Genetic diversity in chrysophytes: Comparison of different gene markers

Christina Bock¹, Antonis Chatzinotas²,³ & Jens Boenigk¹,⁴

¹Department of Biodiversity, University Duisburg–Essen, Universitaetsstrasse 5, D–45117 Essen, Germany
²Department of Environmental Microbiology, Helmholtz Centre for Environmental Research – UFZ, Permoserstrasse 15, D–04318 Leipzig, Germany
³German Centre for Integrative Biodiversity Research (iDiv) Halle–Jena–Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
⁴Centre for Water and Environmental Research (ZWU), University of Duisburg–Essen, Universitaetsstr. 2, 45117 Essen, Germany

Abstract: Chrysophytes are a diverse group of protists belonging to the Stramenopiles. They encompass a broad range of nutrition strategies (phototrophy, mixotrophy and heterotrophy) and morphological forms (unicellular, colonial, branched, non–scaled to silica–scaled). However, taxonomy, identification and delineation of genera and species are hampered by the polyphyletic origin of several morphotypes and the low resolution of morphological features; moreover, description of these features requires in many cases special equipment like TEM or SEM. To test the congruent phylogeny, genetic resolution and to reveal the genetic diversity of chrysophytes, we evaluated in this study different gene markers, in particular SSU and LSU rRNA genes, ITS, and COI in terms of phylogenetic clustering, separation of phylogenetic lineages and genetic distances within and between clades. Our results showed that the general topology of the SSU rRNA gene phylogeny is similar to that of the LSU rRNA gene phylogeny, with a higher genetic divergence for the latter. The topology of the COI phylogeny differed from the ribosomal data, and showed the highest genetic divergence within groups and between groups for all compared markers. The 5.8S rRNA gene offered only a limited phylogenetic resolution with low genetic divergence. ITS showed a high genetic divergence, but phylogenetic analyses were hampered by high variance in sequence lengths between clades.

Key words: Chrysophyceae, SSU, LSU, COI, ITS

Introduction

Chrysophyceae sensu lato (chrysophytes), the so called “golden algae”, comprise the groups of Synurales, Ochromonadales, Chromulinales and others within the Stramenopiles (Andersen et al. 1999; Kristiansen & Preisig 2007; Jordan & Iwataki 2012). They represent an abundant and in some environments even dominating group of planktonic protists (Rott 1988; Tolotti et al. 2003, 2006; Kammerlander et al. 2015). They occur in a broad range of ecotypes and comprise heterokont biflagellate eukaryotes, which received their common name due to the golden–brown color of the accessory pigment fucoxanthin present in some of their members (Bellinger & Sigee 2013). The morphology of chrysophytes ranges from unicellular (e.g. Ochromonas) to colonial (e.g. Synura) or branching forms (e.g. Hydrurus). Non–scaled as well as silica–scaled taxa (belonging either to the Synurales [or Synurophyceae], Paraphysomonadida or Chrysosphaerella) belong to this diverse group of protists (Škaloud et al. 2013; Škaloudová & Škaloud 2013). They encompass varying nutritional strategies, including phototrophy, mixotrophy and heterotrophy with both phagotrophy by ingesting bacteria (Thingstad et al. 1996; Kamjunke et al. 2007; Unrein et al. 2007) or osmotrophy (Pringsheim 1952). Despite the broad range of morphological forms, identification and thus distinction of members belonging to the same genus or even to closely related genera is often problematic. Species delineations are usually based on small morphological differences. These features as for example the scale morphology in Synura or Paraphysomonas can often only be detected with special equipment like transmission electron microscopy (TEM) or scanning electron microscopy (SEM) (Škaloud et al. 2012; ScoBle & CaVaLier–Smith 2014). The fact that morphology may change under different environmental conditions because of phenotypic plasticity adds to the difficulties of species delimitation (Luo et al. 2006). With the advent of marker–based phylogenetic ap-
proaches, new light was shed into the taxonomy and phylogenetic relationship of this group (Andersen et al. 1999). Molecular analyses revealed that some morphotypes are in fact polyphyletic and form separate clusters within the chrysophytes, showing that classification solely based on morphology is sometimes misleading (Grossmann et al. 2016). One good example are colorless, nonscaled chrysophytes which are genetically far more diverse than morphology reflects (Boenigk et al. 2005; Pfandl et al. 2009; Findenig et al. 2010). Several new genera were recently erected mainly based on molecular identification, since light microscopy could not distinguish between isolates belonging to different phylogenetic lineages (Grossmann et al. 2016). Even TEM and SEM offer only a limited amount of information within this species complex (Findenig et al. 2010). A huge hidden diversity was also observed within species complexes of Paraphysomonas, which lead to the establishment of several new species; it is very likely, that even more species exist since environmental surveys have detected additional undescribed phylogenetic lineages within the Paraphysomonadidae (Scoble & Cavalier-Smith 2014).

In the past, biodiversity and monitoring questions in freshwater systems were mainly addressed based on morphological data (e.g. Padisák et al. 1998). With the new knowledge gained from molecular biological approaches, environmental surveys can overcome the risks associated with campaigns solely based on morphology of organisms, namely to miss cryptic species complexes and to lose information regarding their geographical and/or seasonal distribution.

Many studies base their results on SSU rRNA gene data. Even if the SSU rRNA gene offers a good backbone for the phylogeny within protist groups, it was shown in several studies that this genetic marker might be too conserved to reveal the occurring genetic diversity within related lineages (Bock et al. 2010; Nassonova et al. 2010; Skaloud et al. 2012). Within the Synura petersenii complex, six genetic lineages were revealed by multiple genetic markers (internal transcribed spacer rDNA, psaA, rbcL and COI) which were further supported by morphological data based on TEM and SEM of silica scales (Skaloud et al. 2010). This combination of morphological and molecular data allowed the authors an additional interpretation on the diversity and distribution patterns. These results demonstrate the need of phylogenetic markers with high resolution and reliable stable topology to distinguish between closely related species and thus to obtain reliable information about phylogenetic relationships of chrysophytes and their distribution.

Here, we present comparative analyses of the SSU rRNA gene, the ITS1–5.8S–ITS2 sequence region, the D1–D2 region of the LSU rRNA gene and COI sequences from a huge range of different chrysophytes. We analyzed phylogeny, genetic distances within and between clades with respect to reliable and stable topology and phylogenetic lineage distinction. The main aim of this paper was to find a suitable marker region for a stable phylogeny, which further allows a good separation between closely related lineages.

**Material and Methods**

**Isolates and culture conditions.** Most strains were isolated during sampling campaigns in 2002 and 2006. Isolation and maintenance of strains were conducted as already described in Boenigk et al. (2005). For details on origin of the isolates see table S1. Isolates were discarded after the DNA-extraction.

**DNA isolation, amplification and sequencing.** For DNA isolation, cells were harvested by centrifugation at 14000×g and the resulting pellet was used for DNA extraction. DNA was extracted using the DNeasy Tissue kit (Qiagen Gmbh Hilden, Germany) and amplified separately for the different gene regions (details on used primers see table S2).

**Phylogenetic analyses.** Four different alignments were constructed for the phylogenetic analyses with the newly obtained sequences and additional sequences obtained from NCBI (http://www.ncbi.nlm.nih.gov/). Reference sequences representing the different clades of the Chrysophyceae were selected on the basis of previous publications (e.g. Medinger et al. 2010; Silverfeld et al. 2010; Klaveness et al. 2011; Bock et al. 2014). The sequences were initially aligned using ClustalW (Thompson et al. 1994) and adjusted by eye using the SequentiX Alignment Editor (Heppelle 2004). Only highly ambitious regions were excluded from the analyses. The phylogeny of the SSU rRNA gene (Fig 1–2, Fig S1) was inferred from an aligned dataset including 238 sequences based on 1206 characters. The sequences Synchronema grande DQ788730, Leukarachnion sp. FJ356265 were chosen as outgroup according to Bock et al. (2014). LSU rRNA gene phylogeny was inferred from 175 sequences with 615 characters (Fig 3, Fig S2). COI phylogeny was inferred from 139 sequences with 421 characters (Fig 4, Fig S4). ITS phylogeny was inferred from 17 sequences with 902 characters (Fig 5). The 5.8S rRNA gene phylogeny was inferred with 95 sequences and 132 characters (Fig S3). All phylogenetic trees (except for ITS) presented here were inferred by maximum likelihood settings with RAXML 8.2.9 (Stamatakis 2014) using the Cipres gateway (Miller et al. 2010). Models for ITS were used as proposed by Treefinder under AICc criteria: partitioned dataset (ITS1: HKY[Optimum,Empirical];G[Optimum]:5, 5.8S: HKY[3,1,1,1,1,3,Empirical], ITS2: TVM[Optimum,Empirical];G[Optimum]:5]; 5.8S: HKY[3,1,1,1,1,3,Empirical]). To test the confidence of the tree topologies, bootstrap analyses were carried out for distance (neighbor-joining, NJ; 1000 pseudoreplicates) and maximum parsimony analyses (MP, 1000 pseudoreplicates; with heuristic search options based on random taxon addition, tree-bisection-reconnection (TBR) branch swapping algorithm and Multrees option enabled) using PAUP*, portable version 4.0b10 (Swoford 1993), and for maximum likelihood (ML, 1000 pseudoreplicates; settings as described above) respectively. Bayesian analyses were carried out using MrBayes 3.2.6 via the Cipres portal (Miller et al. 2010). For each presented tree, two runs with four chains of Markov
chain Monte Carlo (MCMC) iterations were performed with 10,000,000 generations with tree sampling every 100 generations. The GTR+G model with gamma shape parameter was chosen for the datasets. Tracer V1.4 (RAMBAUT & DRUMMOND 2007) was used to check the stationary phase and to identify an appropriate burn in value. The first 25% of the trees were discarded as burn-in and 50% majority-rule consensus trees were calculated for posterior probabilities.

Average genetic distances in different clades. Average mean distances based on the K2P option were calculated for different clades and sequences by using Mega 5.2.2. All ambiguous positions were removed for each sequence pair. The sequences were group based on their phylogenetic position within the SSU rRNA gene phylogeny: Ochromonadales (containing the genera Poteriospumella, Spumella, Pedospumella and Dinobryon); Synurales, Paraphysomonadida, Apoikiida, Chromulinales (details about assignment to groups see table S1). Average mean divergences within groups (Fig 6) and between groups (Fig 7) were compared for the gene regions: partial SSU rRNA gene (1165 characters), V4 region of the SSU rRNA gene (197 characters), 5.8S rRNA gene (132 characters), D1–D2 region of LSU rRNA gene (615 characters) and COI (420 characters). Amount of sequences compared are noted in Tables S3, S4 and correspond to the sequences within the phylogenies. The Ochromonadales were separated into the genera Poteriospumella, Pedospumella and Spumella for the ITS1 and ITS2 divergence since suitable alignment over the whole clade was not possible.

RESULTS

In a first step, phylogenies were calculated to compare the cluster pattern of the isolates for the different gene regions. Subsequently, average mean distances based on the K2P model within and between different clades were calculated for the different gene markers. All genetic regions, including variable parts, were included in the phylogenies to allow a detailed comparison between genetic distances and corresponding phylogeny. 360 new chrysophyte sequences including LSU rRNA gene (rRNA), SSU rRNA gene (rRNA), COI (including sequences with TGA codon) and ITS were obtained by PCR for the purpose of this study and have been made publicly available (details on accession numbers and analyzed strains see Table S1). Due to amplification problems, not all isolates could be sequenced for all mentioned marker regions. Selected sequences of the occurring clades available from GenBank were downloaded and used as references within the phylogenies (Figs 1–5, S1–4).

The general topology of SSU rRNA gene, LSU rRNA gene, COI

The general topology of the SSU phylogeny is in congruence with previously published studies. Labeling of clades was done according to PFANDEL et al. (2009) and GROSSMANN et al. (2016). Clades including the Chromulinales, Paraphysomonadida, Hydrurales, Hibberdiales, Synurales and Ochromonadales were resolved in our study with various statistical supports (Figs 1–2, Fig S1). The unrooted LSU rRNA gene phylogeny (Fig 3, Fig S2) showed a similar topology compared to the SSU phylogeny. Several clades could only be represented with one sequence in the LSU rRNA gene phylogeny as for example Hydrurus and Segregatospumella. The order of the main groups (e.g. Apoikiida, Chromulinales, Synurales, Hydrurales, Ochromonadales)...

Table 1. Estimates of average evolutionary divergence between sequences for the ITS1-5.8S-ITS2. All positions containing gaps and missing data were eliminated. Analyses were conducted using the Kimura 2-parameter model. Evolutionary analyses were conducted in MEGA5 (TAMURA et al. 2011). The number of base substitutions per site from averaging over all sequence pairs within each group are shown.

<table>
<thead>
<tr>
<th>Region</th>
<th>194f</th>
<th>1305</th>
<th>1058</th>
<th>1031</th>
<th>186hm</th>
<th>199hm</th>
<th>367hm</th>
<th>JBL14</th>
<th>JBM09</th>
<th>391f</th>
<th>AR4A6</th>
<th>JRNA45</th>
<th>9-10-C1</th>
<th>1020</th>
<th>1026</th>
</tr>
</thead>
<tbody>
<tr>
<td>194f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1813</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0462</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0462</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>186hm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>199hm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>367hm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0126</td>
<td>0.0126</td>
<td>0.0126</td>
<td>0.0126</td>
<td>0.0126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JBL14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1302</td>
<td>0.1302</td>
<td>0.1302</td>
<td>0.1302</td>
<td>0.1302</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JBM09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1349</td>
<td>0.1349</td>
<td>0.1349</td>
<td>0.1349</td>
<td>0.1349</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>391f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0126</td>
<td>0.0126</td>
<td>0.0126</td>
<td>0.0126</td>
<td>0.0126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR4A6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2051</td>
<td>0.2051</td>
<td>0.2051</td>
<td>0.2051</td>
<td>0.2051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JRNA45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2125</td>
<td>0.2125</td>
<td>0.2125</td>
<td>0.2125</td>
<td>0.2125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-10-C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3177</td>
<td>0.3177</td>
<td>0.3177</td>
<td>0.3177</td>
<td>0.3177</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2409</td>
<td>0.2409</td>
<td>0.2409</td>
<td>0.2409</td>
<td>0.2409</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2482</td>
<td>0.2482</td>
<td>0.2482</td>
<td>0.2482</td>
<td>0.2482</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JBNZ39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2188</td>
<td>0.2188</td>
<td>0.2188</td>
<td>0.2188</td>
<td>0.2188</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig 1. Phylogeny of the chrysophytes obtained by maximum likelihood inference of the SSU rDNA dataset. Values at the nodes indicate statistical support estimated by four methods—Bayesian inference (MB), maximum likelihood (ML), maximum parsimony (MP), distance (neighbor-joining; NJ). Hyphens indicate support below 50% for ML, MP, NJ and below 0.95 for MB. Scale bar represents the expected number of substitutions per site. Some phylogenetic lineages are collapsed and only the number of sequences within the lineage is shown. Ochromonadales are omitted and shown in Fig 2. The complete phylogeny including all used sequences is shown in Fig S1.

Fig 2. The complete phylogeny including all used sequences is shown in Fig S1.

adale) were not statistically supported in either SSU or in LSU phylogeny. In the SSU phylogeny the clade containing Segregatospumella clusters outside the Ochromonadales forming an own clade (without statistical support for the actual placement) (Figs 1–2). Based on LSU rRNA gene data, Segregatospumella clusters with a long branch within the Ochromonadales without statistical support (Fig 3). The 5.8S rRNA gene phylogeny showed only a limited resolution for different phylogenetic lineages (Fig S3).

The COI phylogeny (Fig 4, Fig S4) shows a different topology and clustering of the strains as compared to the ribosomal data. Support for main phylogenetic lineages and groups was not achieved. Even supported groups based on SSU have no support based on COI e.g. Synurales, Paraphysomonadida or Ochro-

Average genetic distances in different clades
Average mean divergences within groups were compared for the gene regions: partial 18S rRNA gene, V4 region of the SSU rRNA gene, 5.8S rRNA gene, D1–
Fig 2. Phylogeny of the Ochromonadales obtained by maximum likelihood inference of the SSU rDNA dataset. Values at the nodes indicate statistical support estimated by four methods—Bayesian inference (MB), maximum likelihood (ML), maximum parsimony (MP), distance (neighbor joining; NJ). Hyphens indicate support below 50% for ML, MP, NJ and below 0.95 for MB. Scale bar represents the expected number of substitutions per site. Some phylogenetic lineages are collapsed and only number of sequences within the lineage is shown.
D2 region of LSU rRNA gene and COI (Fig 6). The highest divergence within groups was observed for the COI region (0.21–0.28), followed by LSU rRNA gene (0.22–0.11) and the V4 rRNA gene region (0.04–0.11). 5.8S rRNA gene and SSU show the smallest variation within groups (0.01–0.06 respectively 0.02–0.04) (details see table S3). The mean divergence between groups for the analyzed markers can be observed in Fig 7 and table S4. Highest average divergences between groups were observed for the COI gene, LSU rRNA gene shows bit lower divergences. SSU divergences ranges between 0.05–0.06; V4 varies between 0.11–0.17; 5.8S rRNA gene between 0.07–0.11. The ITS was exemplarily analyzed for divergences within groups of the C2–clade of the Ochromonadales (see table 1, Fig 5). The distances vary between 0.0–0.34.
Fig 4. Phylogeny of the chrysophytes obtained by maximum likelihood inference of the COI dataset. Values at the nodes indicate statistical support estimated by four methods—Bayesian inference (MB), maximum likelihood (ML), maximum parsimony (MP), distance (neighbor joining; NJ). Hyphens indicate support below 50% for ML, MP, NJ and below 0.95 for MB. Scale bar represents the expected number of substitutions per site. Some phylogenetic lineages are collapsed and only number of sequences within the lineage is shown. Complete phylogeny including all sequences used is shown in figure S4.
Figure 5. Phylogeny of the C2-clade obtained by maximum likelihood inference of the ITS rDNA dataset. Values at the nodes indicate statistical support estimated by four methods—Bayesian inference (MB), maximum likelihood (ML), maximum parsimony (MP), distance (neighbor joining; NJ). Hyphens indicate support below 50% for ML, MP, NJ and below 0.95 for MB. Scale bar represents the expected number of substitutions per site.

Figure 6. Average mean divergences within groups of chrysophytes. Estimates were conducted using the Kimura 2-parameter model using MEGA5 (Tamura et al. 2011). All ambiguous positions were removed for each sequence pair. Amount of compared sequences (n) varied between genes and groups: Ochromonadales (Poteriospumella (18S n=14, V4 n=15, 5.8S n=15, LSU n=16, COI n=24), Spumella (18S n=24, V4 n=25, 5.8S n= 22, LSU n=33, COI n=11, ITS1 n=17, ITS2 n=17), Pedospumella (18S n=13, V4 n=13, 5.8S n=8, LSU n=11, COI n=10, ITS1 n=9, ITS2 n=9), Dinochrysophyceae (18S n=38, V4 n=38, 5.8S n=7, LSU n=35, COI n=20, ITS1 n=7, ITS2 n=7)), Synurales (18S n=24, V4 n=24, 5.8S n=13, LSU n=27, COI n=10, ITS1 n=11, ITS2 n=11), Paraphysomonadida (18S n=15, V4 n=16, 5.8S n=7, LSU n=12, COI n=14, ITS1 n=0, ITS2 n=0), Apokida (18S n=3, V4 n=3, 5.8S n=2, LSU n=4, COI n=3, ITS1 n=2, ITS2 n=2), Chromulinales (18S n=3, V4 n=4, 5.8S n=1, LSU n=4, COI n=2, ITS1 n=0, ITS2 n=0). The Ochromonadales were separated into the genera Poteriospumella, Pedospumella and Spumella for the ITS1 and ITS2 divergence since suitable alignment over the whole clade was not possible. Details shown in table S3.
Analyses between clades for the ITS region were not possible due to the high variance in lengths.

DISCUSSION

The application of molecular markers for distinguishing protist species and/or assessing diversity patterns in different environments is nowadays an established method (Krienitz & Bock 2012; Leliaert et al. 2014). In phylogenetic studies, gene markers like the SSU rRNA gene, sometimes in combination with other markers, are regularly used to reveal phylogenetic relationships (Jo et al. 2013; Škaloud et al. 2014). Our results confirm previous results that the SSU offers good phylogenetic resolution on genus level, but is in part too conserved to differentiate between closely related lineages as already reported from different other organisms (Piganeau et al. 2011; Leliaert et al. 2014; Annenкова et al. 2015). More rarely is the LSU used to show phylogenetic relationships. In our study the D1–D2 region showed high potential as a phylogenetic marker. The phylogeny is in general congruent with phylogenies obtained from SSU data (see Fig 1, Fig 3; Grossmann et al. 2016) even if statistical support for main groups is not as high. The LSU shows high genetic variance within clades and between clades. Similar good results were already obtained from studies on cercozoa (Wylezich et al. 2010) and ciliates (Stoeck et al. 2014). In addition, the analyzed D1–D2 is a relative short region which makes it suitable for NGS sequencing techniques. In how far the primers might be suitable for universal amplification of protists needs still to be addressed.

Several studies on chrysophytes have evaluated the COI as a marker region for phylogenetic analyses. It was shown, for instance, that it is indeed a highly variable region, suitable for distinguishing closely related lineages (Jost et al. 2010; Bock et al. 2014). Jost...
et al. (2010) showed that the COI region is suitable to distinguish *Dinobryon divergens* from *D. bavaricum* strains, which is nearly impossible based on SSU rRNA gene analyses. Our analyses confirmed the high degree of variation for chrysophytes. Nevertheless, we came across several problems. Even if the average divergence between groups is high, the evolutionary divergence within groups is nearly as high (Figs 6,7). The appearance of TGA codons in the C1-C2 kinship is also striking. The extent to which TGA serves as a stop codon is controversially discussed. Studies on closely related groups suggest that the TGA codon in COI is assigned for Trp (Ehara et al. 2000; Inagah et al. 1998; Knight et al. 2001 and citations within). The clustering of several sequences outside the Chrysophyceae (however without support) lead to the assumption of the occurrence of multiple copies, homologous or paralogous genes (see Fig 3). This is a known problem and was e.g. found in naked lobose amoebae as well (Nassonova et al. 2010). Even if paralogous genes are distinguishable in studies based on clonal isolates, they are difficult to recognize in NGS studies. Furthermore, we encountered problems in the amplification of the marker gene. Our chosen primers seemed to be too constricted, and marker genes of several strains could not be amplified. In diatoms, different COI primers were used for raphid pennates and centric diatoms (Davis et al. 2007); in fact, different COI primers were even needed for different ciliate genera (Barth et al. 2006; Lynn et al. 2006).

Massive parallel sequencing techniques provide a sufficient sequence depth for environmental studies, yet, most results are based on short sequencing reads only (Massana et al. 2015; De Vargas et al. 2015; Le Bescot et al. 2016). To receive reliable information about species compositions and biodiversity patterns it is important to use a well suited genetic marker. The marker should offer the advantage of a highly variable region, which can be amplified by universal primers and offers good background information in existing databases to identify the reads preferable down to species level (Bock et al. 2014; Zimmermann et al. 2014). Until now, the discussion on the most suitable marker gene for protists is still ongoing (Pawlowski et al. 2012). A first step towards a uniform protocol for protists was proposed by the CBOL protist working group (initiated by the consortium for the barcode of life; http://www.barcodeoflife.org/content/community/projects?page=1). They suggested a two-step barcoding approach to address protist biodiversity (Pawlowski et al. 2012). In a nested strategy, the variable V4 region of the SSU rRNA gene is proposed as a pre-barcode, followed by a group specific barcode. Our analyses showed that the V4 region is a variable marker for chrysophytes and shows a moderate sequencing distance gap between average within group variance and between groups variance. One exception is the *Paraphysomonas* clade, where the variance within the clade is sometimes higher than the variance to the other analyzed clades. This is due to the relative high distance of the lineage containing *Paraphysomonas vestita* and relatives (our analyzed strains JBAS 37, JBAF/S47, JBAF49, JBAF/S48, JBAU44).

Adjoining the SSU rRNA gene is the ITS region, comprising the ITS1, 5.8S rRNA gene and ITS2. The ITS is often used to base taxonomic measures on the occurrence of complementary base changes within the secondary structure of either ITS1 or ITS2 or both (Luo et al. 2006; Bock et al. 2011a,b; da Silva et al. 2016). Darienko et al. (2015) even recommended the use of the V9 and ITS region as barcode for the green algae *Coccomyxa*. Our analyses of the ITS showed a high degree of genetic diversity (see Fig 6, table 1) and variance in length (data not shown), hampering the unambiguously aligning of sequences between clades. Similar difficulties were already reported from Škaloudová & Škaloud (2013). The secondary structure may help by aligning diverse sequences, but so far different models exist for members of the Chrysophyceae (Škaloudová & Škaloud 2013; Ankenbrand et al. 2015). As we could show with the Ochromonadales, an alignment between clades of this highly variable region is not possible without a general model. Nevertheless, even without secondary structure information, the ITS offers high resolution to distinguish related species within clades as already demonstrated (Jost et al. 2010; Škaloud et al. 2012).

To conclude, our results showed the importance of a detailed evaluation of phylogenetic markers for addressing taxonomic and biodiversity questions. Nearly all markers analyzed in this study show a promising variable region for distinguishing phylogenetic lineages. The ITS, LSU rRNA gene and COI are the most diverse marker in terms of genetic distances. In addition, also the V4 region of the SSU rRNA gene shows a promising high variability to distinguish phylogenetic lineages. The use of the SSU–ITS region or the D1–D2 region of the LSU rRNA gene is to be preferred to COI in order to avoid the potentially misleading clustering of some strains.

**Acknowledgements**

We thank Verena Jaschik and Karin Pfandl for technical assistance and sequencing of strains. Funding of the study by the DFG Projekt BO 3245/17–2 and DFG Projekt BO 3245/2–1 is gratefully acknowledged.

**References**


KNIGHT, RD; FREELAND, SJ & LANDWEBER LF. (2001): Rewir-
ing the keyboard: evolvability of the genetic code. – Nat Rev Genet: 49–58.


Škaloud, P.; Kynčlová, A.; Benada, O.; Kofronová, O. & Škaloudová, M. (2012): Toward a revision of the genus Synura, section Peterseniaceae (Synurophyceae,