A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes

Jana Fulnečková[1], Tereza Ševčíková[1,2,3], Jiří Fajkus[1,2], Alena Lukešová[4], Martin Lukeš[5], Čestmír Vlček[6], B. Franz Lang[7], Eunsoo Kim[8], Marek Eliáš[3,9] and Eva Sýkorová[1,2,*]

[1] Institute of Biophysics, Academy of Sciences of the Czech Republic, 612 65 Brno, Czech Republic

[2] Faculty of Science, and CEITEC – Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

[3] University of Ostrava, Faculty of Science, Department of Biology and Ecology, Chittussiho 10, 710 00 Ostrava, Czech Republic

[4] Institute of Soil Biology, Biology Centre Academy of Sciences of the Czech Republic, Na Sádkách 7, 370 05 České Budějovice, Czech Republic

[5] Laboratory of Photosynthesis, Institute of Microbiology, Academy of Sciences of the Czech Republic, Opatovický mlýn, Třeboň, CZ-379 81, Czech Republic

[6] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Prague, Czech Republic

[7] Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, Québec, Canada

[8] Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY, 10024, USA

[9] Charles University in Prague, Faculty of Science, Department of Botany, Benátská 2, 128 01 Praha 2, Czech Republic

*Author for Correspondence: Eva Sýkorová, Institute of Biophysics, Academy of Sciences of the Czech Republic, v.v.i., Královopolská 135, CZ-61265 Brno, Czech Republic, Phone: +420-541517199 FAX: +420-541211293, E-mail: evin@ibp.cz

Abstract

Telomeres, ubiquitous and essential structures of eukaryotic chromosomes, are known to come in a variety of forms, but knowledge about their actual diversity and evolution across the whole phylogenetic breadth of the eukaryotic life remains fragmentary. To fill this gap we employed a complex experimental approach to probe telomeric minisatellites in various phylogenetically diverse groups of algae. Our most remarkable results include the following findings: (1) algae of the streptophyte class Klebsormidiophyceae possess the *Chlamydomonas*-type telomeric repeat (TTTTAGGG) or, in at least one species, a novel TTTTAGG repeat, indicating an evolutionary transition from the *Arabidopsis*-type repeat (TTTAGGG) ancestral for Chloroplastida; (2) the *Arabidopsis*-type repeat is also present in telomeres of Xanthophyceae, in contrast to the presence of the human-type repeat (TTAGGG) in other ochrophytes studied, and of the photosynthetic alveolate *Chromera velia*, consistent with its phylogenetic position close to apicomplexans and dinoflagellates; (3) glaucophytes and haptophytes exhibit the human-type repeat in their telomeres; (4) ulvophytes and rhodophytes have unusual telomere structures recalcitrant to standard analysis. To obtain additional details on the distribution of different telomere types in eukaryotes, we performed *in silico* analyses of genomic data from major eukaryotic lineages, utilizing also genome assemblies from our on-going genome projects for representatives of three hitherto unsampled lineages (jakobids, malawimonads, and goniomonads). These analyses confirm the human-type repeat as the most common and possibly ancestral in eukaryotes, but alternative motifs replaced it along the phylogeny of diverse eukaryotic lineages, some of them several times independently.

Abbreviations: rDNA – ribosomal DNA; LSU – large subunit; PCR – polymerase chain reaction; PEG – polyethylene glycol; SDS – sodium dodecylsulphate; SRA - Sequence Read Archive; SSC – saline sodium citrate; SSU – small subunit; TAE – Tris-acetate-EDTA; TRAP – telomere repeat amplification protocol; TRF – terminal restriction fragment; WGS – whole genome sequence

Introduction

The termini of linear eukaryotic chromosomes are protected by telomeres. Their DNA part is formed typically by a long array of conserved minisatellite sequences that tend to be conserved in particular groups of organisms, e.g. TTAGGG in vertebrates and fungi (named here as the human-type; Meyne et al. 1989), TTTAGGG in most plants (*Arabidopsis*-type; Richards and Ausubel 1988), or TTAGG in insects (Frydrychova et al. 2004; Okazaki et al. 1993; Vitkova et al. 2005). These telomeres are maintained by a special reverse transcriptase, telomerase, which elongates telomeres by addition of telomeric repeats and thus solves the so-called end replication problem (reviewed in Chan and Blackburn 2004). However, many exceptions to these rules are known, e.g. telomeres maintained by retrotransposons in *Drosophila melanogaster* (Biessmann and Mason 2003), diverse telomeric minisatellite sequences in yeasts (Teixeira and Gilson 2005), plants with the human-type or unknown telomeric sequences (Sýkorová et al. 2006a; Sýkorová et al. 2003a; Sýkorová et al. 2003b), or novel telomeric sequence in Arthropoda (Mravinac et al. 2011; Vitkova et al. 2005). An interesting diversity of telomeres was recently described also in the green algal group Chlamydomonadales, where at least two evolutionary transitions from the ancestral TTTAGGG type to the TTTTAGGG (*Chlamydomonas*-type) occurred in the clade Chloromonadinia and independently in a subclade of the Reinhardtinia clade; moreover, the human-type telomeric sequence was found in some green algal species that fall within the Dunallielinia and Stephanosphaeria clades (Fulnečková et al. 2012).

Eukaryotic algae are a polyphyletic assemblage of phylogenetically diverse organisms with different life styles and strategies and are thus of interest for telomere biology, since they represent a substantial portion of the eukaryotic phylogenetic diversity. Three algal lineages Glaucophyta, Rhodophyta, and Chloroplastida (the later including also land plants) represent direct descendants of an ancestral alga with a cyanobacterium-derived plastid and are thought to form a monophyletic "supergroup" called Archaeplastida, whereas other algal groups obtained their plastids from red or green algae through a process called secondary or tertiary endosymbiosis (Archibald 2009).

Different telomere types were described in chromosomes of the nucleus and the nucleomorph (vestigial endosymbiont-derived nucleus) of cryptophytes or chlorarachniophytes, with the nucleomorph telomeres presumably descending from original telomeres of the ancestral engulfed algal endosymbiont (Gilson and McFadden 1995; Zauner et al. 2000). Previous genome sequencing projects reported human-type telomeric sequences in the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Armbrust et al. 2004; Bowler et al. 2008) and an unusual $AATG_6$ sequence in the red alga *Cyanidioschyzon merolae* (Nozaki et al. 2007). Synthesis of *Arabidopsis*-type telomeric repeats by telomerase was observed in dinoflagellates (Fojtová et al. 2010; Zielke and Bodnar 2010), illustrating the diversity of telomeric sequences in Alveolata, where ciliates possess telomeres with TTGGGG or TTTTGGGG repeats (Prescott 1994) and apicomplexan taxa display several related telomere types (see Kissinger and DeBarry 2011 for review).

Excavata is a potentially monophyletic "supergroup" of unicellular eukaryotes that ancestrally share a characteristic ventral feeding groove and an associated specifically organized microtubular cytoskeleton (Hampl et al. 2009; Simpson 2003). Excavates may occupy a key position in the eukaryotic phylogeny and include many important or biologically interesting species, yet they remain among the most poorly explored eukaryotic supergroups with regard to their molecular genetic and genomic features. This holds true also with regard to telomere biology, as telomeres have been characterized only in very few excavates, including the parasitic trypanosomatids (Lira et al. 2007; Van der Ploeg et al. 1984) and the diplomonad *Giardia intestinalis* (= *Giardia lamblia*; Le Blancq et al. 1991). Genome sequences have been additionally reported for the parabasalid *Trichomonas vaginalis* (Carlton et al. 2007) and the heterolobosean *Naegleria gruberi* (Fritz-Laylin et al. 2010), but no information is available on their telomeres. For other deep excavate lineages, such as Jakobida, Preaxostyla, or Malawimonadida, no knowledge about telomeres and no representative genome sequences are available.

Here we tested the activity of telomerase and investigated the presence of minisatellite repeats in diverse algal lineages using experimental approaches. In addition, our on-going genome sequencing

projects enabled us to obtain the first data about telomeres in jakobids, malawimonads, and goniomonads (the latter group representing the closest heterotrophic relatives of the cryptophyte algae). Finally, we surveyed diverse eukaryotic genome sequences available in public databases and inferred their telomeric sequences by *in silico* analyses. This combination of experimental and bioinformatic analyses allowed us to describe the diversity of telomeres across the eukaryotic phylogeny (Fig. 1), to confirm a predominant occurrence of the human-type telomeric sequence in basal lineages, and to demonstrate independent acquisition of the same telomeric repeats in various phylogenetic lineages.

Materials and methods

Algal cultures and DNA extraction

The algal material used in this study originated from culture collections as specified in supplementary table S1 (Supplementary Material online). Algae were grown in the recommended liquid media BBM or MASM (www.ccap.ac.uk/media/pdfrecipes.htm), or on nutrient agar plates. C*hromera velia* was grown in a modified f/2 medium (www.ccap.ac.uk/media/pdfrecipes.htm) in which natural seawater was replaced by seawater prepared by dissolving 23.38 g of "Red Sea" salt (Red Sea U.S.A) in 1 l of distilled water. *Phaeodactylum tricornutum* was cultivated in MASM medium supplemented with 30 mg of $Na_2SiO_3.9H_2O$ per litre.

The absence/presence of eukaryotic contaminants was monitored microscopically and using algal cultures grown on BBM and bacterial LB agar plates. The identity of most of algal samples was verified by PCR amplification (see below) and sequencing of the internal transcribed spacer and/or SSU ribosomal DNA (rDNA) regions, and when necessary (in some green algae), phylogenetic analyses were conducted to confirm the assignment into specific algal classes (data not shown). Genomic DNA for PCR amplification was isolated using the "modified IRRI" method (Collard et al. 2007). Genomic DNA from a control alga *Chlorella vulgaris* (TEL01, supplementary table S1, Supplementary Material online) was isolated according to protocol described by Saghai-Maroof et al. (1984). Isolation of DNA from other algal samples was performed according to a previously

described protocol (Fulnečková et al. 2012) which involves the use of proteinase K during the lysis step. The DNA sample concentrations were estimated from agarose gels.

Dot-blot hybridization

Genomic DNA samples (ca. 1 μg per sample) were dot-blotted onto Amersham Hybond-XL nylon membrane (GE Healthcare) and hybridized with radioactively end-labelled oligonucleotide probes (ATSB, CHSB, HUSB, TTCAGGG-SB, TTTAGGC-SB, T4AG2-SB, T3G3-SB, T2CG3-SB, Red alga-SB, supplementary table S2, Supplementary Material online) as described in Sýkorová et al. (2003b) with minor modifications according to Neplechová et al. (2005). Briefly, membranes were hybridized at 55ºC for 16 h and washed at 55ºC under low stringency conditions (2xSSC, 0.1xSDS); the final wash for the ATSB and HUSB oligonucleotides was done using a high-stringency washing buffer (0.6xSSC, 0.1xSDS) to avoid cross-hybridization. Membranes for rehybridization with another probe were gently washed three times in 0.5% SDS at 80ºC. A control probe of mixed SSU and LSU rDNA fragments was prepared by mixing an equal amount of PCR products from several phylogenetically diverse algae (TEL213 *Rhodella maculata*, TEL97 *Klebsormidium subtilissimum*, TEL211 *Tetraselmis chui*, TEL207 *Euglena geniculata*, TEL01 *Chlorella vulgaris*) obtained by amplification using a combination of gene-specific primers (18SrDNA-F and 18SrDNA-R (Katana et al. 2001) for TEL207, TEL211, TEL97, TEL01; p4 and p23 (Van der Auwera et al. 1994) for TEL 97, TEL211; ITS-A (Blattner 1999) and p23 for TEL213); the mixture was labelled by DecaLabel™ DNA Labeling Kit (Fermentas, Thermo Scientific). The probe was used for final rehybridization overnight at 62ºC and low stringency conditions (2xSSC, 0.1xSDS) or at 65ºC and high stringency conditions (0.2xSSC, 0.1% SDS). Membranes were exposed to autoradiography screens and signals were visualized using a phosphoimager FLA5000 (FujiFilm) and evaluated by the Multigauge software (FujiFilm).

Restriction digestion, pulsed field gel electrophoresis and Southern hybridization

Genomic DNA samples (1-5 μg) were digested with restriction endonucleases *Rsa*I, *Alu*I or *Taq*I (NEB) and run on an 0.9 % agarose gel in TAE buffer, the DNA fragments were alkali blotted and

hybridized using the same hybridization and washing conditions as for dot-blots (for exceptions see Supplementary Material online). Agarose plugs with high-molecular-mass DNA samples for pulse field electrophoresis were prepared from lyophilized algal samples, and BAL-31 and restriction enzyme digestion was performed as described in Sýkorová et al. (2006a). Briefly, agarose plugs with high molecular weight DNA (TEL206 *Euglena stellata*, TEL207 *Euglena geniculata*, TEL133 *Eustigmatos polyphem*, TEL201 *Vischeria punctata*, TEL103 *Klebsormidium nitens*, TEL97 *Klebsormidium subtilissimum*, TEL131 *Porphyridium purpureum*) were digested with BAL-31 nuclease for 15, 45 (or 60) min, and then by the restriction endonucleases *Sma*I (TEL97, TEL131) or *Hin*dIII (TEL206, 207, 133, 201, 103) (all enzymes from NEB). The DNA was then analyzed by pulse-field gel electrophoresis using a CHEF Mapper (BioRad) under the following conditions: 1% agarose (BioRad) gel in 0.5x TBE buffer, 6V/cm, pulses 0.5-35 s for 20 h at 13°C. Gels were alkali blotted and hybridized with the telomere probes.

Telomere repeat amplification protocol (TRAP assay)

Telomerase activity was investigated using a protocol originally developed for plant telomerases (Fitzgerald et al. 1996; Sýkorová et al. 2003b) and applied with modifications to dinoflagellates (Fojtová et al. 2010) and green algae (Fulnečková et al. 2012)(see supplementary fig. S2A, Supplementary Material online). Briefly, 35-100 mg of lyophilized algal samples ground in liquid nitrogen were incubated in 1 ml of telomerase extraction buffer (Fitzgerald et al. 1996) and total proteins were recovered in supernatant ("crude extract") after centrifugation at 17,000 g for 15 min. The telomerase-enriched fraction was purified from the supernatant by precipitation with 10% PEG 8000 and after centrifugation at 17,000 g for 5 min the pellet was dissolved in one quarter of the original volume of telomerase extraction buffer. Alternatively, the samples of crude protein extracts (without PEG precipitation) and the fraction of proteins not precipitated by PEG were used as specified in Results. A control telomerase extract was prepared from *Arabidopsis thaliana* Col-0 cell culture. The amount of total protein in extracts was determined using the Bradford method (Bradford 1976). The TRAP assay was performed as described in Sýkorová et al. (2003b) using a

substrate primer 47F (Fojtová et al. 2002) and a reverse primer TELPR30-3A (Fulnečková et al. 2012)(see Supplementary Material online). Alternatively, different combinations of substrate primers, to cope with possible enzyme preference for the substrate primer sequence, and reverse primers representing different telomere variants (supplementary table S2, Supplementary Material online) were used. Products were analyzed by polyacrylamide gel electrophoresis (PAGE), stained by GelStar(R) Nucleic Acid Gel Stain (LONZA) and visualized on a LAS3000 Imager (FujiFilm). TRAP products from selected algal species were cloned into the pCRIITOPO vector (Invitrogen) according to the manufacturer's recommendations and sequenced (Macrogen).

Gathering genomic data for jakobids, malawimonads, and goniomonads

Completed draft genome sequences and their systematic analyses of the jakobid *Andalucia godoyi*, the malawimonad *Malawimonas californiana*, and the goniomonad *Goniomonas avonlea* will be published elsewhere together with details on DNA isolation, sequencing, and assembly protocols. Briefly, *A. godoyi* (ATCC PRA-185) and *M. californiana* (ATCC 50740) were sequenced by GS FLX Titanium platform (454 Life Sciences/Roche) employing both shotgun and pair-end libraries. Draft assemblies were generated using Newbler 2.6. Genome sequence data of the recently described species *G. avonlea* (Kim and Archibald 2012) were generated using the Illumina sequencing platform from multiple libraries, including two standard short insert libraries (300 bp) and mate pair libraries (2 kbp and 6 kbp) (McGill University and Genome Quebec Innovation Centre). Errors in the raw Illumina data were corrected using ALLPATHS-LG (Gnerre et al. 2011), and the corrected reads were assembled using the ABySS *de novo* assembler (Simpson et al. 2009).

Bioinformatic analyses of telomeric sequences

Candidate telomeric sequences were searched in our genome assemblies (see above) and in sequenced genomes available in various databases, including GenBank (www.ncbi.nih.gov), TraceArchive and Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/Traces/home/), Joint Genome Institute (www.jgi.doe.gov/), Broad Institute (http://www.broad.mit.edu/), University of Tokyo (http://merolae.biol.s.u-tokyo.ac.jp/) and EMBL (http://ct.bork.embl.de). Datasets of WGS

assemblies were downloaded in the FASTA format and searched using the BioEdit software and a string search for TTAGGG, TTTAGGG, TTGGGG, TTTTGGGG, TTTTAGGG and TTAGG types of a telomeric sequence. If unsuccessful, a search was then performed manually (by eye) at 5'- and 3'-ends of scaffolds looking for a TG/CA-rich telomere-like repetitive minisatellites. The position and distribution of candidate telomeric sequences (terminal and/or internal) in the genome assembly was subsequently assessed by an automatic search for the respective strings. In addition, unassembled genomic reads from *Phaeocystis antarctica* and *Porphyra umbilicalis* (Sanger reads in the TraceArchive and Illumina reads in the SRA archive, respectively), were searched using as a query a trimer of candidate sequences or one repeat of published telomere sequence of *Cyanidioschyzon merolae* (Nozaki et al. 2007).

Results

Sample collection and analyses of algal telomeres

To cover a phylogenetically wide sample of algae we cultivated 48 algal strains from culture collections (supplementary table S1, Supplementary Material online), including members of Chloroplastida (Chlorophyta, Streptophyta), Rhodophyta, Glaucophyta, Haptophyta, Alveolata, Ochrophyta (Bacillariophyceae, Xanthophyceae, Eustigmatophyceae), and Euglenozoa (supplementary table S1, Supplementary Material online). We examined all algal strains for telomerase activity by the TRAP assay (supplementary fig. S2A, Supplementary Material online) and cloned the TRAP products from 31 strains to determine what DNA sequence forms the ends of chromosomes (i.e. what sequence is synthesized by telomerase) (table 1, figs. 2, 3, supplementary fig. S2, Supplementary Material online). Algal strains used in this study came both from groups where the telomeric sequence could be presumed from published data and from groups where the telomeric sequence has not been described yet. In the latter cases, a set of alternative reverse primer sequences were used in combination with three different substrate primers to avoid false negative results. In a subset of algal strains (32 in total), the occurrence of variant minisatellite telomeric repeats was examined by Southern hybridization (dot-blot hybridization and/or terminal restriction

10

fragment analysis (TRF); fig. 4, supplementary fig. S3, Supplementary Material online) using telomeric oligonucleotide probes (supplementary table S2, Supplementary Material online). A terminal position of candidate telomeric sequences was tested by BAL31 nuclease digestion and Southern hybridization in eight algal species (representative samples shown on fig. 5, supplementary figs. S3, S4, Supplementary Material online).

Telomerase activity screening in Archaeplastida using TRAP assay

We investigated 23 and eight algal strains from Chlorophyta and Streptophyta, respectively, for the presence of a telomerase activity using the reverse primer TELPR30-3A with the *Arabidopsis*-type telomeric sequence. The *Arabidopsis*-type sequence was presumed as an ancestral telomere type for this group based on our previous results (Fulnečková et al. 2012). Algal strains from the chlorophyte classes Chlorophyceae, Trebouxiophyceae, and Chlorodendrophyceae showed positive telomerase activity with products of a 7-nucleotide periodicity (fig. 2) and cloned TRAP products confirmed synthesis of the *Arabidopsis*-type telomeric sequence (table 1). A comparison of telomerase activity in the "crude" telomerase extracts, the PEG-purified extracts, and the PEG-non-precipitated protein fraction revealed that the telomerase activity is present also in the PEG-non-precipitated fraction in all these telomerase-positive algal strains (supplementary table S3, Supplementary Material online). However, algal strains from the class Ulvophyceae failed to show a reproducible telomerase activity. Testing of reverse primers with sequences corresponding to alternative telomere types or minisatellite variants and/or using different substrate primers to cope with possible telomerase substrate preference did not produce positive results (supplementary table S4, Supplementary Material online). A control experiment excluded the presence of telomerase inhibitors in algal extracts (supplementary fig. S1, Supplementary Material online, see below). In two cases, we experienced a very weak ladder of TRAP products using human-type reverse primer; however, we identified fungal contaminants in the respective two algal cultures by PCR (see Materials and Methods), which might be responsible for this residual activity in the samples tested. Three algal strains representing different branches of the streptophyte class Zygnematophyceae

(TEL181 *Zygnema circumcarinatum*, TEL196 *Micrasterias crux-melitensis*, TEL198 *Mesotaenium endlicherianum*) displayed a positive telomerase activity and synthesis of the *Arabidopsis*-type telomeric repeat (fig. 2, table 1). In contrast, two different telomere types could be demonstrated in the class Klebsormidiophyceae (fig. 2, table 1). Four strains (TEL100 *Klebsormidium dissectum*, TEL101 *Klebsormidium flaccidum*, TEL103 *Klebsormidium nitens*, TEL187 *Klebsormidium crenulatum*) showed a pattern with 8-nt periodicity and cloning of TRAP-products revealed synthesis of the *Chlamydomonas*-type telomeric sequences. Interestingly, the TRAP assay failed with the *Chlamydomonas*-type reverse primer (fig. 2) in *Klebsormidium subtilissimum* (TEL97) and showed a 7-nt periodicity that resulted from synthesis of a variant TTTTAGG minisatellite repeat (table 1). Investigation of telomerase activity using various substrate primers showed a difference in substrate usage between telomerases from Zygnematophyceae, Chlorophyceae, Trebouxiophyceae, and *Arabidopsis*, resulting in a changed pattern of the TRAP products (supplementary fig. S2, Supplementary Material online). A similar difference in substrate primer usage was observed in plants from the monocotyledonous order Asparagales, which possesses telomerase synthesizing human-type telomere repeats (Sýkorová et al. 2006b).

Three algal strains covering different branches of rhodophytes failed to show telomerase activity when investigated using six variants of a reverse oligonucleotide primer derived from the telomeric sequence known in *Cyanidioschyzon merolae* (supplementary tables S2, S4; Supplementary Material online) or four alternative reverse primers (supplementary table S4, Supplementary Material online) including the *Arabidopsis* type (fig. 2, table 1). To check if the failure of the TRAP assay could be caused by the presence of inhibitors in algal telomerase extracts, we performed a control experiment, in which the negative red algal telomerase extracts was added to a positive control extract from *Arabidopsis thaliana* in ratio 1:1 or 3:1. Neither of the red algal extracts posed a clear inhibitory effect (supplementary fig. S1, Supplementary Material online). Further analyses of telomerase activity using three different substrate primers with the human-type reverse primer showed a positive result in all three red algal samples, but only using the substrate primer 47F

12

(supplementary fig. S5, Supplementary Material online). To verify the identity of the amplified minisatellite repeat, we performed TRAP reactions also with an alternative reverse primer (T3AG2-C) that was able to amplify successfully the TRAP products containing human-type repeats from other algal species (see below) and could thus serve as a control. Sequencing of cloned TRAP products from these control reactions revealed that only repeats of the reverse primer sequence were amplified, suggesting a false positive result.

The glaucophyte alga *Glaucocystis nostochinearum* (TEL195) showed a telomerase activity (fig. 2) with a 6-nt periodicity of the TRAP products using an *Arabidopsis*- or a human-type reverse primer. Cloning of the TRAP products from both primer combinations verified synthesis of the human-type telomeric sequence by the *G. nostochinearum* telomerase (table 1). Interestingly, algal strains from both streptophyte classes showed a substantial enrichment of telomerase in protein extract after PEG purification, in contrast to chlorophytes and the glaucophyte alga, which displayed similar activity both in purified and non-purified samples (see supplementary table S3, Supplementary Material online).

Telomerase activity in other algal groups

We investigated telomerase activity also in representatives of other algal groups. According to published data, diatoms and euglenophytes should possess telomeres formed by the human type of telomeric minisatellite (Armbrust et al. 2004; Dooijes et al. 2000), but we were not able to detect any telomerase activity (fig. 3) in the diatom TEL231 *Phaeodactylum tricornutum*. In contrast, the three *Euglena* species tested (Euglenophyceae) and the haptophyte TEL210 *Pavlova lutheri* displayed high telomerase activity and synthesis of the human-type telomeric repeats (table 1). Control experiments designed to investigate possible preferences in substrate primer sequence or presence of inhibitors in diatom telomerase extract excluded these technical reasons of the TRAP assay failure (supplementary table S4, supplementary fig. S1, Supplementary Material online; see also above). An observed 7-nt periodicity and cloning of the TRAP products confirmed synthesis of the expected *Arabidopsis*-telomeric type in *Chromera velia* (fig. 3, table 1), which is in agreement

with its phylogenetic position within Alveolata close to dinoflagellates known from previous experiments to possess telomeres formed by the *Arabidopsis*-type sequence (Fojtová et al. 2010). Samples from the classes Xanthophyceae and Eustigmatophyceae showed very different results, despite the fact that they both belong to the algal phylum Ochrophyta within Stramenopiles. Whereas all five of xanthophytes (TEL95 *Xanthonema* cf. *hormidioides*, TEL202 *Pleurochloris meiringensis*, TEL203 *Xanthonema debile*, TEL204 *Heterococcus protonematoides*, TEL205 *Botrydiopsis intercedens*) showed telomerase synthesizing the *Arabidopsis*-type sequence, the two eustigmatophyte strains investigated (TEL133 *Eustigmatos polyphem*, TEL201 *Vischeria punctata*) did not reveal any reproducible telomerase activity (fig. 3). Similarly to diatoms, control experiments using different combinations of substrate and reverse primers showed negative result (supplementary table S4, Supplementary Material online) while a presence of inhibitors was excluded (see above, supplementary fig. S1, Supplementary Material online). A PEG-purification step was successful in telomerase enrichment of protein extracts from algal strains of Euglenophyceae, Haptophyta, Xanthophyceae, and *C. velia* (supplementary table S3, Supplementary Material online).

Dot-blot hybridization screening and testing for telomeric localization of minisatellite repeats using BAL 31 digestion

We screened samples of algal genomic DNA by Southern hybridization using radioactively labelled oligonucleotides as probes (fig. 4) to unveil a possible occurrence of other telomere-like minisatellites in the respective genomes and to possibly identify candidate telomeric sequences in samples with no detected telomerase activity. We experienced difficulties in DNA extraction from several algal strains, mainly from Zygnematophyceae and rhodophytes, which showed the presence of coloured substances and a low DNA yield; moreover, genomic DNA extraction was not successful for TEL196 *Micrasterias crux-melitensis* and TEL198 *Mesotaenium endlicherianum*. For the remaining samples from Chlorophyta, Streptophyta, Xanthophyceae, Euglenophyceae, Haptophyta, Glaucophyta, and *C. velia*, the dot-blot hybridization confirmed the presence of

telomeric minisatellites identified as "true" telomeric types synthesized by telomerase in the respective algal strains. However, dot-blot hybridization of genomic DNA from telomerase-negative strains did not suggest any other candidate telomeric sequence and in general, dot-blot hybridization signals were much weaker than we experienced in our previous study (Fulnečková et al. 2012). A weak signal of control hybridization with a mixed probe consisting of a mixture of LSU and SSU rDNA sequences (see Material and Methods) may be caused by a wide phylogenetic span of our algal collection and a limited similarity among rDNA sequences or by a low quality of genomic DNA prepared by the proteinase K-based method, because we observed difficulties in PCR amplification of control rDNA sequences and other Southern hybridization experiments (see Supplementary Material online). The terminal position of a candidate human-type telomeric sequence in euglenophytes (TEL206, TEL207) and the terminal position of the *Chlamydomonas*-type or the TTTTAGG-type of a telomeric sequence in Klebsormidiophyceae (TEL103, TEL187, and TEL97) were verified using BAL 31 nuclease digestion (fig. 5, supplementary figs. S3, S4, Supplementary Material online). Subsequent rehybridization of BAL31-digested samples of TEL97 (*Klebsormidium subtilissimum*) with the *Chlamydomonas*-type sequence probe confirmed the presence of both sequence types in terminal restriction fragments (TRF)(supplementary fig. S3 B – the bottom panel, Supplementary Material online). Investigation of the TRF lengths showed that the TTTTAGG-type sequences hybridize with 0.7-1.5 kb long fragments (supplementary fig. S3A, right panel, Supplementary Material online), suggesting short telomeres similar to Chlamydomonadales. Correspondingly, digestion with *Sma*I digestion produced longer restriction fragments and the signal of both TTTTAGG- and *Chlamydomonas*-type probes was distributed among multiple BAL31-sensitive fragments of 2.5 - 23 kb length (supplementary fig. S3 B – the bottom panel, Supplementary Material online). Besides these, short BAL31-resistant fragments (1.3-2.3 kb) representing interstitial telomeric sequences could also be seen in the hybridization patterns of both probes. The presence of internal telomere repeats is also apparent in *Klebsormidium crenulatum* (supplementary fig. S4A, Supplementary Material online). While high molecular weight restriction

fragments hybridizing with *Chlamydomonas*-type probe shortened upon BAL31 treatment (supplementary fig. S4A, the left panel, Supplementary Material online), the low molecular weight fragments were resistant to BAL31, which reflects their internal (non-telomeric) positions (supplementary fig. S4A, the right panel, Supplementary Material online). We also performed BAL31 digestion on both strains of Eustigmatophyceae to check if the quality of the genomic DNA could be the reason for the failure of dot-blot hybridization. Probing with the *Arabidopsis*-type or the human-type telomeric sequence, which are expected as candidate telomere types due to the phylogenetic position of Eustigmatophyceae in Ochrophyta, did not produce any specific signal (supplementary fig. S4, Supplementary Material online), confirming thus the negative results of the TRAP assay and dot-blot hybridization. BAL 31 nuclease digestion was performed also in TEL131 *Porphyridium purpureum* samples, but both investigated probes (*Cyanidioschyzon*-type and human-type) failed to show any specific signal.

Identification of candidate telomere sequences in genome sequences of phylogenetically diverse eukaryotes

To cover a wider spectrum of phylogenetic lineages across the tree of eukaryotic life, we coupled our experimental investigations with *in silico* searches for candidate telomeric sequences in published or publicly available genome sequences, focusing on groups that have been ignored or poorly studied with regard to their telomeres. In addition, we take advantage of the genome data yielded by our on-going genome sequencing projects for three phylogenetically unique organisms, the jakobid *Andalucia godoyi*, the malawimonad *Malawimonas californiana*, and the goniomonad *Goniomonas avonlea*. We also used available genomic sequences to verify the presence of telomeric sequences that have been described previously for the respective organisms by methods in telomere biology. We searched the genome assemblies for stretches consisting of repeated units of the major known types of telomeric sequences ($T_nA_mG_o$) and assessed them as candidate telomeric minisatellites by taking into account their position and orientation with respect to adjacent sequences. Our simple database search could not uncover degenerated telomere types, like those

known from yeasts, and experimental tests are also required to confirm the terminal position of the candidate sequences. We searched 143 genomes (including 32 from species where the telomeric sequence has been published before) and 80 of them showed a convincing pattern of telomeric sequence (supplementary table S5, Supplementary Material online). A majority of the genomes that showed a dominant presence of the candidate sequence in terminal regions also exhibited internal telomeric repeats occurring in short stretches or in large blocks (> 100 bp of un-interrupted minisatellite). The genomes where we found only short or occasional repeats positioned terminally and/or in large internal blocks were considered inconclusive and ignored for the summary of the phyletic distribution of telomeric sequences in eukaryotes shown in fig. 1 (except species with previously published telomeric sequences; supplementary table S5, Supplementary Material online). In several cases, we identified unexpected candidate telomeric repeats in genomes representing hitherto unstudied key phylogenetic groups, e.g. TTTCGGG in the parasitic relative of dinoflagellates *Perkinsus marinus*, TTTGGG in the heterolobosean *Naegleria gruberi*, TTAGG in the labyrinthulid *Aurantiochytrium limacinum*, and the highly unusual 10-11 nucleotide repeat unit TTTATT(T)AGGG in the rhodophyte *Galdieria sulphuraria* (fig. 1 and supplementary Table S5, Supplementary Material online). In addition, minisatellites that differed from the types "canonical" for the respective organismal groups were found in fungi and stramenopiles, indicating the evolutionary flexibility of the telomeric sequence at various phylogenetic scales. Our database searches corroborated the experimental results from Haptophyta, Glaucophyta, and Chlorophyta, whereas no genome assemblies from Xanthophyceae, *Chromera*, Ulvophyceae, Euglenophyceae, or dinoflagellates were available for analysis. The genome assemblies of two *Nannochloropsis* species (Eustigmatophyceae) displayed the presence of large internal blocks of TTAGGG-type repeats in addition to several terminally positioned stretches; and without experimental evidence, these repeats should be taken as a candidate telomere sequence. Searches of unassembled genomic reads available for the red alga *Porphyra umbilicalis* did not identify the telomere types described in *C. merolae* genome or predicted in the *G. sulphuraria* genome (see above), but revealed a large

number of reads containing a TTAGGG-type minisatellite. In several cases these repeats could be assessed as internal, but whether any of the other sequences represent the true telomere cannot be verified without a full genome assembly or an experiment. A similar result was achieved when we searched unassembled Sanger reads from an on-going genome project for the haptophyte *Phaeocystis antarctica*.

Discussion

People have long been fascinated by the question how mechanisms of linear chromosome maintenance might have originated. It is believed that recombination-based pathways, which today serve mostly as a backup mechanism (Fajkus et al. 2005), were original and were subsequently replaced by a more successful, steady and efficient synthesis of telomeres consisting of minisatellite sequences by telomerase. Telomeres of most investigated organismal groups are generally conserved within the groups and conform to a limited number of minisatellite types. The repeat units of these minisatellites are mainly variants of the $T_nA_mG_o$ sequence and their evolutionary success presumably depends on their properties stemming from the G-rich sequence and their capacity to form alternative DNA structures like the G-quartet or the T-loop typical for telomere function (de Lange 2005). The telomere sequence of the red alga *Cyanidioschyzon merolae* is somewhat atypical, but it is a G-rich sequence and genes for telomerase subunits were predicted in the genomic sequence (Nozaki et al. 2007). It is generally assumed that telomeres are formed by T/G-rich minisatellites and maintained by telomerase until alternative telomere and maintenance mechanisms are shown for a given organism. The results of our new experimental and *in silico* analyses have substantially expanded the sampling of telomere structures across the eukaryotic phylogeny. In addition, telomerase activity has also now been tested against a number of evolutionarily distant groups. The data enable us to paint a picture of telomere evolution across eukaryotes with an unprecedented level of detail (fig. 1), although many aspects of the scheme remain to be clarified.

Characterization of telomeres in many algal groups remains technically challenging

In algae, telomerase activity has been previously experimentally proven in photosynthetic and non-photosynthetic dinoflagellates (Fojtová et al. 2010; Zielke and Bodnar 2010) and in chlorophytes (Fulnečková et al. 2012). Despite a published completed genome sequence including predicted telomere sequences and a telomerase gene (Bowler et al. 2008), our TRAP assay was unsuccessful in the diatom *Phaeodactylum tricornutum*. This experimental failure could be for some unknown technical reason or for the telomerase being active only in specific developmental stages (e.g. during auxospore formation) that were not presented in our samples at a detectable level. Another possible explanation comes from experiments in the silkworm *Bombyx mori*, where telomeres are maintained by telomerase with a very low processivity and thus difficult to be detected (Sasaki and Fujiwara 2000). Active telomerase synthesizing the *Arabidopsis*-type telomeric sequences was demonstrated by our experiments in *Chromera velia* (Alveolata), Xanthophyceae, chlorophytes, and streptophyte algae. In contrast, we demonstrated the human-type telomeric sequence synthesized by telomerase in euglenophytes, haptophytes (*Pavlova lutheri*), and glaucophytes (*Glaucocystis nostochinearum*).

Unexpectedly, we did not detect telomerase activity in the classes Eustigmatophyceae (Ochrophyta) and Ulvophyceae (Chlorophyta), and also Southern hybridization of genomic DNA with other minisatellite telomeric probes failed to resolve telomeric sequences, as was observed also in rhodophytes. A low signal intensity by Southern hybridization could be caused by impurities in the genomic DNA, which in our case had to be isolated by the proteinase K-based method (see Supplementary Material online). Control TRAP experiments excluded the presence of telomerase inhibitors in protein extracts of all telomerase-negative algal strains (a diatom, Eustigmatophyceae, Ulvophyceae, Rhodophyta). Our use of a different substrate primers should accommodate variable telomerase preferences for substrate sequence. Despite these controls, it remains possible that there are technical reasons that prevent the telomere motifs from being resolved. Remarkably, we found a similar behaviour of telomerases during PEG purification from protein extracts in various algal

groups, suggesting their similar biochemical properties. In most cases, telomerase was enriched in a purified fraction, with the exception of protein extracts from chlorophytes and a glaucophyte alga, which showed telomerase activity also in the non-precipitated fraction (supplementary table S3, Supplementary Material online). However, PEG purification led to removal of compounds inhibiting TRAP assay from these extracts (Ševčíková et al. 2013).

We did not detect telomerase activity in three red algal representatives investigated for synthesis of *Cyanidioschyzon*-type telomeric sequences. This may be due to specific problems with PCR-based TRAP assay (a difficult G-rich template of a candidate sequence) or to occurrence of different telomeric sequences, because the investigated red algal species are only distantly related to *C. merolae*. The behaviour of the three rhodophyte strains investigated here does not seem to be a peculiarity of one particular lineage, because they represent three different deeply diverged rhodophyte classes (fig. 1; supplementary table S1, Supplementary Material online). Searches for telomeric motifs in databases suggest that some rhodophytes might have the human-type telomeric sequence, because a great portion of genomic reads from *Porphyra umbilicalis* (representing an additional class Bangiophyceae) contain this sequence type. However, many of them could be scored as internal sequences, and other telomere-like minisatellites also occur in the genomic reads, making it difficult to distinguish the "true" telomeric type without an experiment. Moreover, the assembled genome sequence of *Galdieria sulphuraria* (a relative of *C. merolae* also belonging to Cyanidiophyceae) showed the presence of an unusual minisatellite repeat (TTTATTAGGG or TTTATTTAGGG) predominantly at the ends of scaffolds, suggesting a putative telomeric position. The sequence of this repeat unit seems to be an AT-rich variant derived from the typical $T_nA_mG_o$ telomeric minisatellite. Neither the human-type nor other investigated telomere types (including the *Cyanidioschyzon*-type) were found in the *Galdieria* genome assembly. The path of the telomere evolution in rhodophytes thus remains unclear. Assuming that Archaeplastida are monophyletic (but see, e.g., Burki et al. 2012) and that the human-type telomeric repeat occurs in the presumably basal archaeplastid lineage Glaucophyta as well as in most other major eukaryotic lineages (see below), it

is possible that the last common ancestor of Archaeplastida had the TTAGGG telomeric sequence. This would then mean that this telomere type could be retained by some red algal lineages, but further testing of archaeplastid monophyly, and more evidence for the presence of the human-type telomere in rhodophytes, are needed to corroborate this scenario.

Novel telomere forms seem to have evolved in some algal groups

In plants and algae, two categories of evolutionary changes of telomere sequences have been described previously: (1) change of the telomeric sequence synthesized by telomerase to a related minisatellite sequence, i.e. from the *Arabidopsis* type to the human type in Asparagales (Sýkorová et al. 2003b) and from the *Arabidopsis* type to the *Chlamydomonas* type or the human type in Chlamydomonadales (Fulnečková et al. 2012); and (2) loss of a typical minisatellite and of telomerase activity reported in the genus *Allium* (Asparagales) (Sýkorová et al. 2006a) and three genera of Solanaceae (Sýkorová et al. 2003a). Our results bring evidence for further examples following the first and possibly also the second category.

The first is exemplified by switches to variant minisatellite telomeric sequences in Klebsormidiophyceae, specifically a one-nucleotide addition resulting in a change from the ancestral TTTAGGG (*Arabidopsis* type) to TTTTAGGG (*Chlamydomonas* type) early in the evolution of Klebsormidiophyceae, followed by a one-nucleotide deletion in the lineage leading to TEL97 *Klebsormidium subtilissimum*, which resulted in the unusual TTTTAGG sequence (this interpretation is based on the nested phylogenetic position of the latter *K. subtilissimum* within the genus *Klebsormidium*; Rindi et al. 2011). The TEL97 strain still possesses a large amount of ancestral *Chlamydomonas*-type sequence in its genome, which is similar to the situation in Hyacinthaceae (Asparagales) that have telomerase synthesizing the human-type sequence, but the ancestral telomere type still occurs in the genome at a level detectable by fluorescent *in situ* hybridization (Adams et al. 2001). The negative result of the telomerase assay using typical telomeric minisatellites as reverse primers in Eustigmatophyceae and Ulvophyceae suggests a change of the second type, but a more detailed study is needed to pinpoint the evolutionary

transition in the telomere sequence and/or maintenance mechanism. Similarities between land plants and algae are also seen in different substrate primer usage (see Supplementary Material online) that was demonstrated in Asparagales, and in that case its phyletic pattern did not show a simple correspondence to the phylogeny of this group (Sýkorová et al. 2006b). It seems that the substrate primer usage (supplementary fig. S2, Supplementary Material online) is linked to enzyme properties of telomerases in different species, which generally show a varying accuracy in minisatellite repeat synthesis (Fitzgerald et al. 2001; Sýkorová et al. 2003b).

Telomere evolution in the context of the phylogeny of eukaryotes

Although the telomere structure remains unknown or enigmatic for many deeply branching eukaryotic lineages, including those with reference genome sequences available (e.g. Parabasalia, Archaemobae, or Apusomonadida), a combination of previously published data with the results of our experiment and *in silico* searches allows to paint at least the major outlines of telomere evolution in eukaryotes (fig. 1). Metazoa (Traut et al. 2007; Gomes et al. 2010) and Fungi (Teixeira and Gilson 2005) show the TTAGGG-type as ancestral to both groups and all changes in their telomeric sequences seem to have happened later in their evolution. The same evolutionary path, i.e. from the ancestral TTAGGG type to secondarily derived alternative telomeric repeats, can now be deduced for Amoebozoa, Excavata, Stramenopiles, and probably also Archaeplastida (fig. 1). The TTAGGG motif remains the only telomeric repeat known in Choanoflagellata, Haptophyta, and Rhizaria (excluding nucleomorph genomes), hence it is likely that it is also ancestral for these groups, but a much better sampling (especially for Rhizaria) is needed to confirm this. On the other hand, the plant type of telomeric repeats (TTTAGGG) may be ancestral in cryptomonads, which include cryptophyte algae and their heterotrophic sister lineage goniomonads. Inference on the type of the telomeric repeat is difficult for the last common ancestor of Alveolata, since the two principal lineages, i.e. Myzozoa (including dinoflagellates, perkinsids, apicomplexans, and chromerids) and ciliates, have different telomeres.

Considering the wide occurrence of the TTAGGG telomeric repeat and its inferred ancestral

presence in most major eukaryotic groups, it is tempting to suggest that the human-type telomeric repeat was ancestral for eukaryotes as a whole. Deducing the ancestral state of any character for any taxon requires knowledge on the position of the taxon's root (the deepest branching point). The question about the root of the eukaryotic phylogeny has not yet settled, and at least four contradictory hypotheses, supported by different sources of evidence, have been suggested recently (fig. 1). Interestingly, all those root positions are compatible with the idea of the TTAGGG being the ancestral telomeric repeat for eukaryotes.

Another important aspect to consider is the frequency of homoplasy in the evolution in telomeric sequences. Indeed, assuming the TTAGGG motif as ancestral, different phylogenetic lineages evolved independently on each other to the same alternative motifs. A most notable case concerns the "plant" type TTTAGGG motif, which seems to have evolved independently in Chloroplastida, cryptomonads, fungi, oomycetes, xanthophytes, and alveolates (or at least their subgroup Myzozoa; fig. 1 and supplementary table S5, Supplementary Material online). Additional changes of secondarily evolved telomeric motifs then occurred many times, including shifts to novel motif (e.g. the TTTTAGGG motif in some chlamydomonadalean and klebsormidiophyte green algae and in the apicomplexan *Theileria*; (Sohanpal et al. 1995), or the TTTCGGG motif in *Perkinsus marinus* probably derived by a A-to-C substitution in the "plant" motif ancestral for Myzozoa) and reversions to the ancestral "human" type in some members of Chlamydomonadales (Fulnečková et al. 2012).

In summary, our experiments and analyses substantially expand the sampling of telomere diversity in eukaryotes and strengthen the view of telomeres as evolutionarily flexible structures of eukaryotic genomes. The much more comprehensive picture of the phyletic distribution of various telomere types should facilitate future studies on mechanistic causes of the evolutionary changes in telomeres and on underlying driving forces.

Acknowledgements

Literature cited

Adams SP, et al. 2001. Loss and recovery of Arabidopsis-type telomere repeat sequences 5'-(TTTAGGG)(n)-3' in the evolution of a major radiation of flowering plants. Proceedings of the Royal Society London Biological Sciences. 268: 1541-1546.

Archibald JM. 2009. The puzzle of plastid evolution. Curr Biol. 19: R81-88.

Armbrust EV, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science. 306: 79-86.

Biessmann H, Mason JM. 2003. Telomerase-independent mechanisms of telomere elongation. Cell Mol Life Sci. 60: 2325-2333.

Blattner FR. 1999. Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. Biotechniques. 27: 1180-1186.

Bowler C, et al. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature. 456: 239-244.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 72: 248-254.

Burki F, Okamoto N, Pombert JF, Keeling PJ. 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc Biol Sci. 279: 2246-2254.

Carlton JM, et al. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science. 315: 207-212.

Cavalier-Smith T. 2010. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol Lett. 6: 342-345.

Collard BCY, Das A, Virk PS, Mackill DJ. 2007. Evaluation of "quick" and "dirty" DNA extraction methods for marker-assisted selection in rice (*Oryza sativa* L.). Plant Breeding. 126: 47-50.

de Lange T. 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 19: 2100-2110.

Derelle R, Lang BF. 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. Mol Biol Evol. 29: 1277-1289.

Dooijes D, et al. 2000. Base J originally found in kinetoplastida is also a minor constituent of nuclear DNA of Euglena gracilis. Nucleic Acids Res. 28: 3017-3021.

Fajkus J, Sýkorová E, Leitch AR. 2005. Telomeres in evolution and evolution of telomeres. Chromosome Res. 13: 469-479.

Fitzgerald MS, McKnight TD, Shippen DE. 1996. Characterization and developmental patterns of telomerase expression in plants. Proceedings of the National Academy of Sciences of the United States of America. 93: 14422-14427.

Fitzgerald MS, et al. 2001. Different modes of de novo telomere formation by plant telomerases. Plant J. 26: 77-87.

Fojtová M, Fulnečková J, Fajkus J, Kovarik A. 2002. Recovery of tobacco cells from cadmium stress is accompanied by DNA repair and increased telomerase activity. J Exp Bot. 53: 2151-2158.

Fojtová M, et al. 2010. Telomere maintenance in liquid crystalline chromosomes of dinoflagellates. Chromosoma. 119: 485-493.

Fritz-Laylin LK, et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell. 140: 631-642.

Frydrychova R, et al. 2004. Phylogenetic distribution of TTAGG telomeric repeats in insects. Genome. 47: 163-178.

Fulnečková J, et al. 2012. Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. Genome Biol Evol. 4: 248-264.

Gilson P, McFadden GI. 1995. The chlorarachniophyte: a cell with two different nuclei and two different telomeres. Chromosoma. 103: 635-641.

Gnerre S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 108: 1513-1518.

Gomes NM, Shay JW, Wright WE. 2010. Telomere biology in Metazoa. FEBS Lett. 584: 3741-3751.

Hampl V, et al. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc Natl Acad Sci U S A. 106: 3859-3864.

Chan SR, Blackburn EH. 2004. Telomeres and telomerase. Philos Trans R Soc Lond B Biol Sci. 359: 109-121.

Katana A, et al. 2001. Phylogenetic position of *Koliella* (Chlorophyta) as inferred from nuclear and chloroplast small subunit rDNA. Journal of Phycology. 37: 443-451.

Katz LA, Grant JR, Parfrey LW, Burleigh JG. 2012. Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. Syst Biol. 61: 653-660.

Kim E, Archibald JM. 2012. Ultrastructure and Molecular Phylogeny of the Cryptomonad *Goniomonas avonlea* sp. nov. Protist.

Kissinger JC, DeBarry J. 2011. Genome cartography: charting the apicomplexan genome. Trends Parasitol. 27: 345-354.

Laurin-Lemay S, Brinkmann H, Philippe H. 2012. Origin of land plants revisited in the light of sequence contamination and missing data. Curr Biol. 22: R593-594.

Le Blancq SM, Kase RS, Van der Ploeg LH. 1991. Analysis of a *Giardia lamblia* rRNA encoding telomere with [TAGGG]n as the telomere repeat. Nucleic Acids Res. 19: 5790.

Lira CB, et al. 2007. Telomere biology of trypanosomatids: beginning to answer some questions. Trends Parasitol. 23: 357-362.

Meyne J, Ratliff RL, Moyzis RK. 1989. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci U S A. 86: 7049-7053.

Mravinac B, Mestrovic N, Cavrak VV, Plohl M. 2011. TCAGG, an alternative telomeric sequence in insects. Chromosoma. 120: 367-376.

Neplechová K, Sýkorová E, Fajkus J. 2005. Comparison of different kinds of probes used for analysis of variant telomeric sequences. Biophys Chem. 117: 225-231.

Nozaki H, et al. 2007. A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. BMC Biol. 5: 28.

Okazaki S, et al. 1993. Identification of a pentanucleotide telomeric sequence, (TTAGG)n, in the silkworm *Bombyx mori* and in other insects. Mol Cell Biol. 13: 1424-1432.

Paps J, et al. 2013. Molecular Phylogeny of Unikonts: New Insights into the Position of Apusomonads and Ancyromonads and the Internal Relationships of Opisthokonts. Protist. 164: 2-12.

Parfrey LW, et al. 2010. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. Syst Biol. 59: 518-533.

Prescott DM. 1994. The DNA of ciliated protozoa. Microbiol Rev. 58: 233-267.

Richards EJ, Ausubel FM. 1988. Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. Cell. 53: 127-136.

Rindi F, et al. 2011. Phylogenetic relationships in *Interfilum* and *Klebsormidium* (Klebsormidiophyceae, Streptophyta). Mol Phylogenet Evol. 58: 218-231.

Rogozin IB, Basu MK, Csuros M, Koonin EV. 2009. Analysis of rare genomic changes does not support the unikont-bikont phylogeny and suggests cyanobacterial symbiosis as the point of primary radiation of eukaryotes. Genome Biol Evol. 1: 99-113.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length

polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci U S A. 81: 8014-8018.

Sasaki T, Fujiwara H. 2000. Detection and distribution patterns of telomerase activity in insects. Eur J Biochem. 267: 3025-3031.

Simpson AG. 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol. 53: 1759-1777.

Simpson JT, et al. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19: 1117-1123.

Sohanpal BK, Morzaria SP, Gobright EI, Bishop RP. 1995. Characterisation of the telomeres at opposite ends of a 3 Mb *Theileria parva* chromosome. Nucleic Acids Res. 23: 1942-1947.

Sýkorová E, et al. 2006a. Minisatellite telomeres occur in the family Alliaceae but are lost in *Allium*. American Journal of Botany. 93: 814-823.

Sýkorová E, Leitch AR, Fajkus J. 2006b. Asparagales telomerases which synthesize the human type of telomeres. Plant Mol Biol. 60: 633-646.

Sýkorová E, et al. 2003a. The absence of Arabidopsis-type telomeres in *Cestrum* and closely related genera *Vestia* and *Sessea* (Solanaceae): first evidence from eudicots. Plant Journal. 34: 283-291.

Sýkorová E, et al. 2003b. Telomere variability in the monocotyledonous plant order Asparagales. Proceedings. Biological sciences. 270: 1893-1904.

Ševčíková T, Bišová K, Fojtová M, Lukešová A, Hrčková K, Sýkorová E. 2013. Completion of cell division is associated with maximum telomerase activity in naturally synchronized cultures of the green alga *Desmodesmus quadricauda*. FEBS Letters, in press.

Teixeira MT, Gilson E. 2005. Telomere maintenance, function and evolution: the yeast paradigm. Chromosome Res. 13: 535-548.

Traut W, et al. 2007. The telomere repeat motif of basal Metazoa. Chromosome Res. 15: 371-382.

Van der Auwera G, Chapelle S, De Wachter R. 1994. Structure of the large ribosomal subunit RNA

of *Phytophthora megasperma*, and phylogeny of the oomycetes. FEBS Lett. 338: 133-136.

Van der Ploeg LH, Liu AY, Borst P. 1984. Structure of the growing telomeres of Trypanosomes.
Cell. 36: 459-468.

Vitkova M, et al. 2005. The evolutionary origin of insect telomeric repeats, (TTAGG)n.
Chromosome Res. 13: 145-156.

Zauner S, et al. 2000. Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres
in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. Proc Natl Acad
Sci U S A. 97: 200-205.

Zhao S, et al. 2012. *Collodictyon*--an ancient lineage in the tree of eukaryotes. Mol Biol Evol. 29:
1557-1568.

Zielke S, Bodnar A. 2010. Telomeres and telomerase activity in scleractinian corals and
Symbiodinium spp. Biol Bull. 218: 113-121.

Figure legends

Fig. 1 Telomeres in a framework of the eukaryotic phylogeny. The schematic phylogenetic tree
of eukaryotes has been drawn based on recent phylogenomic analyses (Burki et al. 2012; Derelle
and Lang 2012; Hampl et al. 2009; Laurin-Lemay et al. 2012; Paps et al. 2013; Parfrey et al. 2010;
Zhao et al. 2012). Unresolved or contentious regions of the eukaryotic phylogeny are shown as
polytomies; note that the supergroups "Excavata" and "Archaeplastida" are depicted as
monophyletic for convenience, but monophyly of all their constituent lineages (specifically
concerning the position of malawimonads and glaucophytes) remains uncertain. Lineages with no
information available about their telomeres have been omitted for simplicity. Alternative, recently
suggested positions of the root of the eukaryotic phylogeny are marked: root 1 – (Rogozin et al.
2009); root 2 – (Cavalier-Smith 2010); root 3 – (Derelle and Lang 2012); root 4 – (Katz et al. 2012).
Telomere types documented for individual lineages are indicated on the left, with the various
(putative) telomeric repeats indicated in different colours; NCT, non-canonical telomere (e.g.

transposon-based). Telomeric sequences confirmed by experiments or by a sufficiently complete genome sequence assembly are indicated in upper case, candidate telomeric sequences deduced from available draft genome sequences are indicated in lower case. When there is an apparently dominant telomere type known for a given taxon, the minor (typically secondarily derived) variants are shown in parentheses. For the sake of simplicity, the various unique telomeric sequences known in Saccharomycotina are indicated only by the collective label "various yeast repeats" and some of the additional unique fungal telomeric sequences are omitted (but presented in supplementary Table S5, Supplementary Material online). Lineages that were targeted by experiments in this study yet failed to reveal their telomere structure are indicated with question marks. Details on the species representing the lineages in the phylogeny and the type of evidence for their telomere types are provided in table 1 and supplementary table S5 (Supplementary Material online).

Fig. 2. Telomerase activity in Archaeplastida investigated by TRAP assay.

Telomerase activity in representative algal strains of Glaucophyta (A, TEL195 *Glaucocystis nostochinearum*), Rhodophyta (B, TEL213 *Rhodella maculata*), Chlorophyta (C, TEL211 *Tetraselmis chui*; D, TEL121 *Dictyochloropsis irregularis*; E, TEL94 *Pseudendocloniopsis botryoides* and TEL124 *Pseudendoclonium basiliense*), and Streptophyta (F, TEL198 *Mesotaenium endlicherianum*; G, TEL97 *Klebsormidium subtilissimum*, TEL100 *Klebsormidium dissectum*, TEL101 *Klebsormidium flaccidum*, TEL103 *Klebsormidium nitens*, TEL187 *Klebsormidium crenulatum*) grouped according to their phylogenetic provenance (indicated above panels); the activity is shown using combinations of substrate and reverse primers - GG(21) and HUTC (human-type primer)(A), 47F and TELPR30-3A (*Arabidopsis*-type)(B, C, F), or pSSyF and TELPR30-3A (D, E). Synthesis of telomeric repeats corresponding to the human-type and the *Arabidopsis*-type sequence (compare to table 1) was observed in Glaucophyta (A) and three green algal classes (C, Chlorodendrophyceae; D, Trebouxiophyceae; F, Zygnematophyceae), respectively. Negative results were obtained in Rhodophyta (B) and Ulvophyceae (E). The Klebsormidiophyceae samples (G) showed synthesis of two different telomere types; alternative combinations of the substrate primer

TS21 and the *Chlamydomonas*-type repeat reverse primer (CHTTRAPRev1), or of the substrate

primer pSSyF and the TTTTAGG-type repeat reverse primer (T4AG2-C), displayed synthesis of a

seven or an eight nucleotide periodicity of TRAP products (arrows) by telomerase of *K.*

subtilissimum (TEL97) or other *Klebsormidium* spp., respectively (see table 1). Differences in

efficiency of telomerase purification during preparation from protein extracts are documented in

samples shown on C, D, F, and G (summarized in supplementary table S3, Supplementary Material

online). Triangles indicate different amounts of total protein (0.1 and 1 µg) in protein extract

without PEG precipitation (crude, cr.), in fractions non-precipitated (supernatant, sup) and

precipitated by PEG (telomerase extract, ex), except TEL94 (E; 0.1, 0.2 µg), TEL124 (E; 0.1, 0.5

µg), TEL97 (G; 0.1, 0.8 µg), TEL100 (G; 0.1, 0.3 µg), and TEL101 (G; 0.1, 0.4 µg). When one

sample is indicated, 1 µg or a higher amount of total protein mentioned above was used, except

TEL121 (D; all 0.5 µg), TEL211 (C; sup 0.5 µg), TEL94 (E; cr. 0.1, sup 0.3 µg), TEL101 (G; sup

0.4 µg). Telomerase-enriched extracts (50 ng of total protein) from *Chlamydomonas hydra*

(TTTTAGGG), *Arabidopsis thaliana* seedlings (TTTAGGG), and *Euglena stellata* (TTAGGG)

were used as a pattern control of an eight-, a seven-, and a six-nucleotide periodicity ladder,

respectively; negative control (-), no extract.

Fig. 3. Telomerase activity in algae outside Archaeplastida investigated by TRAP assay.

Results of telomerase activity assay from representative samples of algal strains with telomerase

synthesizing the human-type (A, Euglenophyceae, Haptophyta) and the *Arabidopsis*-type (B,

Xanthophyceae, Alveolata) telomeric sequence, and those with negative telomerase activity (C,

Bacillariophyceae, Eustigmatophyceae), are shown. The ladder of positive TRAP products (A, B)

corresponds to 6-nt or 7-nt periodicity of control samples (human- and *Arabidopsis*-type,

respectively). The efficiency of telomerase purification during preparation in protein extract

(summarized in supplementary table S3, Supplementary Material online) was monitored (A, B)

without PEG precipitation (crude, cr.), and in fractions non-precipitated (supernatant, sup) and

precipitated by PEG (telomerase extract, ex), protein extracts containing 100 ng and/or 1 µg of total

31

protein were used. Negative results (C) were obtained using different amounts of total protein (indicated by triangle) from *Phaeodactylum tricornutum* (TEL231; 1, 0.5, 0.1 μg), *Vischeria punctata* (TEL201, 0.5 μg) and *Eustigmatos polyphem* (TEL133, 1, 0.1 μg on left and 0.5 μg on middle panel) and different primer combinations (indicated under panels). Combinations of the substrate primer GG(21) and the human-type repeat reverse primer HUTC (A, C) or of the substrate primer 47F and the *Arabidopsis*-type repeat reverse primer TELPR30-3A (B, C) were used. Telomerase-enriched extracts (50 ng of total protein) from *Chlamydomonas hydra* (TTTTAGGG), *Arabidopsis thaliana* seedlings (TTTAGGG), and *Euglena stellata* (TTAGGG) were used as a pattern control of an eight-, a seven-, and a six-nucleotide periodicity ladder, respectively; negative control (-), no extract.

Fig. 4. Dot-blot hybridization of genomic DNA with telomere and telomere-like minisatellite probes. Genomic DNA samples (1-2 μg per dot) from algal strains and control samples (listed according phylogeny position on left) were blotted and hybridized with radioactively labelled oligonucleotide probes representing different telomere types and derived sequences (indicated above dot columns). Samples in which the telomere type was revealed in a telomerase analysis (figs. 2, 3, table 1) hybridized with the corresponding oligonucleotide, but occurrence of other minisatellites was also indicated. The T4AG2 and CHSB probes cross-hybridized and could not be distinguished by Southern hybridization (for details see Supplementary Material online). The T2CG3 minisatellite showed a signal across algal samples, but not in a terminal position (see Supplementary Material online); a similar situation was described in *Allium* (Sýkorová et al. 2006a). Control DNA samples represent the *Arabidopsis*-type (*Chlorella vulgaris*) and the human-type telomeres (human and *Ipheion uniflorum*, Alliaceae); note that the plant DNA contains also a portion of the ancestral *Arabidopsis*-type minisatellite. Control rehybridization of membranes was done with a mixed rDNA probe (see Materials and Methods for details); n.a. – not analyzed.

Fig. 5. Analysis of telomeres in *Euglena stellata*. Samples of high molecular weight DNA in agarose plugs were digested with BAL31 nuclease for indicated times (in minutes) and then with

restriction enzyme *Hin*dIII. DNA remaining in the plugs was analyzed by pulsed field electrophoresis (left panel) and the low molecular fraction of DNA after cleavage (diffused into reaction solution) was analyzed by conventional agarose gel electrophoresis (right panel). Terminal restriction fragments detected with a human-type telomere probe (without BAL31 digestion) range between 20 and 145 kb and are sensitive to BAL31 digestion. A fraction of TRFs shortened after 15 and 50 min of digestion with BAL31 shifts to the low-molecular-weight fragments detected on conventional gel hybridization. The ladder is in Kbs.

Table 1. Results of TRAP assay and cloning

Phylum/Class	TEL	Species	telomere type	No. of clones	No. of repeats	activity (primers)	minisatellite variant				T/G-slippage	mismatches	other combinations
							TTAGGG	TTTAGGG	TTTTAGGG	TTTTAGG			
Glaucophyta	195	*Glaucocystis nostochinearum*	TTAGGG	2	39	+++ (a,f)	38	1	0	0	0	0	+++ (b,f)
				4	67	+++ (c,h)	65	1	0	0	0	0	
Rhodophyta	131	*Porphyridium purpureum*	unknown	N.A.									- (d,g); Table S4
	213	*Rhodella maculata*	unknown	N.A.									- (b,f); Table S4
	214	*Rhodosorus marinus*	unknown	N.A.									- (b,f); Table S4
Chlorophyta/ Chlorophyceae	87	*Scenedesmus vacuolatus*	TTTAGGG	2	17	++ (e,g)	0	17	0	0	0	0	
	89	*Muriella decolor*	TTTAGGG	4	78	+++ (c,f)	0	77	1	0	0	0	
	91	*Mychonastes homosphaera*	TTTAGGG	3	33	++ (e,g)	0	33	0	0	0	0	
	98	*Chromochloris zofingiensis*	TTTAGGG	6	55	+++ (a,f)	0	50	1	0	4	0	+++ (c,f)
	108	*Neochloris conjuncta*	TTTAGGG	3	24	+++ (b,f)	0	24	0	0	0	0	
	123	*Chlorococcum hypnosporum*	TTTAGGG	5	38	+++ (b,f)	1	38	0	0	1	1	+++ (c,f)
	138	*Pseudomuriella aurantiaca*	TTTAGGG	2	18	+++ (b,f)	2	13	1	0	1	0	
	140	*Follicularia paradoxalis*	TTTAGGG	3	21	+++ (b,f)	1	20	0	0	0	0	+++ (c,f)
	188	*Bracteacoccus cohaerens*	7-nt (tttaggg)	N.A.									+++ (b,f)
Chlorophyta/ Trebouxiophyceae	84	*Jaagiella alpicola*	TTTAGGG	2	30	+++ (e,g)	3	24	2	0	1	0	+++ (c,f)
				2	4	+++ (a,f)	0	3	1	0	0	0	
	85	*Asterochloris phycobiontica*	TTTAGGG	5	67	+++ (b,f)	0	64	1	0	2	0	
	88	*Heterochlorella luteoviridis*	TTTAGGG	3	32	+++ (a,f)	0	31	0	0	0	1	+++ (c,f)
	90	*Auxenochlorella prototothecoides*	TTTAGGG	4	30	+++(d,g)	0	30	0	0	0	0	+++ (c,f)
				3	54	+++ (a,f)	0	53	1	0	0	0	
	121	*Dictyochloropsis irregularis*	TTTAGGG	4	61	+++ (a,f)	0	58	0	0	2	1	+++ (c,f)
	134	„*Heterotrichella gracilis*"	7-nt (tttaggg)	N.A.		+++ (c,f)							
Chlorophyta/ Chlorodendrophyceae	211	*Tetraselmis chui*	TTTAGGG	4	38	+++ (b,f)	0	22	9	0	5	3	
	212	*Tetraselmis striata*	TTTAGGG	4	18	++ (b,f)	0	9	7	0	2	0	
Chlorophyta/ Ulvophyceae	86	*Planophila laetevirens*	unknown	N.A.									+/- (a,f); Table S4
	94	*Pseudendocloniopsis botryoides*	unknown	N.A.									- (a,f); +/- (c,h); Table S4

	111	*Pseudendoclonium printzii*	**unknown**	N.A.								+/- (a,f); Table S4
	124	*Pseudendoclonium basiliense*	**unknown**	N.A.								- (a,f); +/- (c,h); Table S4
	137	*Desmochloris halophila*	**unknown**	N.A.								- (a,f); +/- (b,f); +/- (c,h); Table S4
	139	*Pirula salina*	**unknown**	N.A.								+/- (a,f); Table S4
Streptophyta/ Klebsormidiophyceae	97	*Klebsormidium subtilissimum*	**TTTTAGG**	3	53	++ (a,i)	0	0	0	**53**	0	0	+/- (c,l); ++ (d,i); +/- (d,j); - (d,k); +/- (b,f); - (d,h)
	100	*Klebsormidium dissectum*	**8-nt (ttttaggg)**	N.A.		+ (d,k)							+/- (d,j); + (c,k); - (d,h)
	101	*Klebsormidium flaccidum*	**TTTTAGGG**	5	55	++ (a.i)	0	0	**53**	0	1	1	+++ (d,j); - (d,h); +++ (c,k); ++ (d,k)
	103	*Klebsormidium nitens*	**TTTTAGGG**	2	25	+ (a,i)	0	0	**23**	0	1	1	++(d,j); ++(c,k); ++ (d,k); - (d,h)
				1	20	++ c.i	0	0	**19**	0	1	0	
				3	37	+ (c,l)	0	0	**33**	0	4	0	
	187	*Klebsormidium crenulatum*	**8-nt (ttttaggg)**	N.A.		++ (d,k)							- (b,f); - (a,f); - (a,i); + (c,i); + (c,l); +/- (d,h)
Streptophyta/ Zygnematophyceae	181	*Zygnema circumcarinatum*	**TTTAGGG**	5	38	+ (d,f)	4	**29**	0	0	1		+ (b,f)
	196	*Micrasterias crux-melitensis*	**TTTAGGG**	4	38	+++ (d,f)	0	**35**	0	0	1	1	
	198	*Mesotaenium endlicherianum*	**TTTAGGG**	4	18	+++ (b,f)	0	**15**	0	0	3	0	+++ (d,f)
Ochrophyta/ Xanthophyceae	95	*Xanthonema cf. hormidioides*	**TTTAGGG**	1	13	+ (b,f)	2	**6**	2	0	2	0	+/- (a,f); +/- (c,h)
	202	*Pleurochloris meiringensis*	**TTTAGGG**	4	39	+++ (b,f)	0	**39**	0	0	0	0	+++ (a,f);, +/- (c,h)
	203	*Xanthonema hormidioides*	**TTTAGGG**	3	35	+++ (a,f)	7	**10**	0	1	4	0	
	204	*Heterococcus protonematoide*	**TTTAGGG**	5	85	++ (b,f)	0	**83**	0	0	1	1	– (a,f)
	205	*Botrydiopsis intercedens*	**TTTAGGG**	3	53	+++ (a,f)	6	**42**	2	0	3	0	+++ (b,f); ++ (c,f); +/- (c,h)
Ochrophyta/ Eustigmatophyceae	133	*Eustigmatos polyphem*	**unknown**	N.A.									- (b,f); - (c,h); Table S4
	201	*Vischeria punctata*	**unknown**	N.A.									- (c,h); Table S4

Group	No.	Species	Telomere type	Minisat. variants	n	Signal							
Ochrophyta/ Bacillariophyceae	**231**	*Phaeodactylum tricornutum*	**unknown**	N.A.									- (c,f); - (c,h);
Haptophyta	**210**	*Pavlova lutheri*	**TTAGGG**	7	216	+++ (c,h)	**215**	0	0	0	0	1	
Alveolata	**233**	*Chromera velia*	**TTTAGGG**	4	90	++ (b,f)	0	**79**	0	0	11	0	
Euglenophyceae	**185**	*Euglena anabaena*	**TTAGGG**	2	39	+ (b,f)	**33**	6	0	0	0	0	
				2	37	+++ (c,h)	**35**	2	0	0	0	0	
	206	*Euglena stellata*	**6-nt (ttaggg)**	N.A.									+++ (c,h)
	207	*Euglena geniculata*	**TTAGGG**	3	49	++ (c,h)	**49**	0	0	0	0	0	- (b,f)
				2	39	+++ (a,f)	**38**	1	0	0	0	0	

Telomere type was determined from the nucleotide periodicity of the TRAP products (lower case letters) or from cloned TRAP products (upper case letters). Primers – a, pSSyF; b, 47F; c, GG(21); d, TS21; e, CAMV; f, TELPR30-3A; g, TELPR; h, HUTC; i, T4AG2-C; j, T4AG2-PR; k, CHTRTTRAPRev1; l, TTATAG3-C. N.A. not analyzed. Minisatellite variants refers to number of indicated telomere motifs detected among telomerase products, the major telomere type in bold. Errors classified as T- or G-slippage are additional T or G nucleotides incorporated into the reiterated unit. Other errors, mostly nucleotide deletion or A/G substitution, are considered nucleotide misincorporations (mismatches).

Taxon	Telomere repeat
Euglenophyceae	**TTAGGG**
Kinetoplastida	**TTAGGG**
Heterolobosea	tttggg
Jakobida	ttaggg
Fornicata	TAGGG
Malawimonadida	ttaggg
Lobosea	ttaggg
Myxogastria	**TTAGGG**
Dictyostelida	**TTAGGG**, **TAAGGG**, **NCT**
Metazoa	**TTAGGG** (TTAGG, **TCAGG**, **TTAGGC**, **TTGCA**, **NCT**)
Choanoflagellata	**TTAGGG**
Microsporidia	**UNUSUAL MIXED REPEATS**
Fungi	**TTAGGG** (**VARIOUS YEAST REPEATS**, **TTGTGG**, **TTAGGGGG**, tttaggg, tttagg, **etc.**)
Haptophyta	**TTAGGG**
Rhizaria	**TTAGGG**
Ciliata	**TTTTGGGG**, **TTGGGG**, **TTTGGG**
Dinoflagellata	**TTTAGGG**
Perkisea	tttcggg
Chromerida	**TTTAGGG**
Apicomplexa	**TTTAGGG**, **TTTTAGGG**, TT(T/C)AGGG, TTTAGG
Labyrinthulida	ttagg
Opalinata	ttaggg
Oomycetes	tttaggg
Bacillariophyceae	ttaggg
Eustigmatophyceae	ttaggg?
Pelagophyceae	ttaggg
Xanthophyceae	**TTTAGGG**
Phaeophyceae	ttaggg
Cryptophyceae	**TTTAGGG**
Goniomonadida	tttaggg
Glaucophyta	ttaggg
Cyanidiophyceae	**AATGGGGGG**, tttatt(t)aggg
Porphyridiophyceae	???
Stylonematophyceae	???
Rhodellophyceae	???
Bangiophyceae	ttaggg?
Klebsormidiophyceae	**TTTTAGGG**, **TTTTAGG**
Zygnematophyceae	**TTTAGGG**
Embryophyta	**TTTAGGG** (**TTAGGG**, **NCT**)
Mamiellophyceae	tttaggg
Chlorodendrophyceae	**TTTAGGG**
Ulvophyceae	???
Trebouxiophyceae	**TTTAGGG**
Chlorophyceae	**TTTAGGG** (**TTTTAGGG**, **TTAGGG**)

Clade labels: Euglenozoa, root 2, Amoebozoa, root 3, root 4, Opisthokonta, Alveolata, Myzozoa, Stramenopiles, Ochrophyta, Cryptomonadea, Rhodophyta, Streptophyta, Chlorophyta

A Glaucophyta
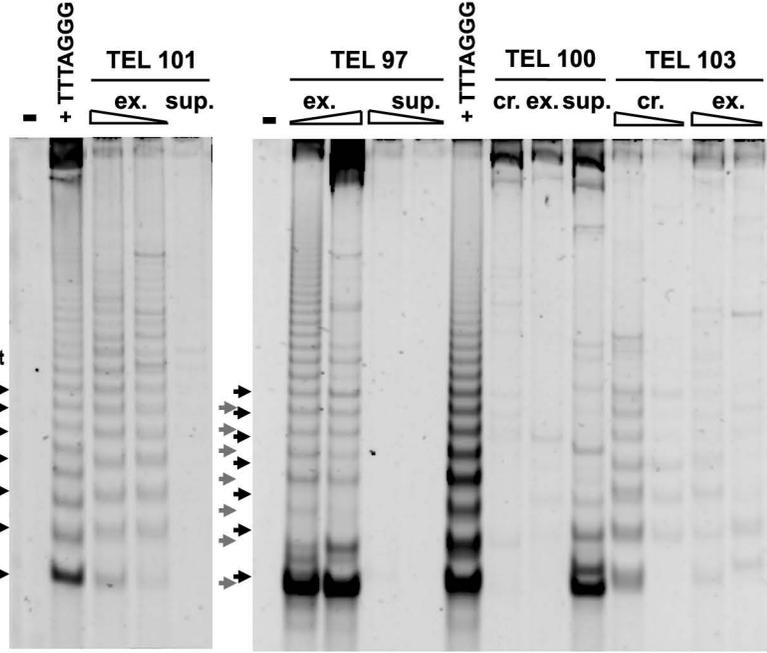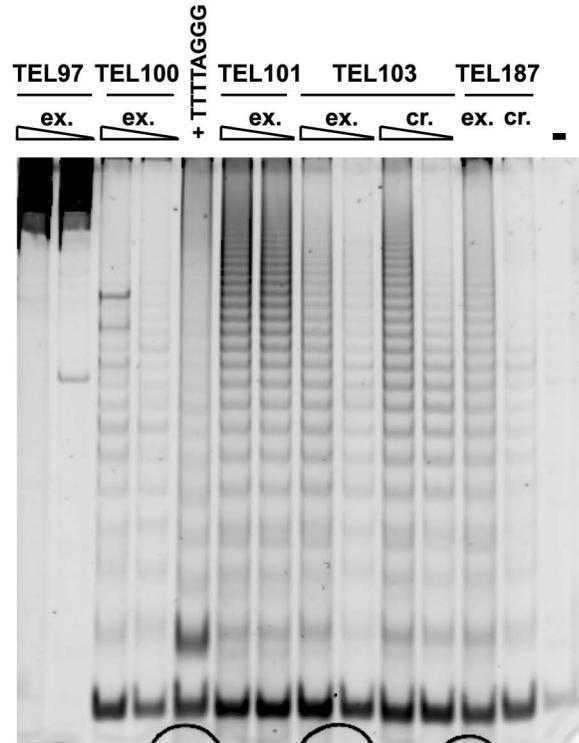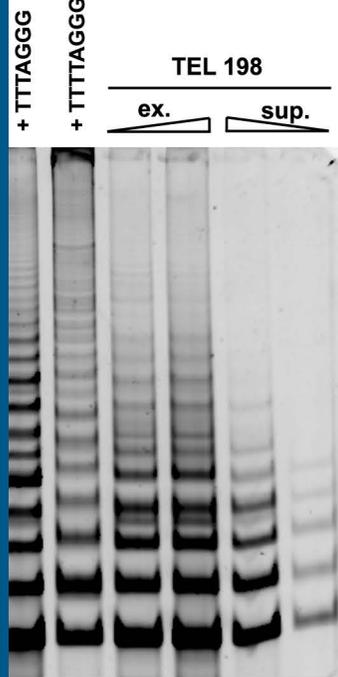
B Rhodophyta

C Chlorodendrophyceae

D Trebouxiophyceae

E Ulvophyceae

Zygnematophyceae

G Klebsormidiophyceae

TS21 X CHTRTTTRAPRev1

pSSyF X T4AG2-C

8-nt
7-nt

glenophyceae

Euglena stellata cr. ex. sup. + TTAGGG
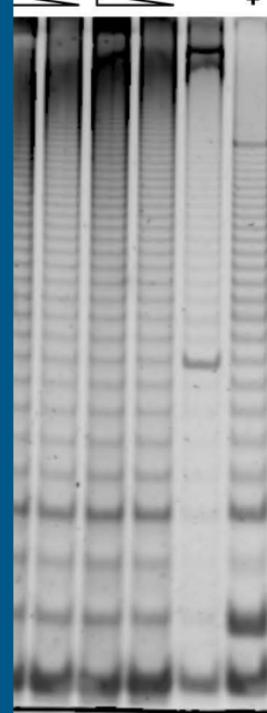
Haptophyta

Pavlova lutheri + TTAGGG cr. ex. sup.
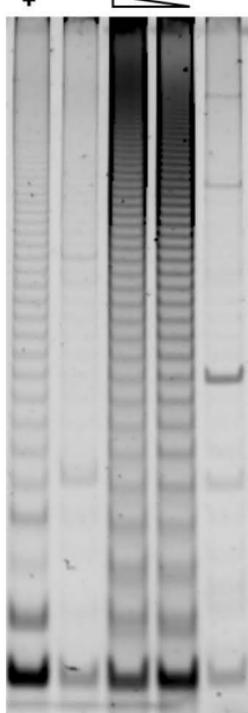
B

Alveolata

Chromera velia sup. ex. + TTTAGGG + TTTTAGGG
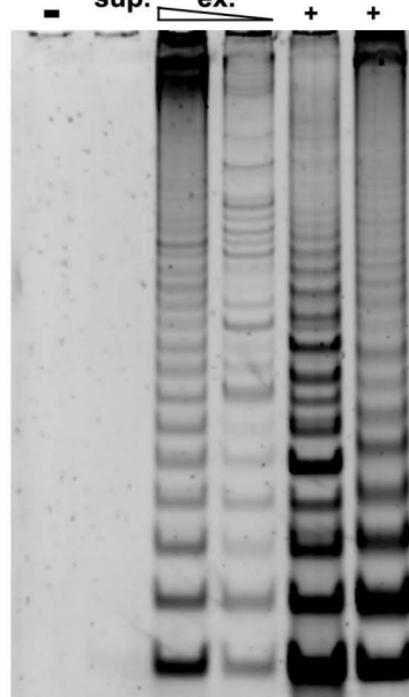
Xanthophyceae

+ TTTAGGG *Pleurochloris meiringensis* ex. sup. −

C
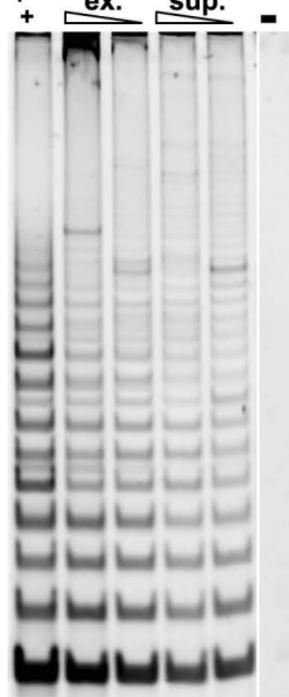
Bacillariophyceae

Phaeodactylum tricornutum − + TTAGGG

Eustigmatophyceae

+ TTAGGG − TEL 201 TEL 133 TEL 133 + TTTAGGG + TTTTAGGG −
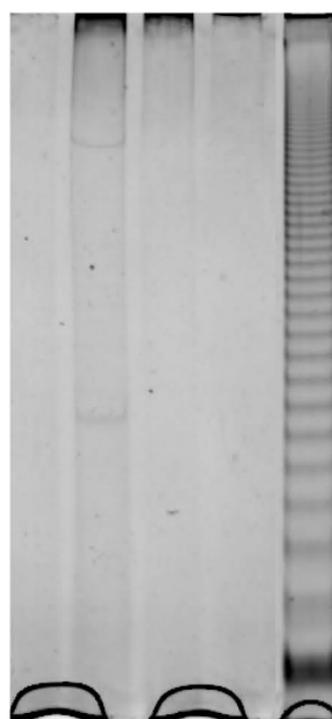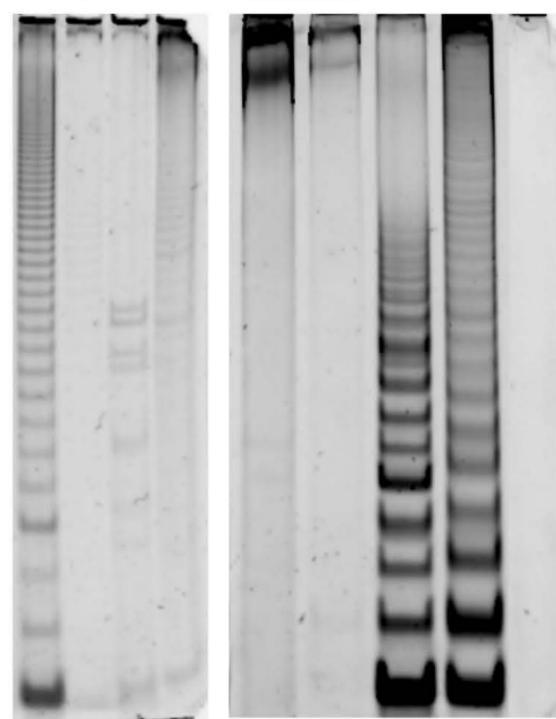
GG(21) x HUTC

GG(21) x HUTC 47F X TELPR30-3A

	ATSB	CHSB	HUSB	T4AG2-SB	T2CG3-SB	Red alga-SB	T3G3-SB	TTCAG3-SB	T3AGGC-SB	rDNA

...phyta
195 *Glaucocystis nostochinearum*

...yta
131 *Porphyridium purpureum*
213 *Rhodella maculata*

...
233 *Chromera velia*

...ophyceae
206 *Euglena stellata*
207 *Euglena geniculata*

...atophyceae
133 *Eustigmatos polyphem*
201 *Vischeria punctata*

...hyceae
202 *Pleurochloris meiringensis*
203 *Xanthonema debile*
204 *Heterococcus protonematoides*
205 *Botrydiopsis intercedens*

...idiophyceae
97 *Klebsormidium subtilissimum*
100 *Klebsormidium dissectum*
101 *Klebsormidium flaccidum* (n.a. ... n.a. ... n.a. n.a.)
103 *Klebsormidium nitens*
187 *Klebsormidium crenulatum*

...atophyceae
181 *Zygnema circumcarinatum*

...ceae
86 *Planophila laetevirens*
94 *Pseudendocloniopsis botryoides*
111 *Pseudendoclonium printzii*
124 *Pseudendoclonium basiliense*
139 *Pirula salina*

...ophyceae
88 *Heterochlorella luteoviridis*
90 *Auxenochlorella protothecoides*
121 *Dictyochloropsis irregularis*

...yceae
89 *Muriella decolor*
98 *Chromochloris zofingiensis*
108 *Neochloris conjuncta* (n.a. n.a. n.a. ... n.a. n.a.)
140 *Follicularia paradoxalis* (n.a. n.a. n.a. ... n.a. n.a.)

...endrophyceae
212 *Tetraselmis striata*
211 *Tetraselmis chui*

...NA
101 *Chlorella vulgaris*
...mo sapiens
Ipheion uniflorum

SMBE

Euglena stellata