Diversity in the Era of Genome-Resolved Metagenomics.” *Nature Reviews. Microbiology* 20 (12): 721–36.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” *Bioinformatics* 31 (19): 3210–12.

Suzuki, Shigekatsu, Shu Shirato, Yoshihisa Hirakawa, and Ken-Ichiro Ishida. 2015. “Nucleomorph Genome Sequences of Two Chlorarachniophytes, Amorphochlora Amoebiformis and Lotharella Vacuolata.” *Genome Biology and Evolution* 7 (6): 1533–45.

Tashiro, Sanki, Yuki Nishihara, Kazuto Kugou, Kunihiro Ohta, and Junko Kanoh. 2017. “Subtelomeres Constitute a Safeguard for Gene Expression and Chromosome Homeostasis.” *Nucleic Acids Research* 45 (18): 10333–49.

Torres-Machorro, Ana Lilia, Roberto Hernández, John F. Alderete, and Imelda López-Villaseñor. 2009. “Comparative Analyses among the Trichomonas Vaginalis, Trichomonas Tenax, and Tritrichomonas Foetus 5S Ribosomal RNA Genes.” *Current Genetics* 55 (2): 199–210.

Torres-Machorro, Ana Lilia, Roberto Hernández, Ana María Cevallos, and Imelda López-Villaseñor. 2010. “Ribosomal RNA Genes in Eukaryotic Microorganisms: Witnesses of Phylogeny?” *FEMS Microbiology Reviews* 34 (1): 59–86.

Tůmová, Pavla, Magdalena Uzlíková, Gerhard Wanner, and Eva Nohýnková. 2015. “Structural Organization of Very Small Chromosomes: Study on a Single-Celled Evolutionary Distant Eukaryote Giardia Intestinalis.” *Chromosoma* 124 (1): 81–94.

Wang, Feng, Ming-Liang Tang, Zhi-Xiong Zeng, Ren-Yi Wu, Yong Xue, Yu-Hua Hao, Dai-Wen Pang, Yong Zhao, and Zheng Tan. 2012. “Telomere- and Telomerase-Interacting Protein That Unfolds Telomere G-Quadruplex and Promotes Telomere Extension in Mammalian Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (50): 20413–18.

Xu, Feifei, Aaron Jex, and Staffan G. Svärd. 2020. “A Chromosome-Scale Reference Genome for Giardia Intestinalis WB.” *Scientific Data* 7 (1): 38.

Xu, Xiaodan, Changyi Huang, Zhexian Xu, Huixia Xu, Zhao Wang, and Xinjun Yu. 2020. “The Strategies to Reduce Cost and Improve Productivity in DHA Production by Aurantiochytrium Sp.: From Biochemical to Genetic Respects.” *Applied Microbiology and Biotechnology* 104 (22): 9433–47.

Yokoyama, Rinka, and Daiske Honda. 2007. “Taxonomic Rearrangement of the Genus Schizochytrium Sensu Lato Based on Morphology, Chemotaxonomic Characteristics, and 18S rRNA Gene Phylogeny (Thraustochytriaceae, Labyrinthulomycetes): Emendation for Schizochytrium and Erection of Aurantiochytrium and Oblongichytrium Gen. Nov.” *Mycoscience* 48 (4): 199–211.