

RESEARCH ARTICLE

Combining and Comparing Coalescent, Distance and Character-Based Approaches for Barcoding Microalgae: A Test with *Chlorella*-Like Species (Chlorophyta)

Shanmei Zou, Cong Fei, Jiameng Song, Yachao Bao, Meilin He, Changhai Wang*

Jiangsu Provincial Key Laboratory of Marine Biology, College of Resources and Environmental Science, Nanjing Agricultural University, Nanjing 210095, PR China

* chwang@njau.edu.cn



OPEN ACCESS

Citation: Zou S, Fei C, Song J, Bao Y, He M, Wang C (2016) Combining and Comparing Coalescent, Distance and Character-Based Approaches for Barcoding Microalgae: A Test with *Chlorella*-Like Species (Chlorophyta). PLoS ONE 11(4): e0153833. doi:10.1371/journal.pone.0153833

Editor: Peter Prentis, Queensland University of Technology, AUSTRALIA

Received: November 17, 2015

Accepted: March 13, 2016

Published: April 19, 2016

Copyright: © 2016 Zou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The financial support was from the China Postdoctoral Science Foundation (2014M561661, 2015T80558), and Natural Science Fund project in Jiangsu Province (BK20150680). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Several different barcoding methods of distinguishing species have been advanced, but which method is the best is still controversial. *Chlorella* is becoming particularly promising in the development of second-generation biofuels. However, the taxonomy of *Chlorella*-like organisms is easily confused. Here we report a comprehensive barcoding analysis of *Chlorella*-like species from *Chlorella*, *Chloroidium*, *Dictyosphaerium* and *Actinastrum* based on *rbcl*, ITS, *tufA* and 16S sequences to test the efficiency of traditional barcoding, GMYC, ABGD, PTP, P ID and character-based barcoding methods. First of all, the barcoding results gave new insights into the taxonomic assessment of *Chlorella*-like organisms studied, including the clear species discrimination and resolution of potentially cryptic species complexes in *C. sorokiniana*, *D. ehrenbergianum* and *C. vulgaris*. The *tufA* proved to be the most efficient barcoding locus, which thus could be as potential “specific barcode” for *Chlorella*-like species. The 16S failed in discriminating most closely related species. The resolution of GMYC, PTP, P ID, ABGD and character-based barcoding methods were variable among *rbcl*, ITS and *tufA* genes. The best resolution for species differentiation appeared in *tufA* analysis where GMYC, PTP, ABGD and character-based approaches produced consistent groups while the PTP method over-split the taxa. The character analysis of *rbcl*, ITS and *tufA* sequences could clearly distinguish all taxonomic groups respectively, including the potentially cryptic lineages, with many character attributes. Thus, the character-based barcoding provides an attractive complement to coalescent and distance-based barcoding. Our study represents the test that proves the efficiency of multiple DNA barcoding in species discrimination of microalgae.

Introduction

Chlorella (Trebouxiophyceae, Chlorophyta), single-celled green algae, is one of the most famous microalgae genus worldwide that grow in marine, freshwater or edaphic habitats. *Chlorella* could be used as powerful ‘superfoods’, and is significant in the development of second-

generation biofuels and medical treatments [1–4]. Nevertheless, the taxonomic assignment of *Chlorella* is easily confused since there are no obvious structural features among species or some of observable characteristics are variable within species. After the type species are identified, more than 100 *Chlorella* species have been described [5–9]. For a long time, numerous studies focusing on morphological characters, ultrastructural composition of the cell wall, biochemical and physiological characters and molecular phylogenetic characteristics have been carried out to revise the system of *Chlorella* [6,9–25]. New *Chlorella*-related species and genera are often discovered in recent studies [4,26–30]. Most notably, *Chlorella* species are recognized as members of Chlorophyceae and Trebouxiophyceae. Based on the biochemical and molecular data, *Chlorella* is shown that it consists of only five “true” *Chlorella* species [9,18,20]. Darienko et al. [22] propose to transfer all *Chlorella*-like strains that have been identified as *Chlorella saccharophila* and *Chlorella ellipsoidea* to the genus *Chloroidium* in the so-called *Watanabea* clade. According to recent molecular studies, species that have typical *Chlorella* morphology are assigned to the family Chlorellaceae (Trebouxiophyceae) that is divided into *Chlorella*-clade and the *Parachlorella*-clade [20,31,32]. Based on SSU- and ITS rDNA sequences and light microscopic observations, Bock et al. [25] detect six lineages of *Dictyosphaerium*-like strains that are closely related to *Chlorella vulgaris* and describe several new species. Consequently, high levels of cryptic diversity found within *Chlorella* and the polyphyletic characters between *Chlorella* and *Dictyosphaerium* results in fundamental taxonomic revision of these organisms, e.g. the description of many new species and genera [20,23–25,32,33]. All these studies indicate that the classification of *Chlorella* is still very confused and it is urgent to revise the genus. However, the identification of *Chlorella*-like taxa in the species level is still unclear. It is still unknown how many species are actually included in *Chlorella*. Moreover, most previous molecular studies generally focus on the phylogenetic analysis of *Chlorella*-related clades based on 18S, ITS or SSU gene data [21,31–33]. The molecular taxonomic identification of *Chlorella*-like microalgae often analyzes limited gene loci or samples. For example, the taxonomic reassessment of *Chlorella* by molecular signatures (barcodes) [24] is just based on the ITS. Therefore, numerous cryptic species within *Chlorella*-like organisms may be overlooked. To recover the hidden diversity, more molecular markers that have sufficient nucleotide diversity, low saturation and a simple alignment process should be used for taxonomic identification [32].

DNA barcoding is the most promising approach for species identification and detection of cryptic species and potentially new species, particularly for the microbial communities [34–44]. The traditional DNA barcoding [34], including the monophyly and distance-based methods, are originally used for DNA barcoding. The distance method relies on the ‘barcoding gap’ and the monophyly method can reconstruct the evolutionary histories of character traits [34,35,45]. However, for some taxonomic groups, it is impossible to identify the specimens based on intraspecific variation vs. interspecific divergence. For example, in plants, the species discrimination is not always accurate using the ‘barcoding gap’, e.g. [46–49], and in some cases, species identification is possible even if a ‘barcoding gap’ is absent [50]. The monophyly-based DNA barcoding approach has the drawback that relates to the use of hierarchical methods for identification [51–60]. On the other hand, the phylogenetic tree is often used for flagging species.

Recently, several different methods have been put forward for distinguishing species [61]. The generalized mixed Yule–coalescent model sets a threshold to delineate evolutionary significant units (ESUs) akin to species [62–64]. The P ID (Liberal) method of species delimitation is advanced for the exploration of species boundaries [65], which allows differing species boundary hypotheses to be investigated by enabling the user to a priori assign taxa to putative species groups on a phylogenetic tree. The poisson tree process (PTP) model is another tree-based

method that distinguishes specimens in both populations and species level using coalescence theory [66]. Automatic Barcode Gap Discovery (ABGD), a new distance method, can assign the sequences into potential species based on the barcode gap whenever the divergence within the same species is smaller than that among organisms from different species [67].

The character-based barcoding approach has recently been proved useful in species identification and cryptic species revelation of some organisms (including some plants) [39,40–41,50,68–72]. It is based on the concept that members of a given taxonomic group have the same diagnostic characters that are absent from comparable groups [69,73]. The character-based approach has the logical advantage that it will fail to diagnose the specimens when diagnostic character data are lacking, in comparison with using distances. With the development of DNA barcoding, it seems that combination of improved species-level phylogenetic trees and new statistical methods that evaluate quantitative character states will greatly help us to understand the biodiversity patterns [74,75]. In this context, combination of multiple DNA barcoding approaches may be more effective to reveal cryptic biodiversity. However, the character-based barcoding approach is not yet commonplace in barcoding practice. So far, few studies about barcoding microalgae have been performed using character-based methods. On the contrary, most molecular taxonomic identification of algae is just based on the phylogenetic trees or genetic distance, including DNA barcoding of marine green macroalgae [76], freshwater green algae [77] and some *Chlorella*-related samples [24].

Besides the barcoding approaches, the efficient ‘DNA barcodes’ also play an important role in successful species identification across a wide range of taxonomic groups. Due to a much slower mutation rate the cytochrome *c* oxidase 1 (COI) sequence which has been proved efficient in barcoding animals does not discriminate most plants [78,79]. Although in red and brown algae and some diatoms, the 5’ end of COI (COI-5P) provides resolution at the species level [79–84], it is often unsuccessfully amplified in green algae, despite extensive primer testing [76,77]. The presence of introns within COI [85–88] may be the largest obstacle to developing the COI-5P as a suitable DNA barcode marker for green algae. For land plants the core DNA barcodes are portions of two plastid coding genes (*rbcl* and *matk*) [89]. However, the candidate loci fail in eliminating the disadvantages of current DNA barcoding for plants [90]. Efficient ‘DNA barcodes’ for plants are still unknown. Moreover, since *matK* is absent in algae it is urgently needed to select the specific barcodes for taxonomic group of algae, especially for microalgae. Recently, a new concept, the ‘specific barcode’ is proposed, which refers to a fragment of DNA sequence that can enable species identification within a given taxonomic group (e.g. a genus or family) by sufficiently high mutation rate [91]. ‘Specific barcodes’ for plants can assist species-level identifications. Presently the *rbcl* gene (encodes the large subunit of Rubisco), *tufA* gene (encoding elongation factor) and ITS (internal transcribed spacer region) have been proved useful in discriminating some microalgae species, e.g. [76,77,92]. Thus, they could be candidates as ‘specific barcodes’ for green algae.

In this study we present a comprehensive DNA barcode analysis (traditional barcoding, GMYC, P ID, PTP, ABGD and character-based approaches) of *Chlorella*-like species from *Chlorella*, *Chloroidium*, *Actinastrum* and *Dictyosphaerium*, based on four gene loci *rbcl*, *tufA*, ITS (ITS1–5.8S–ITS2) and 16S. Publicly available sequences were added to the newly obtained sequences from this study to better evaluate identification success among the organisms. Sequences from all the genes are used: (i) to identify *Chlorella*-like taxa and reveal the possible existence of cryptic species (ii) to evaluate the efficiency of coalescent, distance and character-based barcoding approaches in retrieving the taxon identities of this morphologically complex microalgae.

Materials and Methods

Ethics Statement

No specific permits were required for the described field studies. The field studies did not involve endangered or protected species. No specific permissions were required for the locations. The locations are not privately-owned or protected in any way.

Algal sampling, culturing and morphological identification

The *Chlorella*-like green microalgae studied were from the genera *Chlorella*, *Chloroidium*, *Actinastrum* and *Dictyosphaerium*, most of which were *Chlorella* strains. A total of 176 *Chlorella*-like samples were analyzed. The collection spots covered marine, freshwater, north pole and terrestrial areas. The procedure for clone isolation followed Andersen [93]. The nonaxenic strains were grown in 250 mL flask containing 200 mL liquid at an irradiance of $40 \text{ umol m}^{-2} \text{ s}^{-1}$ with 14:10 h light: dark cycle at 20°C. Some strains were obtained from different Culture Collections, e.g. Austin, Texas, Waller Creek at University Campus, USA. A detailed list of taxa studied, was provided in [S1 Table](#).

The samples collected in this study were first identified based on the available morphological characters. Specimens that could not to be assigned to binomial names were just labeled as unknowns.

DNA Extraction, Amplification and Sequencing

DNA was extracted using the Qiagen DNEasy Plant Extraction kit (Qiagen Inc., Valencia, CA, USA) following the instructions given by the manufacturer. The *rbcL*, *tufA*, ITS and 16S barcode regions were amplified and sequenced from most species using universal primers or primers designed in this study ([S2 Table](#)) [24, 94–96]. PCR reactions for all barcode regions were carried out in a total volume of 25 μL , using 2 \times Taqman PCR MasterMix. PCR conditions for all primer sets were as follows: 95°C for 3 min, primer-specific annealing temperatures for 45s, 72°C for 1 min; 35 cycles of 95°C for 30 s, primer-specific annealing temperatures for 45s, 72°C for 1 m, with a final extension of 72°C for 1 min. Then the PCR products were sequenced on an ABI 3730XL (Applied Biosystems).

Sequence alignment

Forward and reverse sequences of each region were edited in Sequencher (Gene Codes Corporation), and a set of publicly available sequences from Genbank was added. All *rbcL*, ITS, 16S and *tufA* sequences were aligned using MAFFT 6.717 [97] and trimmed to a region 1158 nucleotides, 1016–1300 nucleotides, 315–429 nucleotides and 783 nucleotides in length respectively.

Traditional barcoding analysis

Phylogenetic reconstruction. Neighbour joining trees of *rbcL*, *tufA*, ITS and 16S sequences were constructed based on Kimura 2-parameter (K2P) distance model as recommended by Hebert et al. [34–35] in MEGA 5.0 [98] with bootstrap values (1000 replications). Since the identification of *Chlorella*-like species was difficult by morphological characters, the NJ trees were first used to flag species for character-based barcoding analysis.

For Bayesian analysis of each gene, the jModeltest v.0.1.1 [99] was used to estimate the best substitution model using Akaike Information Criterion (AIC). The most appropriate models for *rbcL*, ITS, 16S and *tufA* were GTR+G, GTR+ G, TVMef+I+G and GTR +G respectively. As described in detail previously [41], the Bayesian analysis were conducted in MRBAYES 3.1.2 [100]. The maximum-likelihood (ML) search was performed using PHYML 3.0 [101].

Distance analyses. Genetic distances were ascertained using MEGA 5.0 [98] and the Distance Summary applications of the BOLD website, with the K2P model [37]. To assess the barcoding gap effectiveness, the analyses of intra- and interspecific divergences were conducted among the taxa assignments based on multiple method-based barcoding analyses.

GMYC species delimitation

Using BEAST [102,103], a linearised Bayesian phylogenetic tree was first calculated employing a Yule pure birth model [104] (Gernhard 2008) tree prior. Settings in BEAUTi v. 1.7.1 were: substitution models for each gene, empirical base frequencies, four gamma categories, all codon positions partitioned with unlinked base frequencies and substitution rates. An uncorrelated relaxed lognormal clock model was used with rate estimated from the data and uclmean parameter with uniform prior to value 0 as a lower and 10 as an upper boundary. All other settings were left as defaults. The length of MCMC chain was 40 000 000 sampling every 4000. All BEAST runs were executed in Bioportal [105], and the ESS values and trace files of runs were evaluated in Tracer v1.5.0. Two independent runs were merged using Log-Combiner v1.7.1 with 20% burn-in. Maximum clade credibility trees with a 0.5 posterior probability limit, and node heights of target tree were constructed in TreeAnnotator v1.7.1. Single-threshold GMYC analyses was conducted in R [106] using the APE [107] and SPLITS [108] packages.

Poisson tree process model (PTP)

Since the ultrametric trees are not required as input this coalescent-based method is very fast. This method is implemented in a web server (<http://species.h-its.org/>).

P ID (Liberal) species boundary delimitation

The Species Delimitation plugin [65] within Geneious Pro v5.5.4 (Biomatters; <http://www.geneious.com>) was investigated to assess species boundary hypotheses across the Bayesian gene tree. Geneious is a bioinformatics desktop software package produced by Biomatters Ltd (<http://www.biomatters.com>). P ID(Liberal) in Geneious, represents the probability of making a correct identification of an unknown specimen by measuring the genetic variation found within its putative species group and comparing that to the species group with which it is most likely to be confused [109]. Maximum Likelihood trees were inferred from *rbcl*, *tufA* and ITS datasets by employing PhyML 3.0 [101].

Automatic Barcode Gap Discovery

The ABGD method is available at <http://www.wabi.snv.jussieu.fr/public/abgd/>. The *rbcl*, 16S, ITS and *tufA* complete sequence data were processed in ABGD using the K2P nucleotide substitution model. Prior for the maximum value of intraspecific divergence was set between 0.001 and 0.1, and 10 recursive steps within the primary partitions was defined by the first estimated gap. The gap width was set 1.0.

Character-based DNA barcode analyses

Pure unique identifying characters, termed diagnostic characters or “characteristic attributes” (CAs) that distinguish a species from others, were determined using characteristic attribute organization system (CAOS) which comprises P-Gnome and P-Elf programs [66,110]. The CAOS algorithm could extract CAs for each clade at branching node within a guide tree [69]. In this study, the guide trees inferred from *rbcl*, ITS and *tufA* sequences were first produced using the programs PAUP v4.0b10 [111], and were incorporated into a NEXUS file containing

rbcl, ITS and *tufA* sequence data respectively in MacClade [112]. Then the incorporated NEXUS datasets were conducted in CAOS system where the P-Gnome script was used to identify characters. Finally, the most variable character states were listed.

Results

A total of 176 *Chlorella*-like samples from this study and publicly available data were analyzed. PCR amplification and sequencing were successful with the ITS locus in most of the samples (S1 Table). With the *rbcl*, 16S and *tufA* loci, there were more difficulties in amplification and in sequencing, especially for *tufA*. A total of 96 *rbcl*, 76 ITS, 86 16S and 66 *tufA* sequences of *Chlorella*-like samples and outgroups were analyzed (S1 Table). The accession numbers of newly obtained sequences submitted to the GenBank Barcode database were: KM514738-KM514804 for 16S, KM514805-KM514860 for ITS, KM514861-KM514917 for *rbcl* and KR154236-KR154291 for *tufA*.

Traditional DNA barcoding

Phylogenetic analyses. Generally, the NJ, Bayesian and Maximum Likelihood trees of *rbcl*, ITS and *tufA* recovered consistent groups respectively (Figs 1–3 and S1 Fig, S2 Fig and S3 Fig). A total of 31, 20 and 14 monophyletic *Chlorella*-like clades were recovered in *rbcl*, ITS and *tufA* bayesian trees respectively (Figs 1–3), including the potentially cryptic lineages in *Chlorella sorokiniana*, *Dictyosphaerium ehrenbergianum* and *Chlorella vulgaris*. These recovered lineages would be further analyzed by GMYC, PTP, P ID, ABGD and character-based barcoding. Sequences of *C. sorokiniana*, *D. ehrenbergianum* and *C. vulgaris* fell into several distinct clades respectively in *rbcl*, ITS and *tufA* trees, which might be indicating potentially cryptic lineages (Figs 1–3). Thereinto, *C. sorokiniana* was divided into: five clades (I), (II), (KC810315, JQ415926), (JQ415921) and (HM101339) in *rbcl* tree (Fig 1); five clades (I), (II), (III), (KJ676111, KJ676109) and (KJ676113) in ITS tree (Fig 2); and four clades (I), (II), (III) and (KJ742376, KJ397925) in *tufA* tree (Fig 3). *C. vulgaris* was also divided into: five clades (I), (II), (EU038286, JQ315474, EU038284), (KC810313, JQ717305, AB240145) and (JQ415915) in *rbcl* tree (Fig 1); three clades (I), (FR865683) and (FM205832, KC517115, JX185298) in ITS tree (Fig 2); and three clades (I), (II) and (III) in *tufA* tree (Fig 3). *D. ehrenbergianum* was divided into three clades (I), (II) and (III) in *rbcl* tree (Fig 1). It was worth noting that all the samples collected from Arctic pole grouped together as a separate clade, but they still could not be identified to the specific taxa in the species level. In addition, most unknown samples were recovered as separate clades that did not group together with other species (Figs 1–3). The 16S NJ, Bayesian and Maximum trees, however, could not separate the closely related *Chlorella*-like samples, and the supports for the monophyletic clades were very low (S4 and S5 Figs). Therefore, the coalescent, distance and character assignments of 16S sequences were not analyzed in this study. It was apparent that the *tufA* phylogenetic trees recovered more well-supported monophyletic taxa in comparison with the *rbcl*, ITS and 16S phylogenies.

Distance analyses. Based on the phylogenetic, GMYC, PTP, P ID, ABGD and character-based barcoding assignments of *rbcl*, ITS and *tufA* sequences, intra- and interspecific variation of above defined *Chlorella*-like assignments was conducted respectively (Figs 1–3), and the existence of DNA barcoding gap was tested. The results showed that the pairwise intraspecific distance of *rbcl* were from 0% to 4.2% with a mean of 0.51% while the pairwise interspecific distances was from 0% to 21.70%. The mean *rbcl* divergence among the possible cryptic lineages within *C. sorokiniana*, *D. ehrenbergianum* and *C. vulgaris* were from 1.3% to 9.8%, 2.3% to 5.4% and 0.9% to 9.6% respectively, which was higher than 0.51% (the mean intraspecific distance) (S3 Table). However, no clear barcoding-gap was found between the intra- and

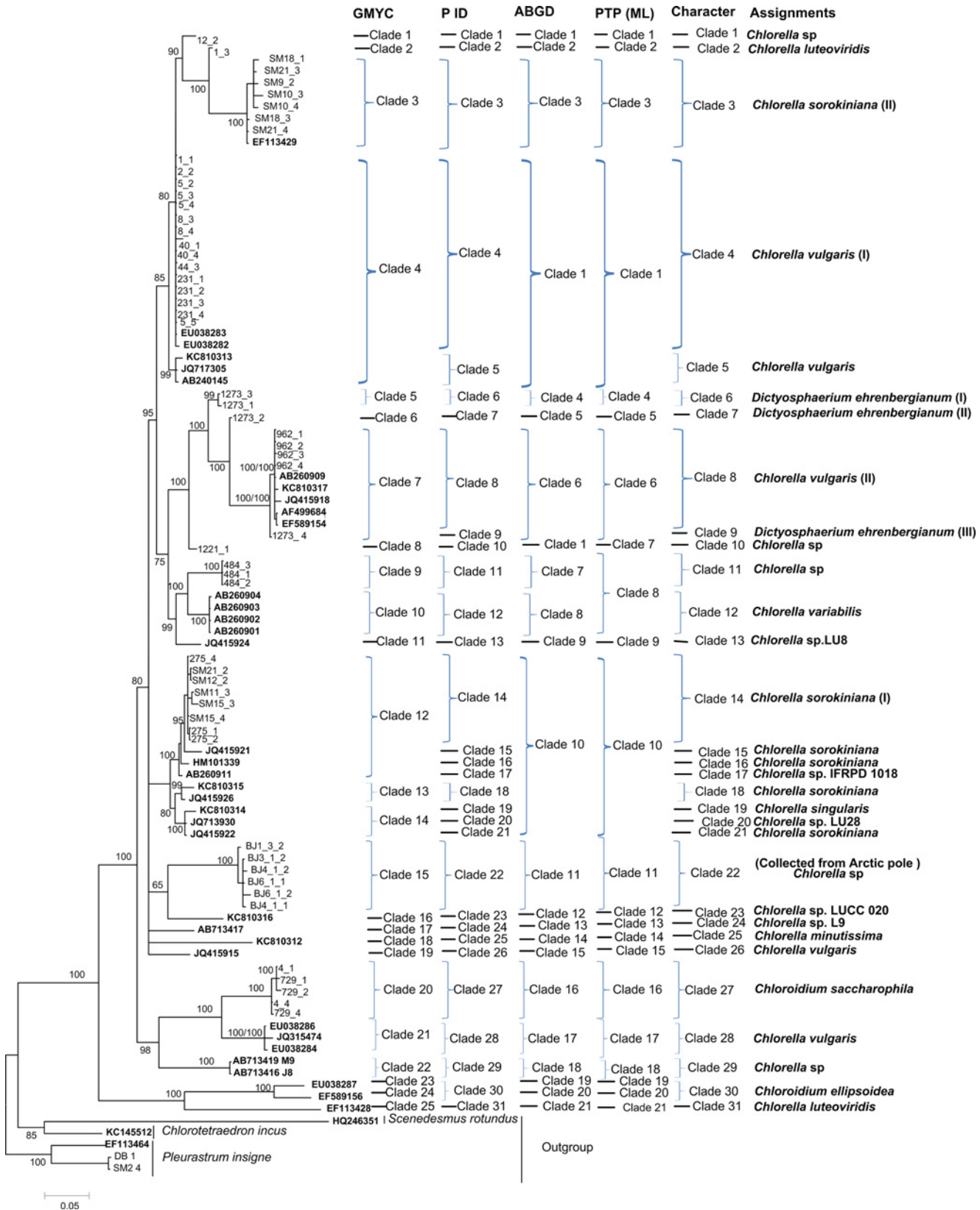


Fig 1. Bayesian phylogenetic tree for the *rbcL* gene. Vertical bars on the right indicate the clades detected by the coalescent-based GMYC, PID, PTP, the distance-based ABGD approach, the character-based CAOS and the final assignment. Posterior probabilities and NJ bootstrap values were included.

doi:10.1371/journal.pone.0153833.g001

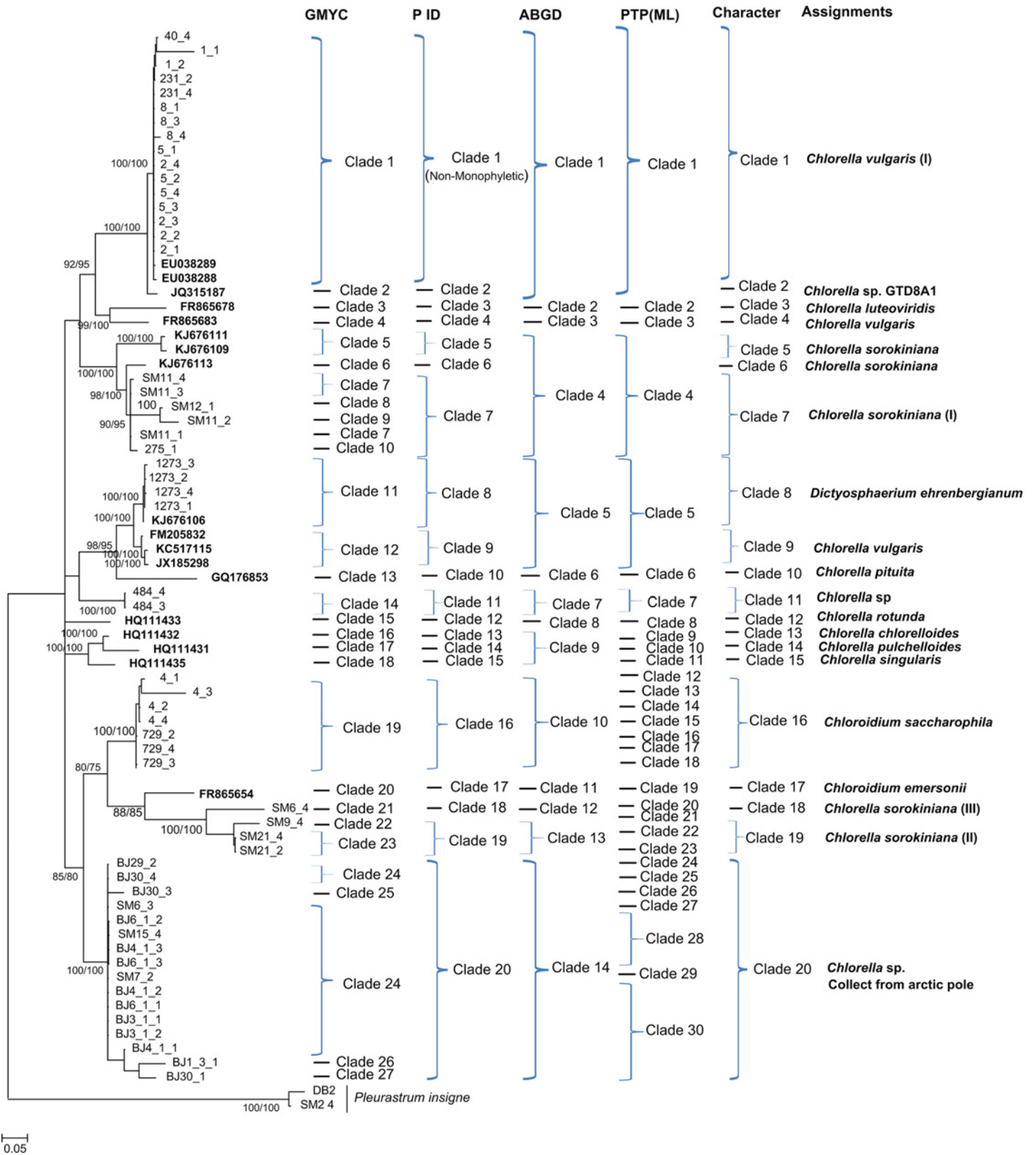


Fig 2. Bayesian phylogenetic tree for the ITS gene. Vertical bars on the right indicate the clades detected by the coalescent-based GMYC, PID, PTP, the distance-based ABGD approach, the character-based CAOS and the final assignment. Posterior probabilities and NJ bootstrap values were included.

doi:10.1371/journal.pone.0153833.g002

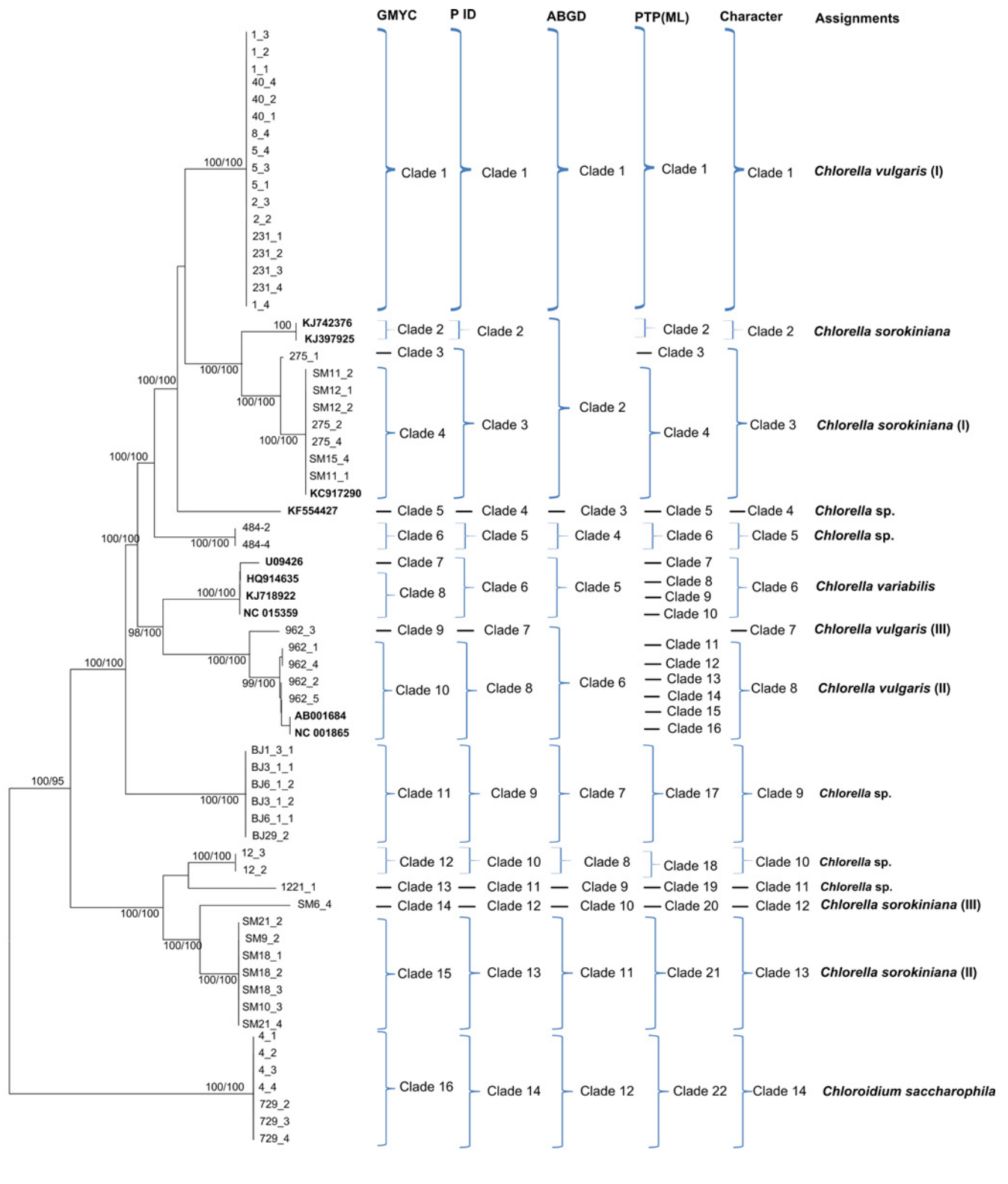


Fig 3. Bayesian phylogenetic tree for the *tufA* gene. Vertical bars on the right indicate the clades detected by the coalescent-based GMYC, PID, PTP, the distance-based ABGD approach, the character-based CAOS and the final assignment. Posterior probabilities and NJ bootstrap values were included.

doi:10.1371/journal.pone.0153833.g003

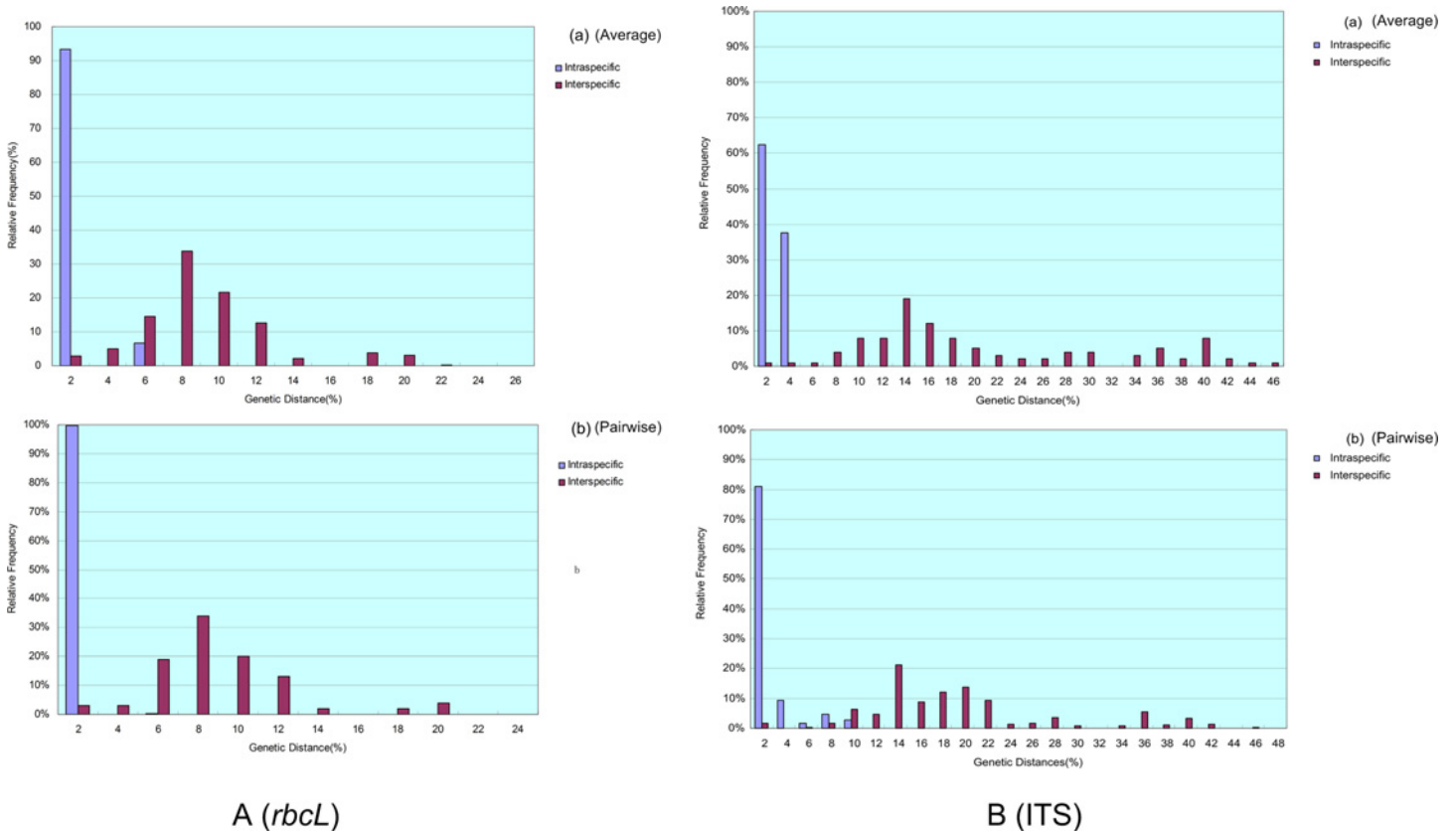


Fig 4. A (a): Histograms of intra- (in blue) and inter-specific (in red) average distances between *rbcL* sequences; (b): Histograms of intra- (in blue) and inter-specific (in red) pairwise distances between *rbcL* sequences. **B** (a): Histograms of intra- (in blue) and inter-specific (in red) average distances between ITS sequences; (b) Histograms of intra- (in blue) and inter-specific (in red) pairwise distances between ITS sequences.

doi:10.1371/journal.pone.0153833.g004

interspecific distances of *rbcL* sequences (Fig 4A). The ITS pairwise intraspecific distance ranged from 0% to 9.3% with a mean of 1.6% while the pairwise interspecific distances was from 0% to 45.1%. The mean ITS divergence among the separate cryptic lineages within *C. sorokiniana* and *C. vulgaris* ranged from 1.7%-36.6% and 13.4%-18.4% respectively, which were also higher than 1.6% (the mean intraspecific distance) (S4 Table). However, as with *rbcL*, there was apparent overlap between the intra- and interspecific distances of ITS sequences (Fig 4B). The *tufA* divergences among 14 *Chlorella*-like taxa in phylogenetic and character analysis were analyzed. The *tufA* pairwise intraspecific distance ranged from 0% to 1% with a mean of 0.1% while the pairwise interspecific distances was from 0% to 28.6%. The mean *tufA* divergence among the cryptic lineages *C. sorokiniana* (I),(II),(III),(KJ742376,KJ397925) and *C. vulgaris* (I),(II),(III), ranged from 7.9%-28.6% and 10.7%-16.1% respectively, which was also greatly higher than the mean intraspecific distance (0.1%) (S5 Table). Since all the intraspecific distance was lower than 2% (proposed as 10× rule by Hebert et al. [35]) and all the interspecific distance was higher than 2%, there was a small barcoding gap between the intra- and interspecific variation of *tufA* sequences (Fig 5).

ABGD analysis

Based on the distance-based approach as implemented in the software ABGD, different groups as candidate species were produced for *rbcL*, ITS and *tufA* gene sequences. Generally, the ABGD analysis of *rbcL*, ITS and *tufA* produced fewer genetic groups than other barcoding

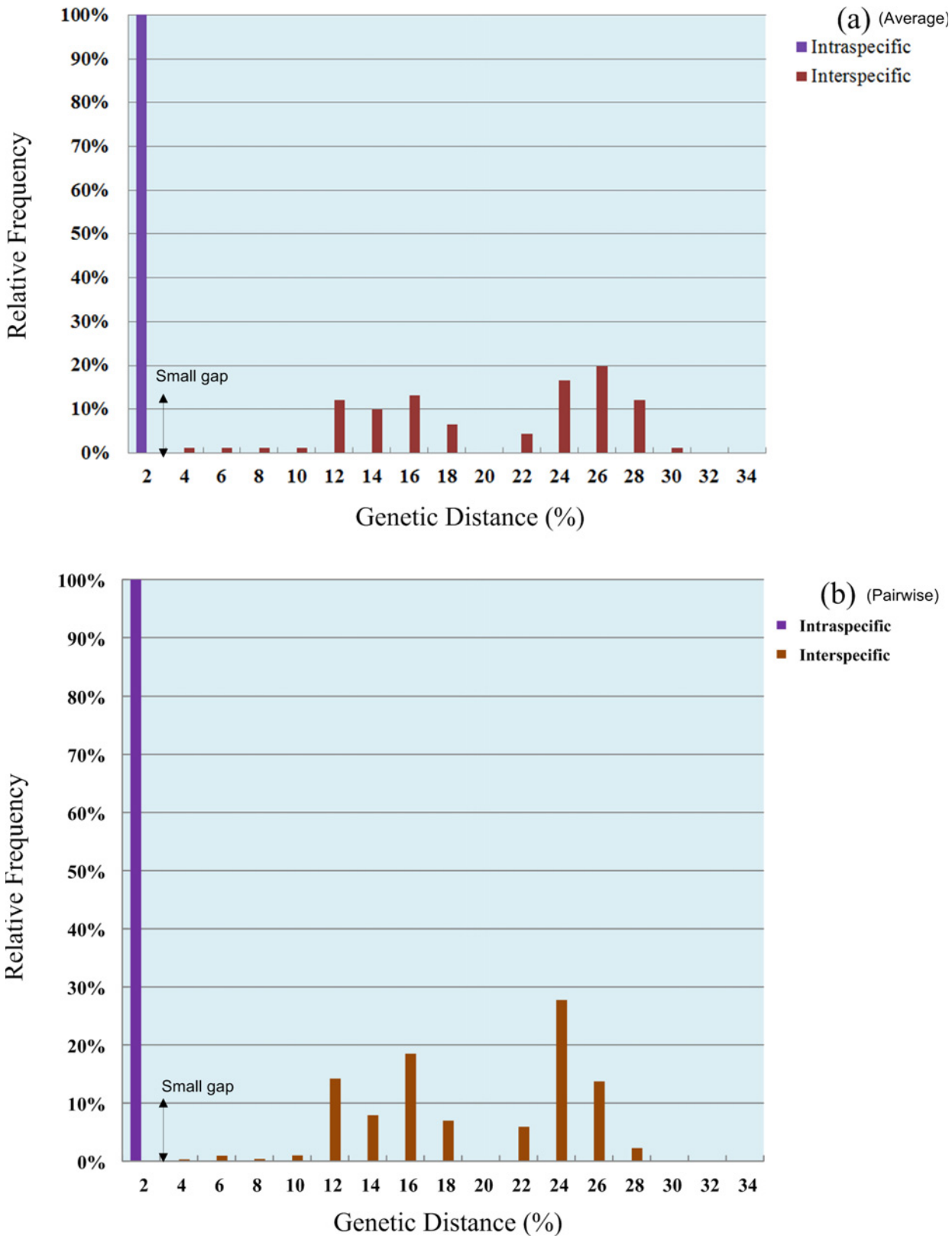


Fig 5. (a) Histograms of intra- (in blue) and inter-specific (in red) average distances between *tufA* sequences; (b) Histograms of intra- (in blue) and inter-specific (in red) pairwise distances between *tufA* sequences.

doi:10.1371/journal.pone.0153833.g005

methods (Figs 1–3). For *rbcL*, the ABGD analysis revealed 21 genetic groups when using restrictive values with priori genetic distance thresholds 0.77% (Fig 1 and S6 Fig). Most of the 21 groups were consistent with the ABGD analysis of ITS and *tufA* in which 14 groups and 12 groups were produced at a priori genetic distance thresholds of 1% and 3.59% respectively (Figs 2 and 3, S7 and S8 Figs). In all analyses, *C. vulgaris* and *C. sorokiniana* were split into several groups.

GMYC species delimitation

The optimal threshold points obtained by the GMYC model for *rbcL*, ITS and *tufA* genes were shown in Figs 1–3, S9–S11 Figs, respectively. As a whole, the specimens studied were oversplit by the GMYC model for *rbcL*, ITS and *tufA* genes in comparison with ABGD analysis (Figs 1–3). The results of the single threshold analysis for the *rbcL*, ITS and *tufA* gene suggested 25, 27 and 16 groups respectively, some of which consisted of single specimens. Especially in ITS analysis, the *C. sorokiniana* (I) and *C. sp.* Collect from Arctic pole were split into several groups.

PTP-based identification

The resolution produced by bPTP approach was variable among *rbcL*, ITS and *tufA* genes (Figs 1–3). The maximum-likelihood identification produced better resolution than bayesian identification. For *rbcL*, it recognized 21 independent entities which were consistent with the groups revealed by ABGD analyses (Fig 1). However, for ITS and *tufA*, the taxa were all over split by PTP analysis than other methods (Figs 2 and 3).

P ID-based identification

Based on the bayesian analysis, the tree-based hypotheses were reevaluated for species hypothesis testing. Most candidate species were recovered as monophyletic clades in P ID species boundary delimitation of *rbcL*, ITS and *tufA* genes except *Chlorella vulgaris* (I) which was not monophyletic in *rbcL* analysis (Figs 1–3). The resolution produced by P ID method was generally consistent with the Character analysis. All delimited species of *rbcL*, ITS and *tufA* possessed a P ID (Liberal) value $P > 0.7$ (S6–S8 Tables).

Character-based identification

Based on the morphological identification, traditional barcoding, GMYC, PTP, P ID and ABGD analysis, above 31, 20 and 14 defined *Chlorella*-like clades recovered by *rbcL*, ITS and *tufA* sequences (Figs 1–3) were analyzed respectively for searching for diagnostic characters. It was shown that all the *Chlorella*-like species including the possible cryptic lineages and unknowns were clearly distinguished in the character-based DNA barcoding. In the *rbcL* gene region of 31 *Chlorella*-like taxa recovered in Fig 1, 45 character states were detected (Fig 6), in which all the 31 clades revealed a unique combination of character states at 45 nucleotide positions with more than three CAs. The possible cryptic lineages within *C. sorokiniana*, *D. ehrenbergianum* and *C. vulgaris*, e.g. *C. sorokiniana* (I),(II), *C. vulgaris* (I),(II) and *D. ehrenbergianum* (I),(II),(III) were all clearly separated with many diagnostic characters (Fig 1, Fig 6). The ITS character-based DNA barcode were shown in Fig 7, in which 20 defined *Chlorella*-like clades recovered in Fig 2 revealed a unique combination of character states, including the possible cryptic lineages *C. sorokiniana* (I),(II),(III) and *C. vulgaris* (I),(FR865683),(FM205832,KC517115,JX185298). The *tufA* character states for 14 *Chlorella*-like clades recovered in *tufA* NJ tree (Fig 3) were shown in Fig 8. At 30 nucleotide positions of the *tufA* gene region more than five CAs were revealed for each clade, also including the cryptic lineages *C. sorokiniana* (I),(II),(III) and *C. vulgaris* (I),(II),(III). In comparison with *rbcL* and ITS, *tufA* detected the most diagnostic characters in the fewest

Species (cryptic lineage number, strain designation or Genbank number)	Positions																																																	
	151	166	175	190	196	239	255	260	267	287	290	299	308	311	378	395	404	425	482	614	626	632	658	660	702	705	706	714	718	720	730	738	745	760	792	803	804	810	812	840	864	885	914	920	998					
<i>Chlorella sorokiniana</i> (II)	A	T	T	T	C	C	A	A	T	T	T	C	T	G	C	T	T	A	A	T	A	C	A	T	T	A	G	A	C	G	G	A	T	A	A	G	T	G	A	T	T	T	A	T	T					
<i>Chlorella sorokiniana</i> (I)	G	C	T	T	T	C	A	A	T	T	T	C	A	T	C	A	T	A	A	C	T	T	T	G	A	T	A	G	T	T	A	A	G	T	T	A	A	G	T	A	A	G	C	G	T	A				
<i>Chlorella sorokiniana</i> (JQ415921)	-	C	T	T	T	C	A	A	T	T	T	C	A	T	C	A	T	A	A	C	T	T	G	A	T	A	G	T	T	G	A	A	G	T	T	A	A	G	T	A	A	C	C	G	T	A				
<i>Chlorella sorokiniana</i> (HM101339)	G	C	T	T	T	C	A	A	T	T	T	C	A	T	C	A	T	A	A	C	T	T	T	G	A	T	A	G	T	T	A	A	G	T	T	A	A	G	T	T	A	A	G	C	G	T	A			
<i>Chlorella sorokiniana</i> (KC810315, JQ415926)	-	T	T	T	T	C	A	A	T	T	T	T	C	A	T	C	A	A	C	A	C	A	T	T	G	A	T	A	G	T	T	A	A	G	T	T	A	A	G	T	T	A	A	C	G	T	A			
<i>Chlorella sorokiniana</i> (KC810314)																																																		
<i>Chlorella</i> sp. (LU28)	-	C	T	T	T	C	A	A	T	T	T	C	T	A	T	C	A	C	A	A	C	G	C	A	T	T	T	G	A	T	A	G	G	T	A	T	G	T	T	A	A	G	C	A	T	A				
<i>Chlorella singularis</i> (JQ415922)	-	C	T	T	T	C	A	A	T	T	T	C	T	A	T	C	A	C	A	A	C	G	C	A	T	T	T	G	A	T	A	G	G	T	A	T	G	T	T	A	A	G	C	A	T	A				
<i>Chlorella vulgaris</i> (I)	G	C	T	C	T	C	A	A	T	T	T	C	T	A	C	T	T	A	A	T	T	T	A	T	A	G	A	C	T	G	T	T	A	A	G	T	T	A	A	G	T	T	A	A	T	A	T	A		
<i>Chlorella vulgaris</i> (II)	G	T	T	C	T	T	A	A	T	A	T	C	T	A	C	T	T	A	A	T	T	C	A	T	A	G	A	A	G	G	G	T	A	A	G	T	T	A	A	G	T	T	A	A	T	A	T	A		
<i>Chlorella vulgaris</i> (KC810313, JQ717305, AB240145)	G	C	T	C	T	C	A	A	T	T	T	C	T	A	C	T	C	A	C	T	T	A	A	C	T	A	T	A	G	A	C	T	G	T	T	A	A	G	T	T	A	A	T	C	A	T	T			
<i>Chlorella vulgaris</i> (EU038286, JQ315474, EU038284)	-	-	T	C	T	C	A	A	T	A	T	C	G	A	C	T	T	A	T	C	G	C	A	T	T	T	G	A	C	G	G	A	T	A	A	G	C	T	A	A	T	A	T	A	T	A				
<i>Chlorella vulgaris</i> (JQ415915)	A	C	T	T	T	T	A	T	T	T	T	T	A	C	T	A	T	A	C	A	T	A	T	T	A	G	A	C	T	C	A	T	G	A	G	T	T	C	A	T	G	T	T	G	T	T				
<i>Chlorella luteoviridis</i> (EF113428)	T	T	A	C	A	G	G	A	C	T	A	C	T	A	C	A	T	A	A	T	T	C	G	A	T	T	T	T	A	G	G	A	C	A	T	C	G	A	A	A	C	C	A	A	T	A				
<i>Chlorella luteoviridis</i> (1_3)	G	C	T	T	T	C	A	A	T	T	T	C	T	A	C	T	A	C	T	T	A	A	T	T	C	A	T	T	A	G	A	C	G	G	A	T	A	A	G	T	T	A	A	T	T	A	T	A		
<i>Dictyosphaerium ehrenbergianum</i> (I)	A	C	A	T	T	C	A	A	T	A	T	C	T	A	C	T	T	A	A	T	T	A	T	T	A	G	A	A	G	G	T	A	A	G	T	T	A	A	G	T	A	A	T	T	A	T	A			
<i>Dictyosphaerium ehrenbergianum</i> (II)	A	C	A	T	T	C	A	A	T	A	T	C	T	A	C	T	T	A	A	T	T	A	T	T	A	G	A	A	G	G	T	A	A	G	T	T	A	A	G	T	T	A	A	T	T	A	T	A		
<i>Dictyosphaerium ehrenbergianum</i> (III)	G	T	T	C	T	C	A	A	T	A	T	C	T	A	C	T	T	A	A	T	T	C	A	T	T	A	G	A	A	G	G	T	A	A	G	T	T	A	A	G	T	T	A	A	T	A	T	A		
<i>Chloroidium ellipsoidea</i>	A	T	A	C	A	A	G	A	C	T	A	C	T	A	C	A	T	A	A	T	T	C	G	A	T	T	G	A	T	G	A	C	A	T	C	A	A	G	T	A	A	G	C	A	A	T	A			
<i>Chlorella variabilis</i>	A	T	T	C	C	A	A	T	A	T	T	A	C	T	T	A	A	C	T	T	A	A	C	T	A	T	G	A	C	T	G	A	T	G	A	T	G	A	T	G	T	A	A	T	T	A	T	G		
<i>Chlorella minutissima</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
<i>Chloroidium saccharophila</i>	T	C	T	T	T	T	T	T	T	T	T	T	T	C	T	A	C	A	G	T	T	T	T	T	T	T	T	A	A	C	A	C	A	G	A	T	T	C	A	C	T	T	C	A	T	T	T	A		
<i>Chlorella</i> sp. (12_2)	A	C	T	T	C	C	A	A	T	T	T	C	T	T	A	C	T	T	A	A	T	T	A	T	T	A	G	A	C	T	G	T	T	A	A	G	T	T	A	A	G	T	T	A	A	T	A	T	A	
<i>Chlorella</i> sp. (1221_1)	A	C	A	T	T	C	A	A	T	T	T	C	T	A	C	T	T	A	A	T	T	T	T	A	T	T	A	G	A	C	T	G	T	T	A	A	G	T	T	A	A	G	T	T	A	A	T	A	T	A
<i>Chlorella</i> sp. (484)	A	T	T	C	T	C	A	A	T	A	T	C	A	A	C	T	C	A	A	C	G	T	A	T	G	T	G	A	C	A	G	T	T	A	A	G	T	T	A	A	G	T	T	A	G	T	T	G	T	T
<i>Chlorella</i> sp. (LU8)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
<i>Chlorella</i> sp. (BJ)	T	T	T	T	T	T	A	T	T	T	T	C	T	A	C	T	T	A	G	T	T	T	A	T	T	T	A	A	G	T	T	A	A	G	C	T	A	A	G	C	T	A	A	T	T	T	T	T		
<i>Chlorella</i> sp. (LUCC 020)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
<i>Chlorella</i> sp. (IFRPD 1018)	G	C	T	T	T	C	A	A	T	T	T	C	A	T	C	A	T	A	A	C	T	C	A	T	T	G	A	T	A	G	T	T	A	A	G	T	T	A	A	G	T	T	A	A	G	C	G	C	A	
<i>Chlorella</i> sp. (L9)	G	C	T	C	T	C	A	T	T	T	T	C	A	T	C	T	T	A	A	C	T	T	A	T	A	G	C	T	A	G	A	T	A	A	G	T	T	A	A	G	T	T	A	A	T	C	T	T	T	
<i>Chlorella</i> sp. (J8,M9)	T	C	T	T	T	C	G	A	T	C	T	C	T	T	T	T	T	T	A	C	A	C	A	T	A	G	A	C	A	G	A	C	A	T	G	T	T	A	A	T	A	T	A	T	A	T	T	T		

Fig 6. Combinations of diagnostic nucleotides for each of the 31 *Chlorella*-like taxa recovered in Fig 1. Nucleotide numbers refer to 45 selected positions on the *rbcl* sequences (positions 151 to 998).

doi:10.1371/journal.pone.0153833.g006

nucleotide positions. Therefore the discrimination of taxa of all clades, including cryptic and unknown taxa, could be resolved by character-based DNA barcoding.

Discussion

This study showed that DNA barcoding based on multiple barcoding approaches was useful in species identification and cryptic species revelation of *Chlorella*-like green microalgae. Actually, like *Chlorella*, the identification of microalgae is often difficult due to their morphological

Species (cryptic lineage number, strain designation or Genbank number)	Positions																																														
	229	327	382	411	447	475	477	481	503	505	506	508	511	512	560	569	570	573	592	593	599	600	610	614	615	801	808	809	811	825	826	840	844	857	877	950	968	1025	1036	1042	1062	1064	1068	1069			
<i>Chlorella vulgaris</i> (I)	C	G	T	G	C	T	G	G	A	T	T	G	T	C	C	T	A	G	C	A	G	G	T	C	T	T	T	C	C	G	T	G	A	T	G	G	G	C	C	G	C	G	G	C			
<i>Chlorella vulgaris</i> (FR865683)	C	G	-	A	C	C	A	G	T	G	C	G	T	C	A	A	A	A	T	T	G	A	T	-	T	C	T	T	G	G	T	A	T	C	A	-	T	T	A	T	T	T	T	G			
<i>Chlorella vulgaris</i> (FM205832, KC517115, JX185298)	C	T	G	A	T	C	C	T	A	C	C	A	T	C	A	T	T	A/G	A	G	T	G	A	C	A	C	A	T	T	-	-	G	T	T	G	T	T	C	A	A	C	A	C				
<i>Chlorella luteoviridis</i> (FR865678)	C	G	G	G	C	G	C	T	G	T	G	A	C	T	C	C	T	G	G	C	C	G	A	C	A	C	G	A	C	A	-	-	G	C	T	G	C	C	C	T	C	G	G	C			
<i>Chlorella sorokiniana</i> (I)	C	G	T	T	G	C	T	G	T	G	C	T	G	T	C	C	T	T	G	G	T	C	C	C	A	C	C	C	C	G	T	T	G	C	-	G	C	C	C	C	C	G	T	A			
<i>Chlorella sorokiniana</i> (II)	G	-	-	T	A	T	A	G	T	C	T	G	T	C	A	A	A	A	A	T	T	A	A	T	T	C	T	C	C	G	G	T	G	A	A	C	A	A	C	-	-	G	G	T			
<i>Chlorella sorokiniana</i> (III)	G	-	-	A	T	G	-	T	C	T	A	A	C	A	A	A	G	C	T	A	A	T	T	C	C	C	C	C	G	G	T	G	G	A	A	C	A	A	G	C	T	G	G	T			
<i>Chlorella sorokiniana</i> (KJ676111, KJ676109)	C	G	A	G	C	T	G	T	C	A	A	A	C	C	C	C	T	G	G	G	T	G	A	C	A	C	C	C	C	G	T	T	G	C	-	G	C	C	C	C	A	T	A	C			
<i>Chlorella sorokiniana</i> (KJ676113)	C	G	T	A	C	T	G	T	G	C	T	T	T	C	C	T	G	G	G	T	C	C	A	C	C	C	-	G	T	C	G	C	-	G	C	C	C	C	C	C	C	G	T	A			
<i>Chlorella pituita</i>	A	T	C	A	C	C	C	C	A	C	A	G	T	C	A	C	A	G	G	G	T	T	A	C	-	C	A	T	T	T	T	-	G	C	T	G	T	C	T	C	T	C	C	A	C		
<i>Chlorella</i> sp. (484)	C	G	-	A	C	C	T	A	A	A	A	A	T	C	C	C	A	G	A	C	G	T	C	A	C	A	C	C	T	C	G	T	C	G	-	A	T	C	C	G	C	C	C	C			
<i>Chlorella rotunda</i>	C	G	G	A	C	C	G																																								

Species (cryptic lineage number, strain designation or Genbank number)	Positions																													
<i>Chlorella vulgaris</i> (I)	T	C	C	T	C	G	G	G	G	C	A	T	A	A	T	C	G	A	G	G	A	C	C	T	A	A	A	G	A	A
<i>Chlorella vulgaris</i> (II)	T	C	A	T	A	A	A	A	A	C	A	A	A	G	T	T	T	T	G	T	A	A	A	A	A	T	A	T	A	G
<i>Chlorella vulgaris</i> (III)	T	C	C	T	C	A	G	A	A	C	A	T	A	G	T	T	T	G	G	T	A	A	T	A	T	T	A	G	A	G
<i>Chlorella sorokiniana</i> (I)	T	C	A	A	A	A	G	G	G	C	A	A	C	G	T/C	T	A	A/T	G	T	A	T	A	T	T	A	A	T	A	
<i>Chlorella sorokiniana</i> (II)	A	T	C	A	C	A	T	A	A	G	T	A	T	A	T	T	G	A	A	A	T	A	T	T	A	T	T	T	A	C
<i>Chlorella sorokiniana</i> (III)	T	C	C	A	C	G	G	G	G	C	A	T	C	G	T	T	A	A	G	T	A	T	C	T	T	A	A	A	T	A
<i>Chlorella sorokiniana</i> (KJ742376, KJ397925)	T	C	T	A	T	A	G	G	G	C	A	A	T	A	T	C	A	A	G	T	A	T	A	C	T	A	A	A	T	A
<i>Chlorella variabilis</i>	G	C	A	A	A	A	A	G	G	C	A	T	G	T	T	T	C	T	A	A	G	A	A	T	A	A	A	T	A	A
<i>Chloroidium saccharophila</i>	A	C	A	T	A	A	G	A	A	C	T	A	T	T	C	T	A	A	G	A	A	T	A	T	A	A	T	A	A	A
<i>Chlorella</i> sp. (484_2, 484_4)	T	C	C	T	C	A	G	G	G	C	A	T	G	T	T	C	T	G	G	C	A	A	T	C	T	T	A	G	G	A
<i>Chlorella</i> sp. (BJ)	T	C	C	G	C	A	A	G	G	C	A	A	C	A	T	T	T	A	G	T	C	A	G	T	A	T	A	T	T	A
<i>Chlorella</i> sp. (1221_1)	A	T	A	C	A	T	A	A	A	T	A	A	T	C	T	A	A	A	T	A	T	A	T	T	A	A	T	A	A	C
<i>Chlorella</i> sp (12_2,12-3)	A	G	A	A	A	A	T	A	A	G	T	A	A	T	C	T	A	A	A	T	A	T	T	A	T	T	A	T	T	C
<i>Chlorella</i> sp. (ArM0029B)	T	C	T	A	T	A	A	G	G	C	A	T	A	A	T	T	A	C	T	A	A	A	C	G	A	A	T	A	A	

Fig 8. Combinations of diagnostic nucleotides for each of the 14 *Chlorella*-like taxa recovered in Fig 3. Nucleotide numbers refer to 30 selected positions on the *tufA* sequences (positions 19–673).

doi:10.1371/journal.pone.0153833.g008

plasticity or tiny body, which enables DNA barcoding as a powerful tool for revealing microalgal biodiversity, particularly with the combination of different analytical approaches.

Chlorella primers for barcoding

DNA barcoding of plants has struggled to seek universal DNA barcodes that not only have the discriminatory power but also are easily amplified across Plantae as the COI gene chosen for animals. However, none of the available barcode loci worked well across the kingdom Plantae that includes land plants and algae [113]. For example, in molecular identification of freshwater green algae [77], the ITS1 and ITS2 regions were successfully amplified in only partial samples, which impeded the gene regions for barcoding green algae. For the barcoding of marine green macroalgae, low amplification success of *rbcL* reduced the utility of this marker as a universal barcode system [76]. In this study, the *rbcL*, 16S and *tufA* genes were successfully amplified in only a small part of *Chlorella*-like samples with the available primers. Amplification failure might be because of the primer site incompatibility in them since we have amplified the gene ITS from the same DNA templates. Thus, the *rbcL* and 16S specific primers for *Chlorella*-like strains were developed in this study based on known sequences from this study and Genbank, which turned out to be effective for amplifying and sequencing some *Chlorella*-like samples. For *tufA*, however, due to its highly variable sites and very limited available sequences, it was not possible to design the specific primers at present.

Barcoding identification and unveiling cryptic diversity of *Chlorella*-like taxa

Identification of *Chlorella*-like organisms has long been problematic on the basis of morphological characteristics. Since most molecular studies of *Chlorella* focus on the phylogenetic analysis, the identification of *Chlorella*-like green microalgae at species level often analyzed limited gene loci or samples. For example, for *rbcL*, a commonly used molecular marker for algae identification, prior to the initiation of this research, only six sequences of *Chlorella* species had been deposited in Genbank, and most of them were identified as unknowns. For both *Chloroidium* and *Dictyosphaerium*, only three *rbcL* sequences had been deposited in Genbank. For *tufA*, another commonly used marker, only 27 *Chlorella* sequences had been deposited in the Genbank database, and 16 of them were not identified to the species level. Moreover, the previous molecular taxonomic identification of *Chlorella*-like algae, including the use of DNA barcoding, was generally based on the phylogenetic trees or genetic distance that has often been disputed for barcoding. Therefore, the taxonomy of *Chlorella*-like species is still very

complicated, and numerous cryptic species within *Chlorella*-like taxa may be still not revealed. In the present study, the barcoding analysis based on distance and character-based approaches indicate that the sequences analyzed can give useful information into the taxonomic assessment of *Chlorella*-like strains, including the species discrimination and the revelation of unveiling complexes of possible cryptic species.

Firstly, the comprehensive barcoding analysis enabled the separation of all the specimens studied. The NJ, Bayesian and Maximum Likelihood trees produced with *rbcL*, ITS, *tufA* and 16S generally revealed congruent species delineation topologies, which revealed distinct and deeply diverged lineages. Based on the integrated analysis of GMYC, PTP, P ID and ABGD methods, the character-based barcoding of *rbcL*, ITS and *tufA* all showed clear taxa assignments that corresponded to the diverged lineages in the phylogenetic trees. Secondly, the barcoding analysis led to the revelation of potentially cryptic species in *C. sorokiniana*, *C. vulgaris* and *D. Ehrenbergianum*. These potentially cryptic species were all separated in all GMYC, PTP, ABGD, P ID and character-based barcoding analysis of *rbcL*, ITS and *tufA* genes. Particularly, they are clearly recovered with many diagnostic characters in character-based barcoding. In addition, to some extent, the interspecific genetic variation of *rbcL*, ITS and *tufA* among the possible cryptic lineages were higher than the intraspecific divergence (see the intra- and interspecific variation results in [S3–S5](#) Tables). Thus, all the barcoding results revealed the potentially cryptic species complexes in *C. sorokiniana*, *C. vulgaris* and *D. Ehrenbergianum*. In short, the genetic data in this study indicate that there is extraordinary cryptic diversity in *Chlorella*-like taxa and further taxonomic re-evaluation of these possible cryptic species should be performed. Finally, many unknown *Chlorella*-like samples in this study did not match well with sequences published in BOLD and Genbank database. Our DNA barcoding analysis did not allow the identification of unknown specimens at the species-level. Researchers argue that barcoding is helpful in species discovery by evaluating their sequence divergence [[114,115](#)]. That is, if a matching target sequence in a barcoding database is absent the novelty of the species is generated. Thus, the unknown *Chlorella*-like specimens in this study need to be further studied. At present the DNA barcoding databases contain a limited number of reference sequences (targeted barcode) for microalgae. More available microalgae target barcoding sequences will be greatly helpful to understand the microalgae diversity since only with an increase of accurate barcode sequences in the target databases is how DNA barcoding methods can help to produce reliable assignments of unknown species (query sequences). Also it is becoming apparent that an increased sampling may be needed to ensure the presence of a discernable barcoding gap between interspecific divergence and intraspecific variation in any given taxon and to confirm the existence of diagnostic molecular characters [[50,116](#)].

‘Specific barcode’ for *Chlorella*-like green microalgae

COI was suggested as the locus that could provide recognition tags for all animals [[34,35, 45](#)]. However, COI along with other mitochondrial genes are not suitable for barcoding plantae due to their very low rates of substitution [[117](#)]. Thus, the search for plant barcodes shifted to chloroplast and nuclear genomes with high substitution rates. Despite some arguments, the most viable candidates as DNA barcode loci for plants are *rbcL* and *matK*. However, previous findings show that the *matK* or *rbcL* gene alone can not be used as a suitable universal barcode [[89,91,117](#)]. Moreover, *matK* is absent in algae. In this context, the concept of ‘specific barcode’ for plants is put forward, which involves a trade-off between single-locus barcodes and superbarcodes [[91](#)]. The ‘specific barcodes’ for different plant groups will resolve better resolution for DNA barcoding of Plantae. The *rbcL*, ITS and *tufA* have been recommended as the most promising DNA barcodes for some green algae [[76,77](#)]. In the present study, the 16S gene

which has been used for identifying *C. vulgaris* [95] failed in discriminating most *Chlorella*-like strains, especially for the closely related species, which corresponds with the arguments that mitochondrial genes are not suitable for plant barcoding. Both of *rbcL* and ITS proved useful in distinguishing most *Chlorella*-like taxa. Yet a much higher proportion of resolution success was shown by *rbcL*, in comparison with *tufA* and ITS, including the existence of a small barcoding gap, the consistent groups among GMYC, ABGD and P ID methods, and many more diagnostic characters. Therefore, the *tufA* could be as potentially suitable 'specific barcode' for *Chlorella*-like taxa, which of course needs to be further compared with other gene sequences.

Efficiency of distance and character-based DNA barcoding

Several different methods of distinguishing species have been advanced by members of the barcoding community, but which method is the best is still in debate, especially for plants [50,68,69,89,117–121]. Although the phylogenetic or distance trees in traditional barcoding approaches are informative about the genetic affinities, they are arbitrary as criteria for species identification [50,68,69]. Recently, it is proposed that incorporation of multiple lines of methodologies should be used for understanding species boundaries framework to develop the initial species hypotheses where distinct clades are defined as those that do not share haplotypes between populations and can be identified as divergent ones, especially with the methods of GMYC, ABGD, PTP, P IN and CAOS [109, 122–124]. It has also been proposed that an optimal path to understand species boundaries is starting with a tree or distance-based population clusters [109]. Then the character-based approach is employed to confirm the initial identification. Our study represents one of the first efforts to test the congruence of barcoding results from multiple delimitation methods.

For traditional barcoding, generally, the NJ, Bayesian and Maximum Likelihood analysis recovered consistent topology for each gene of *rbcL*, ITS, 16S and *tufA*. However, due to the shortcoming of tree-based species identification [55, 60] the phylogenetic trees are more likely to be used initially to identify putative independently-evolving lineages. The intra and interspecific distance of traditional barcoding was also analyzed in this study. To some degree, the distance method was helpful in species discrimination. For example, as a whole, the interspecific variation of *rbcL*, ITS and *tufA* sequences among the potentially cryptic species complexes in *C. sorokiniana*, *C. vulgaris* and *D. Ehrenbergianum* was higher than the intraspecific variation. Nevertheless, for all of *rbcL*, ITS and *tufA* sequences, although the interspecific genetic variation was generally higher than the intraspecific genetic variation, there was no apparent barcoding gap between them (seen S3–S5 Tables). That is, the minimum interspecific distance is smaller than the maximum intraspecific distance, which contradicts the criterion of species identification with sequences distance [89].

The resolution produced by GMYC, PTP, P ID, ABGD and character-based barcoding methods were variable in each of *rbcL*, ITS and *tufA* genes. In *rbcL* analysis, the groups recovered by GMYC, ABGD and PTP methods were consistent while the groups recovered by P ID and character methods were consistent. In ITS analysis, the groups produced by GMYC, ABGD and PTP methods were all different from each other while P ID and character methods produced same groups. The best resolution for species differentiation appeared in *tufA* analysis where GMYC, PTP, ABGD and character-based approaches produced consistent groups while the PTP method over-split the taxa. Similar to previous studies [122, 125–129], GMYC typically generates more OTUs (operational taxonomic units) than other approaches for *rbcL* sequences and errors in the ultrametric gene tree will influence final results. The PTP, however, generate more OTUs than other methods in both ITS and *tufA* genes. Generally, the P ID and character-based methods produced consistent groups in all *rbcL*, ITS and *tufA* genes.

Based on the integrated analysis of traditional barcoding, GMYC, ABGD, PTP and P ID methods, the putative species recovered were confirmed by character-based barcoding. The character-based DNA barcoding showed more advantages, particularly for revealing the possible cryptic lineages. For example, as expected, the character-based analysis generated relatively congruent results in *rbcL*, ITS and *tufA* genes, and most taxonomic groups analyzed by *rbcL*, ITS, 16S and *tufA* genes, including the potentially cryptic species, possessed unique simple identifying character states in character-based barcoding. Some species that could not be discriminated with traditional barcoding, GMYC, PTP or ABGD methods could be detected by character-based method, e.g. *Chlorella vulgaris* (I) and *Chlorella sorokiniana* (I) in *rbcL* barcoding analysis, and *Chlorella sorokiniana* (I) in ITS barcoding analysis (Figs 1–3). In addition, if one species is represented with only a single individual or not all closely related species are sampled it is not possible to determine the correct intra- and interspecific divergences, which may hinder the presence of a discernable barcoding gap. Nevertheless, a single individual can be still assigned to a distinct clade in character-based DNA barcode. In this study, quite a few *Chlorella*-like taxa represented with only a single individual were clearly distinguished with unique combination of character attributes, especially for the unknowns. This is particularly useful for flagging hidden new species. Thus, a character-based discrimination criterion can maximize the success rate of molecular identification in *Chlorella*-like organism, which can resolve cases that the coalescent and distance-based barcoding does not. It may be an optimal option to first combine multiple barcoding approaches to test primary species hypotheses species and then confirm the taxonomic assignments by the character-based method. Future DNA barcoding of comprehensive *Chlorella*-like green microalgae with character-based analysis may move towards a better understanding of this morphologically complex microalgae.

Conclusion

This study indicates that the combination of *rbcL*, ITS and *tufA* sequence data analyzed by combination of GMYC, ABGD, PTP, P ID and character-based barcoding is very useful to discriminate the *Chlorella*-like samples and reveal the complexes of potentially cryptic species that merit further study. The resolution produced by GMYC, PTP, P ID, ABGD and character-based barcoding methods were variable in each of *rbcL*, ITS and *tufA* genes. The *tufA* produced consistent groups among GMYC, ABGD, P ID and character-based methods and also offered many more diagnostic characters than *rbcL* and ITS. The *tufA* region thus could be as potentially suitable 'specific barcode' for *Chlorella*-like taxa. On the other hand, all the character analysis of *rbcL*, ITS and *tufA* sequence could clearly distinguish all taxonomic groups, including the potentially cryptic lineages, with many character attributes. In comparison with other barcoding methods, the character-based discrimination criterion can maximize the success rate of molecular identification in *Chlorella*-like organisms, which can resolve cases that the distance and coalescent-based criterion does not. The character-based barcoding could be used as an attractive complement to coalescent and distance-based barcoding. It could be an optimal option to first combine multiple barcoding approaches to test primary species hypotheses species and then confirm the taxonomic assignments by the character-based method. Further DNA barcoding of comprehensive *Chlorella*-like green microalgae with character-based analysis may move towards a better understanding of this morphologically complex microalgae.

Supporting Information

S1 Fig. Maximum Likelihood tree for the *rbcL* gene.
(TIF)

S2 Fig. Maximum Likelihood tree for the ITS gene.
(TIF)

S3 Fig. Maximum Likelihood tree for the *tufA* gene.
(TIF)

S4 Fig. Bayesian phylogenetic tree for the 16S gene. Posterior probabilities and NJ bootstrap values were included.
(TIF)

S5 Fig. Maximum Likelihood tree for the 16S gene.
(TIF)

S6 Fig. Automatic partition of tellinaceans based on *rbcL* gene. The number of groups inside the partition (initial and recursive) of each given prior intraspecific divergence value were reported.
(JPG)

S7 Fig. Automatic partition of tellinaceans based on ITS gene. The number of groups inside the partition (initial and recursive) of each given prior intraspecific divergence value were reported.
(JPG)

S8 Fig. Automatic partition of tellinaceans based on *tufA* gene. The number of groups inside the partition (initial and recursive) of each given prior intraspecific divergence value were reported.
(JPG)

S9 Fig. GMYC resolution of *rbcL* genes. The red vertical line in the tree was the threshold point obtained from the GMYC model.
(TIF)

S10 Fig. GMYC resolution of ITS genes. The red vertical line in the tree was the threshold point obtained from the GMYC model.
(TIF)

S11 Fig. GMYC resolution of *tufA* genes. The red vertical line in the tree was the threshold point obtained from the GMYC model.
(TIF)

S1 Table. List of specimens with the classification, collection details, and voucher numbers.
(DOC)

S2 Table. Primer sequences and annealing temperatures used to amplify the different regions.
(DOC)

S3 Table. The mean interspecific divergences of *rbcL* sequences for *Chlorella*-like taxa.
(DOC)

S4 Table. The mean interspecific divergences of ITS sequences for *Chlorella*-like taxa.
(DOC)

S5 Table. The mean interspecific divergences of *tufA* sequences for *Chlorella*-like taxa.
(DOC)

S6 Table. Species Delimitation Results of PID for *rbcL*. The species number (clade) corresponds to the P ID clades in [Fig 1](#).
(XLSX)

S7 Table. Species Delimitation Results of PID for ITS. The species number (clade) corresponds to the P ID clades in [Fig 2](#).
(XLSX)

S8 Table. Species Delimitation Results of PID for *tufA*. The species number (clade) corresponds to the P ID clades in [Fig 3](#).
(XLSX)

Acknowledgments

The financial support from the China Postdoctoral Science Foundation (2014M561661, 2015T80558), and Natural Science Fund project in Jiangsu Province (BK20150680) were gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: SZ CW. Performed the experiments: SZ CF JS YB. Analyzed the data: SZ. Contributed reagents/materials/analysis tools: MH. Wrote the paper: SZ.

References

1. Tokuşoglu Ö, Ünal MK. Biomass Nutrient Profiles of Three Microalgae: *Spirulina platensis*, *Chlorella vulgaris*, and *Isochrysis galbana*. *J Food Sci*. 2003; 68(4):1144–1148.
2. Phukan MM, Chutia RS, Konwar BK, Katak R. Microalgae *Chlorella* as a potential bioenergy feedstock. *Appl Energ*. 2011; 88(10):3307–3312.
3. Xu H, Miao X, Wu Q. High quality biodiesel production from a microalga *Chlorella protothecoides* by heterotrophic growth in fermenters. *J Biotechnol*. 2006; 126(4):499–507. PMID: [16772097](#)
4. Krienitz L, Huss VAR, Bock C. *Chlorella*: 125 years of the green survivalist. *Trends Plant Sci*. 2015; 20: 67–69. doi: [10.1016/j.tplants.2014.11.005](#) PMID: [25500553](#)
5. Komárek J, Fott B. Chlorophyceae (Grünalgen), Ordnung Chlorococcales. *Nord J Bot*. 1983; 5 (1):111–111.
6. Huss VAR, Scharpf TK, Kessler E. Deoxyribonucleic acid reassociation in the taxonomy of the genus *Chlorella*. *Arch Microbiol*. 1989; 152(5):512–514.
7. Summerer M, Sonntag B, Sommaruga R. Ciliate-symbiont specificity of freshwater endosymbiotic *Chlorella* (Trebouxiophyceae, Chlorophyta). *J Phycol*. 2008; 44(1):77–84. doi: [10.1111/j.1529-8817.2007.00455.x](#) PMID: [27041043](#)
8. Khaybullina LS, Gaysina LA, Johansen JR, Krautova M. Examination of the terrestrial algae of the Great Smoky Mountains National Park, USA. *Fottea*. 2010; 10(2):201–215.
9. Pröschold T, Darienko T, Silva PC, Reisser W, Krienitz L. The systematics of *Zoochlorella* revisited employing an integrative approach. *Environ Microb*. 2011; 13(2):350–364.
10. Andreyeva VM. Rod *Chlorella*: Morfologiya, Sistematika, Prinzipy Klassifikaziyi. Nauka, Leningrad. 1975; 88 pp.
11. Kessler E. Comparative physiology, biochemistry, and the taxonomy of *Chlorella* (Chlorophyceae). *Plant Syst Evol*. 1976; 125(3):129–138.
12. Kessler E. Chemotaxonomy in the Chlorococcales. *Phycol Res*. 1982; 111–135.
13. Kessler E. A general review on the contribution of chemotaxonomy to the systematics of green algae. *Systematics of the green algae*. 1984; 27:391–407.
14. Nozaki H, Katagiri M, Nakagawa M, Kenichi A, M WM. Taxonomic Re-examination of the Two Strains labeled "Chlorella" in the Microbial Culture Collection at the National Institute for Environmental Studies (NIES-Collection). *JSMRS*. 1995; 11(1):11–18.

15. Kapaun E, Reisser W. A chitin-like glycan in the cell wall of a *Chlorella* sp. (Chlorococcales, Chlorophyceae). *Planta*. 1995; 197(4):577–582.
16. NEMcová Y, Kalina T. Cell wall development, microfibril and pyrenoid structure in type strains of *Chlorella vulgaris*, *C. kessleri*, *C. sorokiniana* compared with *C. luteoviridis* (Trebouxiophyceae, Chlorophyta). *Arch Hydrobiol*. 2000; 136:95–106.
17. Kessler E, Huss VAR. Comparative physiology and biochemistry and taxonomic assignment of the *chlorella* (chlorophyceae) strains of the culture collection of the university of Texas at Austin. *J Phycol*. 1992; 28(4):550–553.
18. Huss VAR, Frank C, Hartmann EC, Hirmer M, Kloboucek A, Seidel BM, et al. Biochemical taxonomy and molecular phylogeny of the genus *chlorella* sensu lato (Chlorophyta). *J Phycol*. 1999; 35(3):587–598.
19. Huss VAR, Sogin ML. Phylogenetic position of some *Chlorella* species within the chlorococcales based upon complete small subunit ribosomal RNA sequences. *J Mol Evol*. 1990; 31(5):432–442. PMID: [2124631](#)
20. Krienitz L, Hegewald EH, Heppeler D, Huss VAR, Rohr T, Wolf M. Phylogenetic relationship of *Chlorella* and *Parachlorella* gen. nov. (Chlorophyta, Trebouxiophyceae). *Phycologia*. 2004; 43(5):529–542.
21. Luo W, Pflugmacher S, Pröschold T, Walz N, Krienitz L. Genotype versus phenotype variability in *Chlorella* and *Micractinium* (Chlorophyta, Trebouxiophyceae). *Protist*. 2006; 157(3):315–333. PMID: [16843061](#)
22. Darienko T, Gustavs L, Mudimu O, Menendez CR, Schumann R, Karsten U, et al. *Chloroidium*, a common terrestrial coccoid green alga previously assigned to *Chlorella* (Trebouxiophyceae, Chlorophyta). *Eur J Phycol*. 2010; 45(1):79–95.
23. Bock C, Pröschold T, Krienitz L. Two new *Dictyosphaerium*-morphotype lineages of the *Chlorellaceae* (Trebouxiophyceae): *Heynigia* gen. nov. and *Hindakia* gen. nov. *Eur J Phycol*. 2010; 45(3), 267–277.
24. Bock C, Krienitz L, Pröschold T. Taxonomic reassessment of the genus *Chlorella* (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea*. 2011; 11: 293–312.
25. Bock C, Pröschold T, Krienitz L. Updating the genus *Dictyosphaerium* and description of *Mucidosphaerium* gen. nov. (Trebouxiophyceae) based on morphological and molecular data 1. *J Phycol*. 2011; 47(3):638–652. doi: [10.1111/j.1529-8817.2011.00989.x](#) PMID: [27021993](#)
26. Neustupa J, Němcová Y, Veselá J, Steinová J, Škaloud P. *Parachloroidium* gen. nov. (Trebouxiophyceae, Chlorophyta), a novel genus of coccoid green algae from subaerial corticolous biofilms. *Phycologia*. 2013; 52: 411–421.
27. Neustupa J, Němcová Y, Veselá J, Steinová J, Škaloud P. *Leptochlorella corticola* gen. et sp. nov. and *Kalinella apyrenoidosa* sp. nov.: two new *Chlorella*-like green microalgae (Trebouxiophyceae, Chlorophyta) from subaerial habitats. *Int J Syst Evol Micr*. 2013; 63: 377–387.
28. Safi C, Zebib B, Merah O, Pontalier PY, Vaca-Garcia C. Morphology, composition, production, processing and applications of *Chlorella vulgaris*: A review. *Renew Sust Energ Rev*. 2014; 35:265–278.
29. Bashan Y, Lopez BR, Huss VAR. *Chlorella sorokiniana* (formerly *C. vulgaris*) UTEX 2714, a non-thermotolerant microalga useful for biotechnological applications and as a reference strain. *J Appl Phycol*. 2015;1–9.
30. Ma S, Han B, Huss VAR, Hu X, Sun X, Zhang J. *Chlorella thermophila* (Trebouxiophyceae, Chlorophyta), a novel thermo-tolerant *Chlorella* species isolated from an occupied rooftop incubator. *Hydrobiologia*. 2015; 1–9.
31. Luo W, Pröschold T, Bock C, Krienitz L. Generic concept in *Chlorella*-related coccoid green algae (Chlorophyta, Trebouxiophyceae). *Plant Biol*. 2010; 12(3):545–553. doi: [10.1111/j.1438-8677.2009.00221.x](#) PMID: [20522192](#)
32. Škaloud P, Němcová Y, Pytela J, Bogdanov NI, Bock C, Pickinpaugh SH. *Planktochlorella nurekis*, gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a novel coccoid green alga carrying significant biotechnological potential. *Fottea*. 2014; 14:53–62.
33. Krienitz L, Bock C, Kotut K, Pröschold T. Genotypic diversity of *Dictyosphaerium*-morphospecies (*Chlorellaceae*, Trebouxiophyceae) in African inland waters, including the description of four new genera. *Fottea*. 2012; 12:231–253.
34. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc R Soc Lond B*. 2003; 270:313–321.
35. Hebert PDN, Ratnasingham S, de Waard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *P Roy Soc Lond B Bio*. 2003; vol 270 (Suppl 1):S96–S99.

36. Waugh J. DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*. 2007; 29(2):188–197. PMID: [17226815](#)
37. Ratnasingham S, Hebert PDN. Bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes*. 2007; 7(3):355–364. PMID: [18784790](#)
38. Bertolazzi P, Felici G, Weitschek E. Learning to classify species with barcodes. *BMC Bioinformatics*. 2009; 10(Suppl 14):S7. doi: [10.1186/1471-2105-10-S14-S7](#) PMID: [19900303](#)
39. Reid BN, Le M, McCord WP, Iverson JB, Georges A, Bergmann T, et al. Comparing and combining distance-based and character-based approaches for barcoding turtles. *Mol Ecol Resour*. 2011; 11(6):956–967. doi: [10.1111/j.1755-0998.2011.03032.x](#) PMID: [21635698](#)
40. Zou S, Li Q, Kong L, Yu H, Zheng X. Comparing the Usefulness of Distance, Monophyly and Character-Based DNA Barcoding Methods in Species Identification: A Case Study of Neogastropoda. *PLoS ONE*. 2010; 6(10):e26619.
41. Zou S, Li Q, Kong L. Monophyly, Distance and Character-Based Multigene Barcoding Reveal Extraordinary Cryptic Diversity in Nassarius: A Complex and Dangerous Community. *PLoS ONE*. 2012; 7(10):e4727.
42. Dong W, Cheng T, Li C, Xu C, Long P, Chen C, et al. Discriminating plants using the DNA barcode rbcLb: an appraisal based on a large data set. *Mol Ecol Resour*. 2014; 14(2):336–343. doi: [10.1111/1755-0998.12185](#) PMID: [24119263](#)
43. Krawczyk K, Szczecińska M, Sawicki J. Evaluation of 11 single-locus and seven multilocus DNA barcodes in *Lamium* L. (Lamiaceae). *Mol Ecol Resour*. 2014; 14(2):272–285. doi: [10.1111/1755-0998.12175](#) PMID: [24118979](#)
44. Chakraborty C, Doss CGP, Patra BC, Bandyopadhyay S. DNA barcoding to map the microbial communities: current advances and future directions. *Appl Microbiol Biot*. 2014; 98(8): 3425–3436.
45. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. Identification of Birds through DNA Barcodes. *PLoS Biol*. 2004; 2(10):e312. PMID: [15455034](#)
46. Pettengill JB, Neel MC. An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae). *AJB*. 2010; 97(8):1391–1406.
47. Jiang Y, Ding CB, Zhang L, Yang R, Zhou Y, Tang L. Identification of the genus *Epimedium* with DNA barcodes. *J Med Plants Res*. 2011; 5, 6413–6417.
48. Yang JB, Wang YP, MÖLler M, Gao LM, Wu D. Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). *Mol Ecol Resour*. 2012; 12(2):267–275. doi: [10.1111/j.1755-0998.2011.03095.x](#) PMID: [22136257](#)
49. Ashfaq M, Asif M, Anjum ZI, Zafar Y. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). *Mol Ecol Resour*. 2013; 13(4):573–582. doi: [10.1111/1755-0998.12089](#) PMID: [23480447](#)
50. Jaén-Molina R, Marrero-Rodríguez Á, Reyes Betancort- JA, Santos-Guerra A, Naranjo-Suárez J, Caujapé- Castells J. Molecular taxonomic identification in the absence of a 'barcoding gap': a test with the endemic flora of the Canarian oceanic hotspot. *Mol Ecol Resour*. 2015; 15(1):42–56. doi: [10.1111/1755-0998.12292](#) PMID: [24919404](#)
51. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann Rev Ecol Syst*. 1987; 18:489–522.
52. Avise JC. A role for molecular genetic in the recognition and conservation of endangered species. *Trends Ecol Evol*. 1989; 4:279–281. doi: [10.1016/0169-5347\(89\)90203-6](#) PMID: [21227370](#)
53. Avise JC, Ball RM. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford surveys in evolutionary biology*. 1990; 7:45–67.
54. Hudson RR, Coyne JA. Mathematical consequences of the genealogical species concept. *Evolution*. 2002; 56(8):1557–1565. PMID: [12353748](#)
55. Knowles LL, Carstens BC. Delimiting Species without Monophyletic Gene Trees. *Systematic Biol*. 2007; 56(6):887–895.
56. Kizirian D, Donnelly MA. The criterion of reciprocal monophyly and classification of nested diversity at the species level. *Mol Phylogenet Evol*. 2004; 32(3):1072–1076. PMID: [15288076](#)
57. Baum DA. Phylogenetic species concepts. *Trends Ecol Evol*. 1992; 7:2.
58. Baum DA, Donoghue MJ. Choosing among alternative 'Phylogenetic' species concepts. *Syst Bot*. 1995. 20:560–573.
59. Baum DA, Shaw KL. Genealogical perspectives on the species problem. *Experimental and molecular approaches to plant biosystematics*, 1995; 53(289–303):123–124.

60. Goldstein PZ, DeSalle R. Phylogenetic species, nested hierarchies, and character fixation. *Cladistics*. 2000; 16:364–384.
61. Fontaneto D, Flot J F, Tang C Q. Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity*. 2014; 1–19.
62. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A., Duran DP, Hazell S, et al. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol*. 2006; 55:595–609. PMID: [16967577](#)
63. Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Syst Biol*. 2013; 62:707–724. doi: [10.1093/sysbio/syt033](#) PMID: [23681854](#)
64. Simpson GG. The species concept. *Evolution*. 1951; 5:285–298.
65. Masters BC, Fan V, Ross HA. Species delimitation—A Geneious plugin for the exploration of species boundaries. *Mol Ecol Resour*. 2011; 11, 154–157. doi: [10.1111/j.1755-0998.2010.02896.x](#) PMID: [21429114](#)
66. Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*. 2013; 29:2869–2876. doi: [10.1093/bioinformatics/btt499](#) PMID: [23990417](#)
67. Puillandre N, Lambert A, Brouillet S, Achaz G ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol Ecol*. 2012; 21, 1864–1877. doi: [10.1111/j.1365-294X.2011.05239.x](#) PMID: [21883587](#)
68. DeSalle R, Egan MG, Siddall M. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos Trans R Soc Lond B Biol Scivol*. 2005; 360 (1462):1905–1916.
69. Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H. Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proy Soc Lond B Bio*. 2008; vol 275.
70. Damm S, Schierwater B, Hadrys H. An integrative approach to species discovery in odonates: from character-based DNA barcoding to ecology. *Mol Ecol*. 2010; 19(18):3881–3893. doi: [10.1111/j.1365-294X.2010.04720.x](#) PMID: [20701681](#)
71. Yassin A, Markow TA, Narechania A, O’Grady PM, DeSalle R. The genus *Drosophila* as a model for testing tree- and character- based methods of species identification using DNA barcoding. *Mol Phylogenet Evol*. 2010; 57(2):509–517. doi: [10.1016/j.ympev.2010.08.020](#) PMID: [20800099](#)
72. Goldstein PZ, DeSalle R. Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. *BioEssays*. 2011; 33(2):135–147. doi: [10.1002/bies.201000036](#) PMID: [21184470](#)
73. Sarkar IN, Planet PJ, Desalle ROB. Caos software for use in character-based DNA barcoding. *Mol Ecol Resour*. 2008; 8(6):1256–1259 doi: [10.1111/j.1755-0998.2008.02235.x](#) PMID: [21586014](#)
74. FitzJohn RG. Quantitative traits and diversification. *Syst Biol*. 2010; 59:619–633. doi: [10.1093/sysbio/syq053](#) PMID: [20884813](#)
75. Joly S, Davies TJ, Archambault A, Bruneau A, Derry A, Kembel SW, et al. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Mol Ecol Resour*. 2014; 14(2):221–232. doi: [10.1111/1755-0998.12173](#) PMID: [24118947](#)
76. Saunders GW, Kucera H. An evaluation of rbcL, tufA, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. *Cryptogamie Algologie*. 2010; 31(4):487–528.
77. Hall JD, Fucikova K, Lo C, Lewis LA, Karol KG. An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogamie Algol*. 2010; 31(4):529–555.
78. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *P Natl Acad Sci USA*. 2005; 102(23):8369–8374.
79. Li M, Cao H, But PPH, Shaw PC. Identification of herbal medicinal materials using DNA barcodes. *JSE*. 2011; 49(3):271–283.
80. Saunders GW. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos Trans R Soc Lond B Biol Sci*. 2005; 360:1879–1888. PMID: [16214745](#)
81. Evans KM, Wortley AH, Mann DG. An assessment of potentially diatom “barcode genes” (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). *Protist*. 2007; 158 (3):349–364. PMID: [17581782](#)
82. Ferri E, Barbuto M, Bain O, Galimberti A, Uni S, Guerrero R, et al. Integrated taxonomy: traditional approach and DNA Barcoding for the identification of filarioid worms and related parasites (Nematoda). *Front Zool*. 2009; 6:1. doi: [10.1186/1742-9994-6-1](#) PMID: [19128479](#)
83. Mcdevit DC, Saunders GW. A DNA barcode examination of the Laminariaceae (Phaeophyceae) in Canada reveals novel biogeographical and evolutionary insights. *Phycologia*. 2010; 49(3):235–248.

84. Le Gall L, Saunders GW. A nuclear phylogeny of the Florideophyceae (Rhodophyta) inferred from combined EF2, small subunit and large subunit ribosomal DNA: establishing the new red algal subclass Corallinophycidae. *Mol Phylogenet Evol.* 2007; 43:1118–1130. PMID: [17197199](#)
85. Watanabe S, Floyd GL. Consideration on the systematics of coccoid green algae and related organisms based on the ultrastructure of swimmers. In: Chaudhary B. R. & Agrawal S. B. (eds), *Cytology, genetics and molecular biology of algae*. Amsterdam, Netherlands, SPB Academic Publishers.1996; pp. 1–19.
86. Turmel M, Otis C, Lemieux C. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage the led to land plants. *P Natl Acad Sci USA.* 2002; 99(17):11275–11280.
87. Pombert JF, Beauchamp P, Otis C, Lemieux C, Turmel M. The complete mitochondrial DNA sequence of the green alga *Oltmannsiellopsis viridis*: evolutionary trends of the mitochondrial genome in the Ulvophyceae. *Curr Genet.* 2006; 50:137–147. PMID: [16721603](#)
88. Pombert JF, Otis C, Lemieux C, Turmel M. The complete mitochondrial DNA sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. *Mol Biol Evol.* 2004; 21:922–935. PMID: [15014170](#)
89. CBOL Plant Working Group. A DNA barcode for land plants. *PNAS, USA.* 2009; 106:12794–12797.
90. Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *PNAS.* 2011; 108(49):19641–19646. doi: [10.1073/pnas.1104551108](#) PMID: [22100737](#)
91. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev.* 2015; 90(1):157–166. doi: [10.1111/brv.12104](#) PMID: [24666563](#)
92. Fučíková K, Flechtner VR, Lewis LA. Revision of the genus *Bracteacoccus* Tereg (Chlorophyceae, Chlorophyta) based on a phylogenetic approach. *Nova Hedwigia.* 2012; 96:15–59.
93. Andersen RA. *Algal Culturing Techniques*. Elsevier Academic Press, Amsterdam, The Netherlands. 2005.
94. Sun X, Wu X, Li X, Pei L. Molecular identification of *Chlorella* strains based on sequence analysis of nuclear rDNA ITS and chloroplast *rbcL* gene. *Journal of fisheries of China.* 2009; 33(4):565–571.
95. Burja AM, Tamagnini P, Bustard MT, Wright PC. Identification of the green alga, *Chlorella vulgaris* (SDC1) using cyanobacteria derived 16S rDNA primers: targeting the chloroplast. *FEMS Microbiol Lett* vol. 2011; 202 (2):195–203.
96. Famà P, Wysor B, Kooistra WHCF, Zuccarello GC. Molecular phylogeny of the genus *Caulerpa* (Caulerpales, Chlorophyta) inferred from chloroplast *tufA* gene. *J Phycol.* 2002; 38(5):1040–1050.
97. Kato K, Asimenos G, Toh H. Multiple Alignment of DNA Sequences with MAFFT. In: Posada D (ed) *Bioinformatics for DNA Sequence Analysis*. Methods Mol Biol.2009; vol 537. Humana Press, pp 39–64.
98. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28:2731–2739. doi: [10.1093/molbev/msr121](#) PMID: [21546353](#)
99. Posada D. JModelTest: phylogenetic model averaging. *Mol Biol Evol.* 2008; 25:1253–1256. doi: [10.1093/molbev/msn083](#) PMID: [18397919](#)
100. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003; 19:1572–1574. PMID: [12912839](#)
101. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59:307–321. doi: [10.1093/sysbio/syq010](#) PMID: [20525638](#)
102. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007; 7:214. PMID: [17996036](#)
103. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 2006; 4(5):e88. PMID: [16683862](#)
104. Gernhard T. The conditioned reconstructed process. *J Theoret Biol.* 2008; 253:769–778.
105. Kumar S, Skjaeveland A, Orr RJS, Enger P, Ruden T, Mevik B. AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics.* 2009; 10:357. doi: [10.1186/1471-2105-10-357](#) PMID: [19863793](#)
106. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.2012.

107. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004; 20:289–290. PMID: [14734327](#)
108. Ezard T, Fujisawa T, Barraclough TG. splits: SPecies' Limits by Threshold Statistics. R package version.2009; 1.0-14/r31. <http://R-Forge.R-project.org/projects/splits/>.
109. Hamilton C A, Hendrixson B E, Brewer M S, Bond J E. An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: a case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). *Mol Phylogenet Evol*. 2014; 71:79–93. doi: [10.1016/j.ympev.2013.11.007](#) PMID: [24280211](#)
110. Bergmann T, Hadrys H, Breves G, Schierwater B. Character-based DNA barcoding: a superior tool for species classification. *Charakter-basierte DNS Kodierung: ein überlegenes Werkzeug für die Klassifizierung von Arten*. *Berl Münch Tierärztl Wschr*. 2009; 5(11–12):446–450.
111. Swofford DL. PAUP/. *Phylogenetic Analyses Using Parsimony (and Other Methods)*. Version 4. Sinauer Associates, Sunderland, Massachusetts. 2003.
112. Maddison WP, Maddison DR. *MACCLADE: Analysis of Phylogeny and Character Evolution*. Version 3.0. Sinauer Associates, Sunderland, Massachusetts. 2005.
113. Chase MW, Fay MF. Barcoding of plants and fungi. *Science*. 2009; 325(5941):682–683. doi: [10.1126/science.1176906](#) PMID: [19644072](#)
114. Hebert PDN, Gregory TR. The Promise of DNA Barcoding for Taxonomy. *Systematic Biol*. 2005; 54(5):852–859.
115. Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet*. 2007; 23(4):167–172. PMID: [17316886](#)
116. Kvist S. Barcoding in the dark?: A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Mol Phylogenet Evol*. 2003; 69(1):39–45.
117. Meier R, Shiyang K, Vaidya G, Nq PK. DNA barcoding and taxonomy in Diptera: a tale of high intra-specific variability and low identification success. *Syst Biol*. 2006; 55:715–728. PMID: [17060194](#)
118. Little DP, Stevenson DW. A comparison of algorithms for the identification of species using DNA barcodes: examples for gymnosperms. *Cladistics*. 2007; 23:1–21.
119. Meier R. DNA sequences in taxonomy: opportunities and challenges. In: *The New Taxonomy* (ed. Wheeler QD), Chapter.2008; 7, pp. 95–127. CRC Press, New York.
120. Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SC, et al. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol Ecol Resour* 9. 2009; (Suppl. 1.):130–139.
121. Lou M, Golding BG. Assigning sequences to species in the absence of large interspecific differences. *Mol Phylogenet Evol*. 2010; 56:187–194. doi: [10.1016/j.ympev.2010.01.002](#) PMID: [20067837](#)
122. Kekkonen M, Hebert PDN. DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. *Mol Ecol Resour*. 2014; 14:706–715. doi: [10.1111/1755-0998.12233](#) PMID: [24479435](#)
123. Weiss M, Macher JN, Seefeldt MA, Leese F. Molecular evidence for further overlooked species within the *Gammarus fossarum* complex (Crustacea: Amphipoda). *Hydrobiologia*. 2014; 721:165–184.
124. Yu Z, Li Q, Kong L, Yu H. Utility of DNA Barcoding for Tellinoidea: A Comparison of Distance, Coalescent and Character-based Methods on Multiple Genes. *Mar Biotechnol (NY)*. 2015; 17:55–65.
125. Esselstyn JA, Evans BJ, Sedlock JL, Khan FAA, Heaney LR. Single-locus species delimitation: a test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society of London B: Biological Sciences*. 2012; rspb20120705.
126. Paz A, Crawford AJ. Molecular-based rapid inventories of sympatric diversity: a comparison of DNA barcode clustering methods applied to geography-based vs clade-based sampling of amphibians. *J Biosci*. 2012; 37:887–896. PMID: [23107924](#)
127. Sauer J, Hausdorf BA. Comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy. *Cladistics*. 2012; 28(3):300–316.
128. Miralles A, Vences M. New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. *PLoS One*. 2013; 8(7):e68242. doi: [10.1371/journal.pone.0068242](#) PMID: [23874561](#)
129. Talavera G, Dincă V, Vila R. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods Ecol Evol*. 2013; 4(12):1101–1110.