HHS Public Access

Author manuscript

J Appl Phycol. Author manuscript; available in PMC 2021 October 01.

Published in final edited form as:

J Appl Phycol. 2020 October ; 32(5): 2699–2709. doi:10.1007/s10811-020-02190-5.

Identification of Eukaryotic Microalgal Strains

Marvin W. Fawley, **Karen P. Fawley**

Division of Natural Sciences and Mathematics, University of the Ozarks, Clarksville, AR 72830, USA

Abstract

Proper identification and documentation of microalgae is often lacking in publications of applied phycology, algal physiology and biochemistry. Identification of many eukaryotic microalgae can be very daunting to the non-specialist. We present a systematic process for identifying eukaryotic microalgae using morphological evidence and DNA sequence analysis. Our intent was to provide an identification method that could be used by non-taxonomists, but which is grounded in the current techniques used by algal taxonomists. Central to the identification is database searches with DNA sequences of appropriate loci. We provide usable criteria for identification at the genus or species level, depending on the availability of sequence data in curated databases and repositories. Particular attention is paid to dealing with possible misidentifications in DNA databases and utilizing current taxonomy.

Keywords

DNA Barcoding; Microalgae; species concepts

A problem frequently associated with reports of biotechnical applications of microalgae is the identification of the alga. This issue can arise when the authors have isolated a new strain which they may have misidentified or used a strain that was isolated and misidentified by another lab. Even if a strain was obtained from a culture collection it may have been misidentified when placed in the collection or was subsequently contaminated with a different strain which has since overgrown the culture. Cultures may also be misnamed because the taxonomy has changed since the initial identification. A goal of any study in the public venue should be to accurately relate the materials and methods used such that the project could be replicated by other investigators. If the alga used in the study is misidentified, especially if it is not from an openly available collection, then the study would be difficult to replicate. Worse still, if the identity given for the alga places the organism in

marvinfawley@gmail.com.

Authors' contributions: Both authors developed the concepts. MWF wrote the bulk of the manuscript. KPF contributed to manuscript structure and edited the manuscript.

an incorrect order or class the results or the study are of questionable value until the alga is properly identified.

The purpose of this article is to provide a fundamentally sound procedure for the identification of algal strains including step-by-step instructions for the use of this method. Identification is intimately tied to the procedures used to delimit new species. We will briefly review the techniques and philosophies used when naming new taxa, and how those same techniques can then be used to identify algal strains with confidence.

History of microalgal identification techniques

Morphological identification

The methods of algal taxonomy have always been bound by the technology available to the taxonomist, and the history of identification techniques for microalgae follows this pattern. Much of the early history of algal taxonomy that we present below is a brief summary of that of Moestrup (2014). This introduction is intended to demonstrate how taxonomy changes over time and to present possible procedures for an investigator to navigate this changing taxonomic landscape.

The first discoveries of the myriad forms of microalgae came in the late 17th century with the development of light microscopes with 100x or more magnification. These organisms were initially classified and described based on the morphologies and life histories that could be seen with simple microscopes, and sometimes so little detail was provided that the description could cover a wide range of organisms. As microscopic techniques and equipment became more sophisticated, so did the ability to subdivide existing taxa into additional species. Many of these techniques also relied on chemicals that stained certain biomolecules, such as iodine stains for starch. These stains provide more detail of cellular structures that could be used to differentiate among several different groups of microalgae. For example, the early differentiation of the Chlorophyta and the group of algae now known as the Xanthophyceae was indicated by the presence of starch in the Chlorophyta and its absence in the Xanthophyceae, as well as differences in photosynthetic pigments.

Some algae, such as the diatoms and some Chrysophyta, produce silica walls and scales that are highly ornamented and, as a result, many species of these organisms can be delimited based on light microscopy. The coccolithophorids likewise produce calcium carbonate-covered scales (coccoliths) which aid in identification. Organic wall features and a high degree of overall morphological variation were also used to describe many species of other algal groups, such as the desmids, the green algal genus *Scenedesmus* (as originally described), and dinoflagellates. During this early period of algal taxonomy, those groups with highly variable morphologies received the most attention from the taxonomists, whereas those genera with very simple morphologies received little taxonomic interest. As a result, many diatom species were described, though relatively few species of simple coccoid organisms (e.g., *Chlorella*) were named. Usually distributional or environmental data were secondary considerations and not critical features for identification. This was essentially the state of algal taxonomy until the mid 20th century. Because morphological differences had always been used to name new taxa, many taxonomists considered morphological

distinctions paramount and that they should be required for naming new taxa; otherwise, how would anyone be able to tell them apart in natural samples?

Modern approaches

The situation gradually changed with the advent of new technologies. In the mid 20th century, electron microscopy emerged, both TEM and SEM, which expanded the possible morphological characteristics used for describing taxa to include ultrastructure. Scanning electron microscopy was especially important at the species identification level because it revealed even greater detail in the walls of diatoms and Scenedesmaceae, and the scales of synurophytes and coccolithophorids, for example.

Taxonomic research on algae and protists in the 1960s through 1980s typically utilized electron microscopy, photosynthetic pigments and additional biochemical features. A few studies also utilized DNA or amino acid sequence data in the 1980s. However, most early studies that included DNA sequence data focused on higher-level taxonomy and therefore this was a period of many discoveries that formed the basis for new classification systems, some of which are at least partially used today. The higher-level taxonomic work continued even after new technologies, such as the polymerase chain reaction (PCR) and Sanger DNA sequencing, simplified the procurement of DNA sequences to the point that these data were readily available to taxonomists. Especially beginning with the advent of automated sequencing in the late 1990s, DNA sequencing became arguably the most important tool for testing hypotheses of species boundaries and relationships, and this trend has continued to the present day (Leliaert et al. 2014). During this time, the use of sequences of the nuclear ribosomal RNA Internal Transcribed Spacer (ITS) regions, and especially ITS2, emerged as a tool for species delimitation, and hence, identification (Coleman 2003, 2009).

However, initial fervor over the use of DNA sequence data alone to describe new species was strongly questioned and alpha-level taxonomy has continued to advance by incorporating new ideas on species concepts and their applications (Leliaert et al. 2014). After a period of competing taxonomic philosophies, many algal taxonomists have settled on some kind of combined approach, in which both morphology and sequence data are employed and other features, such as biogeography and ecology are often considered. Underlying these approaches is the understanding that the naming of any new taxon is a hypothesis (Pante et al. 2015). As such, the new taxon can be considered the best taxonomic hypothesis that is consistent with existing data. This approach posits that an established species could be broken into two separate species based on just a single morphological distinction, or, for that matter, a single nucleotide difference in the DNA sequences. However, most taxonomists would consider such a new species to be a very weakly supported hypothesis indeed! To strengthen their taxonomic hypotheses beyond morphology, modern taxonomists examine a range of possible characters, such as DNA sequence data from multiple genes (or even genomes), and include biochemical, ecological, distributional, and life history information when possible. The more characters of all kinds that support the description of a new species, the more robust the species hypothesis and the greater the likelihood that the new species will be acceptable to the taxonomic community (De Queiroz 2007; Fawley et al. 2011). These approaches have led to several techniques that attempt to

provide objective methods of delimiting species, rather than the subjective approaches that have long been common in taxonomy (e.g., Leliaert et al. 2014[Herrera and Shank 2016).

The same principles that are used to name new species apply to the identification of an unknown algal strain. Just as the original description of a species is a hypothesis of evolutionary relationships, any identification is actually a hypothesis about the identity of the strain, within the context of the existing taxonomy. The major difference is that identification at the species level should require less breadth of evidence than a species description (Leliaert et al. 2014). However, the more evidence there is to link the strain to a species identification, the more confidence there will be in the identification. Even with the less stringent requirements, the identification of algal species may be very labor intensive, requiring highly trained individuals with excellent LM and possibly SEM equipment. For example, the discussions by Mann (1999) and Mann et al. (2010) succinctly describe the effort and funding required to identify diatoms using morphological criteria. As a result, DNA barcoding techniques are becoming very important for the identification of algae. This technique typically employs a highly variable portion of a specific gene which has been shown to possess differences at the species level. Mann et al. (2010) present a discussion of the pros and cons concerning barcode-based identification, especially for diatoms. They concluded that barcoding is the most efficient way to identify species, but that the choice of gene and region to use is critical, and that the "perfect" gene region for barcoding diatoms has not yet been determined.

We expect that the acceptable level of evidence for species identification will change over time as new technologies emerge, but at the present time, we propose that the minimum data for species identification of cultured strains should include some level of morphological evaluation and DNA sequence for an appropriate gene or region of a gene. The level of importance of these two types of data will vary with the organism that is being evaluated. For example, the basic standard for the identification of many diatom taxa is morphology of the frustule, as observed with light or scanning electron microscopy. However, analyses of DNA sequence data have revealed several cases of cryptic diversity among well-studied diatom taxa (Mann et al. 2010). In such a case, it becomes very important to provide appropriate sequence data for the strain. By appropriate data, we refer to the DNA sequence of a gene or locus that has been used to differentiate among the named species that may otherwise appear identical, based on published taxonomic studies. For the diatoms, the major curated database of DNA barcodes is Diat-barcode, which primarily catalogues data for the plastid *rbc*L gene (Rimet et al. 2019). However, several recent papers have indicated that the mitochondrial cytochrome *c* oxidase subunit one (abbreviated as COI, CO1 or cox1) locus may be a better choice for some diatoms (e.g., Kollár, et al. 2019). The use of cox1 is compromised by PCR amplification difficulties in some diatom lineages. Many other algae, such as some coccoid green algae (e.g., "*Chlorella*"), have simple morphologies that may provide little information at the species level. For such algae, one or more highly variable loci, such as ITS or *rbc*L, should suffice for identification. Once again, the locus or loci examined should be dependent on the breadth of information available in public repositories such as GenBank or curated databases, as they are developed.

Once such sequence data are available for a strain, a BLAST search of the GenBank database will usually provide the necessary identification. However, it must be remembered that the taxonomic identifications associated with GenBank accessions may be flawed. Problems can occur when GenBank records have not been updated to keep pace with taxonomic revisions, when strains used were not properly identified, or through errors during data acquisition, processing or submission. Inaccurate identifications of even well-studied taxa such as insects frequently occur when using both GenBank and the curated database BOLD, and these errors are likely to be more frequent in less-studied organisms (Pentinsaari et al. 2020). Pentinsaari et al. (2020) provide excellent examples illustrating how flawed sequences in these databases or in subsequent studies originated.

Some additional discretion is also necessary when using BLAST results. Unless the new sequence exactly matches the sequence in GenBank there may be some uncertainty as to the species level identity. If there are some differences, then the level of difference should be noted (as, for example, 655 out of 660 total bases are identical for the ITS1/5.8S rDNA/ITS2 region or a table of percent similarities).

Below we present a step-by-step method (Fig 1) that should suffice to identify many algae in culture to the species level, but with several caveats concerning possible uncertainties. This procedure is consistent with the recommendations of the Consortium for the Barcode of Life (Pawlowski et al. 2012). Hadi et al. (2016) demonstrated the utility of similar methods for identifying strains from an algal collection, although they used different criteria for accepting species identifications than we propose.

1) Pure culture, at least unialgal or uni-eukaryotic.—The culture must be free of contaminating organisms that could compromise PCR and DNA sequencing, with special attention to eliminating fungi and oomycetes.

2) Light microscopy.—Strains should always be examined by LM. Experienced individuals can usually place even unfamiliar organisms within a range of higher taxa. Without this experience, it is still important to examine the culture and record images for future reference and publication. Be sure to acquire images that will have enough detail to provide diagnostic information. When several strains are being examined, they can often be grouped as similar by LM, but beware of relying only on LM as many very different organisms look quite similar. In any case, images should be acquired for all strains examined.

DNA sequencing is rapidly becoming the most important component of species identification. However, how to proceed with DNA sequencing is very much organism dependent. If you are unsure about the correct genus or family for your organism, proceed to step 3. If you are confident about the identity at genus or family level, proceed to step 4.

3) Ribosomal RNA gene sequencing for preliminary identification—If in doubt of the affinities of the strain after LM, which can easily be true for the “little green balls” such as *Chlorella* and similar organisms, it is important to start with sequencing the appropriate ribosomal RNA gene, generally 18S rDNA for eukaryotes. The 18S rDNA

sequence is the first choice because the algal data set for this gene is very large. Primers that are “universal” and produce amplicons that include the most variable region of the target gene, such as the V4 region of the 18S rRNA gene should be used. There are several studies that propose optimal universal primers for 18S rRNA (e.g. Wang et al. 2014); however, we have had good success with many algae using standard primers developed years ago (Hamby et al. 1988; White et al. 1990) or our own minor modifications of these primers (Fawley and Fawley 2004). The Consortium for the Barcode of Life has designated the V4 region as the starting point for identification for many protists (Pawlowski et al. 2012). In general, we recommend sequencing at least a large portion of the 18S gene with multiple primers to produce the best results, rather than just the V4 region. Long sequences will provide a more accurate identification and also enable future inclusion of your data in more thorough phylogenetic analyses. Accurate long sequences are easily produced using Sanger sequencing so there is no need to stop with only the V4 region. Once you have produced a high-quality sequence from 18S and used BLAST to search the GenBank database (see below), you can proceed to select the appropriate more variable locus for sequencing.

4) Species-Level identification using a “strong” DNA sequence—If you are confident of the genus or family of your strain as determined by LM or 18S rDNA sequence, you can proceed directly to using a locus (or loci) which is variable enough to provide robust identification at the species level (Mann et al. 2010; Leliaert et al. 2014). Using the terminology of Mann et al. (2010), the structural rRNA loci, such as 18S, are considered “weak” in the context of taxon identification and perform poorly at the level of species. Other loci, mostly protein-coding genes, are considered “strong” because they are more variable than 18S and therefore they perform better for species-level identification. To select the proper locus (loci), search for recent taxonomic studies of the genus or family of interest. These studies will not only provide the information on what locus to sequence, but also the primers and PCR conditions. The internal transcribed spacer (ITS) regions (especially ITS2) of the nuclear rRNA operon or the plastid *rbc*L gene are two regions that are used, but there are other possibilities (Table 1 and Mann et al. 2010). Be careful using ITS with many stramenopile taxa, such as diatoms (Mann et al. 2010) as well as dinoflagellates (Stat et al. 2011) because the multiple copies of the ITS in the genome may have different sequences. For many other algae, and especially the green algae, the ITS region (ITS1, 5.8S rDNA, and ITS2) can be a very important region for species identification. In these groups, ITS can usually be sequenced using “universal” primers. However, it is a good idea to look at published studies of the group and select from among the primers that have been shown to work well within that group of organisms. We recommend that you sequence the full ITS1, 5.8S rDNA and ITS2 region, which can provide additional useful data compared to sequencing only the ITS2 region.

If studies using “strong” loci have not yet been performed for the relatives of your strain, then you may have to resign yourself to using only 18S for your identification; however, if this is the case, you must make clear that the identification is provisional, rather than robust.

Assessing the quality of DNA sequencing results: If a poor sequence with “double peaks” is produced from your sample, it is likely that there are one or more contaminating

organism(s) in your culture, or there may be intragenomic heterogeneity among copies of the locus. A third possibility is poor lab technique resulting in contamination from biological material outside the culture. If you cannot produce a clean culture, it may be possible to design PCR primers that will specifically amplify the target gene from your organism of interest. However, the best approach is to clean up the culture and remove contaminating organisms. If the sequencing does not produce regular evenly spaced peaks it is likely that the primers are not working properly or that the DNA template was not well purified. Once high-quality sequences are obtained, the sequence should be assembled using commercial software or free software such as the Staden Package (http://staden.sourceforge.net). These software packages provide excellent views of the raw sequence data and automate the process of assembling the separate primer sequences into a single continuous sequence. Be sure to check the full sequence by eye, rather than simply accepting the calls of the software that you are using. Pay special attention to the ends of the sequences as these are the regions where the sequence quality is usually lowest. Do not hesitate to remove potentially ambiguous data from your sequence. The quality of your DNA sequence(s) is/are extremely important for the integrity of the GenBank database (see https://www.ncbi.nlm.nih.gov/books/NBK44940/). Also be aware that introns are often found in ribosomal RNA genes from strains in some lineages of the Chlorophyta, but they can also be present in other lineages and loci. The presence of introns can sometimes make sequence assembly more difficult.

BLAST search: A BLAST search of GenBank (see https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs for full information on using BLAST) with a high quality 18S DNA sequence will almost certainly reveal the family-level affinity of your strain, and likely the genus-level. Be cautious, however, as sequences posted to GenBank may be from misidentified strains, or they could be old names that have not been updated based on recent taxonomic studies (see below). Strive to use results from an annotated source or published taxonomy studies for all work involving DNA sequence data to help minimize this problem. In many cases, there will not be an exact match for your DNA sequence in GenBank. The closest match could be the correct species, but you should be wary when taking this approach. Be aware of the meaning of BLAST search results. The default ranking for hits is the “E value” (Expect value). This value, which is shown for each hit, is the number of sequences in the database that, simply by random chance, are expected to have the same similarity to your query sequence (https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ). The lower the E value, the higher the probability that the match is significant (not due to chance alone). Therefore, the best E value is zero. The E value is determined from a combination of the “percent identity” and the “query coverage.” The query coverage may be dependent on the length of the target sequence in GenBank, not the overall similarity of the sequences. Thus, a short sequence in GenBank that has 100% identity to your sequence may have a higher E value (more likely due to chance) than a full-length sequence that has several differences. This result often occurs when there is an intron present in your sequence. The occurrence of short sequences of high identity can be revealed by reordering the hits based on percent identity rather than E value. Always be aware of these factors, rather than just accepting the sequence at the top of the list as the identity of your strain.

5) Assessing the results of a BLAST search—After completing the BLAST searches (or searches of a curated database), you must decide whether or not to accept the closest match as the working identification for your alga. If a thorough taxonomic study of the group of organisms has been performed, there may be several strains with very similar sequences assigned to the same species. In this case, when your sequence is within the intraspecies sequence variation seen for this species using the sequence from a strong locus, there is excellent evidence that your strain is properly placed within that species. This is a practical application of the so-called "barcoding gap" (Leliaert et al. 2014; Zou et al. 2016a, b). This process can also be performed using dedicated software such as ABGD (Puillandre et al. 2011) and the Species Determination plug-in for Geneious 6+ (www.geneious.com), but those applications are not required for accurate identification.

When only a single strain of the species (or very few strains) has been sequenced or when very few sister taxa have been identified and sequenced there are obvious problems with species identification using the barcoding gap (Meyer & Paulay 2005). In either case, both of which are common for many algal taxa, we propose the following more subjective identification technique. If the level of identity (when using *rbc*L or ITS2) between your sequence and a published sequence is 99.0% or greater, the species name can be used. If the identity is less than 99.0%, but equal to or greater than 98.0%, you should indicate that your strain may or may not be the same species by using the formal designation "cf.", as in *Chlorella* cf. *vulgaris*. The abbreviation cf. basically means "compare to" and indicates the uncertainty of the identification. If the sequences are less than 98.0% identical, use the form sp., as in *Chlorella* sp. This designation will indicate that you do not have good evidence to confidently place your strain in or near any named species of the genus. Regardless of the level of sequence difference, that difference should be clearly stated.

In some cases, your sequence may be very different from anything in GenBank or the curated database. For example, there may not be a clear relationship with any established genus. In this case, you would refer to the most appropriate taxon above the genus level, such as Chlorellaceae sp. for a strain that could not be confidently placed in any genus in the family Chlorellaceae. None of these conclusions are infallible; they are all subject to revision in the future. However, your best estimate of the taxonomy, placed in these terms, should be acceptable as a good hypothesis of the identity (see discussion). There is also the possibility that your specimen can be recognized by morphology, at least to the genus level, but no sequences exist for that genus. If this is the case and the morphological identification is clear, you should use that name, but with the "cf." or "sp." included. In our view, it is imperative to have the DNA sequence data from a strong locus to make an unqualified identification to the species level.

6) Examine the taxonomy.—Delve into the taxonomic history of the species you have determined using GenBank sequences or a curated database. Is the species name you have determined a taxonomically accepted species? Or is it synonymous with a different specific epithet that is the accepted taxon? Much information on changes in taxonomy can be obtained from AlgaeBase (http://www.algaebase.org, Guiry and Guiry 2020). Often, recent papers on the taxonomy of your organism will be given in the flat files of the sequences in GenBank. Try to use the most up-to-date accepted name.

This step in the identification process is extremely important, especially when working with GenBank, which is the only workable database for sequences of many algal groups at the present time. GenBank is not a curated database, and, as such, the identifications are sometimes out of date, or the algal strain that was sequenced was misidentified. It is very important that you do not simply accept that taxon that is associated with a sequence in GenBank. Instead, you should seek information from the taxonomic literature that applies to your strain. See the discussion section below for examples.

Discussion of barcode cutoffs

The above levels of sequence differences used to identify species are clearly subjective, but they are often quite workable. The following examples show the percentages of intraspecies similarity for some algal genera:

I. In a comprehensive analysis of several loci for strains identified as *Chlorella* (Trebouxiophyceae, Chlorophyta) and related genera (Zou et al. 2016a), the mean intraspecific variations for *rbc*L, ITS1+5.8S+ITS2, and *tuf*A were 0.51%, 1.6% and 0.10%, respectively. The similarity for ITS is skewed because the comparisons include ITS1 and 5.8S rDNA sequences. If only ITS2 were used, which we highly recommend, the intraspecific variation would likely be much less.

II. Nearly all species of the *Nannochloropsis*/*Microchloropsis* (Eustigmatophyceae) group have *rbc*L intraspecific variation of less than 1%, the exception being *N. limnetica*, which has been subdivided into multiple varieties (Fawley and Fawley 2007). The two species of *Microchloropsis*, *M. granulata* and *M. salina* nearly overlap, with the range of intraspecies similarities of 100–99.86% and 100–99.52%, respectively, whereas the maximum interspecies similarity is 98.98%. The organellar genomes of these two species are so similar that Starkenburg et al. (2014) suggested *M. granulata* and *M. salina* (as *N. granulata* and *N. salina*) might be considered variants of the same species.

III. The well-studied group of *Desmodesmus* (Chlorophyceae, Chlorophyta) species closely related to *D. serratus* (Fawley et al. 2011) also have interspecific variation of less the 1% for both the *rbc*L and ITS2 sequences except *D. serratus*, which has maximum intraspecific variations of 1.20% for *rbc*L and 1.95% for ITS2. However, there are enough published *D. serratus rbc*L and ITS2 sequences that a new *D. serratus* strain would easily be placed in this species without using the 99% subjective cutoff.

For both the *Nannochloropsis*/*Microchlorospsis* and the *D. serratus* group, the original taxonomic studies did not consider barcode gaps.

Although these examples show that the 99% cut-off will often work, it is by no means infallible. For example, the species of *Coccomyxa* (Trebouxiophyceae) studied by Darienko et al. (2015) are often not separated at the 99% similarity level using ITS2 sequence data. Species groups that are rapidly evolving may be difficult to determine by barcode analysis because they may not have acquired many new substitutions in neutral markers (see

discussion in Mann et al. 2010). Species groups that are rather closely related along a grade, rather than monophyletic (such as the *Chlorella rbc*L clades 1, 4, and 5 in Zou et al 2016a), may also be difficult to identify by this technique. However, when few data are available for a species group, an identification made using these parameters should serve the algal research community much better than the unsubstantiated identifications so often seen in applied phycology publications. Overall, it is likely that the 99% similarity requirement for assigning a new strain to a particular species is more stringent than might be required.

Table 2 shows the results of example BLAST searches employing GenBank sequence data from unidentified strains or uncultured DNA clones. These examples provide the proper identifications of these organisms based on our suggested cutoffs. Some BLAST results were unambiguous, such as the sequence AB260902, identified as *Chlorella variabilis*, which has a barcoding gap of over 3% divergence. For KX063741, the most similar sequence was from *Sanguina aurantia*, but the similarity was less than 99%. Thus, the proper identification is *Sanguina* cf. *aurantia* which indicates uncertainty. The *Trebouxia* sp. identification shown in Table 2 is a case where there are multiple *Trebouxia* species with identical or nearly identical similarity to AM158969. No taxonomic study has examined these *Trebouxia* species with the detail necessary to define species boundaries using *rbc*L or other loci. To our knowledge, all of these *Trebouxia* species are valid, so it is not possible to select from among them. An unusual case is MK818479, which was 99.90% similar to sequences of *Chroomonas mesostigmatica* in GenBank. However, the strains of *C. mesostigmatica* that have been sequenced are not authentic strains of the species and should be designated *Chroomonas* cf. *mesostigmatica*.

MF483657 provides an example of the some of the problems that can occur with the taxonomy associated with a sequence in GenBank. MF48367 itself is an uncultured clone assigned to the Trebouxiophyceae. BLAST search of the ITS2 portion of this sequence returned the closest sequenced strain Chlorellales sp. LH08AG1034, assession number KX355550. However, the publication listed with the GenBank file for KX355550, Hodač et al. (2016) indicates that strain LHO8AG1034 is a *Marvania* relative. The phylogenetic analyses of Hodač et al. (2016) place LH08AG1034 as a strongly supported member of a terminal lineage that includes *Marvania geminata*. Thus, this strain could be listed either as *Marvania* relative or *Marvania* sp. The publication associated with the sequence can often provide a much better identification than the GenBank description.

Prospects for the Future

We can expect that species identification of eukaryotic microalgae will become much easier in the future, primarily through advancements in sequencing technology, databasing, new species descriptions, and taxonomic clarifications. Perhaps the biggest challenge involves curated databases, such as BOLD (Ratnasingham and Hebert 2007), which aims to database all living organisms and all marker loci, ITSoneDB (Santamaria et al. 2018), which focuses specifically on the ITS1 region, PR^2, which is limited to 18S rDNA (Guillou et al. 2013; del Campo et al. 2018), and Diat-barcode (Rimet et al. 2019), which houses barcode information for diatoms. Unfortunately, the accuracy of sequences already in GenBank and the identifications attached to them are both highly variable and can result in erroneous

identifications using both the GenBank and BOLD databases (Meiklejohn et al. 2019). Correcting these errors and adding new data are both massive tasks. For now, these databases, and especially the broadly focused databases such as BOLD, do not have adequate coverage of most algal groups to be very useful for algal identifications. However, this situation may change in the future. Until that time, the most useful databases are likely to be focused on specific groups of organisms, such as the Diatbarcode database, or GenBank must be used, which is a sequence repository, rather than a curated database and therefore will likely continue to have some taxonomic errors.

The technology for acquisition of DNA sequence data will continue to improve, resulting in the ability to generate massive amounts of sequence data for individual strains (or single cells) very easily. We expect that, in the very near future, many studies that result in naming new taxa, and especially new genera, will include genomic data. We are already embarking on projects to produce organellar genomes for new genera (e.g., Fawley et al. 2019; Ševcíková et al. 2019) and this effort could easily extend to the species level. High quality genomic data would allow the selection of barcoding loci that resolve species better than the loci in current use, such as the recent study of crytomonad clades using *atp*B and *psa*A (Yang et al. 2020) or the use of cox1 in some diatom groups (Kollár, et al. 2019 and references therein). Transcriptomics data can also be used to differentiate species and identify new barcoding loci (Tekle & Wood 2018). The Barcode Data Standards proposed by the Consortium for the Barcoding of Life (https://sibarcodenetwork.readthedocs.io/en/latest/barcode_data_standard.html) should be followed for all taxonomic and diversity studies in order to assure the reliability of database information.

In addition to the data sets, databases are incorporating tool for identification using barcode data. The BOLD database already includes tools for identification and barcode gap calculations. These tools can be expected to improve in the future, which will result in automated, or nearly automated, identifications from barcode data whenever there are adequate sequence data. However, these identifications will still need to be guided by researchers who are cognizant of the techniques employed in the applications.

Acknowledgments

Funding Portions of this work were funded by the National Science Foundation under Grant Nos. MCB0084188 and DEB1248291. This project was supported by the Arkansas INBRE program, with a grant from the National Institute of General Medical Sciences, P20 GM103429 from the National Institutes of Health.

References

Bendif EM, Probert I, Carmichael M, Romac S, de Hagino K, Vargas C (2014) Genetic delineation between and within the widespread coccolithophore morpho-species Emiliania huxleyi and Gephyrocapsa oceanica (Haptophyta). J Phycol 50:140–148. doi:10.1111/jpy.12147 [PubMed: 26988015]

Bock C, Chatzinotas A, Boenigk J (2017) Genetic diversity in chrysophytes: Comparison of different gene markers. Fottea 17:209–221. doi: 10.5507/fot.2017.005

Buchheim M, Buchheim J, Carlson T, Braband A, Hepperle D, Krienitz L, Wolf M, Hegewald E (2005) Phylogeny of the Hydrodictyaceae (Chlorophyceae): inferences from rDNA data. J Phycol 41:1039–1054

del Campo J, Kolisko M, Boscaro V, Santoferrara LF, Nenarokov S, Massana R, Guillou L, Simpson A, Berney C, de Vargas C, Brown MW, Keeling PJ, Wegener Parfrey L (2018) EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. PLOS Biology 16:e2005849. DOI: 10.1371/journal.pbio.2005849

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet 19:370–375. doi:10.1016/S0168-9525(03)00118-5 [PubMed: 12850441]

Coleman AW (2009) Is there a molecular key to the level of “biological species” in eukaryotes? A DNA guide. Mol Phylogenet Evol 50:197–203. doi: 10.1016/j.ympev.2008.10.008 [PubMed: 18992828]

Darienko T, Gustavs L, Eggert A, Wolf W, Pröschold T (2015) Evaluating the species boundaries of green microalgae (Coccomyxa, Trebouxiophyceae, Chlorophyta) using integrative taxonomy and DNA barcoding with further implications for the species identification in environmental samples. PLoS One 10: e0127838. 10.1371/journal.pone.0127838

Fawley MW, Fawley KP (2004). A simple and rapid technique for the isolation of DNA from microalgae. J Phycol 40:223–225. doi:10.1111/j.0022-3646.2004.03-081.x

Fawley KP, Fawley MW (2007) Observations on the diversity and ecology of freshwater Nannochloropsis (Eustigmatophyceae), with descriptions of new taxa. Protist 158:325–336 doi:10.1016/j.protis.2007.03.003 [PubMed: 17576099]

Fawley MW, Fawley KP, Hegewald E (2011) Taxonomy of Desmodesmus serratus (Chlorophyceae, Chlorophyta) and related taxa on the basis of morphological and DNA sequence data. Phycologia 50:23–56. doi:10.2216/10-16.1

Fawley MW, Němcová Y, Fawley KP (2019) Phylogeny and characterization of Paraeustigmatos columelliferus, gen. et sp. nov., a member of the Eustigmatophyceae that may represent a basal group within the Eustigmatales. Fottea 19:107–114. doi: 10.5507/fot.2019.002

Ghosh S, Love N (2011) Application of rbcL based molecular diversity analysis to algae in wastewater treatment plants. Bioresour Technol 102 3619–22. 10.1016/j.biortech.2010.10.125 [PubMed: 21130646]

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C et al. (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucl Acids Res 41:D597–604 [PubMed: 23193267]

Guiry MD, Guiry GM (2020) AlgaeBase. World-wide electronic publication, National University of Ireland, Galway http://www.algaebase.org

Hadi SI, Santana H, Brunale PP, Gomes TG, Oliveira MD, Matthiensen A, Oliveira ME, Silva FC, Brasil BS (2016) DNA barcoding green microalgae isolated from neotropical inland waters. PLoS One 11: e0149284. doi: 10.1371/journal.pone.0149284

Hamby RK, Sims L, Issel L, Zimmer E (1988) Direct ribosomal RNA sequencing: optimization of extraction and sequencing methods for work with higher plants. Plant Mol Biol Rep 6:175–192. 10.1007/BF02669591

Hamsher S, Evans K, Mann D, Poulíčková A, Saunders G (2011) Barcoding diatoms: exploring alternatives to COI-5P. Protist 162:405–22. 10.1016/j.protis.2010.09.005 [PubMed: 21239228]

Herrera S, Shank TM (2016) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol Phylogenet Evol 100:70–79. DOI: 10.1016/j.ympev.2016.03.010 [PubMed: 26993764]

Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M (2009) 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. Gene 430:50–57. 10.1016/j.gene.2008.10.012 [PubMed: 19026726]

Kollár J, Pinseel E, Vanormelingen P, Poulíčková A, Souffreau C, Dvořák P, Vyverman W (2019) A polyphasic approach to the delimitation of diatom species: a case study for the genus Pinnularia (Bacillariophyta). J Phycol 55:365–379. doi:10.1111/jpy.12825 [PubMed: 30536851]

Kim JI, Moore CE, Archibald JM, Bhattacharya D, Yi G, Yoon HS, Shin W (2017) Evolutionary dynamics of cryptophyte plastid genomes. Genome Biol Evol 9: 1859–1872. 10.1093/gbe/evx123 [PubMed: 28854597]

Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz J (2010) The ITS2 Database III—sequences and structures for phylogeny. Nucl Acids Res 38 suppl_1: D275–D279, 10.1093/nar/gkp966 [PubMed: 19920122]

Kryvenda A, Rybalka N, Wolf M, Friedl T (2018) Species distinctions among closely related strains of Eustigmatophyceae (Stramenopiles) emphasizing ITS2 sequence-structure data: Eustigmatos and Vischeria. Eur J Phycol 53:471–491. doi: 10.1080/09670262.2018.1475015

Hodač L, Hallmann C, Spitzer H, Elster J, Faßhauer F, Brinkmann N, Lepka D, Diwan V, Friedl T (2016) Widespread green algae Chlorella and Stichococcus exhibit polar-temperate and tropical-temperate biogeography. FEMS Microbiol Ecol 92:122, 10.1093/femsec/fiw122

Leliaert L, Verbruggen H, Vanormelingen P, Steen F, López-Bautista JM, Zuccarello JC, De Clerck O (2014) DNA-based species delimitation in algae, Eur J Phycol 49:179–196. doi: 10.1080/09670262.2014.904524

Litaker WR, Vandersea MW, Kibler SR, Reece KS, Stokes NA, Lutzoni FM, Yonish BA, West MA, Black MND, Tester PA (2007) Recognizing dinoflagellate species using ITS rDNA sequences. J Phycol 43:344–355. doi:10.1111/j.1529-8817.2007.00320.x

Lundholm N, Hasle GR, Fryxell GA, Hargraves PE (2002) Morphology, phylogeny and taxonomy of species within the Pseudo-nitzschia americana complex (Bacillariophyceae) with descriptions of two new species, Pseudo-nitzschia brasiliana and Pseudo-nitzschia linea. Phycologia 41:480–497. 10.2216/i0031-8884-41-5-480.1

Maistro S, Broady PA, Andreoli C, Negrisolo E (2007) Molecular phylogeny and evolution of the order Tribonematales (Heterokonta, Xanthophyceae) based on analysis of plastidial genes rbcL and psaA. Mol Phylogenet Evol 43:407–417. doi:10.1016/j.ympev.2007.02.014 [PubMed: 17400001]

Mann DG (1999) The species concept in diatoms. Phycologia 38:437–495. doi: /10.2216/i0031-8884-386-437.1

Mann D, Sato S, Trobajo R, Vanormelingen P, Souffreau C (2010) DNA barcoding for species identification and discovery in diatoms. Cryptogam Algol 31:557–577

McManus H, Lewis L (2005) Molecular phylogenetics, morphological variation and colony-form evolution in the family Hydrodictyaceae (Sphaeropleales, Chlorophyta). Phycologia 44:582–595. 10.2216/0031-8884(2005)44[582:MPMVAC]2.0.CO;2

Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank - Their accuracy and reliability for the identification of biological materials. PLoS One 14:e0217084. doi: 10.1371/journal.pone.0217084

Meyer CP, Paulay G (2005) DNA Barcoding: error rates based on comprehensive sampling. PLoS Biol 3: e422. 10.1371/journal.pbio.0030422

Moestrup Ø (2006) Algal Taxonomy: Historical Overview. In eLS, (Ed.). doi: 10.1002/9780470015902.a0000328.pub2

Ohmura Y, Takeshita S, Kawachi M (2019) Photobiont diversity within populations of a vegetatively reproducing lichen, Parmotrema tinctorum, can be generated by photobiont switching. Symbiosis 77:59–72. doi: 10.1007/s13199-018-0572-1

Pawlowski J, Audic S, Adl S, Bass D, Belbahri L et al. (2012) CBOL Protist Working Group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol 10: e1001419. doi:10.1371/journal.pbio.1001419

Pante E, Puillandre N, Viricel A, Arnaud-Haond S, Aurelle D, Castelin M, Chenuil A, Destombe C, Forcioli D, Valero M, Viard F and Samadi S (2015) Species are hypotheses: avoid connectivity assessments based on pillars of sand. Mol Ecol 24: 525–544. doi:10.1111/mec.13048 [PubMed: 25529046]

Pentinsaari M, Ratnasingham S, Miller SE, Hebert PDN (2020) BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? PLoS One 15(4): e0231814. 10.1371/journal.pone.0231814

Pinseel E, Kulichová J, Scharfen V, Urbánková P, Van de Vijver B, Vyverman W (2019) Extensive cryptic diversity in the terrestrial diatom Pinnularia borealis (Bacillariophyceae). Protist 170:121–140. doi:10.1016/j.protis.2018.10.001 [PubMed: 30954839]

Pochon X, Putnam HM, Burki F, Gates RD (2012) Identifying and characterizing alternative molecular markers for the symbiotic and free-living dinoflagellate genus Symbiodinium. PLoS One 7: e29816. 10.1371/journal.pone.0029816

Procházková L, Leya T, Křížková H, Nedbalová L (2019) Sanguina nivaloides and Sanguina aurantia gen. et spp. nov. (Chlorophyta): the taxonomy, phylogeny, biogeography and ecology of two newly recognised algae causing red and orange snow. FEMS Microbiol Ecol 95 fiz064. doi:10.1093/femsec/fiz064

Pröschold T, Darienko T, Silva PC, Reisser W, Krienitz L (2011) The systematics of Zoochlorella revisited employing an integrative approach. Environ Microbiol 13:350–364. doi:10.1111/j.1462-2920.2010.02333.x [PubMed: 20874732]

Puillandre Nicolas & Lambert A & Brouillet S, Achaz, Guillaume (2011) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21:1864–77. 10.1111/j.1365-294X.2011.05239.x [PubMed: 21883587]

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886. doi:10.1080/10635150701701083 [PubMed: 18027281]

Ratnasingham S, Hebert PD (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364. doi: 10.1111/j.1471-8286.2007.01678.x [PubMed: 18784790]

Rimet F, Gusev E, Kahlert M, Kelly M, Kulikovskiy M, Maltsev Y, Mann D, Pfannkuchen M, Trobajo R (2019) Diat.barcode, an open-access curated barcode library for diatoms. Sci Rep 9:15116 doi:101038/s41598-019-51500-6 [PubMed: 31641158]

Rybalka N, Andersen RA, Kostikov I, Mohr KI, Massalski A, Olech M, Friedl T (2009) Testing for endemism, genotypic diversity and species concepts in Antarctic terrestrial microalgae of the Tribonemataceae (Stramenopiles, Xanthophyceae). Environ Microbiol 11:554–565. doi:10.1111/j.14622920.2008.01787.x [PubMed: 19278444]

Rybalka N, Wolf M, Andersen RA, Friedl T (2013) Congruence of chloroplast- and nuclear-encoded DNA sequence variations used to assess species boundaries in the soil microalga Heterococcus (Stramenopiles, Xanthophyceae). BMC Evol Biol 13:39. doi: 10.1186/1471-2148-13-39 [PubMed: 23402662]

Santamaria M, Fosso B, Licciulli F, et al. (2018) ITSoneDB: a comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences. Nucl Acids Res 46(D1):D127–D132. doi:10.1093/nar/gkx855 [PubMed: 29036529]

Ševcíková T, Yurchenko T, Fawley KP, Amaral R, Strnad H, Santos LMA, Fawley MW, Eliáš M (2019) Plastid genomes and proteins illuminate the evolution of eustigmatophyte algae and their bacterial endosymbionts. Genome Biol Evol. 11:362–379. doi:10.1093/gbe/evz004 [PubMed: 30629162]

Škaloud P, Friedl T, Hallmann C, Beck A & Dal Grande F (2016) Taxonomic revision and species delimitation of coccoid green algae currently assigned to the genus Dictyochloropsis (Trebouxiophyceae, Chlorophyta). J Phycol 52: 599–617. doi: 10.1111/jpy.12422 [PubMed: 27135898]

Stat M, Bird CE, Pochon X, Chasqui L, Chauka LJ, Concepcion GT, Logan D, Takabayashi M, Toonen RJ, Gates RD (2011) Variation in Symbiodinium ITS2 sequence assemblages among coral colonies. PLoS One 6:e15854. doi: 10.1371/journal.pone.0015854

Starkenburg SR, Kwon KJ, Jha RK, McKay C, Jacobs M, Chertkov O, Twary S, Rocap G, Cattolico RA (2014) A pangenomic analysis of the Nannochloropsis organellar genomes reveals novel genetic variations in key metabolic genes. BMC Genom 15:212. doi:10.1186/1471-2164-15-212

Stern RF, Horak A, Andrew RL, Coffroth MA, Andersen RA, Küpper FC, Jameson I, Hoppenrath M, Véron B, Kasai F, Brand J, James ER, Keeling PJ (2010) Environmental barcoding reveals massive dinoflagellate diversity in marine environments. PLoS One 5:e13991. doi: 10.1371/journal.pone.0013991

Tekle YI, Wood FC (2018) A practical implementation of large transcriptomic data analysis to resolve cryptic species diversity problems in microbial eukaryotes. BMC Evol Biol 18:170 10.1186/s12862-018-1283-1 [PubMed: 30445905]

Trobajo R, Mann DG, Clavero E, Evans KM, Vanormelingen P, McGregor RC (2010) The use of partial cox1, rbcL and LSU rDNA sequences for phylogenetics and species identification within the Nitzschia palea species complex (Bacillariophyceae), Eur J Phycol 45:413–425, DOI: 10.1080/09670262.2010.498586

Vanormelingen P, Hegewald E, Braband A, Kitschke M, Friedl T, Sabbe K and Vyverman W (2007) The systematics of a small spineless Desmodesmus species, D. costato-granulatus (Sphaeropleales, Chorophyceae), based on ITS2 rDNA sequence analyses and cell wall morphology. J Phycol 43:378–396. doi:10.1111/j.1529-8817.2007.00325.x

Vieira HH, Bagatini IL, Guinart CM, Vieira AAH (2016) tufa gene as molecular marker for freshwater Chlorophyceae. ALGAE 31:155–165. 10.4490/algae.2016.31.4.14

Wang Y, Tian RM, Gao ZM, Bougouffa S, Qian PY. (2014) Optimal eukaryotic 18S and universal 16S/18S ribosomal RNA primers and their application in a study of symbiosis. PLoS One 9:e90053. doi:10.1371/journal.pone.0090053

White TJ, Bruns TD, Lee SB and Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (Eds) PCR Protocols: A Guide to Methods and Applications, Academic Press, New York pp 315–322. 10.1016/B978-0-12-372180-8.50042-1

Yang EC, Noh JH, Kim S, Choi DH (2020) Plastid-encoded gene comparison reveals usefulness of atpB, psaA, and rbcL for identification and phylogeny of plastid-containing cryptophyte clades. Phycologia doi: 10.1080/00318884.2019.1709145

Zou S, Fei C, Song J, Bao Y, He M, Wang C (2016a) Combining and comparing coalescent, distance and character-based approaches for barcoding microalgaes: a test with Chlorella-like species (Chlorophyta). PLoS One 11: e0153833. 10.1371/journal.pone.0153833

Zou S, Fei C, Wang C, Gao Z, Bao Y, He M, Wang C (2016b) How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in Scenedesmus (Chlorophyceae). Sci Rep 6:36822 [PubMed: 27827440]

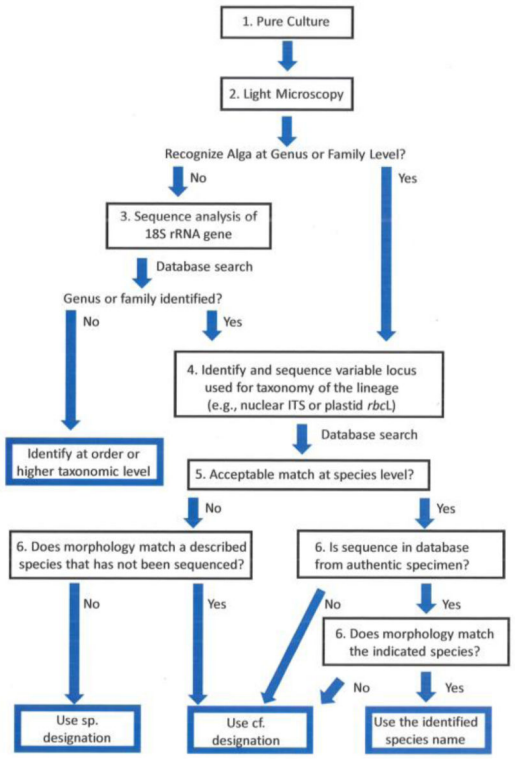

Fig. 1.

Flow chart showing the steps of the identification process. Numbers associated with each step refer to the numbered section on the step-by-step process in the text.

Table 1.

Some strong loci in use for delimiting or identifying species from select eukaryotic algal groups. Example references are not exhaustive lists.

Taxon	Locus	Example Reference(s)
Bacillariophyta		
	*rbc*L	Mann et al. 2010; Rimet et al. 2019
	COI	Trobajo et al. 2010; Kollár et al. 2019
	LSU D2/D3	Hamsher et al. 2011
Chrysophyceae		
	COI	Bock et al. 2017
	LSUD2/D3	Bock et al. 2017
Chlorophyta		
	ITS	Darienko et al. 2015; Hadi et al. 2016
	*rbc*L	Hadi et al. 2016; Zou et al. 2016a, b
	*tuf*A	Vieira et al. 2016
Dinophyceae		
	COI	Stern et al. 2010
	LSU	Pochon et al. 2012
	ITS	Litaker et al. 2007
Eustigmatophyceae		
	*rbc*L	Fawley et al. 2007; Ghosh & Love 2011
	ITS	Kryvenda et al. 2018
Haptophyta		
	COI	Bendif et al. 2014 (also other mitochondrial loci)
Xanthophyceae		
	ITS	Rybalka et al. 2013
	*rbc*L	Maistro et al. 2007; Rybalka et al. 2009; 2013
	*psa*A	Maistro et al. 2007

Table 2.

Examples of using the proposed levels of similarity required for species identification and the barcoding gap with *rbc*L and ITS-2 sequence data. Identifications of several examples selected mostly at random from Genbank using a standard BLAST search. The boundaries of the ITS2 region, when not delimited from within a longer amplicon in the GenBank flat file, were determined using the method of Keller et al. (2009) as implemented in the ITS2-Database (http://its2.bioapps.biozentrum.uniwuerzburg.de, Koetschan et al. 2010). Identifications tagged with an asterisk indicate identity to one section of a broadly defined species. In this context, "gap" indicates the percent similarity of the most similar sister species (or clade). Gap values were not calculated unless a species could be identified.

Accession and original ID	Locus	Blast Search Result	% Similarity	Gap	Taxonomic Reference
KX548261, Uncultured Trebouxiophyceae	*rbc*L	*Chlorella rbc*L Clade 22	99.90	92.39	Zou et al. 2016a
AB260902, uncultured *Chlorella*	*rbc*L	*Chlorella variabilis*	99.23–100	95.85	Pröschold et al. 2011
KF960689, uncultured *Dictyochloropsis*	*rbc*L	*Symbiochloris tschermakiae*	99.72–100	96.38	Škaloud et al. 2016
AM260443, uncultured *Chlorella*	*rbc*L	Prasiolaceae (Trebouxiophyceae)	89.18	n/a	
AM158969, uncultured *Trebouxia*	*rbc*L	*Trebouxia* sp.	99.74	n/a	
MH707957, *Pinnularia* sp.	*rbc*L	*Pinnularia* sp.	97.75	n/a	
KC969810, *Pseudo-Nitzschia* sp.	*rbc*L	*Pseudo-Nitzschia americana* complex	99.48–99.74	95.64	Lundholm et al. 2002
JN418665, *Pinnularia* sp.	*rbc*L	*Pinnularia neglectiformis*	99.57	97.91	Pinseel et al. 2019
KC184837 Xanthophyceae sp.	*rbc*L	*Bumilleriopsis filliformis*	99.75	99.00	Maistro et al. 2007
MK818479, Cryptophyta sp.	*rbc*L	*Chroomonas* cf. *mesostigmatica*	99.90	99.60	Kim et al. 2017
MF483657, uncultured Trebouxiophyceae	ITS2	*Marvania* sp.	99.11	89.90	Hodač et al. 2016
FN298927, *Coccomyxa* sp.	ITS2	*Coccomyxa simplex**	100	97.83	Darienko et al. 2015
KX063741, uncultured *Chlamydomonas*	ITS2	*Sanguina* cf. *aurantia*	98.12–98.59	96.24	Procházková et al. 2019
DQ417533, *Desmodesmus* sp.	ITS2	*Desmodesmus elegans**	99.24	95.82	Vanormelingen et al. 2007
MG266124, uncultured Chlorophyceae	ITS2	Oedogoniales sp.?	80.00	n/a	
KM108768, uncultured *Pediastrum*	ITS2	*Hydrodictyon reticulatum*	99.56	97.38	Buchheim et al. 2005; McManus and Lewis 2005