A case study of species delimitation with molecular methods: the algal genus *Microthamnion* (Microthamniales, Trebouxiophyceae)

Inaugural – Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Tanja Reder

aus Bergisch Gladbach

Köln, 2019

A case study of species delimitation with molecular methods: the algal genus *Microthamnion* (Microthamniales, Trebouxiophyceae)

I n a u g u r a l – D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Tanja Reder

aus Bergisch Gladbach

Köln, 2019

Berichterstatter (Gutachter):	Prof. Dr. Michael Melkonian
	Prof. Dr. Burkhard Becker
Tag der mündlichen Prüfung:	30.10.2019

Abstract

The green algal genus *Microthamnion* (Microthamniales, Trebouxiophyceae) has been extensively studied, but the question of species delimitation remained a matter of controversy. The morphological traits used to discriminate species in *Microthamnion* have shown to be quite polymorphic, and the restriction enzyme analysis performed in a first molecular attempt does no longer meet modern standards.

The present study used a detailed molecular approach and combined several methods to clarify the matter of species delimitation on a molecular level. A multi-gene alignment comprising the nuclear-encoded 18S, and 28S rRNA genes, and the ITS2 molecule, as well as the plastid-encoded *rbc*L gene was assembled and used for concatenated phylogenetic analyses. The 74 *Microthamnion* strains investigated in this study fell into four monophyletic clades (one with a distinct subdivision) and nine long-branched lineages, which are assumed to correspond to species level. An apomorphy analysis was performed in order to find non-homoplasious synapomorphies (NHSs), and thus unique molecular signatures, for the clades and lineages inferred from the phylogeny. In a novel approach that interpreted molecular data in an alternative way, these NHSs and other 'phenotypic molecular characters' were compiled in a data matrix and used for a parsimony tree reconstruction. The clade boundaries and lineages inferred from the *Microthamnion* phylogeny were confirmed by both, unique molecular signatures and the tree based on all phenotypic molecular characters, resulting in 14 putative species delineated by molecular methods.

Zusammenfassung

Die Grünalgengattung *Microthamnion* (Microthamniales, Trebouxiophyceae) war Gegenstand zahlreicher Untersuchungen, jedoch blieb die Frage der Artabgrenzung umstritten. Die morphologischen Merkmale, die zur Artunterscheidung herangezogen wurden, erwiesen sich im Nachhinein als polymorph und eine Restriktionsanalyse, die im Mittelpunkt eines ersten molekularen Ansatzes stand, wird heutigen Standards nicht mehr gerecht.

Die vorliegende Studie verfolgte nun einen detaillierten molekularen Ansatz, der mehrere Methoden kombinierte, um die Problematik der Artabgrenzung auf molekularer Ebene zu klären. Ein Multigen-Alignment, welches die kernkodierten 18S- und 28S-rRNA-Gene und das ITS2-Molekül sowie das plastidär kodierte *rbc*L-Gen umfasste, wurde aufgebaut und für kombinierte phylogenetische Analysen genutzt. Die in dieser Arbeit untersuchten 74 *Microthamnion*-Stämme gliederten sich in vier monophyletische Kladen (eine mit zwei klar abgegrenzten Untergruppen) und neun langästige Linien, von denen angenommen wird, dass sie dem Artlevel entsprechen. Eine Apomorphie-Analyse diente der Identifizierung nicht-homoplasischer Synapomorphien (NHSs), und damit einzigartiger molekularer Signaturen, für die aus der Phylogenie abgeleiteten Kladen und Linien. In einem neuartigen Ansatz, welcher molekulare Daten auf alternative Weise interpretierte, sind diese NHSs und weitere „phänotypische molekulare Merkmale“ in einer Datenmatrix zusammengefasst und für die Rekonstruktion eines Parsimonie-basierten Stammbaumes verwendet worden. Die Abgrenzungen der Kladen und Linien, die sich aus der *Microthamnion*-Phylogenie ergaben, wurden sowohl durch eindeutige molekulare Signaturen, als auch durch den auf allen phänotypischen molekularen Merkmalen basierenden Baum bestätigt. Insgesamt ließen sich so 14 Arten auf molekularer Ebene unterscheiden.

Table of Contents

Abstract ___ *III*

Zusammenfassung ___ *IV*

Table of Contents ___ *V*

List of Tables ___ *VII*

List of Figures ___ *VIII*

List of Abbreviations ___ *IX*

1. Introduction ___ *1*

1.1. Molecular Approaches for Species Delimitation ___ **2**

1.1.1. Phylogenetic Tree Reconstructions ___ 2

1.1.2. Barcoding System ___ 2

1.1.3. CBC Species Concept and Unique Molecular Signatures ___ 3

1.1.4. Recommendation of Holistic Approaches ___ 4

1.2. History of the Genus *Microthamnion* ___ **5**

1.3. Current Situation and Aims of this Study ___ **8**

2. Material and Methods ___ *13*

2.1. Cultures and Culturing Conditions ___ **13**

2.2. DNA Extraction/Preparations for Colony PCR ___ **16**

2.3. Polymerase Chain Reaction ___ **16**

2.4. Agarose Gel Electrophoresis ___ **18**

2.5. Sequencing Reaction ___ **19**

2.6. Sequence Assembly and Correction ___ **19**

2.7. Taxon Sampling and Alignments ___ **19**

2.7.1. *Microthamnion* ___ 19

2.7.2. Trebouxiophyceae ___ 20

2.8. Phylogenetic Tree Reconstructions ___ **21**

2.9. Identification of Phenetic Molecular Characters for Alternate Phylogenetic Tree Reconstructions ___ **21**

2.9.1. *In silico* Restriction Fragment Length Polymorphism Analysis in 18S rDNA Sequences ___ 22

2.9.2. Apomorphy Analysis and Non-Homoplasious Synapomorphies ___ 22

2.9.3. Tree Reconstruction Based on Phenetic Molecular Characters ___ 23

3. Results ___ *25*

3.1. Phylogenetic Analyses ___ **25**

3.1.1. Datasets ___ 25

3.1.2. Single-Gene Analyses ___ 25

3.1.3. Congruent Phylogenies ___ 26

3.1.4. *Microthamnion* Phylogeny 33
3.1.5. Trebouxiophycean Phylogeny 36

3.2. Phenetic Character Identification and Parsimony Tree Reconstruction 38
3.2.1. *In Silico* Analysis of Restriction Enzyme Sites in 18S rDNA Sequences 38
3.2.2. ITS2 Secondary Structure Models 42
3.2.3. Unique Molecular Signatures for *Microthamnion* and the Microthamniales 45
3.2.4. Compensatory Base Changes in the ITS2 Molecule 55
3.2.5. Data Matrix for Alternate Tree Reconstructions 56
3.2.6. Phylogeny Based on Phenetic Molecular Characters 57

4. Discussion 59

4.1. Phylogenetic Analyses 59
4.1.1. ITS2 Secondary Structure 59
4.1.2. Identification of a Promising Barcode Candidate in *Microthamnion* 60
4.1.3. Selection of the Best Marker Combination for a Resolved *Microthamnion* Phylogeny 61
4.1.4. Comparison of the Congruent with the Gapped *Microthamnion* Phylogeny 62
4.1.5. Trebouxiophycean Phylogeny 63
4.1.6. Putative Species Inferred from the *Microthamnion* Phylogeny 65

4.2. Phenotypic Molecular Data 67
4.2.1. NHS and CBC Support for Species and Genus Boundaries 67
4.2.2. Tree Reconstructions Inferred from Phenotypic Molecular Characters 70
4.2.3. *In Silico* RFLP Analysis on 18S rDNA Data 71

4.3. *Excursus* to Habitats and Sampling Strategies 72
4.3.1. Sampling Sites and Habitats 72
4.3.2. Sampling Strategies and Environmental Sequences 74

5. Conclusions and Outlook 77

6. References 79

7. Supplementary Material 91

Acknowledgements 108

Erklärung 109

List of Tables

Table 1. List of strains 13
Table 2. TE Buffer 16
Table 3. PCR and sequencing primers 17
Table 4. Amount of chemicals per PCR reaction using Taq Polymerase 17
Table 5. Amount of chemicals per PCR reaction using Phusion® High-Fidelity DNA Polymerase 17
Table 6. Thermo-cycling conditions PCR (Taq Polymerase) 18
Table 7. Thermo-cycling conditions PCR with Phusion® High-Fidelity DNA Polymerase 18
Table 8. 50 x TAE-Buffer 18
Table 9. Restriction enzymes used for in silico RFLP analysis on 18S rDNA data. 22
Table 10. Total number of restriction enzyme cuts in 18S for all Microthamnion strains 38
Table 11. Unique molecular signatures within Microthamnion and the Microthamniales 48
Table 12. Compensatory Base Changes (CBCs) in the ITS2 molecule 56

Supplementary Table 1. Bold's Basal Medium (BBM, modified) 91
Supplementary Table 2. Isopropanol precipitation protocol 91
Supplementary Table 3. Sequence data (Overview) 92
Supplementary Table 4. Data matrix of phenotypic molecular characters used for phylogenetic tree reconstructions. 93
Supplementary Table 5. List of strains ordered according to the Microthamnion phylogeny 106

List of Figures

Figure 1. Comparison of congruent single-gene phylogenies (ITS2 and 28S) _____ *28*
Figure 2. Comparison of congruent single-gene phylogenies (rbcL and 18S) _____ *29*
Figure 3. Comparison of congruent multi-gene phylogenies 1 _____ *30*
Figure 4. Comparison of congruent multi-gene phylogenies 2 _____ *31*
Figure 5. Molecular phylogeny of all markers based on a congruent dataset _____ *32*
*Figure 6. Molecular phylogeny of Microthamnion*_____ *34*
Figure 7. Molecular phylogeny of the Trebouxiophyceae _____ *36*
*Figure 8. Position of restriction enzyme cutting sites mapped on a schematic of the 18S rRNA molecule*___ *40*
Figure 9. In silico gel electrophoresis analysis of digested Microthamnion 18S data _____ *41*
*Figure 10. Consensus secondary structure model of 37 unique ITS2 Microthamnion sequences*_____ *42*
*Figure 11. Maximum parsimony phylogeny based on phenetic molecular characters*_____ *57*

Supplementary Figure 1. 18S single-gene phylogeny _____ *97*
Supplementary Figure 2. ITS2 'short' single-gene phylogeny _____ *98*
Supplementary Figure 3. ITS2 'long' single-gene phylogeny _____ *99*
*Supplementary Figure 4. 28S 'short' single-gene phylogeny*_____ *100*
Supplementary Figure 5. 28S 'long' single-gene phylogeny _____ *101*
Supplementary Figure 6. RbcL protein single-gene phylogeny _____ *102*
Supplementary Figure 7. RbcL DNA single-gene phylogeny _____ *103*
*Supplementary Figure 8. Consensus secondary structure model of all 56 ITS2 Microthamnion sequences*_ *104*
Supplementary Figure 9. Consensus secondary structure model of 45 ITS2 Microthamnion sequences ___ *105*

List of Abbreviations

°C	degree Celsius
µl	microliter
µm	micrometer
µM	micromolar
µmol	micromole
µS	microsiemens
AA	Amino acid
ACOI	Coimbra Collection of Algae
AG	Arbeitsgruppe
AUT	autapomorphie
BBM	Bold's Basal Medium
Bio-P	biologische Phosphorelimination (i.e. biological phosphorus removal)
BLAST	Basic Local Alignment Search Tool
bp	base pairs
b.t.l.	before terminal loop
CAUP	Culture Collection of Algae of Charles University in Prague
CBC(s)	Compensatory Base Change(s)
CCAC	Culture Collection of Algae at the University of Cologne
CCALA	Culture Collection of Autotrophic Organisms
CCAP	Culturecollection of Algae and Protozoa
cf.	confer/conferatur
Ch. perforatum	*Characium perforatum*
*cox*I	cytochrome c oxidase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
E	East
EDTA	Ethylenediaminetetraacetic acid
e.g.	exempli gratia = for example
ENV	environmental
et al.	et alii (Maskulinum), et aliae (Femininum)
EtBr	ethidium bromide
form.	formerly
g	gram
GmbH	Gesellschaft mit beschränkter Haftung
GPS	Global Positioning System
GTR	General Time Reversible Model
H1, H2, H3, H4	Helix one, Helix two, Helix three, Helix four
Hz	Hertz
I	Invariable sites
i.e.	id est = that is
INA	Index Nominum Algarum

incl.	inclusive
ITS2	Internal Transcribed Spacer 2
IUPAC	International Union of Pure and Applied Chemistry
kb	kilobase
Lat.	Latitude
Long.	Longitude
LSU	large subunit
m	meter
M.	*Microthamnion*
M. sp.	*Microthamnion* species
max.	maximum
MCMC	parallel Markov Chain Monte Carlo
mg	milligram
min.	minutes
ML	Maximum Likelihood
ml	milliliter
mm	millimeter
mM	millimolar
MP	Maximum Parsimony
N	North
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
ng	nanogram
NHS(s)	Non-homoplasious synapomorphy(ies)
NIES	Microbial Culture Collection at the National Institute for Environmental Studies
NJ	Neighbour Joining
nM	nanomole
no.	number
nt(s)	nucleotide(s)
p., pp.	page, pages
PCR	Polymerase Chain Reaction
pos.	position
RaxML	Randomized Accelerated Maximum Likelihood
*rbc*L	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit gene
RuBisCO	ribulose-1,5-bisphosphate carboxylase/oxygenase
rDNA	ribosomal DNA
RFLP	Restriction Fragment Length Polymorphism
RNA	ribonucleic acid
rRNA	ribosomal RNA
RRZK	Regionales Rechenzentrum der Universität zu Köln
SAG	Sammlung von Algenkulturen der Universität Göttingen
scf	shell command file
SFM	Synthetic Freshwater Medium
sp.	species

SSU	small subunit
Taq	*Thermus aquaticus*
term.	terminal
USA	United States of America
UTEX	UTEX The Culture Collection of Algae at The University of Texas at Austin
V	Volt
w/o	without
Γ	Gamma distribution

1. Introduction

A major difficulty regarding species delimitation in (micro) algae is that there is a great abundance of organisms where sexual reproduction is either absent or unknown, and thus the biological species concept (Dobzhansky, 1937; Mayr, 1942) cannot be applied. Instead, classifications were traditionally erected based on morphological characters, which could easily be identified via light microscopy. This approach was later refined as also vegetative stages of complete life cycles were observed. With the development of scanning and transmission electron microscopy, a broad spectrum of ultrastructural features became available which lead to major rearrangements, especially at higher levels of the taxonomic system. Here, the organization of the basal bodies in the flagellar apparatus was of particular interest, but also modes of cell division were brought into focus (Mattox & Stewart, 1984; Melkonian, 1984). Besides, detailed investigations of cell wall structures or biochemical characters, such as pigment compositions, were used for taxonomic purposes.

The introduction of phylogenetic systematics by Willi Hennig (1950) revolutionized the classification system, as it proposed a hierarchic descent approach based on the evolution of heritable traits which are shared by a derived group of organisms. A prerequisite for this phylogeny-based system was the monophyly of taxonomic units consisting of a common ancestor and all of its descendants. With increased feasibility of DNA sequencing, starting in the 1990ies, molecular phylogeny became the state of the art tool for resolving relationships among organisms. Especially the impact on species level was severe, and the need for revision of the previous approaches became evident. Although morphology-based concepts were always in a flux and individual species boundaries a matter of controversial discussions, the full extent of misidentifications and erroneous numbers of species was now revealed. Not only were characters previously used for species delimitation shown to be para- or polyphyletic, but also other peculiarities were discovered. Phenotypic plasticity as a response to different environmental conditions had, of course, been discussed before, but now there was proof that organisms with at times a great range of morphological variety were genetically identical (e.g. Otsuka et al., 1999; Logares et al., 2007; Belton et al., 2014). On the other hand, the existence of cryptic species complexes was detected where organisms formed individual lineages in the phylogenetic tree but were either morphologically indistinguishable or very similar with discriminative features only subsequently recognized (e.g. van der Strate et al., 2002; Lewis & Flechtner, 2004; Šlapeta et al., 2006; Fučíková et al., 2011).

Next to molecular phylogeny, increased availability of molecular data also led to the development of new concepts to address the matter of species delineation, some of which will be described in more detail in the following chapters.

1.1. Molecular Approaches for Species Delimitation

1.1.1. Phylogenetic Tree Reconstructions

From the 1990ies on, molecular phylogenetics skyrocketed due to the introduction of a thermostable DNA polymerase by Saiki and coworkers (Saiki et al., 1988) in the late 1980ies, which facilitated DNA amplification tremendously.

One of the first markers used for molecular phylogenetic tree reconstructions was the 18S rRNA gene, which codes for the small subunit (SSU) of the eukaryotic ribosome and is thus absolutely essential. It is still widely used and a valid marker for higher level phylogenies (e.g. Marin et al., 2003). On species level though, it does often not provide enough phylogenetic resolution (e.g. Bass et al., 2007; Rindi et al., 2007). Instead, the internal transcribed spacer 2 (ITS2) of the nuclear-encoded rRNA operon and the plastid *rbc*L gene, which codes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), have been proposed as suitable candidates for species delimitation in many algal groups (Coleman, 2001; Lewis & Flechtner, 2004; Heesch et al., 2009; Škaloud & Rindi, 2013). However, next to differing resolution capacities, the use of only a single marker in phylogenetic studies may be problematic, since genes from different loci can have conflicting signals (Soltis & Kuzoff, 1995; Huelsenbeck & Bull, 1996; Sang et al., 1997; Okuyama et al., 2005; Gonçalves et al., 2007). Therefore, nowadays usually two or more genes are combined in concatenated analyses both, on higher taxonomic (Marin & Melkonian, 2010; Marin, 2012) as well as species level (Šlapeta et al., 2006; Dupuis et al., 2012), which in turn also has a positive effect on the resolution.

In molecular phylogeny-based species delimitations, the species boundaries are usually determined in regards to the topology. A species is then defined as a monophyletic group of several individuals that is distinguished from its sister species by branch lengths and support values (Leliaert et al., 2014). Ideally, a species can be distinguished by short intraspecific branches and a longer, well supported preceding branch. These results are reckoned to be even more dependable when several markers come to the same conclusion (Dettman et al., 2003).

1.1.2. Barcoding System

Since the assembly of an alignment and the aligning process according to secondary structure information (in rRNA genes) as well as the computation of phylogenetic trees is quite laborious, it was also sought for methods that allowed for a faster and more practical way to distinguish species on a molecular level. The idea of establishing a barcode system where a single, ideally short and variable marker could be used for species identification quickly gained popularity.

The initial barcode proposed for animals (Hebert et al., 2003) was subunit I of the mitochondrial gene cytochrome c oxidase (*cox*I or COI) of the respiratory chain, which had, for example, been used successfully to discriminate between species in butterflies

(Brown et al., 1999), copepods (Bucklin et al., 1999) and velvet worms (Trewick, 2000). The idea was to compile a COI database that linked DNA sequences to known species which could then be used to identify unknown specimens relative to these described taxa – an approach that could ultimately be used for a global bioidentification system in animals (Hebert et al., 2003). Another proposed application was the discovery of new species. A prerequisite for both implementations is a low intra- and high interspecific variation that allows for the definition of a clear threshold. This 'barcode gap' is however not reliably predictable for every group of organisms (e.g. Meyer & Paulay, 2005; Wiemers & Fiedler, 2007; Hendrich et al., 2010).

In plants and green algae, *cox*I was less favorable as a barcode marker due to the presence of introns in several taxa (e.g. Cho et al., 1998; Palmer et al., 2000; Turmel et al., 2002a,b, 2003). Instead, the nuclear-encoded 18S rRNA gene as well as both ITS regions or the plastid-encoded *rbc*L and *tuf*A genes were proposed as suitable candidates in green algae (Hall et al., 2010; Hadi et al., 2016). However, none of the suggested markers was universally applicable since their ability to distinguish species differed between lineages.

Although DNA barcoding can be reliably used for species delimitation in some groups (e.g. (Hebert et al., 2003, 2004) it has, next to the above mentioned, also shown limitations, for example, in the inability to detect recent speciation events, introgressive hybridization, polyploidization or different mutation rates. It was therefore proposed by several authors that a barcode be used rather as a guideline for assessing variability than an actual link to species level and was recommended to be complemented with other methods (Wiemers & Fiedler, 2007; Hoef-Emden, 2012; Zou et al., 2016).

1.1.3. CBC Species Concept and Unique Molecular Signatures

Apart from its use as a marker for molecular phylogenetic tree reconstructions at genus and species levels, and being a putative barcode candidate, the ITS2 molecule was introduced as a tool for species delimitation based on certain features of its secondary structure.

The ITS2 is part of the nuclear-encoded rRNA operon, positioned between the 5.8S and 28S rRNA genes. The operon is transcribed as a whole, and the ITS2 excised from this precursor RNA during ribosome maturation. For this excision process the ITS2 folds into a clover-leave-like structure with typically four paired helices, which are connected by single-stranded spacers – a motif which has been proven to be quite conserved among eukaryotes (Mai & Coleman, 1997; Joseph et al., 1999; Coleman, 2007). The formation of the correct secondary structure is crucial for the excision process and the maintenance of base-pairings in the helices indispensable.

There are two scenarios in which nucleotide changes during evolutionary processes do not disrupt a base pair. A hemi compensatory base change (hemi CBC) refers to a single-sided substitution in a G-C pair to the thermodynamically stable wobble base

pair G-U or vice versa. The other is a compensatory base change (CBC), which refers to a mutation on both sides of the helix which preserves pairing at that position (e.g. C-G to A-U). Coleman (2000) discovered in the green algal order Volvocales a correlation between the presence of CBCs, especially in conserved regions of the helices two (H2) and three (H3) of the ITS2 molecule, and sexual barriers between organisms. A single CBC in such a conserved region was shown to go along with the inability to cross, and consequently two organisms which differed by a CBC in these regions belonged to different biological species. This CBC species concept was further refined as a large-scale analysis on plants and fungi by Müller et al. (2007) showed that any CBC in the paired helix regions of the ITS2 molecule in organisms belonging to the same genus corresponded to distinct species with a probability of 93.11%. A percentage that sufficed for them to conclude that the presence of a single CBC was a sufficient identifier for species delimitation. However, in both studies the reverse conclusion did not apply: the absence of a CBC was no proof for organisms to belong to only a single species (Coleman, 2000; Müller et al., 2007). Caisová and collegues (2011a), on the other hand, who mapped CBCs on a phylogenetic tree, showed for the Ulvales (Ulvophyceae, Chlorophyta) that CBCs did not necessarily correspond to species level but were rather found in deeper nodes corresponding to genus, family or even higher taxonomic levels. These findings were later also reported for the orders Chaetophorales, Chaetopeltidales, Oedogoniales and Sphaeropleales in the Chlorophyceae (Caisová et al., 2013).

Another method using secondary structure information for diagnostic purposes, is the identification of non-homoplasious synapomorphies (NHSs) *sensu* Marin et al. (2003, 2005; Marin & Melkonian, 2010). This *a posteriori* approach screens the secondary structure of rRNA genes (nuclear- or plastid-encoded rRNA operons and spacer regions), or the sequence of a protein coding gene at the amino acid level (e.g. the plastid-encoded *rbc*L gene), for unique molecular signatures for clades in a well-resolved phylogenetic tree with the prerequisite that all members of the investigated group share the derived character state and convergent evolution does not occur outside the investigated clade. These unique features include hemi CBCs, CBCs or disrupted base-pairings in otherwise paired helix regions, nucleotide changes in single-stranded spacers or deviating loop motifs in the rRNA genes as well as deviating residues in protein coding genes.

1.1.4. Recommendation of Holistic Approaches

There is however no universal method that is equally suitable to discriminate between species in all green algal groups. In fact, results can vary rather tremendously and it is not *a priori* predictable which approach will work best for the organisms of interest. Therefore, more and more authors advocate integrated approaches which combine several methods for a more robust prediction of species boundaries (Pröschold & Leliaert, 2007; Saunders, 2008; Marin & Melkonian, 2010; Hoef-Emden, 2012; Neustupa et al., 2013a; Darienko & Pröschold, 2015; Škaloud et al., 2015; Liu et al., 2017). This is explicitly not limited to DNA-based methods but includes, for example,

morphological and ultrastructural data, information on life cycles or ecology and biochemical characteristics. There is no panacea as to which combination works best; the tendency is however that the more methods are used, with ideally congruent outcomes, the more reliable is the allocation on species level.

1.2. History of the Genus *Microthamnion*

The green algal genus *Microthamnion* is a prime example for difficulties arising in morphology-based classifications as both, infrageneric divisions, and its position in a higher taxonomic context, were a matter of controversy since its first description.

Established in 1849 by Nägeli in Species algarum (Kützing, 1849), the genus was introduced with the type species *Microthamnion kuetzingianum* Nägeli, a minuscule green alga forming branched filaments which end in obtuse cells. Cell dimensions in *Microthamnion* are quite variable and range from 1.5 to 5 µm in diameter and can be 2-15 times longer than broad (John & Johnson, 1987). The thin-walled cells are uninucleate (e.g. Greger, 1915) and contain a single parietal chloroplast that contains small starch granules (Prauser, 1957; Watson & Arnott, 1973; Tupa, 1974; Bakker, 1995), but lacks a pyrenoid. Oil droplets have been found in the cytoplasm as an assimilation product (Prauser, 1957; Printz, 1964; Tupa, 1974). The 'plant' is usually attached to a substrate with a basal cell, but also occurs free floating in plankton. *Microthamnion* reproduces through bottle shaped, biflagellate zoospores which are developed in vegetative cells that have undergone differentiation into zoosporangia. Although this differentiation usually starts with the terminal cells, every cell except the basal attachment cell has been shown to be capable of zoospore formation (Greger, 1915; Tupa, 1974; John & Johnson, 1987). Zoospore formation and their ultrastructural features have been extensively studied (Greger, 1915; Watson & Arnott, 1973; Watson, 1975; John & Johnson, 1987), but sexual reproduction is unknown. *Microthamnion* occurs in a wide range of lentic and lotic freshwater habitats, such as rivers, streams, canals, lakes, ponds or peat bogs. But it has also been found in temporary habitats, such as water filled boot prints or smaller puddles, in soil when covered by a water film, in extreme water qualities as present in waste water facilities, or highly acidic mine drainage water (Hargreaves et al., 1975; Foster, 1982). It is globally distributed and does not seem to be restricted to certain degrees of latitude.

Although *Microthamnion* was recognized as a distinct genus right from the beginning, its place in the taxonomic system was frequently discussed and changed repeatedly over time. This was, for one, due to more detailed investigations in *Microthamnion* itself, but was also owed to general taxonomic rearrangements, renamings or the erection of new classes and orders in the Chlorophyta, which were a natural result from advances in laboratory techniques, such as electron microscopy and later, of course, molecular methods.

When first introduced, *Microthamnion* was listed as a genus in the family Ulotrichaceae (Kützing, 1849), an assessment that was only supported by a brief entry in another algal compendium (Reinsch, 1867). Also Hansgirg's (1886) suggestion that *Microthamnion* belonged to the Trentepohliaceae could not persist due to the lack of astaxanthin and zoosporangia not being restricted to terminal cells (Greger, 1915). Instead, an affiliation with the Chaetophoraceae (Chlorophyceae, Chaetophorales), which had been proposed earlier by Rabenhorst (1863), was acknowledged by various members of the research community (e.g. Cooke, 1884; Greger, 1915; Prauser, 1957). In West & West (1907), *Microthamnion* was listed in the Microthamniaceae, a family that was earlier introduced by Heering (1914) as a family in the Ulotrichales, but according to them, formed a sister family to the Chaetophoraceae in the Chaetophorales. Watson & Arnott (1973) also leaned towards the idea to assign *Microthamnion* to the Microthamniaceae, albeit without a distinct position in the systematic system, since *Microthamnion* zoospores possessed unique features among the Chlorophyceae. Another argument they formulated against its integration in the Chaetophoraceae, was the presence of siphonaxanthin (Weber & Czygan, 1972) which other chaetophoralean algae lacked. In 1984, Mattox & Stewart established the new family Pleurastraceae and placed *Microthamnion* in the order Pleurastrales (class Pleurastrophyceae). Melkonian (1990), on the other hand, proposed the erection of a new order, the Microthamniales, yet with uncertain affinities to accommodate chlorophycean algae which shared a distinctive ultrastructure in their zoospores as well as a common type of mitosis and cytokinesis (i. e. closed metacentric spindle/ centripetal furrow with phycoplast). Bakker (1995) saw *Microthamnion* as an intermediate between Pleurastrophyceae and Chlorophyceae due to cell division showing aspects of either groups. Recent molecular phylogenetic studies however clearly confirm the affiliation with the now called Trebouxiophyceae (Friedl, 1995), which cover the former Pleurastrophyceae to the exclusion of the Tetraselmidales, *Pleurastrum* and a few other taxa, and acknowledge the order Microthamniales, albeit now with less members than initially described (e.g. Neustupa et al., 2013a,b; Fučíková et al., 2014; Lemieux et al., 2014; Sanders et al., 2016).

Subgeneric divisions in *Microthamnion* were however even more vividly discussed than its position in the taxonomic system. A recent search of Index Nominum Algarum (INA; Silva, 2019) and AlgaeBase (Guiry & Guiry, 2019), conducted in June 2019, revealed 20 entries on *Microthamnion* species, subspecies and variations in INA, some of them overlapping. AlgaeBase however listed eight species in the genus of which only five had been flagged as accepted taxonomically: the type species *M. kuetzingianum* (Kützing, 1849), *M. strictissimum* (Rabenhorst, 1863), *M. curvatum* (West & West, 1907), *M. exiguum* (Reinsch, 1878) and *M. vexator* (Cooke, 1884). These results give a first insight in how divisive the debate on species delimitation in *Microthamnion* has been since its introduction in the mid-19th century.

Being ubiquitously available, the alga quickly gained popularity in the research community and was subject for multiple studies. The initial species descriptions, based on natural samples or enrichment cultures, were rather simplistic compared to modern standards and consisted mainly of a brief outline of the thallus, some measurements, such as cell diameter and length to breadth ratio of single cells, and modes of branching (Kützing, 1849; Rabenhorst, 1863; Reinsch, 1878; Cooke, 1884; West & West, 1907). However, more detailed studies followed, which went along with revisions and vivid discussions on species boundaries. Especially the two best known species, *M. kuetzingianum* and *M. strictissimum,* were a subject of controversy.

Schmidle (1899) still listed them as two separate species, since he found their habitus to be quite distinct and transient growth forms were unbeknown to him. Besides, he stated that he always found them separately, although it does not become entirely clear whether he referred to the formol material and exsiccated type material (*M. strictissimum*) he investigated in that study, or general observations made on natural samples. But when culture-based investigations, which allowed for more detailed studies and observations over a longer period of time, gained popularity also first doubts on these clear boundaries arose. Greger (1915), for example, who worked with unialgal cultures and gave a detailed description of development and zoospore formation in *Microthamnion*, advised caution regarding species delimitations. He was, for instance, sure that *M. vexator* was identical to *M. strictissimum* and generally suggested that environmental factors, such as nutrient concentrations, could have an effect on the morphology and thus on subgeneric distinctions. This notion was later strongly supported by Prauser (1957), who clearly stated that *M. kuetzingianum* and *M. strictissimum* were mere growth forms of the same species since in his own experiments, which were based on unialgal, axenic (through antibiotic treatment) cultures grown in different media, they were easily transformed into one another. Tupa (1974), who documented the progressive development of *Microthamnion* in axenic culture via photomicrography, agreed with Prauser's findings. She reported plants under favorable growth conditions to resemble the descriptions of *M. strictissimum* with a considerably higher length to breadth ratio, whereas plants in older cultures displayed shorter cells like in *M. kuetzingianum*. She also found curved terminal cells in some plants under less favorable conditions, which resembled those of *M. curvatum*, albeit with a broader cell diameter than described in West & West (1907). She concluded that *M. strictissimum* was a growth form of *M. kuetzingianum* and that *M. curvatum*/*M. exiguum* could be distinct species, but would have to be studied in more detail to confirm this assumption. The work of John & Johnson (1987) further substantiated these findings. Besides giving a detailed review on species erected in the genus so far, they also reported on their own experiments with several monoclonal, yet xenic *Microthamnion* cultures grown under different light and temperature settings. They confirmed the morphological plasticity of *M. kuetzingianum* and *M. strictissimum* in response to more or less favorable conditions. "Optimal" temperatures of 10 to 28°C and light intensities of 3-20 kilolux lead to rapid growth which resulted in longer cells

(i.e. *strictissimum* form), whereas lower temperatures and light intensities at the lower end of the spectrum went along with slower growth and smaller cells (= *kuetzingianum* form). From their observations the authors concluded that *Microthamnion* was either a single, very polymorphic species or consisted of two species complexes. One species complex summarized the type species *M. kuetzingianum* with *M. strictissimum*, *M. vexator* and all previously described variations of foresaid species, and the other consisted of *M. exiguum* and *M. curvatum,* whose sole distinctive feature was a cell diameter below 3 µm.

A first molecular approach by the same authors some years later, seemed to finally end discussions on subgeneric distinctions in *Microthamnion* (John et al., 1993). Amplified nuclear-encoded 18S rDNA of several *Microthamnion* isolates was digested with a set of restriction enzymes and the fragment patterns subsequently analyzed via gel electrophoresis – an approach which at that point had been successfully used to discriminate between closely related species in the red algal genus *Gymnogongrus* (Parsons et al., 1990; Maggs et al., 1992). Since no deviating pattern among the different strains could be detected, the authors assumed a close genetic relationship and saw their previous hypothesis of *Microthamnion* being monospecific confirmed.

1.3. Current Situation and Aims of this Study

Observations on axenic *Microthamnion* cultures in the workgroup Melkonian (Dorothee Langenbach and Veronica Zilz, AG Melkonian, University of Cologne, personal communication) did however question the monospecificy of *Microthamnion* again. Although all cultures were kept under the same laboratory conditions, i.e. culture medium, temperature, light intensities, culture vessels and inoculation date (= same age), the morphology among some of them differed quite remarkably. Initial ITS2 sequencing for three strains also revealed major differences on the molecular level. These findings inspired a Master thesis (Reder, 2015) which was conducted based on 63 *Microthamnion* strains in culture. In that work, a consensus secondary ITS2 structure for 44 *Microthamnion* strains was developed and phylogenetic analyses based on 18S rDNA and ITS2 (inclusive of the flanking regions of 5.8S rDNA and 28S rDNA which were covered by the ITS2 primers) sequences were conducted to address the question of intrageneric variability and putative species number. Also CBCs in the ITS2 were determined and restriction fragment length polymorphism (RFLP) analyses were performed *in silico* on all 26 SSU sequences obtained during that study. Although intrageneric variability could be depicted with combined analyses of those two markers, the phylogeny was not resolved and species numbers could only be estimated by branch lengths and groupings via CBC clades and identical restriction fragment patterns.

A challenge all attempts of species identification in asexual organisms have to face, besides providing criteria for grouping specimens together, is the definition of a clear boundary that discriminates characteristics of a species from inter- and intraspecific variability. Although the individual methods clearly specify their prerequisites, the exact threshold remains somewhat arbitrary as its 'correctness' cannot be proven by mating experiments.

This work used the genus *Microthamnion* for a case study on employing molecular methods for species delimitation. Several approaches were combined to come to a most substantiated conclusion on species boundaries. Other approaches, although undoubtedly important, like investigations on morphology or life cycles et cetera would have exceeded the frame of this study and were thus left for future works.

One aim of this study was the establishment of a well-resolved *Microthamnion* phylogeny, using several molecular markers. The latter were chosen to unite multiple properties. For one, a combination of conserved and variable markers was selected and second, both the nuclear and plastid genomes were covered. The nuclear-encoded rRNA operon offered several useful traits as it unites conserved genes and extremely variable spacer regions which can be used for different phylogenetic purposes. Where the most conserved 18S gene as well as the 3' end of the 28S gene, which codes for the large subunit (LSU) of the eukaryotic ribosome, increase resolution, the spacer regions can give enough variability to discriminate species. Therefore the ITS2 was chosen for its putative power to differentiate on species level and the complete SSU for more stability and increased resolution. The 5' end of the LSU (domains A to D) was included as it unites the variable C-domain with conserved parts and thus holds an intermediate position. The 5.8S gene was not targeted as a whole, but the 3' end was covered by the ITS2 primers and also used for phylogenetic purposes. As described before, the plastid-encoded *rbc*L has been deemed equally suitable as a marker on species level like the ITS2 and as a protein coding gene it offered the possibility to analyze the data on both, DNA and protein level. Single-gene as well as combined analyses were performed to (i) distinguish the marker combination best suitable for the identification of monophyletic, well supported clades and thus putative species and (ii) to find out which marker could best be used for a quick estimation of the intrageneric variability and, in the future, serve as a barcode to assign new *Microthamnion* isolates to existing clades or to reveal new ones.

One problem that presented itself in the beginning of this study was that, besides *Characium perforatum* (also referred to as *Fusochloris perforata*, strain SAG 28.85/ UTEX 2104), no closely related taxa to *Microthamnion* were known. Recent phylogenetic studies covering the Trebouxiophyceae only ever worked with those two taxa, and even the next more distantly related organisms were not unequivocally identified, since the relationship among trebouxiophycean lineages was not resolved (Pröschold et al., 2011; Fučíková et al., 2014; Darienko & Pröschold, 2015; Sanders et al., 2016) or in conflict with previously published phylogenies at positions crucial for the present study (Lemieux et al., 2014; Suzuki et al., 2018).

For sound phylogenetic analyses however, the availability of (preferably closely related) outgroup sequences is essential (Graham et al., 2002). Thus, another aim of this study was to identify additional outgroup candidates. This was achieved by compiling an extensive multi-gene alignment covering the known trebouxiophycean diversity, which was used for a phylogenetic tree reconstruction that allowed for the identification of the best outgroup sequences for phylogenetic analyses in *Microthamnion*. The assembly of the trebouxiophycean alignment included both, exhaustive BLAST (Basic Local Alignment Search Tool; Altschul et al., 1990) searches with *Microthamnion* and *Characium* query sequences, as well as a thorough literature search for trebouxiophycean taxa which could be incorporated in the alignment and also be starting points for continued BLAST searches. Apart from the identification of suitable outgroup candidates, a trebouxiophycean phylogeny with reference strains depicting the *Microthamnion* diversity was needed to confirm the monophyly of the genus.

Another challenge was the lack of secondary structure information for the ITS2 molecule applicable on all trebouxiophycean lineages. Since sound phylogenetic tree reconstructions depend on the comparison of homologous bases, their identification is absolutely essential and can only be achieved by aligning sequences according to secondary structure motifs. Thus, the objective was to establish such a universal secondary structure for the Trebouxiophyceae and descry those positions that could be homologized among all taxa and enable their use in phylogenetic analyses.

Apart from the phylogenetic tree reconstructions, also other methods were applied on the *Microthamnion* sequence data in order to confirm monophyletic clades from the phylogeny and thus reinforce support for species boundaries. In a first step, the ITS2 secondary structure model from Reder (2015) was augmented by sequences of new strains, revised where necessary and searched for putative CBC positions which could then be checked in the alignment file. Besides, the secondary structures of the amplified operon genes as well as the protein alignment of the *rbc*L sequences were screened for NHSs *sensu* Marin et al. (2003) that were in accordance with the monophyletic groups in the previously established phylogeny. Also the fingerprinting method as used in John et al. (1993) was applied *in silico* on the now much enlarged SSU dataset to retrace whether the authors would have come to the same conclusion had they had the same strains at their disposal.

The results from the secondary structure approaches, the restriction fragment length polymorphisms and other 'phenetic' molecular characters (e.g. sequence length of the individual ITS2 helices) were then compiled in a data matrix, similar to those used when calculating trees based on morphological characters, in order to compute a parsimony-based tree that was then compared with the phylogeny inferred from the regular likelihood analyses with an evolutionary model.

It should be noted that own results (mainly sequence data) and templates (e.g. alignment files and figures to be revised) from the above mentioned previous work (Reder, 2015) were incorporated in the present study and labelled accordingly.

2. Material and Methods

2.1. Cultures and Culturing Conditions

The present study was based on 74 *Microthamnion* strains in culture which were all previously identified as *Microthamnion* via light microscopy. Detailed information regarding origins, strain numbers, supposed species name and sample locality are given in Table 1. They were made available through the Culture Collection of Algae at the University of Cologne (CCAC) and were either own isolates (CCAC- and M-numbers) or obtained from the following algal culture collections: ACOI, CAUP, CCALA, CCAP, NIES, SAG and UTEX (for unabridged form consult abbreviations listed on page IX).

The cultures were grown in 100 ml Erlenmeyer flasks with modified Bold's Basal Medium (BBM) which contained the threefold amount of vitamins compared to regular BBM (as of the CCAC homepage; see Supplementary Table 1, p. 91, for recipe). For maintenance, the cultures were kept in a culture chamber by Johnson Controls at 16°C with a photon fluence rate of 10-45 µmol m^{-2} s^{-1} in a light/dark cycle of 14/10 hours.

Table 1. List of strains

Strain numbers in bold refer to axenic cultures; one culture marked with an asterisk (*) is no longer available. The species name is given as listed on the respective culture collection's homepage. Strains with CCAC- or M-numbers not yet publicly available are given with the internal labeling. **ACOI** = Coimbra Collection of Algae, Portugal (http://acoi.ci.uc.pt); **CAUP** = Culture Collection of Algae of Charles University in Prague, Czech Republic (http://botany.natur.cuni.cz/algo/caup-list.html); **CCAC** = Culture Collection of Algae at the University of Cologne, Germany (http://www.ccac.uni-koeln.de); **CCALA** = Culture Collection of Autotrophic Organisms, Třeboň, Czech Republic (http://ccala.butbn.cas.cz); **CCAP** = Culture Collection of Algae and Protozoa, UK (http://www.ccap.ac.uk); **M** = Culture Collection Melkonian, Botanical Institute, University of Cologne, Germany; **NIES** = Microbial Culture Collection at National Institute for Environmental Studies, Tsukuba, Japan (http://mcc.nies.go.jp); **SAG** = Sammlung von Algenkulturen, University of Göttingen, Germany (http://www.epsag.uni-goettingen.de); **UTEX** = Culture Collection of Algae at The University of Texas, Austin, USA (http://www.utex.org).

Strain number	cf. Species	Locality	Isolator	Year
ACOI 140	*M. kuetzingianum*	Portugal, Amieiro, pond near Arazede.	M. F. Santos	1979
ACOI 1447	*M. kuetzingianum*	Portugal, Mina do Vale das Gatas.	P. Ávila	2001
ACOI 1620	*M. kuetzingianum*	Portugal, Abrantes, Campo Militar de Sta Margarida, Barragem do Carvalhoso, canal, plankton.	M. F. Santos	2003
ACOI 1621	*M. strictissimum*	Portugal, Abrantes, Campo Militar de Sta Margarida, lake north of Lagoa da Murta, plankton.	M. F. Santos	2003
ACOI 1817	*M. kuetzingianum*	Portugal, Serra da Peneda.	J. Paiva	2005
ACOI 248	*M. strictissimum*	Portugal, Serra da Estrela, pond near Lagoa Comprida, plankton.	M. F. Santos	1986
ACOI 2656	*Microthamnion sp.*	Portugal, Serra da Estrela, pond near Lagoa Comprida.	O. Lourenço	1991
ACOI 2660	*Microthamnion sp.*	Portugal, Mata Nacional de Foja, canal.	G. Carvalho	2003
ACOI 398	*M. strictissimum*	Portugal, Serra da Estrela, pond near Lagoa Comprida.	O. Lourenço	1990
CAUP J 1201	*M. kuetzingianum*	Czech Republic, Central Bohemia, near Třtice, peat-bog "V Bahnách", soil.	Neustupa	1998
CCAC 0054	*M. kuetzingianum*	England, Cornwall, ford near Bowithick, freshwater.	M. Melkonian	1978
CCAC 0087	*M. kuetzingianum*	Germany, Lohmar near Cologne, Jexmühle, freshwater.	M. Melkonian	2002
CCAC 0539 B	*Microthamnion* sp.	Germany, Harz, freshwater.	M. Melkonian	1979
CCAC 2011	*M.* cf. *strictissimum*	Germany, Cologne, Wahner Heide, location 1, freshwater.	M. Melkonian	2002
CCAC 2081	*Microthamnion* sp.	Germany, Harz, Brunnenbachweg near Braunlage, temporary puddle, freshwater.	M. Melkonian	2002
CCAC 2182	*Microthamnion* sp.	Germany, Cologne, Wahner Heide, freshwater.	M. Melkonian	2003
CCAC 2197 B	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002
CCAC 2198	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002
CCAC 2199 B	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002
CCAC 2223 B	*M. kuetzingianum*	Germany, near Grande, swampy area beside River Bille, freshwater.	L. Kies	1970
CCAC 2224	*M. kuetzingianum*	Germany, near Grande, swampy area beside River Bille, freshwater.	L. Kies	1970
CCAC 2279	*Microthamnion* sp.	Germany, Eifel, Kall, industrial area 1, freshwater.	M. Melkonian	2003
CCAC 2764 B	*Microthamnion* sp.	Austria, Waldviertel. Fuchsteich (near Gmünd), freshwater.	M. Melkonian	2005
CCAC 2771	*Microthamnion* sp.	Austria, Waldviertel, Fuchsteich (near Gmünd), freshwater.	M. Melkonian	2005
CCAC 2804 B	*Microthamnion* sp.	Austria, Waldviertel, peat bog (near Heidenreichstein), freshwater.	M. Melkonian	2006
CCAC 2916	*Microthamnion* sp.	Germany, Cologne, Wahner Heide, Fuchskaule, freshwater.	M. Melkonian	2006
CCAC 2942 B	*Microthamnion* sp.	Austria, Waldviertel, Blockheide, freshwater.	M. Melkonian	2007
CCAC 2943 B	*Microthamnion* sp.	Germany, Cologne, Lindenthal, Stadtwald, freshwater.	M. Melkonian	2006
CCAC 3546	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410), freshwater.	M. Melkonian	2012
CCAC 3547	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410), freshwater.	M. Melkonian	2012
CCAC 3664 B	*Microthamnion* sp.	Germany, Eifel, Oberer Marmagener Stauteich, near shore, plankton.	M. Melkonian	2012
CCAC 3676 B	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012
CCAC 3677	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012
CCAC 3710	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012
CCAC 3838	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012
CCAC 3842 B	*Microthamnion* sp.	Germany, Stallberger Teiche, Lohmar, squeezed Sphagnum, freshwater.	M. Melkonian	2013
CCAC 3843 B	*Microthamnion* sp.	Germany, Stallberger Teiche, Lohmar, squeezed Sphagnum, freshwater.	M. Melkonian	2013
CCAC 4161	*Microthamnion* sp.	Scotland, Longhowe Loch, freshwater.	M. Melkonian	2013
CCAC 4234	*Microthamnion* sp.	Italy, Sardinia, inland pond "Pauli Trottas", near Stagno di Cabras, pipette probe from stone and sediment, freshwater.	S. Wittek	2013
CCAC 4544 B	*Microthamnion sp.*	Austria, Waldviertel (sample 19a), Schremser Hochmoor, Prügelsteg (GPS: 48°47.913' N 15°6.01' E), squeezed Utricularia, freshwater.	M. Melkonian	2014

Strain number	cf. Species	Locality	Isolator	Year
CCAC 4549 B	Microthamnion sp.	Austria, Waldviertel, "Schwarzes Moos", near Brand (GPS: 48°52.33' N 14°58.9' E), squeezed Utricularia, freshwater.	M. Melkonian	2014
CCAC 4558 B	*Microthamnion sp.*	Austria, Waldviertel, "Schwarzes Moos", near Brand (GPS: 48°52.33' N 14°58.9' E), squeezed Utricularia, freshwater.	M. Melkonian	2014
CCAC 4559 B	*Microthamnion sp.*	Austria, Waldviertel, "Schwarzes Moos", near Brand (GPS: 48°52.33' N 14°58.9' E), squeezed Utricularia, freshwater.	M. Melkonian	2014
CCAC 4717 B	*Microthamnion sp.*	Austria, Waldviertel, Schremser Hochmoor, Prügelsteg (GPS: 48°47.913' N 15°6.01' E), squeezed Utricularia, freshwater.	M. Melkonian	2014
CCAC 4818	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410, enrichment SFM), freshwater.	L. Caisová	2012
CCAC 4819	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410, enrichment 3N BBM), freshwater.	L. Caisová	2012
CCAC 4820	*Microthamnion* sp.	Germany, Eifel, Nettersheim (enrichment M7), leaf collected from a shaded pond.	L. Caisová	2012
CCAC 4821	*Microthamnion* sp.	Germany, Eifel, (sample 031, enrichment M7) Dahlemer Binz, Schlenke II (footmark), conductivity: 26,9 µS, temperature: 23,4°C, pH 4-5	L. Caisová	2012
CCAC 4822	*Microthamnion* sp.	Germany, Schwarzwald (sample 031, enrichment M7, clone 2), freshwater.	L. Caisová	2012
CCAC 4853	*Microthamnion sp.*	Germany, Köln, Wahnerheide, roadside ditch, freshwater.	S. Wittek	2014
CCAC 4854	*Microthamnion sp.*	Germany, Köln, Wahnerheide, roadside ditch, freshwater.	S. Wittek	2014
CCAC 4855	*Microthamnion sp.*	Germany, Köln, Wahnerheide, roadside ditch, freshwater.	S. Wittek	2014
CCAC 4856	*Microthamnion sp.*	Germany, Eifel, Dahlemer Binz, freshwater.	S. Wittek	2014
CCAC 4857	*Microthamnion sp.*	Germany, Eifel, Dahlemer Binz, freshwater.	S. Wittek	2014
CCAC 5520 B	*M.* cf. *kuetzingianum*	Germany, Frechen, waste water plant, Bio-P tank, water-body-sample, freshwater.	V. Zilz	2013
CCAC 5521	*M.* cf. *kuetzingianum*	Germany, Frechen, waste water plant, secondary settlement tank, scratch-sample, freshwater.	V. Zilz	2013
CCAC 5530 B	*M.* cf. *kuetzingianum*	Germany, Frechen, waste water plant, secondary settlement tank, scratch-sample, freshwater.	V. Zilz	2013
CCAC 5545	*M.* cf. *strictissimum*	Germany, Frechen, waste water plant, secondary settlement tank, scratch-sample, freshwater.	V. Zilz	2013
CCAC 5547 B	*M.* cf. *curvatum*	Germany, Glessen, waste water plant, aeration tank, scratch-sample, freshwater.	V. Zilz	2013
CCAC 5561	*M.* cf. *kuetzingianum*	Germany, Glessen, waste water plant, aeration tank, scratch-sample, freshwater.	V. Zilz	2013
CCAC 5585	*M.* cf. *strictissimum*	Germany, Villau, waste water plant, aeration tank, water-body-sample, freshwater.	V. Zilz	2013
CCAC 8001 B	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410, enrichment SFM), freshwater.	L. Caisová	2012
CCAC 8002 B	*Microthamnion* sp.	Germany, Schwarzwald (sample 031, enrichment M7, clone 1), freshwater.	L. Caisová	2012
CCALA 368	*M. kuetzingianum*	Slovakia, Orava, peat bog.	Kovacik	1983
CCAP 450/2	*M. kuetzingianum*	Antarctica, South Orkney Islands, Signy Island, freshwater.	Broady	1974
CCAP 450/3	*M. kuetzingianum*	England, Cornwall, River Gannel, freshwater.	Foster	1975
CCAP 450/4	*M. kuetzingianum*	England, Cornwall, by River Hayle, freshwater.	Foster	1976
M 2196/1 A	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/ Ahekapelle, freshwater.	M. Melkonian	2002
M 2412/1 A	*Microthamnion* sp.	Germany, Eifel, Strohner Maarchen, freshwater.	M. Melkonian	2003
M 4555*	*Microthamnion sp.*	Austria, Waldviertel, Schremser Hochmoor, Prügelsteg (GPS: 48°47.913' N 15°6.01' O), squeezed Utricularia, freshwater.	M. Melkonian	2014
NIES 479	*M. kuetzingianum*	Japan, Hokkaido, Sapporo, Toyohira River, freshwater.	F. Kasai	1987
SAG 114.80	*M. kuetzingianum* (form. *M. curvatum*)	Origin unknown, freshwater.	F. Ambard	1968
SAG 115.80	*M. kuetzingianum* (formerly *M. strictissimum*)	Germany, near Hamburg, coordinates: 53.535411/ 10.007172 (Lat./Long.), freshwater.	A. Weber	1969
UTEX LB 237	*Microthamnion* sp.	USA, Indiana, Bloomington, pond, freshwater.	R. C. Starr	1953

2.2. DNA Extraction/Preparations for Colony PCR

Prior to amplification, the DNA was either extracted or the cultures pretreated for colony PCR.

DNA extraction was performed using an E.Z.N.A® Plant DNA Kit by OMEGA according to the manufacturer's instructions for fresh/frozen plant samples. To harvest the cells, 1.5 - 2 ml algal culture were transferred to a 2 ml microcentrifuge tube and, depending on the condition of the culture, either concentrated via sedimentation (older cultures with larger individuals) or centrifuged for 10 minutes at 10,000 x g (younger cultures with smaller specimen/zoospores). The supernatant was discarded afterwards. For mechanical disruption of the cells, they were first frozen in liquid nitrogen with a stainless steel bead (Ø 5 mm) added to the microcentrifuge tube and subsequently disrupted in a tissue lyser (Tissue Lyser II, Qiagen). The tissue lyser adapter set was pre-frozen at -80°C for 15-20 minutes to prevent the pellet from thawing during the first run. The tissue lyser was operated twice for 1.5 minutes at 25 Hz, with the samples thawed before the second run. Lysis buffer was provided with the kit and added immediately after disruption of the cells. All further steps were performed according to the manufacturer's protocol. To achieve higher concentrations, the DNA was eluted in two times 25 µl heated elution buffer (65°C) instead of the recommended 50 µl. Afterwards the DNA concentration was quantified via NanoDrop 2000 spectrophotometer (Thermo Scientific). For long-term storage, the DNA was transferred to -21°C.

When colony PCR was performed, the cells were harvested as described above. After removal of the supernatant, 100 µl of TE buffer (see Table 2 for recipe) were added and the cells disrupted as described earlier. The obtained crude lysate then served directly as a template for PCR. The excess lysate was stored at 21°C.

Table 2. TE Buffer

Amount per 100 ml	Reagent
1 ml	1 M Tris-HCL (pH 8)
400 µl	0.25 M EDTA Titriplex III
98.6 ml	bidistilled water

2.3. Polymerase Chain Reaction

DNA amplification was performed via Polymerase Chain Reaction (PCR) as described by Saiki et al. (1988). In the present study, plastid-encoded as well as nuclear-encoded genes were investigated. The DNA was amplified with either a Taq (*Thermus aquaticus*) polymerase or, in cases of ITS2, the Phusion® High-Fidelity DNA Polymerase by NEB. The primers used for amplification of 18S rDNA, ITS2 and partial 28S rDNA, and the plastid-encoded *rbc*L gene can be retrieved from Table 3.

Table 3. PCR and sequencing primers

PCR and sequencing primers used for amplification of the nuclear-encoded 18S rRNA gene, the ITS2 molecule with flanking regions of 5.8S and 28S rRNA genes (including the D-domain) and the plastid-encoded *rbc*L gene.

PCR/sequencing primer	Sequence (5' to 3')	Published in
EAF3	TCGACAATCTGGTTGATCCTGCCAG	Marin et al. 2003
SSU_BR_49_50	CCTACGGAAACCTTGTTA	unpublished
5.8S_ITS03_forw	CGATGAAGAACGYAGCGA	Marin 2012
LSU_29_forw	TGAACTTAAGCATATCAATAAGCGG	unpublished
28S_ITS055_rev	CTCCTTGGTCCGTGTTTCAAGACGGG	Marin 2012
28S_1491_rev	TGCTGTTCACATGGAACC	Marin 2012
28S_1495_rev	CCAYGTCCAAYTGCTGTTCACRTGG	Marin 2012
BM_rbcL_1a_F	ATGKYWCCACAAACWGARAC	unpublished
BM_rbcL_3R	TCTTTCCAWACTTCACAHGCWGCWG	unpublished

The amount of chemicals used per PCR reaction is given in Table 4 (Taq polymerase) and Table 5 (Phusion® High-Fidelity DNA Polymerase). The volume of DNA template/crude lysate used with Taq polymerase varied between 0.5 and 5 µl, depending on the concentrations achieved during DNA extraction. PCR reactions with Phusion® High-Fidelity DNA Polymerase were prepared with either 1 µl DNA template (concentrations ≥ 50 ng/µl) or 1.5 µl (concentrations < 50 ng/µl). In all reactions the amount of nuclease-free water was adjusted accordingly. DNA was amplified using a Primus 96 plus thermocycler (MWG Biotech) with the program shown in Table 6 and Table 7 respectively. When a low yield of PCR product made reamplification necessary, a nested primer combination was applied and 0.5 µl of the primary PCR product served as a template for secondary PCR.

Table 4. Amount of chemicals per PCR reaction using Taq Polymerase

Amount	Reagent
18.875 µl	Nuclease-free water
2.5 µl	10x DreamTaq™ Buffer
2.5 µl	dNTPs (2 mM)
0.25 µl	Forward Primer (10 mM)
0.25 µl	Reverse Primer (10 mM)
0,5 µl	Template DNA
0.125 µl	DreamTaq™ DNA Polymerase (5 units/µl)
25 µl	Total Volume

Table 5. Amount of chemicals per PCR reaction using Phusion® High-Fidelity DNA Polymerase

Amount	Reagent
10.2 µl	Nuclease-free water
4 µl	5x Phusion GC Buffer
2 µl	dNTPs (2 mM)
1 µl	Forward Primer (10 mM)
1 µl	Reverse Primer (10 mM)
1 µl	Template DNA
0.6 µl	DMSO
0.2 µl	Phusion® High-Fidelity DNA Polymerase (2,000 U/ml)
20 µl	Total Volume

Table 6. Thermo-cycling conditions PCR (Taq Polymerase)

Step	Temperature [°C]	Time [Min.]
Initial Denaturation	95	03:00
30 cycles à		
Denaturation	95	00:45
Annealing	55	01:00
Elongation	72	03:00
Final Elongation	72	05:00
Hold	10	∞

Table 7. Thermo-cycling conditions PCR with Phusion® High-Fidelity DNA Polymerase
The annealing temperature of 67°C refers to the primer combination 5.8S_ITS03_forw + 28S_ITS055_rev.

Step	Temperature [°C]	Time [Min.]
Initial Denaturation	98	00:30
30 cycles à		
Denaturation	98	00:10
Annealing	67	00:30
Elongation	72	00:30
Final Elongation	72	05:00
Hold	10	∞

2.4. Agarose Gel Electrophoresis

The success of DNA amplification was evaluated via agarose gel electrophoresis. 3 µl of the PCR probes were mixed with a drop of loading dye (5 x Green GoTaq® Reaction Buffer, Promega) and loaded on an ethidium bromide stained (2.5 µl EtBr/50 ml gel) 1% agarose gel based on TAE-buffer (see Table 8 for recipe of 50-fold stock solution). For size comparison, 1.5 µl GeneRuler 1 kb DNA ladder (Thermo Scientific) were applied in the outermost slots. Gels were run for 20-30 minutes at 120 V in a VARIA 1 Electrophoresis Unit (Roth), with DNA fragments being visualized afterwards using an ultraviolet transilluminator (INTAS).

Table 8. 50 x TAE-Buffer

Amount	Chemical
242 g	Tris
16.615 g	Na_2 EDTA
57.1 ml	Glacial acetic acid
Up to 1 l final volume	bidistilled water

Adjust pH to 8.0 with glacial acetic acid

2.5. Sequencing Reaction

The PCR products were sequenced commercially by Eurofins Genomics GmbH, Ebersberg, Germany, with the Sanger method (Sanger et al., 1977). Prior to sequencing, the PCR probes were purified via isopropanol precipitation (for protocol see Supplementary Table 2, p. 91) and eluted in 20 µl of nuclease-free water. The DNA concentration was measured with the NanoDrop 2000 spectrophotometer. The sequencing reactions were prepared with the purified PCR product (adjusted to the right concentrations with nuclease-free water) and a primer, according to the requirements of the "PlateSeq Kit Mix for premixed samples" (Eurofins).

2.6. Sequence Assembly and Correction

The results were provided as shell command files (scf), each containing the data for one single reading. All readings belonging to one algal strain were assembled with AlignIR V2.0.48 (Li-COR) in a project file. They were carefully checked and corrected manually where necessary. The consensus sequence of all corrected reads was exported into an alignment file. SeaView 4.4 (Gouy et al., 2010) was used as alignment editor.

2.7. Taxon Sampling and Alignments

2.7.1. *Microthamnion*

A multi-gene alignment template consisting of *Microthamnion* 18S, partial 5.8S, ITS2 and partial 28S rDNA sequences was obtained from a previous work (Reder, 2015) and extended by the C- and D-domain of the nuclear-encoded 28S rRNA and the complete plastid-encoded *rbc*L gene. A total of 17 (18S), 18 (ITS2 with flanking regions of 5.8S and 28S), 29 (28S) and 75 (*rbc*L) sequences were newly generated in this study and incorporated in said alignment (see Supplementary Table 3, p. 92, for details).

18S, 5.8S and 28S sequences were aligned according to conserved secondary structure motifs which were originally obtained from the European Ribosomal RNA Database (Wuyts et al., 2004). Incongruences with other sequences or the secondary structure were rechecked in the alignment files and corrected when necessary. New ITS2 sequences were pre-aligned among each other which allowed an easy recognition of the conserved spacer regions introduced by Caisová et al. (2013). The single helices were then aligned according to the secondary structure information obtained from Reder (2015). Whenever aligning proved difficult or ambiguous versions were possible, the helices were folded via Mfold (http://mfold.rna.albany.edu/?q=mfold). The obtained foldings ran through a manual pairwise comparison of all strains and the secondary structure model was updated where it deemed reasonable.

The *rbc*L gene was aligned manually against the database sequence of *Microthamnion kuetzingianum* UTEX 318 (accession number: KM462876).

In addition to the above mentioned sequences, also database sequences were incorporated in the *Microthamnion* alignment. Therefore, extensive BLAST searches were performed with *Microthamnion* 18S, ITS2, 28S and *rbc*L query sequences.

Datasets for phylogenetic analyses were cleared of all hypervariable regions and those positions which could not be unambiguously aligned.

2.7.2. Trebouxiophyceae

Several publications served as starting points for the set-up of a trebouxiophycean alignment that was meant to cover all known lineages of said class. *Rbc*L data from representatives of the core Trebouxiophyceae and the Chlorellales, based on the work of Lemieux et al. (2014), were obtained from the NCBI (National Center for Biotechnology Information) DNA nucleotide database (https://www.ncbi.nlm.nih.gov) and incorporated in the alignment. The nucleotide sequences of those taxa were also used as query for an extensive BLAST search for additional closely related trebouxiophycean taxa.

Other phylogenetic markers (i.e. 18S, ITS2 and 28S) of the same taxa were targeted with BLAST searches by strain number and, in cases where no hits were found, species name. Since the goal was to sample a congruent multi-gene alignment, only those sequences were summarized, which originated from the same taxon and strain. As an additional precaution to prevent the formation of a chimera in cases of eventually erroneous database information regarding strain numbers and sequence identity, preliminary single-gene phylogenetic analyses (not shown) were performed to confirm the same positioning in the individual marker's phylogeny.

In addition, more publications were screened for representatives of the Trebouxiophyceae (Ueno et al., 2005; Darienko and Pröschold, 2015; Hallmann et al., 2016; Liu et al., 2017; Suzuki et al., 2018) and the data of eligible taxa obtained from the NCBI database.

All significant hits were incorporated in the alignment (for details of the aligning process see above) and preliminary phylogenetic analyses performed (not shown) to find redundant sequences as well as those not belonging to the core-Trebouxiophyceae and Chlorellales and to eliminate those from the alignment. When one group was overrepresented, a selection of taxa was made in order to achieve a well-balanced taxon sampling. Prior to phylogenetic analyses, all hypervariable regions that could not be unambiguously aligned as well as intron sequences were excluded from the alignment.

2.8. Phylogenetic Tree Reconstructions

All tree topologies were calculated on CHEOPS, a mainframe computer of the RRZK (Regionales Rechenzentrum der Universität zu Köln), using RAxML (Randomized Accelerated Maximum Likelihood) version 8.2.9 (Stamatakis, 2014). A total of 100 trees was calculated based on 100 distinct randomized starting MP trees, and the tree with the best topology automatically determined by the program. All multi-gene datasets for ML and Bayesian analyses were partitioned, thus allowing the independent estimation of the evolutionary model parameters (GTR+I+Γ) for each partition separately. When *rbc*L protein data was incorporated, the LG substitution model was chosen after checking with AliView version 1.18 (Larsson, 2014) that there was no translational deviation from the standard code.

The robustness of the trees' branches was tested by bootstrap (Felsenstein, 1985) and Bayesian posterior probabilities analyses. Bootstraps were calculated with maximum likelihood (ML), neighbor joining (NJ) and maximum parsimony (MP) methods. NJ calculations were performed with both, the GTR and the LogDet (Lockhart et al., 1994) model. Bootstrap analyses were performed with 1000 replicates each. ML bootstraps were calculated with RAxML, where the parameters for the evolutionary model (GTR+I+Γ) were automatically determined by the program. NJ and MP bootstraps were calculated with PAUP* 4b10 (Swofford, 2002) with the evolutionary model parameters for NJ (GTR+I+Γ/LogDet) obtained from modeltest. MP bootstrap analyses were performed with 10 heuristic searches per replicate. When gapped alignments were used, the MP analyses had to be constrained towards 1000 rearrangements per replicate, all other analyses were unrestrained. Bayesian posterior probabilities were calculated with MrBayes 3.2.6 (Ronquist et al., 2012). Two parallel Markov chain Monte Carlo (MCMC) runs with four chains each and ten million generations were performed. The covarion option was only used for the congruent dataset since the two MCMC runs did not converge in the gapped alignments with covarion active. Every hundredth tree was sampled and a burn-in between one and five million generations discarded depending on when the two runs converged (i.e. the standard deviation between the two MCMC chains was below 0.10). The remaining trees were summarized in a consensus tree. Support values of bootstrap analyses and posterior probabilities above 50%/0.8 were mapped on the best ML topology tree.

2.9. Identification of Phenetic Molecular Characters for Alternate Phylogenetic Tree Reconstructions

All molecular data was screened for 'phenetic' characters whose information content exceeded that of the mere succession of nucleotides in a linear sequence. This considered, for example, secondary structure information or the results of RFLP analyses (for detailed information see chapters 2.9.1 - 2.9.3). This information was then used for the composition of a data matrix which served as source material for a parsimony tree reconstruction.

2.9.1. *In silico* Restriction Fragment Length Polymorphism Analysis in 18S rDNA Sequences

An *in silico* search for cutting sites in all 43 *Microthamnion* 18S rDNA sequences was performed with a set of restriction enzymes (Table 9) in congruence with those used in John et al. (1993). The sequence data was loaded in the online tool RestrictionMapper Version 3 (http://www.restrictionmapper.org/) and "digested" with each of the enzymes separately.

Table 9. Restriction enzymes used for *in silico* RFLP analysis on 18S rDNA data.
Listed are the recognition- and cutting sites (indicated by three minus signs) of the nine restriction enzymes used for RFLP analysis of *Microthamnion* SSU data.

Restriction enzyme	Recognition site (5' to 3')	Cutting site
*Cfo*I	GCGC	GCG --- C
*Hinf*I	GANTC	G --- ANTC
*Msp*I	CCGG	C --- CGG
*Rsa*I	GTAC	GT --- AC
*Sau*3AI	GATC	--- GATC
*Taq*I	TCGA	T --- CGA
*Bam*HI	GGATCC	G --- GATCC
*Eco*RI	GAATTC	G --- AATTC
*Hin*dIII	AAGCTT	A --- AGCTT

2.9.2. Apomorphy Analysis and Non-Homoplasious Synapomorphies

To find further molecular support for the different *Microthamnion* groups/branches, an extensive apomorphy search on 18S, 5.8S, ITS2 and 28S sequence data was performed with PAUP. An alignment file including an ML tree of all *Microthamnion* strains (gapped alignment, all markers) was imported into PAUP and a search for apomorphies was performed under the parsimony criterion. The *rbc*L protein alignment was screened manually for amino acid changes.

The resulting list of apomorphies was checked for non-homoplasious synapomorphies (NHSs) (i.e. no convergent evolution outside and strict conservation within the investigated group) for previously defined clades and branches in the phylogenetic tree. Besides, positions with 100% sequence conservation within *Microthamnion* and positions 100% conserved in the Microthamniales were checked manually in the alignment against all trebouxiophycean taxa to find unique molecular signatures on genus and order levels.

2.9.3. Tree Reconstruction Based on Phenetic Molecular Characters

A great variety of 'phenetic' molecular characters, that exceeded the mere succession of nucleotides in sequence data, was summarized in a data matrix (Supplementary Table 4, p. 93). This matrix contained 112 characters (93 parsimony-informative, 19 parsimony-uninformative) available for phylogenetic tree reconstructions.
A maximum parsimony topology was calculated with PAUP, using the heuristic search option with a branch swapping algorithm (tree-bisection-reconnection).
A total of 100 replicates was computed with starting trees obtained via stepwise sequence addition ('MaxTrees' constrained to 10000). 3800 'best' trees with a score of 315 were found and a 50% majority rule consensus of all 'best' trees calculated.

3. Results

3.1. Phylogenetic Analyses

3.1.1. Datasets

A multi-gene dataset was assembled for *Microthamnion*, which comprised 18S rDNA (1588 unambiguously aligned positions), partial 5.8S rDNA (82), ITS2 (214), partial 28S rDNA (1446) and *rbc*L (1275 DNA, 425 AA) data. Those analyses containing the ITS2 plus flanking regions, operated with the 411 aligned characters of the 28S which were covered with the ITS2 primers. Thus, the combined analyses of all markers consisted of maximal 4605 aligned positions when the *rbc*L was included as DNA and 3755 when included based on the protein alignment.

Preliminary analyses showed that the topology within *Microthamnion* was not resolved in the basal part of the trees, thus the outgroup's impact on the topology was tested by calculating all analyses both, with and without *Characium perforatum.* Since there was no significant difference in the outcome, only those results with *Ch. perforatum* are shown.

The large trebouxiophycean dataset had a reduced number of unambiguously alignable characters in 18S (1570) and, especially, ITS2 (122). Due to the primer combinations apparently used for the database sequences, the flanking 28S region of ITS2 comprised only 22 aligned positions. All other markers had the same numbers as described above, resulting in 4495 available characters for the combined analysis (*rbc*L included as DNA).

3.1.2. Single-Gene Analyses

Single-gene analyses were performed with all *Microthamnion* sequences available for each marker (Supplementary Figure 7, pp. 97-103) to obtain a first overview of the individual markers' impact on the phylogeny as well as their resolution capacities, and to determine the one best suited for a quick assessment of subdivisions within the genus.

The 18S analysis featured 42 own sequences and was complemented by five *Microthamnion* database sequences (one culture, four environmental) and five environmental sequences forming a hitherto unknown sister group to *Characium perforatum*. ITS2 and '28S short' (i.e. the 411 positions covered with the ITS2 primers) analyses were performed with 56 and '28S long' (i.e. 1446 positions) with 27 own *Microthamnion* sequences. The *rbc*L data were completed for all 74 *Microthamnion* strains, and were augmented by one database sequence (UTEX 318, KM462876). The analyses of the *rbc*L were calculated based on both, the DNA and the protein alignment. Additional analyses discriminating between the different codon positions offered no exceeding information compared to the DNA results and are thus not shown.

The *rbc*L topology based on the DNA alignment (Supplementary Figure 7, p. 103) offered the best results from all single-gene analyses: several groups could be clearly distinguished based on branch lengths and support values, albeit the topology in the basal part of the tree was not resolved. Based on this tree, 27 strains (Supplementary Figure 7, written in bold) were selected for a congruent dataset, where sequence data were available for all markers. A well-balanced taxon sampling was ensured by covering all detected groups and long branches. This dataset was used to compare a great variety of single-gene as well as combined analyses and thus to determine the best combination of markers to resolve the *Microthamnion* phylogeny.

3.1.3. Congruent Phylogenies

Figure 1 and Figure 2 give the single-gene analyses with the congruent dataset. 'ITS2 short' refers to only the ITS2 molecule, whereas in 'ITS2 long' also the flanking regions of the 5.8S and 28S rRNA genes were included. These flanking regions were computed separately in '28S short'; '28S long' covered the 5' end of the 28S, starting with the reverse strand of helix B9 up until the beginning of the E-domain (Helix E2 forward). The *rbc*L gene was analyzed both, based on the DNA and the protein alignment, and the 18S analysis spanned the complete SSU. The topologies of several gene combinations were compiled in Figure 3 and Figure 4. In the analyses with two genes, the combination of a conserved with a variable marker (Figure 3 A+B) and the combination of two variable markers was tested (Figure 3 C). The 'w/o' (i.e. without) topologies were based on datasets where only one gene was excluded. For example, the tree in Figure 3 D was calculated with all markers except the 18S rRNA gene. In the tree reconstructions excluding ITS2 (Figure 4 A and B) or 28S (Figure 4 C+D), two variants were calculated each. One included the *rbc*L based on the nucleotide sequence and the other based on the amino acid sequence. Figure 4 E shows the concatenated analysis of all markers with the *rbc*L protein alignment. The 'best' phylogeny, i.e. the one based on all markers with *rbc*L included as DNA (for details see below), was highlighted in Figure 5. In addition to the ML bootstrap percentages, also NJ (GTR+I+Γ and LogDet) and MP support values as well as Bayesian posterior probabilities were given.

The single-gene analyses with the congruent dataset were naturally in accordance with the results of the single-gene analyses with all available sequences. They did however have the advantage of a maximized validity regarding comparability, since the same strains could be observed throughout all analyses. Although all markers showed the same evolutionary direction, i.e. did not contradict each other, there were significant differences between their ability to resolve the topology within *Microthamnion*. As expected, the most conserved markers 18S (Figure 2 C) and '28S short' (Figure 1 C) were not variable enough to resolve the topology. A bit more variability as in the '28S long' (Figure 1 D) and the *rbc*L protein dataset (Figure 2 B) allowed for a better delimitation of different groups within *Microthamnion* although they were generally low supported. Exceptions were two groups in the basal part of the trees: CCAC 2279/4234 and CCAC 4161/3677 which were well resolved with all markers except '28S short'. The comparison of both ITS2 datasets (Figure 1 A+B) indicated, that the

topology was determined by the ITS2 molecule, not the flanking regions of 5.8S and 28S. The support values remained quite similar between both sets, there was however a significant improvement of the blue strains' support from 64 to 80. The ITS2 worked out a fine distinction between the individual strains and also grouped some of them together in the tip of the tree. In the basal part, groups were clearly distinguishable by branch length but the relationship among the individual groups were largely unresolved in the whole tree. As in the single-gene analysis with all available strains, the *rbc*L based on the DNA alignment (Figure 2 A) offered the best result concerning distinction of different groups as well as support of the topology in the tip of the tree. The basal part however remained unresolved.

A combination of several genes (Figure 3 and Figure 4) though, improved the support values and also clarified the topology in the tip of the tree. A closer look at the multi-gene phylogenies revealed a generally inferior result in those without *rbc*L (Figure 3 A+E) compared to those containing *rbc*L data. The combination of *rbc*L, ITS2 and 18S data was necessary to work out a monophyly of the purple strains, albeit with a very short branch (Figure 4 C+D), which was not achieved with any other two- or three-marker-combination (compare Figure 3 B to E and Figure 4 A+B). The addition of 28S to this three-marker-combination (Figure 4 E and Figure 5) did not give any new insights regarding the topology, but increased the support values of the purple strain node and the branch indicating a sister relationship of strain CCAC 4857 to the colored strains significantly. Whenever the *rbc*L was included based on the protein alignment (Figure 4 B, D and E), the topology among the red, purple and blue groups was either disrupted or support values decreased significantly compared to the same analyses using the nucleotide sequences (Figure 4, A+C and Figure 5).

Figure 5 highlights the tree inferred from all markers with the *rbc*L included with the nucleotide sequence. In the tip of the tree, the topology was well-resolved with most branches displaying high support values. Four monophyletic groups could be distinguished which were highlighted in red, purple, blue and green font (in accordance with the clades introduced in Figure 6, p. 34). The purple group was only moderately supported, whereas the other three received high support with all methods. The node establishing a sister relationship between the colored groups and CCAC 4857, was almost maximally supported and the next descending node also had high support values. The topology in the basal part of the tree was not resolved, although two internal branches received maximal support with all methods.

None of the single-gene or combined analyses were able to clarify the relationship among the basal strains. There were however phylogenies that held statistical support with values between 70 and 80 (Figure 1 A+B, Figure 3 A, Figure 4 D+E) for the assignment of strain CCAC 4855 to the most basal position.

The colors in all congruent phylogenies refer to the clades defined in Figure 6. They were subsequently applied to facilitate better tracking of both, individual sequences and clades throughout different marker combinations.

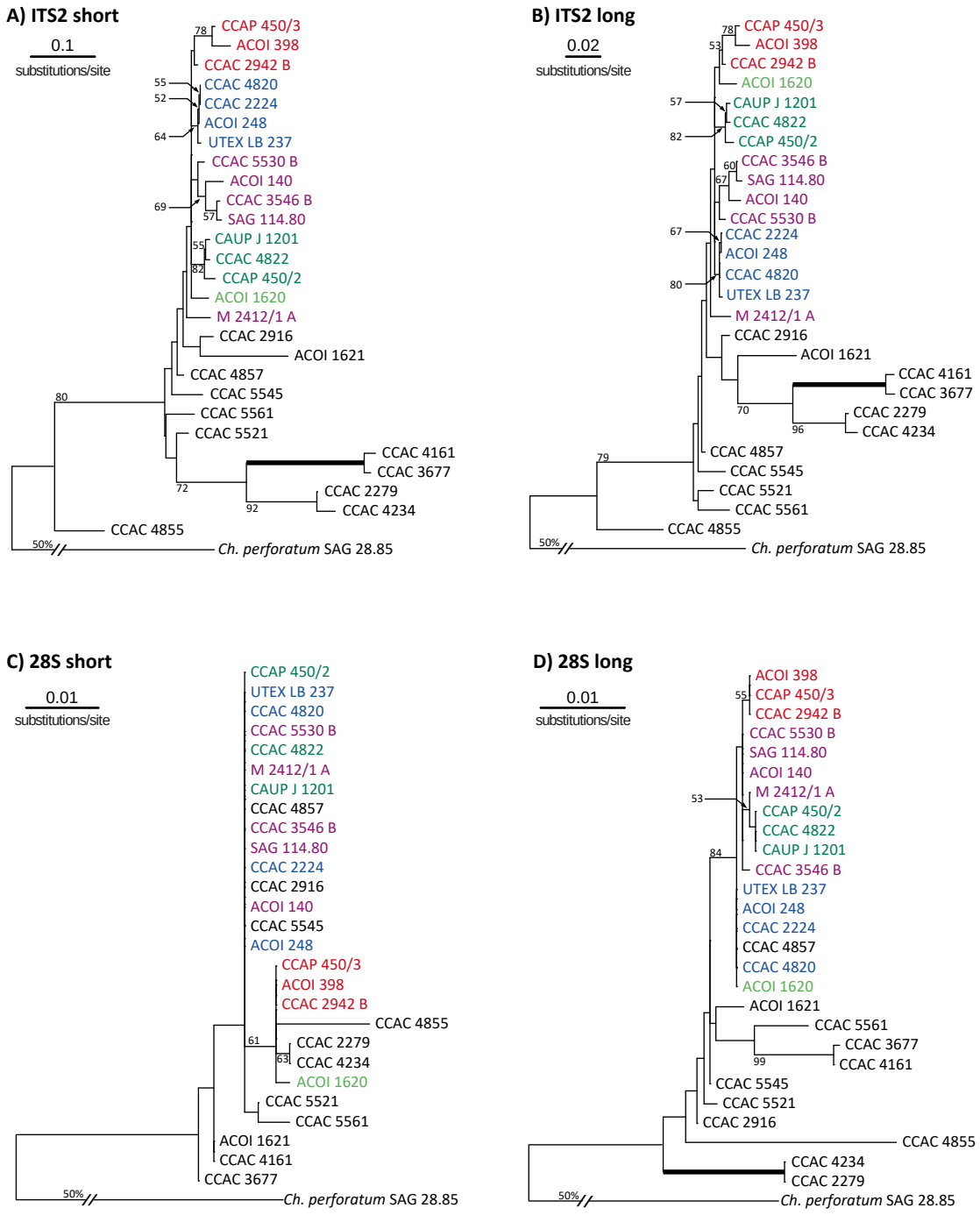


Figure 1. Comparison of congruent single-gene phylogenies (ITS2 and 28S)
ML topologies based on a congruent dataset of 27 *Microthamnion* strains with *Characium perforatum* used as outgroup. **A** ITS2 molecule (214 aligned positions), **B** ITS2 with flanking regions of 5.8S and 28S (707 pos.), **C** flanking regions of 5.8S and 28S without ITS2 (493 pos.), **D** 28S (1446 pos.). Bootstrap percentages: ML (maximal support indicated by bold branches). Very long branches were graphically reduced to 50% (50%//). Colors refer to the clades defined in Figure 6.

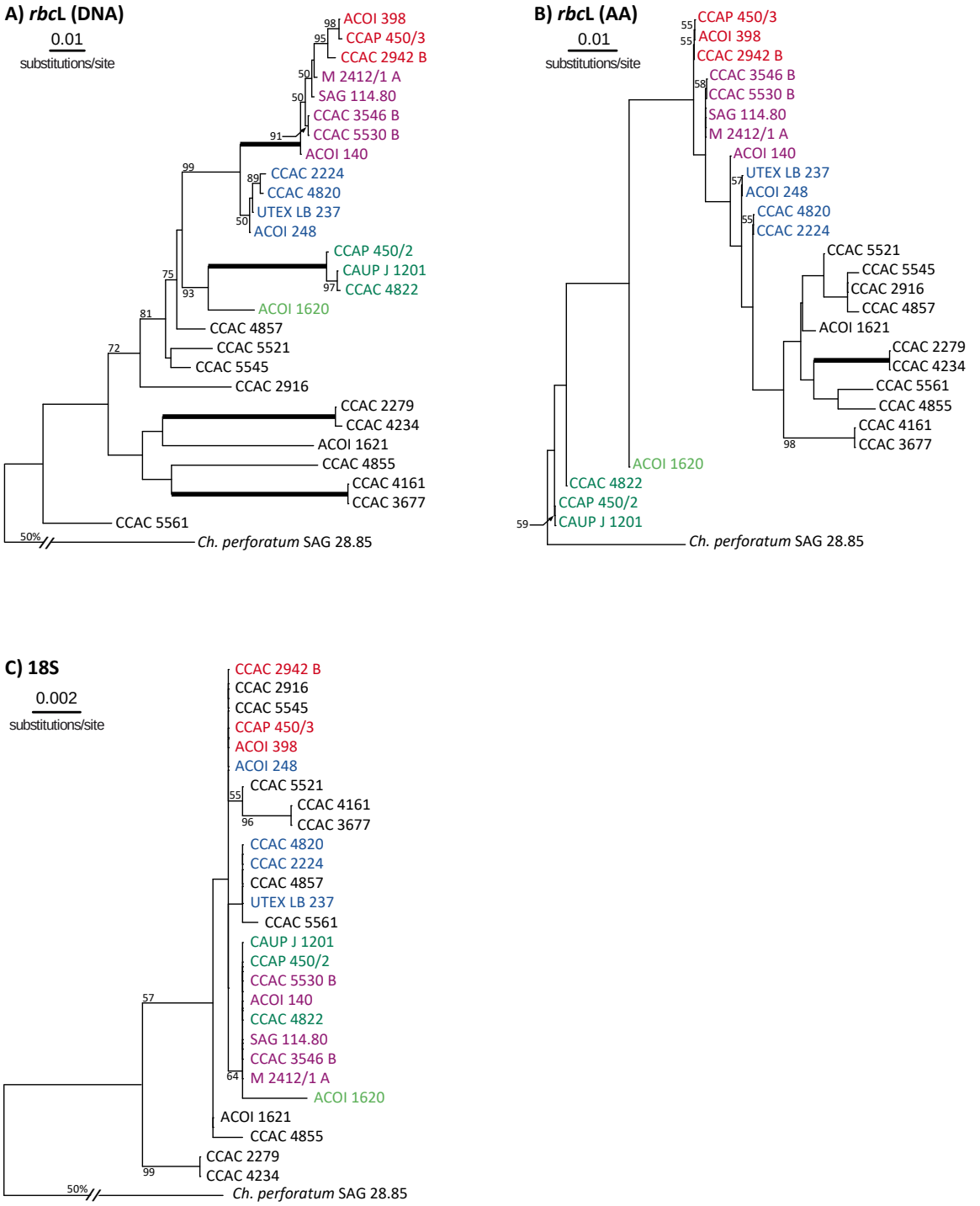


Figure 2. Comparison of congruent single-gene phylogenies (*rbc*L and 18S)
ML topologies based on a congruent dataset of 27 *Microthamnion* strains with *Characium perforatum* used as outgroup. **A** *rbc*L based on the DNA alignment (1275 aligned positions), **B** *rbc*L based on the protein alignment (425 pos.), **C** 18S (1588 pos.). Bootstrap percentages: ML (maximal support indicated by bold branches). Very long branches were graphically reduced to 50% (50%//). Colors refer to the clades defined in Figure 6.

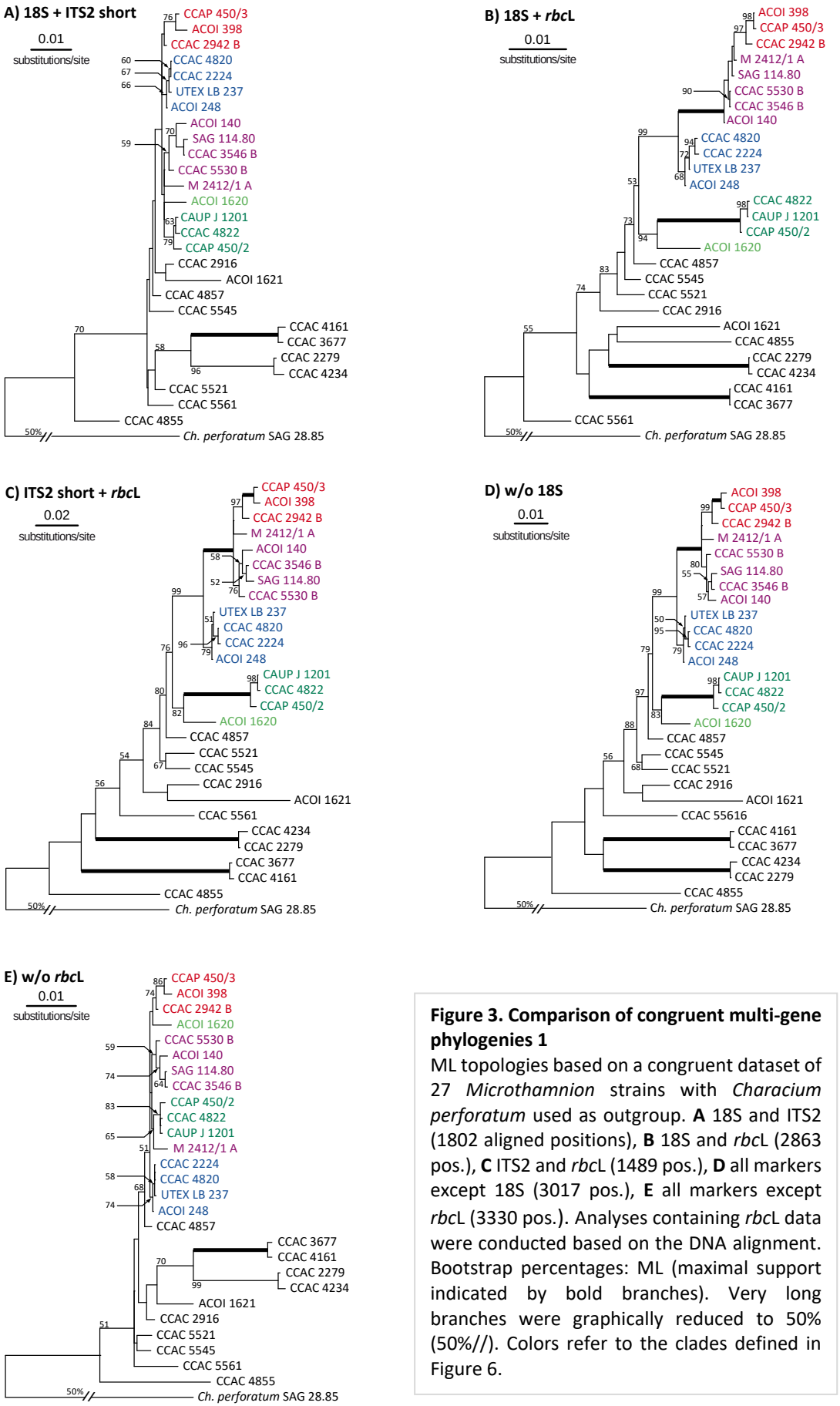


Figure 3. Comparison of congruent multi-gene phylogenies 1

ML topologies based on a congruent dataset of 27 *Microthamnion* strains with *Characium perforatum* used as outgroup. **A** 18S and ITS2 (1802 aligned positions), **B** 18S and *rbc*L (2863 pos.), **C** ITS2 and *rbc*L (1489 pos.), **D** all markers except 18S (3017 pos.), **E** all markers except *rbc*L (3330 pos.). Analyses containing *rbc*L data were conducted based on the DNA alignment. Bootstrap percentages: ML (maximal support indicated by bold branches). Very long branches were graphically reduced to 50% (50%//). Colors refer to the clades defined in Figure 6.

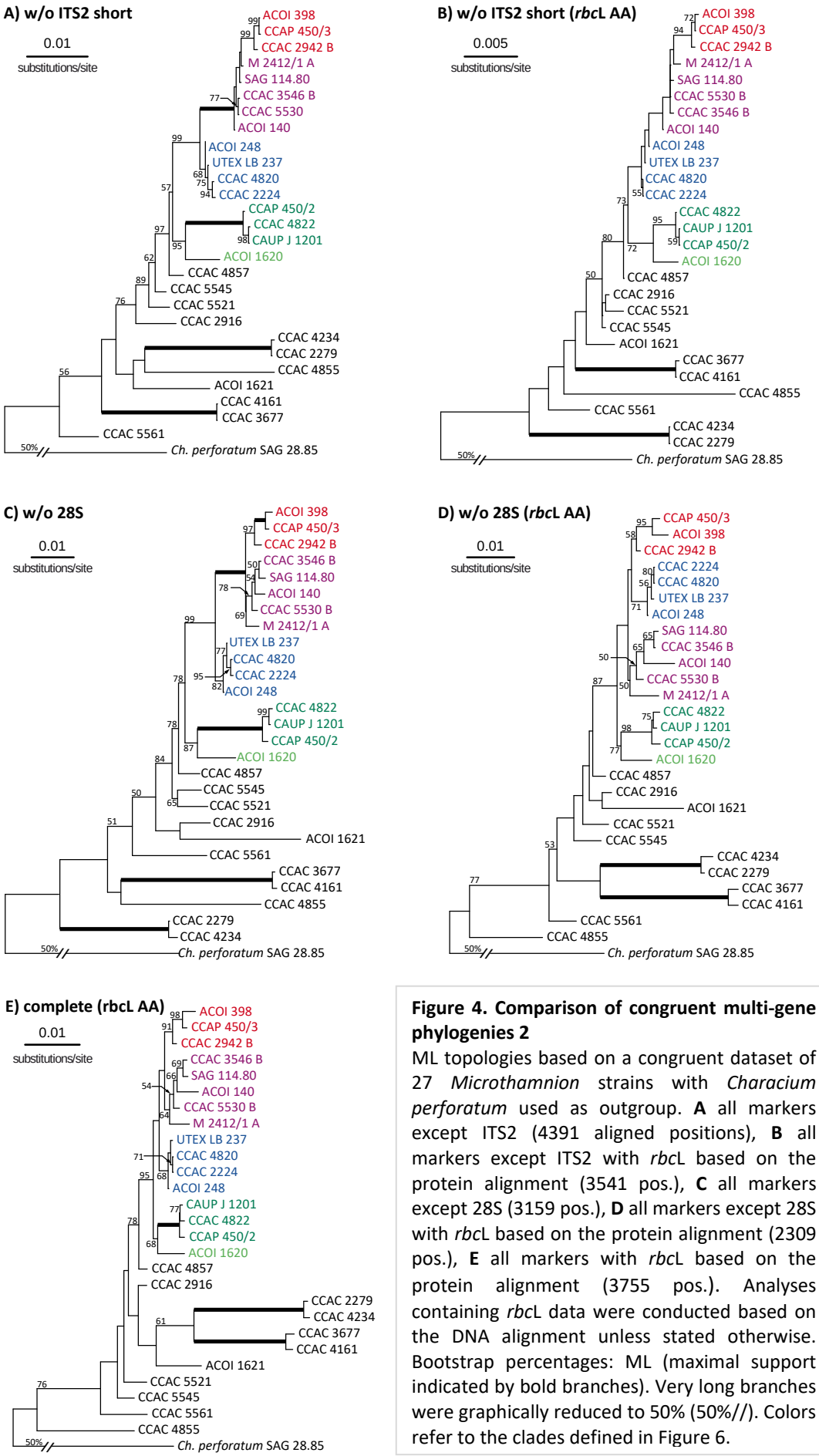


Figure 4. Comparison of congruent multi-gene phylogenies 2

ML topologies based on a congruent dataset of 27 *Microthamnion* strains with *Characium perforatum* used as outgroup. **A** all markers except ITS2 (4391 aligned positions), **B** all markers except ITS2 with *rbc*L based on the protein alignment (3541 pos.), **C** all markers except 28S (3159 pos.), **D** all markers except 28S with *rbc*L based on the protein alignment (2309 pos.), **E** all markers with *rbc*L based on the protein alignment (3755 pos.). Analyses containing *rbc*L data were conducted based on the DNA alignment unless stated otherwise. Bootstrap percentages: ML (maximal support indicated by bold branches). Very long branches were graphically reduced to 50% (50%//). Colors refer to the clades defined in Figure 6.

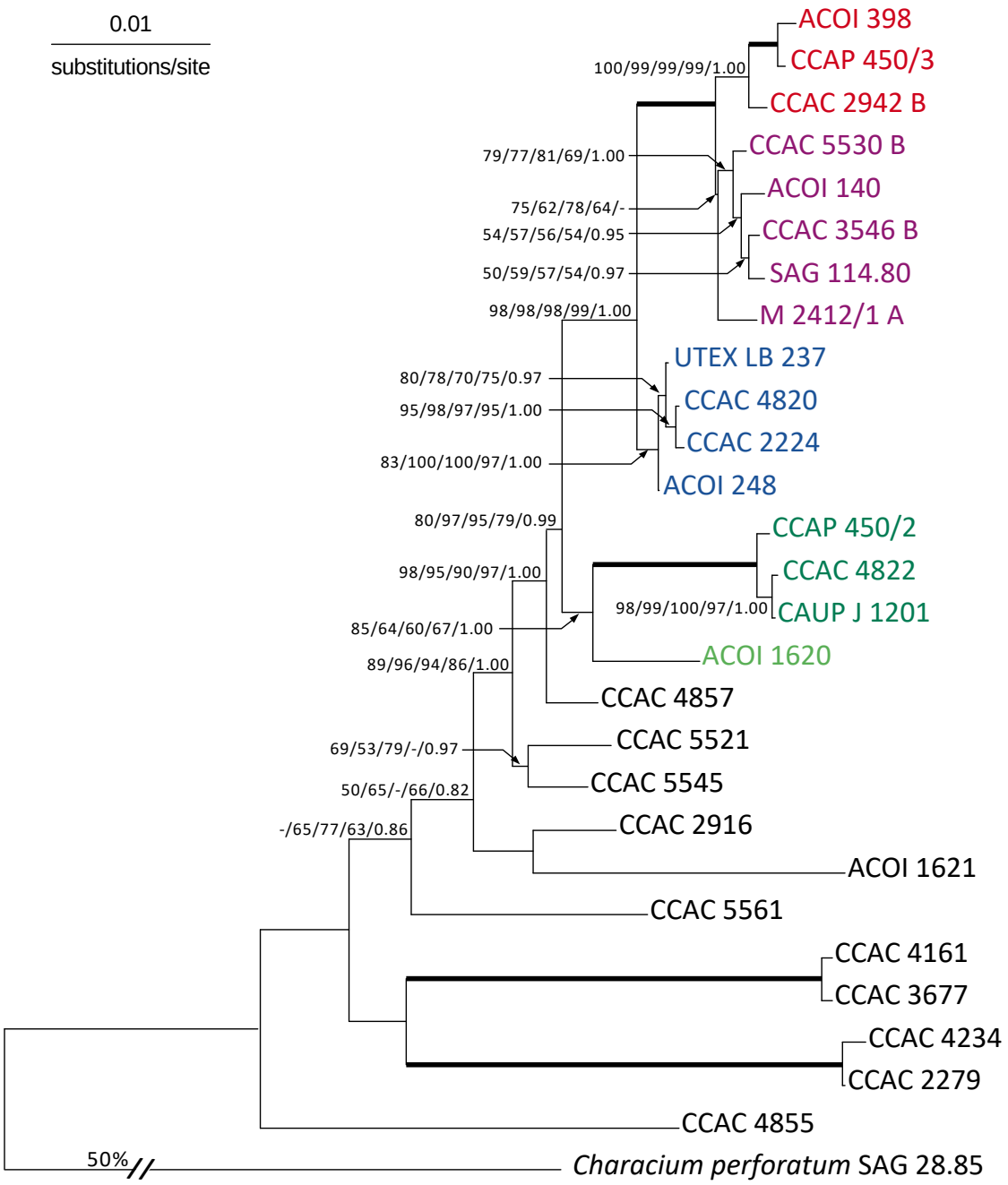


Figure 5. Molecular phylogeny of all markers based on a congruent dataset
ML topology based on a congruent dataset of 27 *Microthamnion* strains with *Characium perforatum* used as outgroup. Combined analyses of all markers with the *rbc*L included as DNA, i.e. 4605 analyzed positions. Support values are given in the following order: ML/NJ(GTR+I+Γ)/NJ(LogDet)/MP/MrBayes. Branches in bold received maximal support with all methods.). A very long branch was graphically reduced to 50% (50%//). Colors refer to the clades defined in Figure 6.

3.1.4. *Microthamnion* Phylogeny

Figure 6 shows the phylogeny of all obtained *Microthamnion* sequence data (i.e. gapped alignment). *Characium perforatum* and five environmental sequences served as outgroup. Before the final analysis, the alignment was however cleared of those *Microthamnion* sequences where only 18S data was available (= some database sequences). Due to the overall sequence similarity in the *Microthamnion* SSU these sequences were positioned randomly throughout the upper part of the tree (clade 1-4 and grade 1) in preliminary analyses (not shown), thus decreasing the support values tremendously.

Four monophyletic clades could be defined (highlighted in red, purple, blue and green) with a subdivision in clade 4 (light and dark green), and the remaining strains were summarized in two grades (different shades of grey). The phylogeny was generally well resolved in the tip; in the basal part however, although internal branches received high support, the topology remained unclear.

Clade 1 showed a high support with all methods except NJ, whereas clade 2 yielded only low support at the short 'clade branch'. Through the well-supported node defining the sister relationship of clades 1 and 2 and the clear distinction of clade 1, it was nonetheless readily identified. Clade 3 was moderately supported and formed the sister group to clades 1 and 2. The branch separating clades 1-3 from the remaining strains, received quite high support values throughout all methods. For clade 4, ML bootstraps and Bayesian posterior probabilities achieved high values, yet NJ methods were below the 50% threshold. The clade could be divided in two subclades (one consisting only of one single strain), depicted in light and dark green, the latter being maximally supported. The node corresponding to the divergence of clades 1-4 from the remaining *Microthamnion* strains (thus defining a sister relationship of clade 4 to clades 1-3) received low support with ML and MP methods, but was almost maximally supported by Bayesian posterior probabilities.

The basal part of the tree was characterized by very long branches with individual high support, but an unresolved topology. Some branches referred to only a single strain or strains originating from the same natural sample. A subdivision in two grades was performed based on a well-supported node that grouped the strains of grade 1 with clades 1-4. The remaining strains were summarized to grade 2.

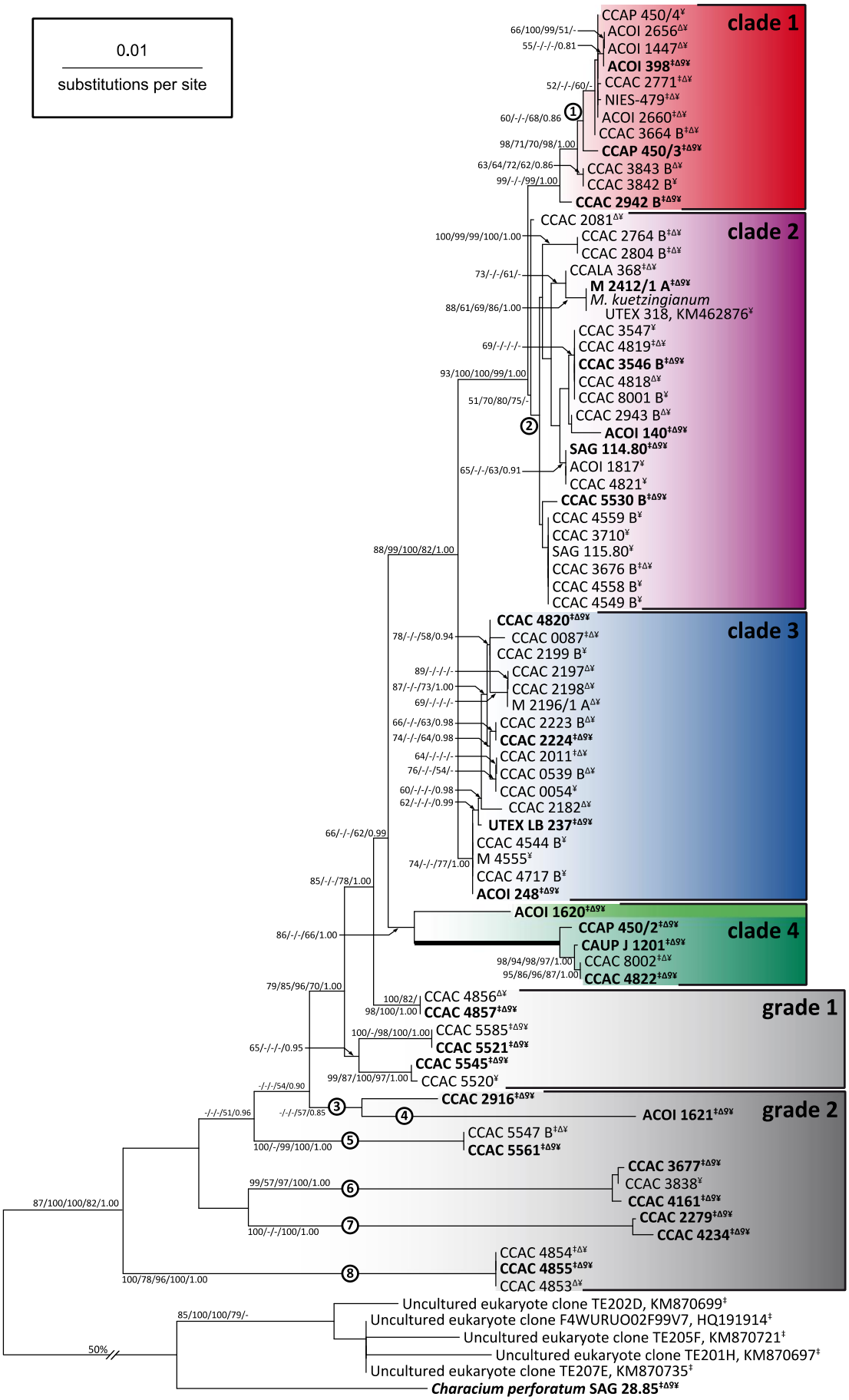
A peculiarity revealed by this phylogeny, was the high abundance of strains allocated to clades 1-3 with at times large clusters of identical or highly similar genotypes. This circumstance, although interesting for conclusions regarding habitat preferences and especially sampling methods (more details in the Discussion chapter), is also tantamount to an uneven taxon sampling.

A comparison of this tree, which is based on a gapped alignment, with the phylogeny inferred from the congruent alignment with a well-balanced taxon sampling (Figure 5) revealed a generally lower statistical support in the tip of the tree, the overall topology was however not affected.

The most dramatic effect was perhaps at the node of clade 2, where ML bootstrap values decreased by 24 points compared with those in Figure 5. Clade 3 also received significantly lower support with ML, distance and parsimony methods. The node defining a sister relationship between clades 1+2 and clade 3 was almost maximally supported in the congruent phylogeny, whereas here the values of ML and MP bootstraps were 10 and 15 points lower respectively. The branch separating clades 1-4 from the remaining strains received almost maximal support with both distance methods in Figure 5, which were below the 50% threshold in the present phylogeny. Also ML and MP bootstraps were significantly lower here. Statistical support for both next descendant nodes deteriorated considerably too. With the transition to grade 2, support was generally weak in both trees with the exceptions of two internal branches (i.e. CCAC 3677/CCAC 4161 and CCAC 2279/CCAC 4234) that received maximum support with the congruent dataset. In Figure 6 on the other hand, distance methods either failed or indicated low support. The relationship among the groups of grade 2 however remained unresolved in both trees.

Figure 6. Molecular phylogeny of *Microthamnion*

Phylogeny of all available *Microthamnion* data, except solely 18S environmental and database sequences, highlighting four monophyletic clades depicted in red, purple, blue and green, with a subdivision in clade four (light and dark green) and two grades (light and dark grey). *Characium perforatum* and five hitherto unidentified environmental sequences used as outgroup. Gapped alignment based on 4605 aligned positions, markers available for each strain are indicated by the following symbols: ‡=18S (1588 pos.), Δ=**ITS2** (∑ 707 pos.: 82/**214**/411), ϑ=28S (1446 pos.) and ¥=*rbc*L (DNA: 1275 pos.). Strains in bold were used for the congruent analyses (Figure 1 to Figure 5). Topology obtained by ML analysis. Support values are given in the following order: ML/NJ (GTR+I+Γ)/NJ (LogDet)/MP/MrBayes. The branch in bold was maximally supported with all methods, a very long branch was graphically reduced to 50% (50%//). Encircled numbers refer to those branches where CBCs occurred (for details see Table 12. P. 56).


0.01
substitutions per site
clade 1
clade 2
clade 3
clade 4
grade 1
grade 2
CCAP 450/4¥
ACOI 2656Δ¥
ACOI 1447Δ¥
ACOI 398‡Δ9¥
CCAC 2771‡Δ¥
NIES-479‡Δ¥
ACOI 2660‡Δ¥
CCAC 3664 B‡Δ¥
CCAP 450/3‡Δ9¥
CCAC 3843 BΔ¥
CCAC 3842 B¥
CCAC 2942 B‡Δ9¥
CCAC 2081Δ¥
CCAC 2764 B‡Δ¥
CCAC 2804 B‡Δ¥
CCALA 368‡Δ¥
M 2412/1 A‡Δ9¥
M. kuetzingianum
UTEX 318, KM462876¥
CCAC 3547¥
CCAC 4819‡Δ¥
CCAC 3546 B‡Δ9¥
CCAC 4818Δ¥
CCAC 8001 B¥
CCAC 2943 BΔ¥
ACOI 140‡Δ9¥
SAG 114.80‡Δ9¥
ACOI 1817¥
CCAC 4821¥
CCAC 5530 B‡Δ9¥
CCAC 4559 B¥
CCAC 3710¥
SAG 115.80¥
CCAC 3676 B‡Δ¥
CCAC 4558 B¥
CCAC 4549 B¥
CCAC 4820‡Δ9¥
CCAC 0087‡Δ¥
CCAC 2199 B¥
CCAC 2197Δ¥
CCAC 2198Δ¥
M 2196/1 AΔ¥
CCAC 2223 BΔ¥
CCAC 2224‡Δ9¥
CCAC 2011‡Δ¥
CCAC 0539 BΔ¥
CCAC 0054¥
CCAC 2182Δ¥
UTEX LB 237‡Δ9¥
CCAC 4544 B¥
M 4555¥
CCAC 4717 B¥
ACOI 248‡Δ9¥
ACOI 1620‡Δ9¥
CCAP 450/2‡Δ9¥
CAUP J 1201‡Δ9¥
CCAC 8002‡Δ¥
CCAC 4822‡Δ9¥
CCAC 4856Δ¥
CCAC 4857‡Δ9¥
CCAC 5585‡Δ9¥
CCAC 5521‡Δ9¥
CCAC 5545‡Δ9¥
CCAC 5520¥
CCAC 2916‡Δ9¥
ACOI 1621‡Δ9¥
CCAC 5547 B‡Δ¥
CCAC 5561‡Δ9¥
CCAC 3677‡Δ9¥
CCAC 3838¥
CCAC 4161‡Δ9¥
CCAC 2279‡Δ9¥
CCAC 4234‡Δ9¥
CCAC 4854‡Δ¥
CCAC 4855‡Δ9¥
CCAC 4853Δ¥
Uncultured eukaryote clone TE202D, KM870699‡
Uncultured eukaryote clone F4WURUO02F99V7, HQ191914‡
Uncultured eukaryote clone TE205F, KM870721‡
Uncultured eukaryote clone TE201H, KM870697‡
Uncultured eukaryote clone TE207E, KM870735‡
Characium perforatum SAG 28.85‡Δ9¥
66/100/99/51/-
55/-/-/-/0.81
52/-/-/60/-
60/-/-/68/0.86
98/71/70/98/1.00
63/64/72/62/0.86
99/-/-/99/1.00
100/99/99/100/1.00
73/-/-/61/-
88/61/69/86/1.00
69/-/-/-/-
93/100/100/99/1.00
51/70/80/75/-
65/-/-/63/0.91
88/99/100/82/1.00
78/-/-/58/0.94
89/-/-/-/-
87/-/-/73/1.00
69/-/-/-/-
66/-/-/63/0.98
74/-/-/64/0.98
64/-/-/-/-
76/-/-/54/-
60/-/-/-/0.98
62/-/-/-/0.99
66/-/-/62/0.99
74/-/-/77/1.00
85/-/-/78/1.00
86/-/-/66/1.00
98/94/98/97/1.00
95/86/96/87/1.00
79/85/96/70/1.00
100/82/
98/100/1.00
100/-/98/100/1.00
65/-/-/-/0.95
99/87/100/97/1.00
-/-/-/54/0.90
-/-/-/51/0.96
-/-/-/57/0.85
100/-/99/100/1.00
99/57/97/100/1.00
87/100/100/82/1.00
100/-/-/100/1.00
100/78/96/100/1.00
85/100/100/79/-
50%
1
2
3
4
5
6
7
8

3.1.5. Trebouxiophycean Phylogeny

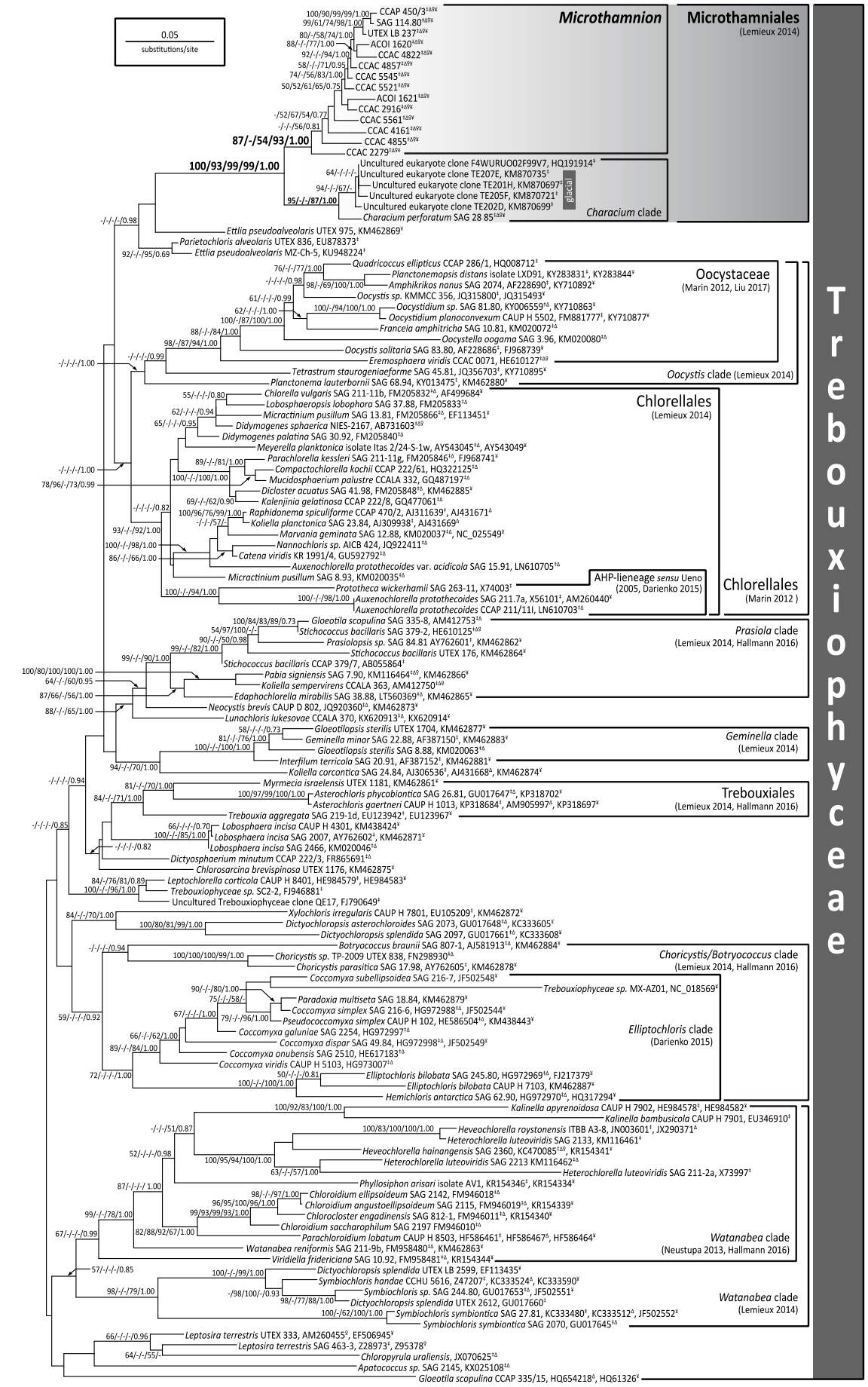
The phylogeny in Figure 7 covers all hitherto known lineages of the Trebouxiophyceae (as of Marin 2012) which are also referred to as 'Core Trebouxiophyceae and Chlorellales' in the literature (for example Lemieux et al., 2014; Suzuki et al., 2018). The alignment was constricted to photoautotrophic organisms, so the non-photosynthetic representatives of the 'AHP-lineage', as described in Suzuki et al. (2018), were excluded. Extensive BLAST searches and preliminary phylogenetic analyses (not shown) were performed to ensure both, the complete coverage of trebouxiophycean data available as well as a balanced taxon sampling with no group being overrepresented.

The tree covered 127 trebouxiophycean strains, 14 of which were representatives of *Microthamnion* selected to cover clades 1-4 and both grades evenly. Five environmental sequences were found to form a sister group to *Characium perforatum* and were summarized with the latter to the newly defined *Characium* clade. The monophyly of *Microthamnion* as well as the *Characium* clade was confirmed with high support values except the NJ bootstraps, and the Microthamniales were almost maximally supported with all methods. The branch separating the Microthamniales from the remaining Trebouxiophyceae was very long and also the discriminating branch between *Microthamnion* and the *Characium* clade was quite distinct. The next relatives to the Microthamniales could not be unequivocally identified, as two *Ettlia* and one *Parietochloris* strains, though placed in proximity by the topology, were not confirmed by statistical support except with Bayesian posterior probabilities.

Most of the trebouxiophycean lineages, and also internal branches, were well supported to the exclusion of the distance methods which were mainly below the 50% threshold. Some groups, like the Chlorellales, could be well-sampled and had shorter internal branches than other groups where sequences for only few members were available (e.g. the 'upper' part of the *Watanabea* clade comprising *Kalinella*, *Heveochlorella* and *Heterochlorella*). As the relationship among the trebouxiophycean lineages remained unclear, midpoint rooting was chosen.

Figure 7. Molecular phylogeny of the Trebouxiophyceae

Molecular phylogeny of 127 Trebouxiophyceae (*Microthamnion*: 14 strains, *Characium* clade: 6) confirming the monophyly of *Microthamnion* and resolving the Microthamniales as a monophyletic clade. Analysis based on 4495 aligned positions, using 18S, ITS2 (with flanking regions of 5.8S and 28S), 28S and *rbcL* (DNA) data. Gapped alignment; markers available for each taxon/strain are indicated by the following symbols: ‡=18S (1570 pos.), Δ=ITS2 (82/**122**/22 pos.), ϑ=28S (1446 pos.) and ¥=*rbc*L (DNA, 1275 pos.). ML topology, midpoint rooting; order of support values is ML/NJ(GTR+I+Γ)/NJ(LogDet)/MP/MrBayes. Taxonomic designations refer to recent phylogenetic studies: Lemieux et al. (2014), Marin (2012), Liu et al. (2017), Ueno et al. (2005), Darienko & Pröschold (2015), Hallmann et al. (2016), Neustupa et al. (2013) and Suzuki et al. (2018).


0.05
substitutions/site
Microthamnion
Microthamniales
(Lemieux 2014)
Trebouxiophyceae
Characium clade
glacial
Oocystaceae
(Marin 2012, Liu 2017)
Oocystis clade (Lemieux 2014)
Chlorellales
(Lemieux 2014)
AHP-lieneage sensu Ueno
(2005, Darienko 2015)
Chlorellales
(Marin 2012)
Prasiola clade
(Lemieux 2014, Hallmann 2016)
Geminella clade
(Lemieux 2014)
Trebouxiales
(Lemieux 2014, Hallmann 2016)
Choricystis/Botryococcus clade
(Lemieux 2014, Hallmann 2016)
Elliptochloris clade
(Darienko 2015)
Watanabea clade
(Neustupa 2013, Hallmann 2016)
Watanabea clade
(Lemieux 2014)

3.2. Phenetic Character Identification and Parsimony Tree Reconstruction

3.2.1. *In Silico* Analysis of Restriction Enzyme Sites in 18S rDNA Sequences

Table 10 presents the total number of cutting sites per enzyme when applied on *Microthamnion* 18S rDNA sequences. The database sequence of *Microthamnion kuetzingianum* UTEX 1914 (accession number Z28974) was given as reference for the majority of *Microthamnion* strains. Strains with divergent sequences that resulted in differing numbers of cuts were listed separately. There was a total of six strains with divergent numbers of cutting sites, in which strains CCAC 2279/4234, CCAC 3677/4161 and CCAC 5561/5547 B had identical sequences and such the same number of cuts each. Enzymes with six-base-pair recognition sites (*Bam*HI, *Eco*RI, *Hind*III) had no hits at all in *Microthamnion* 18S sequences. Digests with *Hinf*I and *Sau3*AI resulted in 12 and four cuts respectively across all *Microthamnion* strains.

Table 10. Total number of restriction enzyme cuts in 18S for all *Microthamnion* strains
The total number of cuts per enzyme in the 18S molecule is given for the majority of *Microthamnion* strains (represented by Z28974) as well as the divergent strains CCAC 5561/5547 B, CCAC 2279/4234 and CCAC 3677/4161.

Restriction enzyme	Total number of cutting sites			
	Z28974	CCAC 5561/ CCAC 5547 B	CCAC 2279/ CCAC 4234	CCAC 3677/ CCAC 4161
*Cfo*I	10	9	7	10
*Hinf*I	12	12	12	12
*Msp*I	7	7	6	7
*Rsa*I	4	4	4	3
*Sau*3AI	4	4	4	4
*Taq*I	5	5	6	4
*Bam*HI	0	0	0	0
*Eco*RI	0	0	0	0
*Hin*dIII	0	0	0	0

Figure 8 gives an overview of the cutting sites' positions within the 18S rRNA molecule. A template of this figure was obtained from a previous work of Reder (2015) and updated with new information now available due to new sequence data. A schematic drawing of the 18S molecule, based on database information for *Chlamydomonas reinhardtii* (originally obtained from The European Ribosomal RNA database homepage: http://bioinformatics.psb.ugent.be/webtools/rRNA/secmodel/Crei_SSU.html), depicts the arrangement and positions of the different helices.

***Microthamnion kuetzingianum* Z28974**

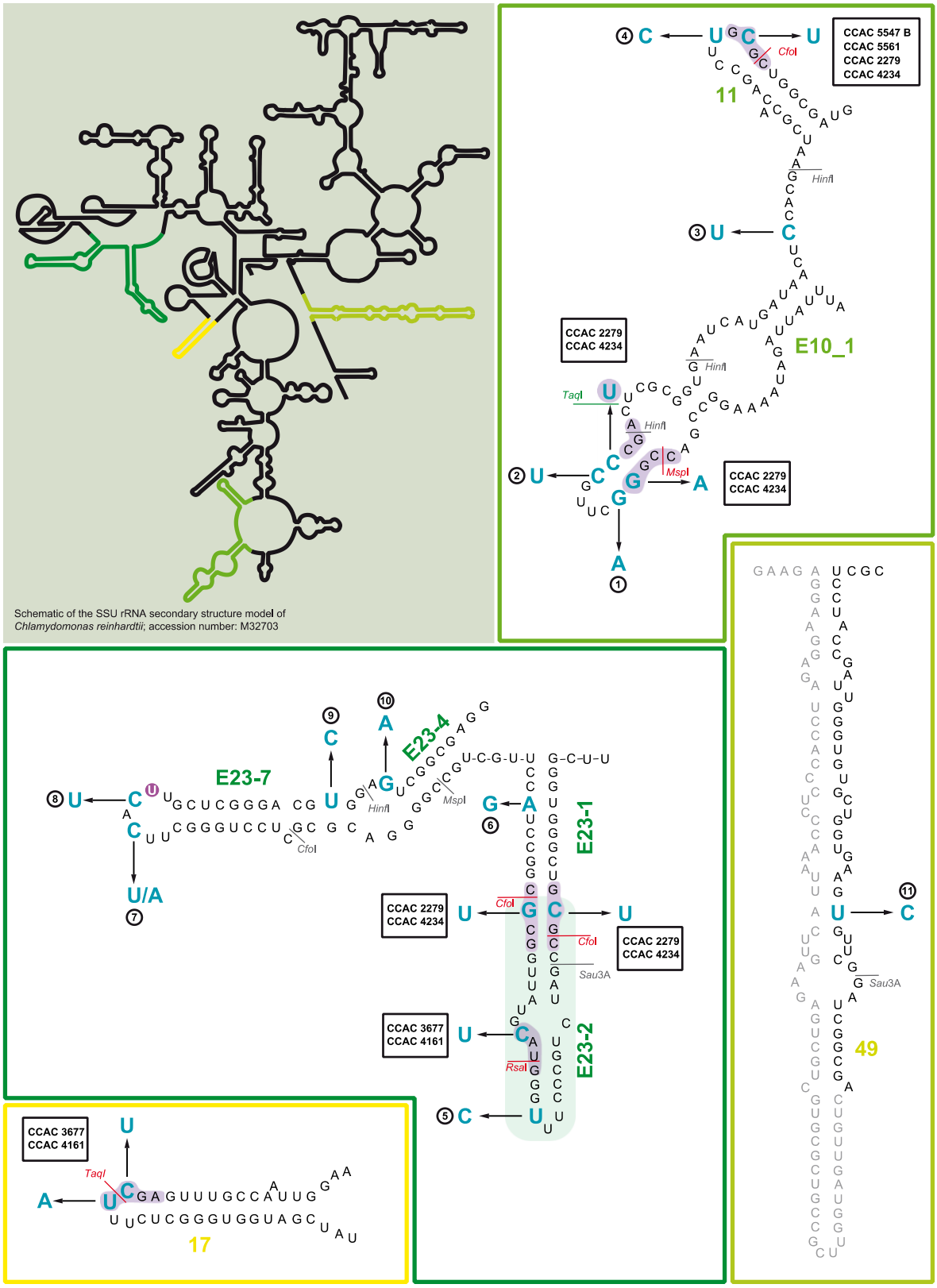


Figure 8. Position of restriction enzyme cutting sites mapped on a schematic of the 18S rRNA molecule
According to Reder (2015), modified. Overview sketch of the 18S molecule based on a *Chlamydomonas reinhardtii* sequence (green background), giving the positions of the different helices within the 18S molecule. Regions with sequence variability in *Microthamnion* are highlighted in different shades of green and yellow which are given in detail (for *Microthamnion kuetzingianum* Z28974) in boxes encircled in the same color. Grey nts in helix 49 were not analyzed (outside the amplified region), helix E23-2 is distinguished from helix E23-1 by a light green background, additional U in loop of helix E23-7 is solely present in Z28974 (purple circle). Helices are numbered according to Van de Peer et al. (1999) and Wuyts et al. (2000). Restriction enzyme sites are given in grey (cuts in all strains), red (prevented cut) or green (additional cut). For the latter two, recognition sites are highlighted in purple, with nt changes given in large blue font. Strains to which changes apply are given in black boxes for each position. Nt changes without effect on number of cuts are marked with encircled numbers that refer to the following strains: 1, 2, 10: ACOI 1620 (i.e. clade 4 light green); 3, 11: CCAC 2279/CCAC 4234; 4, 6: CCAC 4854/CCAC 4855; 5: clades 2 and 4; 7: A in ACOI 1620, U in CCAC 5521/CCAC 5585 and CCAC 3677/CCAC 4161; 8: clade 3 (except ACOI 248), CCAC 4857, CCAC 5561/CCAC 5547 B; 9: ACOI 1621, CCAC 4854/CCAC 4855, CCAC 2279/CCAC 4234.

The four regions with sequence variability within *Microthamnion* were highlighted in different shades of green and yellow. They were given in detail in the other boxes of the figure outlined with the same color. The database sequence of *Microthamnion kuetzingianum* Z28974 served, again, as reference for the majority of the *Microthamnion* strains. Helices were labelled according to Van de Peer et al. (1999) and Wuyts et al. (2000). The regions with sequence variability found in *Microthamnion* were in congruence with the variable regions V2, V3, V4 and V9 according to Lee & Gutell (2012).

All restriction enzyme sites within the variable regions were marked in the figure: those with cuts in all *Microthamnion* strains in grey, those that lead to an additional cut in green, and those resulting in a cut prevented in red. The recognition sites of the latter two were highlighted by a purple background. The nucleotide positions in which different character states occurred were depicted in larger blue font.

There was a total of seven positions in the *Microthamnion* 18S molecule, where sequence variability among strains lead to altered recognition sites and thus to additional or prevented cuts: three in V2 (helix E23-2) and V4 (helices E10_1 and 11) each, and one in V3 (helix 17). The consequent effect on fragment lengths, and thus the pattern of DNA fragments after *in silico* gel electrophoresis, is given in Figure 9. All other positions with changes in character states without effect on the number of cuts were marked by encircled numbers (details of strain information are given in the legend of Figure 8).

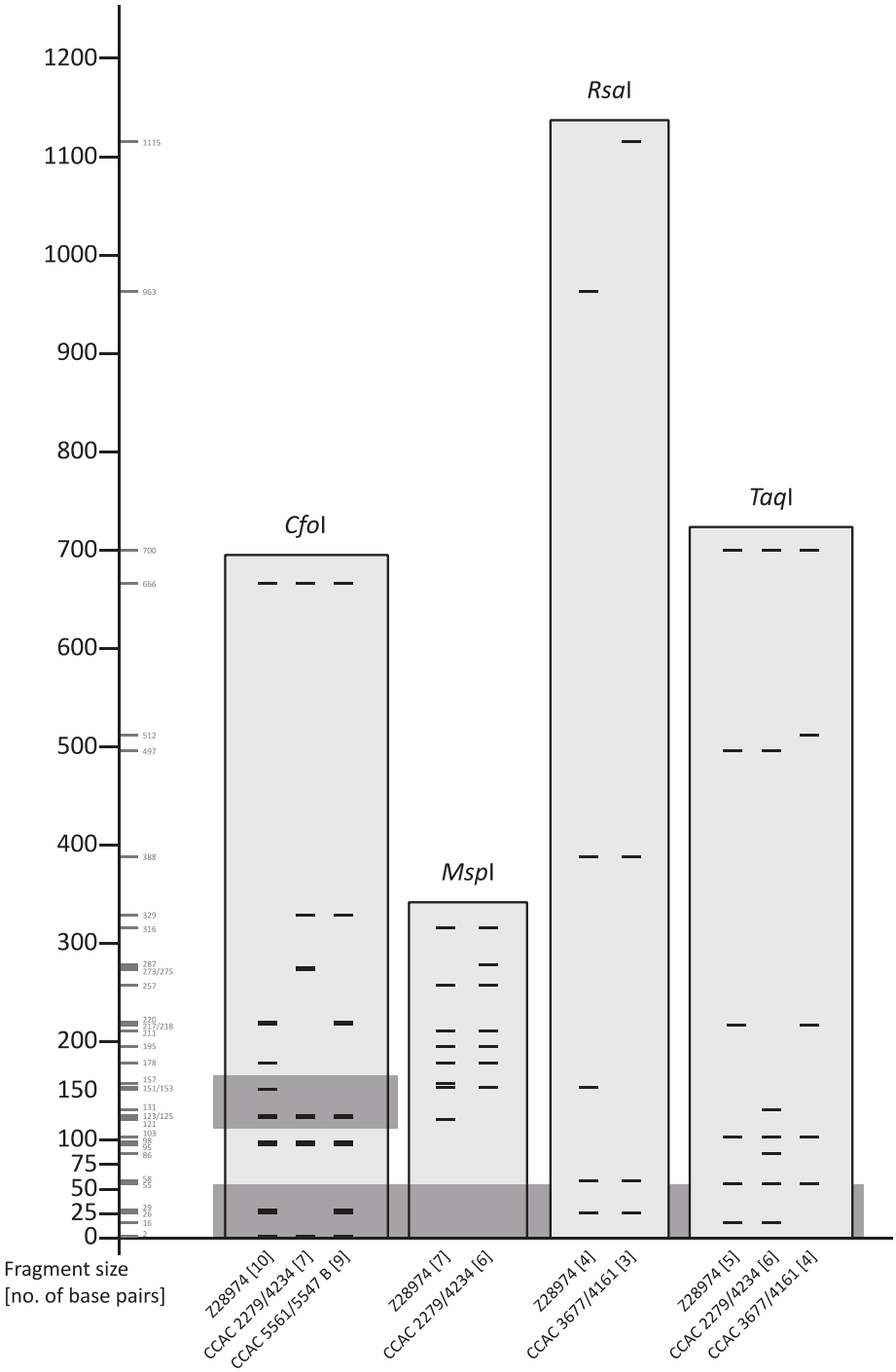


Figure 9. *In silico* gel electrophoresis analysis of digested *Microthamnion* 18S data
Hypothetical gel showing RFLP-patterns of strains CCAC 2279/CCAC 4234, CCAC 3677/CCAC 4161 and CCAC 5561/CCAC 5547 B, when digested *in silico* with enzymes *Cfo*I, *Msp*I, *Rsa*I and *Taq*I. The database sequence of *Microthamnion kuetzingianum* Z28974 is given as reference for all other *Microthamnion* strains. The total number of cutting sites per enzyme follows the strain names in squared brackets. Fragment size is given on the left axis in black (exact numbers for each fragment in small grey font). The lower dark grey bar indicates the resolution limit of the acrylamide gels shown in John et al. (1993) (for *Rsa*I and *Taq*I only assumed and transferred in the background since gels were not shown in the publication). The upper dark grey bar in *Cfo*I indicates an unreadable region in the gel published in John et al. (1993).

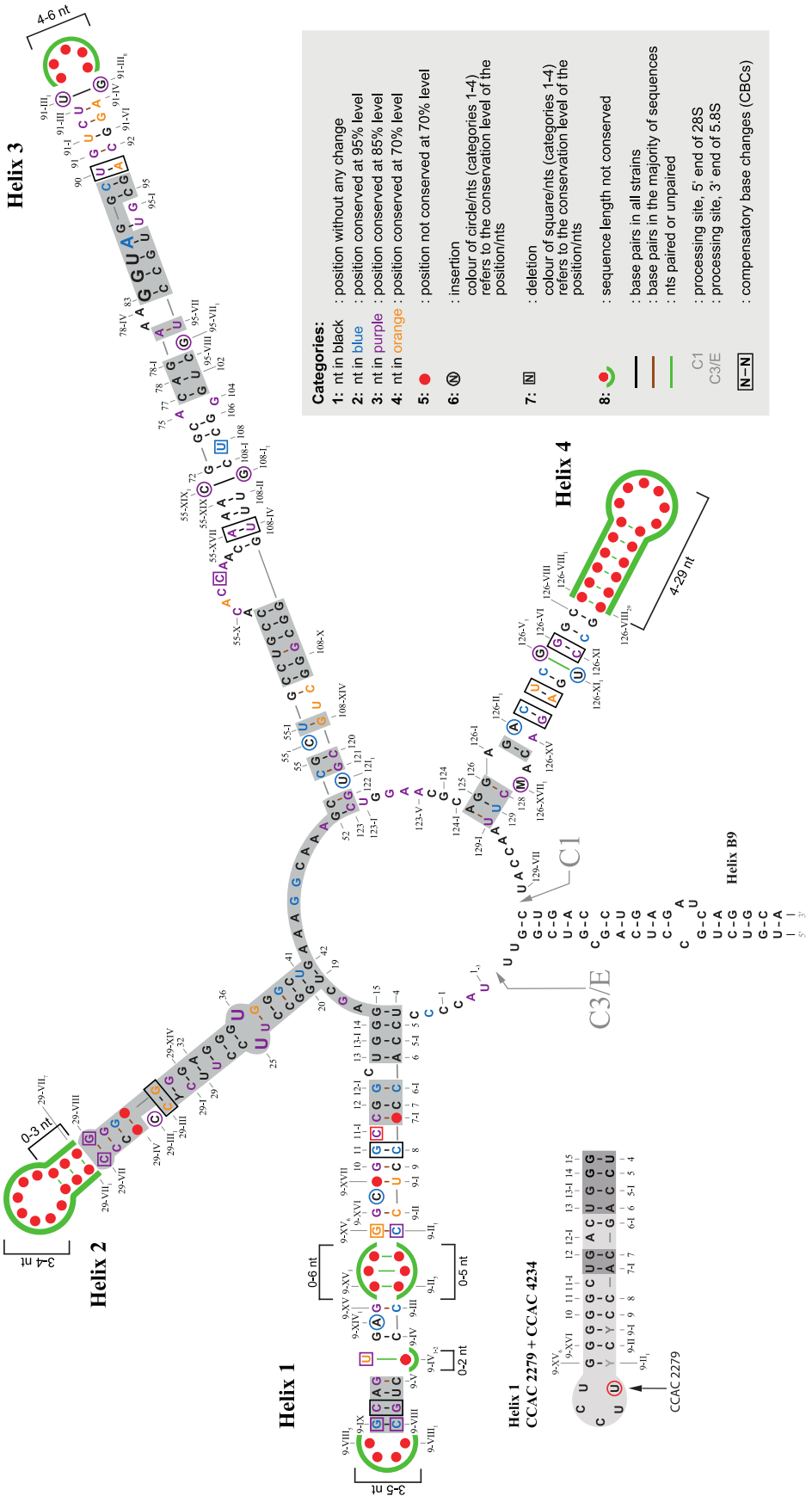
3.2.2. ITS2 Secondary Structure Models

In a total number of 56 *Microthamnion* strains, the ITS2 molecule was sequenced as described above. 37 of these strains featured a unique ITS2 sequence and were incorporated in the general ITS2 secondary structure model for *Microthamnion* given in Figure 10. A figure template was obtained from Reder (2015), updated and reworked according to new information gained in the present study.

For comparison, the secondary structure model with conservation levels adjusted to all 56 *Microthamnion* strains as well as one adjusted to a reduced number of 45 strains (i.e. two identical sequences were only allowed when the strains originated from different natural samples) are given in Supplementary Figure 8 and Supplementary Figure 9 on pages 104 and 105.

A variability map was constructed to depict the conservation levels of the nucleotides within the ITS2 molecule as well as the B9 helix, which is formed by the 3' end of the 5.8S and the 5' end of the 28S rRNA molecules. A position without any change was represented by a black letter, those positions with a conservation at the 95% level were depicted in blue. Nucleotides conserved at the 85% level were given in purple and those conserved at the 70% level in orange. Positions not conserved at the 70% level were displayed as red filled circles. An additional framing thick green line was applied where length variabilities occurred. Insertions were given as encircled letters, deletions as letters surrounded by a box. Here, the color of the surrounding circle/box referred to the conservation of the position, whereas the color of the letter reflected the character state of the inserted/deleted nucleotide. For example, a purple U with a blue circle corresponded to a position where a maximum of 15% of all strains had an inserted nucleotide (= position conserved at 85% level) which was a U in 95% of these cases. A base pairing was indicated by a connecting line in either black (= base pairing in all strains) or brown (= base pairing in > 50% of the sequences). In the variable parts of the helices, the base-connecting green line stood for nucleotides that occurred either paired or unpaired. The eight positions where CBCs occurred were highlighted by black boxes.

Figure 10. Consensus secondary structure model of 37 unique ITS2 *Microthamnion* sequences
The figure spans the ITS2 molecule, consisting of four paired helices (labelled Helix 1 – Helix 4) connected by single-stranded spacers and the B9 helix (site of interaction between 5.8S and 28S). The processing sites C1 and C3/E according to Côté et al. (2002), Granneman et al. (2011) and Schillewaert et al. (2012) are depicted in grey. An alternate, very short Helix 1, unique for strains CCAC 2279 and CCAC 4234 is displayed separately. Different colors of nucleotides and boxes/circles refer to levels of conservation (details are given under "Categories" in the figure). Variable parts that differed in character states and length (range in no. of nts. given for each region, indicated by black brackets) are given as red dots with a framing green line. CBC positions are highlighted by a black box. A dark grey background highlights the positions used for the trebouxiophycean phylogeny. The characteristic U-U mismatch in H2 is accentuated in larger font and set apart from the helix outline. Positions of nucleotides are numbered according to Caisová et al. (2011a, 2013).


Helix 3
Helix 2
Helix 1
Helix 4
C1
C3/E
Helix 1
CCAC 2279 + CCAC 4234
CCAC 2279
Helix B9
4-6 nt
0-3 nt
3-4 nt
0-6 nt
0-5 nt
0-2 nt
3-5 nt
4-29 nt
Categories:
1: nt in black : position without any change
2: nt in blue : position conserved at 95% level
3: nt in purple : position conserved at 85% level
4: nt in orange : position conserved at 70% level
5: : position not conserved at 70% level
6: : insertion
colour of circle/nts (categories 1-4)
refers to the conservation level of the position/nts
7: : deletion
colour of square/nts (categories 1-4)
refers to the conservation level of the position/nts
8: : sequence length not conserved
: base pairs in all strains
: base pairs in the majority of sequences
: nts paired or unpaired
C1 : processing site, 5' end of 28S
C3/E : processing site, 3' end of 5.8S
N–N : compensatory base changes (CBCs)

A dark grey background highlighted those positions which could be unambiguously aligned in all trebouxiophycean taxa investigated in this study, and were thus incorporated in the phylogenetic analysis for the tree given in Figure 7 (p. 36).

The numbering system in *Microthamnion* followed the universal numbering system introduced in Caisová et al. (2011a, 2013). Those characters which could be unambiguously identified as homologous to those in the Ulvales were given the 'universal' Arabic numeral. Positions non-homologous to the Ulvales but homologous within *Microthamnion* were indicated by additional Roman numerals. Insertions and positions with variability in sequence length were labelled with subscript numbers. Of the 129 'universal' positions proposed for the Ulvales, 87 could be applied in *Microthamnion*.

The ITS2 secondary structure model of *Microthamnion* showed an overall high degree of conservation in the basal part of the helices. This included the first six base pairs in helix 1 (H1), the first 11 base pairs and the pyrimidine-pyrimidine mismatch (pos. 25/36) in helix 2 (H2), and the basal part of helix 4 (H4). Helix 3 (H3) showed a high level of conservation throughout the complete paired helix region up until position 91-I/91-VI. Spacers and single-stranded regions between the processing sites and H1/H4 respectively were also quite conserved in character states and 100% conserved in length.

Variations in length as well as character states did however occur in the apical loops of each helix as well as the adjacent paired regions in some of them. An internal loop (5' 9-II$_1$ to 9-II$_5$, 3' 9-XV$_1$ to 9-XV$_6$) in H1 varied not only in length, but in character states as well. There were also several other positions in this helix without conservation at the 70% level as well as some insertions and deletions. In H2 the apical loop region consisted of 3-10 nucleotides, of which up to 6 could be paired and thus extended the paired helix region. H4 was the least conserved helix regarding sequence length, with the apical part of the helix as well as the apical loop being extremely variable (4-29 nts).

The overall sequence length of the ITS2 molecule in *Microthamnion* ranged between 233 (CCAC 5561) and 257 (ACOI 1621) nucleotides. Helix 1 spanned between 40 (CCAC 4855) and 57 (ACOI 1620) nucleotides with an exception of two strains which displayed a very short H1 (CCAC 4234: 30 nts, CCAC 2279: 31 nts) which was depicted separately in Figure 10 (labelled "Helix 1 CCAC 2279 + CCAC 4234", highlighted with grey background).

Helix two had 38 (CCAC 3677/CCAC 4161 and ACOI 1621) up to 48 (SAG 114.80) nucleotides. The reduction in length in strains CCAC 3677/CCAC 4161 and ACOI 1621 was located in the helix' apical part: the base pair at position 29-VII/29-VIII was already part of the loop in these strains (thus the positions were marked as 'deletion'). The pyrimidine-pyrimidine mismatch at position 25/36 was a U-U in all *Microthamnion* strains except CCAC 3677/CCAC 4161 which had a C-C.

Helix three was the longest helix in all *Microthamnion* strains as well as the one most conserved in length. It comprised 88 or 89 nucleotides in all strains except CCAC 3677/CCAC 4161 which had 94 nucleotides each. The additional nucleotides in those two strains were an additional base pair at position 55-XIX$_1$/108-I$_1$ and one before the terminal loop (positions 91-III$_1$ and 91-III$_8$), followed by an additional single nucleotide within the terminal loop. Most of the 88/89 nucleotide strains had the length variation in the terminal loop. There were however small variations: ACOI 1621 had an insertion at position 55$_1$ and a deletion at position 108; strains CCAC 2279/CCAC 4234 had a deletion in the internal loop (position 55-XIII) and an insertion at position 95-VII$_1$; CCAC 4855 an insertion at position 121$_1$.

Helix four was the shortest helix in all *Microthamnion* strains. The majority of strains consisted of 27-30 nucleotides (stem of the helix: 23 nts) with length variation only in the terminal loop (4-7 nts). Strains CCAC 2279/CCAC 4234 had 47 nts, the additional nucleotides being located either paired at the apical part of the helix or unpaired in a larger terminal loop compared to the majority of strains. The longest sequence appeared in ACOI 1621 with 53 nucleotides. The additional bases were also located at the apical part of the helix, where they were either paired or unpaired, and in an enlarged terminal loop. Strain CCAC 4855 however comprised 36 nucleotides with the additional paired bases located in the apical part of the helix. In addition this was the only strain with an insertion in the basal part of the helix' 5'-end (position 126-II$_1$).

3.2.3. Unique Molecular Signatures for *Microthamnion* and the Microthamniales

The thorough search for non-homoplasious synapomorphies and autapomorphies (for important single-strain branches) within *Microthamnion* revealed unique characters for the four clades inferred from phylogenetic analyses, as well as the topology among them and most lineages in grade 2 (Table 11). Besides, the subdivision in clade 4 was acknowledged by several unique molecular signatures, and also in clades 1 and 2 internal branches gained NHS support. Also on genus and order levels, *Microthamnion* and the Microthamniales featured unique characters among the Trebouxiophyceae. This uniqueness could be tested against 93 trebouxiophycean taxa outside the Microthamniales in 18S, 77 in 5.8S and ITS2, and 62 in the *rbc*L. For the beginning of 28S only 23 taxa were available which even decreased to three available strains from Helix C1_E1 on. Consequently, unique 28S characters detected within *Microthamnion*, especially from Helix C1_E1 on, labelled 'NHS in Trebouxiophyceae', should thus only be regarded as confirmed within the Microthamniales, but a mere indication for possible unique features within the Trebouxiophyceae. Accordingly, the two NHSs in 28S found on genus level have to be seen with great caution, since they could only be compared with three other trebouxiophycean taxa. The two positions detected for the Microthamniales however were found in Helix B18 where 23 sequences were available and were thus more likely to effectively display unique features within the Trebouxiophyceae.

Non-homoplasious synapomorphies had to meet the following criteria: the character change for an examined group or branch in *Microthamnion* had to be conserved within, and the same change was not allowed to occur in any other *Microthamnion* group/branch. When NHSs were even unique within the order or class, they were labelled "NHS in Microthamniales" or "NHS in Trebouxiophyceae" respectively. On the genus level, character states had to be consistent within *Microthamnion* and were not permitted to occur in any other member of the Microthamniales, and consequently NHSs listed for the Microthamniales did not appear in any other trebouxiophycean taxa. Uniqueness for higher taxonomic levels were labelled accordingly.

Besides, only those NHSs were listed, which occurred in unambiguously alignable regions. *Rbc*L amino acid changes were given regardless of the specificity of side chain properties as long as they met the NHS criteria. To distinguish the latter, conservative residue changes were given with their number only, whereas changes resulting in different side chain properties were highlighted with an asterisk. Since also single strain branches in the basal part of the *Microthamnion* phylogeny were investigated, they were consequently referred to as autapomorphies (AUTs).

A single exception regarding the NHS criteria was made for residue 54 in the *rbc*L which was a hundred percent conserved among all investigated Trebouxiophyceae (glycine) and uniquely distinguished *Microthamnion* by an alanine except the dark green group of clade 4 which reversed to the original character state. It was highlighted as 'almost unique' in Table 11 (bold print). Clades and branches were given in the order of the phylogeny of Figure 6, starting from top to bottom. Also the subdivisions in clade 1 followed this system: "Strains CCAP 450/4 to CCAC 3664 B", for example, referred to the node that summarized the 12 strains from CCAP 450/4, ACOI 2656, ACOI 1447 down to strain CCAC 3664 B.

As before mentioned, all clades were defined by unique molecular characters and also the subdivision in clade 4 offered many unique features. Clades 1-3, although showing quite remarkable differences regarding their support values in the phylogenetic analyses, did not differ in number or kind of NHS. In fact, they were characterized by the exact same amino acid (residue 282) in different character states. The substitution at residue 282 was conservative for clades 1 and 2, exchanging one basic residue for another (H for R and K respectively). In clade 3 the residue substitution resulted in different amino acid properties as glutamine is uncharged polar. Residue 282 was apparently not fixed to certain properties since it also displayed hydrophobic (isoleucine in branch of strains CCAC 3677/3838/4161) and acidic (asparagine in CCAC 5561/5547 B and CCAC 4853/4854/4855) variants.

NHSs for subdivisions did however occur in clades 1 and 2 which were exclusively positioned in the ITS2 molecule. In clade 1 the branch summarizing strains CCAP 450/4 to CCAP 450/3 (highlighted in Figure 6 by an encircled 1) was characterized by a CBC in H4, and also the next ascending (hemi CBC in H3) as well as the next descending node (unique mismatch pattern in H2) offered unique features. The single strain

CCAC 2942 B was also characterized by a distinctive spacer motif between H1 and H2. In clade 2 NHSs only occurred at two terminal branches, one encompassing the genetically identical strains CCAC 2764 B/CCAC 2804 B which originated from different sampling locations. The other was the node of strains M 2412/1 A and UTEX 318, but as no ITS2 sequence data for UTEX 318 was available, unique signatures could only be documented for strain 2412/1 A and are thus termed 'AUT' in Table 11.

The branch summarizing clades 1-3 together was again characterized by a residue substitution (Cys-284 to Ser-284) in the *rbc*L, resulting in different side chain properties (hydrophobic to uncharged polar). Clade 4, being further apart from clades 1-3, offered a single residue substitution (no. 248) which had no impact on amino acid properties. The subclades however, and here especially the dark green group, displayed several NHSs each. Interestingly, for ACOI 1620 (light green) they were positioned in the conserved SSU and Helix B8 of 5.8S, albeit in variable regions (i.e. a CBC in V4 of 18S and a loop motif in B8), whereas unique characters for the dark green group were found exclusively in the ITS2 molecule (mismatch) and the *rbc*L (two conservative residue changes and two with different properties).

There were no NHSs found for the individual lineages in grade 1. All branches displayed synapomorphies grouping their individual strains though, but these were shared with other lineages of grades 1 and 2 and thus not unique (and not listed). Although the branch for strains CCAC 4856/4857 offered no NHS in itself, two unique features in 28S, one CBC and one distinct mismatch pattern, characterized the node establishing their sister relationship with clades 1 to 4.

All branches in grade 2 except that of CCAC 2916, which was assigned to grade 1 in the tree inferred from phenetic molecular characters (see Figure 11, p. 57), displayed an astonishing number of unique characters for their members. NHSs to group branches together were albeit rarely found: only one hemi CBC in H3 of ITS2 supported the sister relationship between groups CCAC 2279/4234 and CCAC 3677/4161. With the topology unresolved, this result did not come as a surprise. That is why also the alternate topology for grades 1 and 2, as proposed in the tree based on phenotypic characters (Figure 11), was checked for unique synapomorphies in order to clarify the relationship among groups. But also this alternative evolutionary scenario had no hits regarding synapomorphy support.

Table 11. Unique molecular signatures within *Microthamnion* and the Microthamniales
Listed are all unique synapomorphies detected in the nuclear-encoded rRNA operon and *rbc*L protein data which refer to ingroup branches in the *Microthamnion* phylogeny of Figure 6 (marked as '**NHS**' = **n**on-**h**omoplasious **s**ynapomorphies or **'AUT'** = **aut**apomorphies). Besides, also unique signatures on genus (*Microthamnion*) and order (Microthamniales) levels are given. A single exception regarding NHS criteria was made for residue 54 in the *rbc*L which defined *Microthamnion* except for the dark green group of clade 4 (**'almost unique'** in bold fond). Clades (incl. branches in subdivisions) and branches of grades 1 and 2 are given in order of their appearance in the phylogeny of Figure 6, starting from top to bottom. The subdivision in Clade 1 marked with two asterisks (**) refers to branch ① in the *Microthamnion* phylogeny. Positions in 18S, 5.8S and 28S are labelled according to Van de Peer et al. (1999) and Wuyts et al. (2000), positions in ITS2 according to Caisová et al. (2011a, 2013). Amino acid positions (IUPAC code used) are given in reference to the database sequence of UTEX 318 (accession number KM462876). In AAs labelled with an asterisk (*) the change resulted in different side chain properties. The number of investigated trebouxiophycean outgroup sequences for the different markers are as follows: 18S ⇒ 93 plus *Characium* clade (6 sequences), 5.8S/ITS2 ⇒ 77+ *Ch. perforatum*, 28S ⇒ 23 + *Ch. perforatum*. (only 3 from C1_E1 on), *rbc*L ⇒ 62 + *Ch. perforatum*.

Taxon/character	Evolutionary change	Characterization
Clade 1		
rbc*L - AA 282	H ⇒ **R**	- **NHS** in Microthamniales - R also in *Prasiolopsis* sp. - H in clade 4 + CCAC 4856/4857 + CCAC 5545/5520 B + CCAC 5521/5585 + CCAC 2916 + ACOI 1621 + CCAC 2279/4234 and majority of Trebouxiophyceae - N in CCAC 5561/5547 B + CCAC 4853/4854/4855 + *Characium perforatum*; also in one branch of the Oocystaceae (*Planctonemopsis distans*, *Amphikrikos nanus*, *Oocystis* sp. KMMCC 356, *Oocystidium* sp. SAG 81.80, *Oocystidium planoconvexum*) and *Botryococcus braunii* - K in *Planctonema lauterbornii* and *Tetrastrum staurogeniaeforme* - Q in *Dictyochloropsis asterochloroides* and *Xylochloris irregularis* - I in *Parachloroidium lobatum*
Subdivisions in Clade 1		
Strains CCAP 450/4 to CCAC 3664 B		
ITS2 - H3 bp 23 (78-II/95-VII)	A-U ⇒ **G**-U	- **NHS** in Microthamniales - AxC in CCAC 2916 and ACOI 1620 - C-G in majority of Trebouxiophyceae
Strains CCAP 450/4 to CCAP 450/3**		
ITS2 - H4 3rd bp before term. Loop (126-VI/126-XI)	G-C ⇒ **A-U**	- **NHS** in Microthamniales - G-U in ACOI 1621 - AxC in *Characium perforatum* - position not unambiguously alignable in other Trebouxiophyceae
Strains CCAP 450/4 to CCAC 3843 B		
ITS2 - H2 bp 6 (24/37)	U-G ⇒ Ux**U**	- **NHS** in Microthamniales - CxU in CCAC 2764 B and CCAC 2804 B - UxC in CCAC 3677/3838/4161 and *Characium perforatum* - G-C in CCAC 4853/4854/4855 and majority of Trebouxiophyceae
CCAC 2942 B		
ITS2 - Spacer H1-H2	AGC ⇒ A**U**C	- **AUT** in Microthamniales - AAC in M 2412/1 A - position variable in other Trebouxiophyceae
Clade 2		
rbc*L - AA 282	H ⇒ **K**	- **NHS** in Microthamniales - **K** also in *Planctonema lauterbornii* and *Tetrastrum staurogeniaeforme* - H in clade 4 + CCAC 4856/4857 + CCAC 5545/5520 B + CCAC 521/5585 + CCAC 2916 + ACOI 1621 + CCAC 2279/4234 and majority of Trebouxiophyceae - N in CCAC 5561/5547 B + CCAC 4853/4854/4855 + *Characium perforatum*; also in one branch of the Oocystaceae (*Planctonemopsis distans*, *Amphikrikos nanus*, *Oocystis* sp. KMMCC 356, *Oocystidium* sp. SAG 81.80, *Oocystidium planoconvexum*) and *Botryococcus braunii*

Taxon/character	Evolutionary change	Characterization
***rbc*L - AA 282* (continued)**	H ⇒ **K**	- R in *Prasiolopsis* sp. - Q in *Dictyochloropsis asterochloroides* amd *Xylochloris irregularis* - I in *Parachloroidium lobatum*
Subdivisions in Clade 2		
CCAC 2764 B/CCAC 2804 B		
ITS2 - H2 bp 6 (24/37)	U-G ⇒ **CxU**	- **NHS** in Microthamniales - UxU in strains CCAP 450/4 to CCAC 3843 B - UxC in CCAC 3677/3838/4161 and *Characium perforatum* - G-C in CCAC 4853/4854/4855 and majority of Trebouxiophyceae
ITS2 - H4 bp 3 (126/128)	G-C ⇒ G-U	- **NHS** in Microthamniales - GxG in CCAC 5561/CCAC 5547 B - A-U in most other Trebouxiophyceae
ITS2 - H4 2nd bp before term. Loop (126-VII/126-X)	G-C ⇒ G-U	- **NHS** in Microthamniales - U-G in *Characium perforatum* - position not unambiguously alignable in other Trebouxiophyceae
M 2412/1 A		
ITS2 - Spacer H1-H2	AGC ⇒ A**A**C	- **AUT** in Microthamniales - AUC in CCAC 2942 B - position variable in other Trebouxiophyceae
ITS2 - H3 bp 27(86/95-III)	A-U ⇒ **G**-U	- **AUT** in Microthamniales - A is conserved in almost all Trebouxiophyceae. Exceptions: both *Symbiochloris symbiontica* strains (⇒ C), *Symbiochloris* sp. and *Lobosphaera incisa* SAG 2466 (both ⇒ G)
Clade 3		
rbc*L - AA 282	H ⇒ **Q**	- **NHS** in Microthamniales - **Q** also in *Dictyochloropsis asterochloroides* and *Xylochloris irregularis* - H in clade 4 + CCAC 4856/4857 + CCAC 5545/5520 B + CCAC 5521/5585 + CCAC 2916 + ACOI 1621 + CCAC 2279/4234 and majority of Trebouxiophyceae - R in *Prasiolopsis* sp. - K in *Planctonema lauterbornii* and *Tetrastrum staurogeniaeforme* - N in CCAC 5561/5547 B + CCAC 4853/4854/4855 + *Characium perforatum*; also in one branch of the Oocystaceae (*Planctonemopsis distans*, *Amphikrikos nanus*, *Oocystis* sp. KMMCC 356, *Oocystidium* sp. SAG 81.80, *Oocystidium planoconvexum*) and *Botryococcus braunii* - I in *Parachloroidium lobatum*
Clade 1-3		
rbc*L - AA 284	C ⇒ **S**	- **NHS** in Trebouxiophyceae - C in *Characium perforatum* and all other Trebouxiophyceae
Clade 4		
***rbc*L - AA 248**	E ⇒ **D**	- **NHS** in Microthamniales - **D** also in *Planctonema lauterbornii*, *Stichococcus bacillaris* UTEX 176, *Koliella corcontica*, *Lobosphaera incisa* CAUP H 4301 and SAG 2007, *Leptochlorella corticola*, Trebouxiophyceae sp. MX-AZ01, *Kalinella apyrenoidosa*, *Phyllosiphon arisari*, *Viridiella fridericiana*, *Dictyochloropsis splendida* UTEX LB 2599, *Symbiochloris handae*, *Symbiochloris* sp. - E in all other Trebouxiophyceae
Clade 4 (light green)		
18S - Helix E10_1 last bp before terminal loop	G-C ⇒ **A-U**	- **AUT** in Microthamniales - G-C in majority of Trebouxiophyceae - **A-U** also in *Coccomyxa galuniae*, *Coccomyxa onubensis*, *Coccomyxa viridis*, *Elliptochloris bilobata* SAG 245.80, *Hemichloris antarctica*, *Chloroidium ellipsoideum*, *Chloroidium angustoellipsoideum*, *Chlorocloster engadinensis*, *Chloroidium saccharophilum* - C-G in *Quadricoccus ellipticus*, *Amphikrikos nanus*, *Oocystis* sp., *Oocystidium planoconvexum*, *Franceia amphitricha*, *Oocystella oogama*

Taxon/character	Evolutionary change	Characterization
5.8S - Loop of Helix B8 1st nt	C ⇒ **U** (CUCG ⇒ **U**UCG)	- **AUT** in Microthamniales - **U** also in majority of Trebouxiophyceae - C in *Oocystella oogama*, *Auxenochlorella protothecoides* var. *acidicola*, the Prasiola-clade, *Neocystis brevis*, *Lunachloris lukesovae*, the Geminella clade, *Koliella corcontica*, the Trebouxiales, *Dictyosphaerium minutum*, *Dictyochloropsis asterochloroides*, *Dictyochloropsis splendida* SAG 2097, *Choricystis* sp. TP-2009, *Coccomyxa viridis*, *Hemichloris antarctica*, *Heveochlorella roystonensis*, *Heveochlorella hainangensis*, *Heterochlorella luteoviridis*, *Viridiella fridericiana*, *Symbiochloris handae*, *Symbiochloris symbiontica* SAG 27.81, *Apatococcus* sp.
Clade 4 (dark green)		
ITS2 - H3 bp 5 (55-I/108-XIV)	U-G ⇒ **XxX** **(UxC/UxU)**	- mismatch **NHS** in Microthamniales - **UxC** in CCAP 450/2 - **UxU** in CAUP J 1201, CCAC 8002 B, CCAC 4822 - C-G in CCAC 4853/4854/4855 and *Characium perforatum* - U-A in CCAC 3677/3838/4161 and CCAC 2279/4234 - **mismatches** also in other Trebouxiophyceae
***rbc*L - AA 54**	A ⇒ **G**	- **NHS** in *Microthamnion* - **G** also in *Characium perforatum* and all other Trebouxiophyceae
***rbc*L - AA 235**	I ⇒ **V**	- **NHS** in Microthamniales - **V** also in Watanabea-clade, one branch in Choricystis/Botryococcus clade (*Coccomyxa subellipsoidea*, Trebouxiophyceae sp. MX-AZ01, *Paradoxia multiseta*, *Coccomyxa simplex* SAG 216-6, *Pseudococcomyxa simplex* CAUP H 102), one branch in the Oocystaceae (*Planctonemopsis distans*, *Amphikrikos nanus*, *Oocystis* sp. KMMCC 356, *Oocystidium* sp. SAG 81.80, *Oocystidium planoconvexum*) and *Tetrastrum staurogeniaeforme* - I in all other Trebouxiophyceae
rbc*L - AA 287	N ⇒ **H**	- **NHS** in *Microthamnion* - **H** also in *Characium perforatum* and *Ettlia pseudoalveolaris* - N, H or E in other Trebouxiophyceae
rbc*L - AA 428	T ⇒ **I**	- **NHS** in Microthamniales - V in *Characium perforatum* - **I** also in *Neocystis brevis*, *Lobosphaera incisa* SAG 2007, *Paradoxia multiseta*, *Kalinella apyrenoidosa* - T or V in all other Trebouxiophyceae
Clade 1-4 and CCAC 4856/4857		
28S - Helix C1_E3 bp 9	X-X ⇒ **GxG**	- **NHS** in Trebouxiophyceae - C-G in CCAC 5545/5520 B, CCAC 5521/5585, CCAC 2916, ACOI 1621, CCAC 5561, CCAC 3677/4161 - T-G in CCAC 2279/4234 - T-A in CCAC 4853/4854/4855 - G-T in *Characium perforatum* - G-C in other Trebouxiophyceae
28S - Helix C1_E3 bp 10	G-C ⇒ **C-G**	- **NHS** in Trebouxiophyceae - G-C in *Caharacium perforatum* and *Eremosphaera viridis* - G-U in *Leptosira terrestris*, *Stichococcus bacillaris*
ACOI 1621		
5.8S - Spacer Helix B9-ITS2 H1 nt 5	C ⇒ **U** (UACCCC ⇒ (UACC**U**C)	- **AUT** in Trebouxiophyceae - UACUCC in *Characium perforatum* - UACCCC also in Trebouxiales
ITS2 - H3 bp 3 (54/121)	C-G ⇒ **U-G**	- **AUT** in Microthamniales - CxA in CCAC 3677/3838/4161 and NIES-479 - **U-G** also in *Oocystidium* sp., *Franceia amphitricha*, *Oocystella oogama*, *Marvania geminata*, *Asterochloris phycobiontica*, *Asterochloris gaertneri* - C-G in majority of Trebouxiophyceae
ITS2 - H3 4th bp before terminal loop (91/92)	G-C ⇒ **U-G**	- **AUT** in Microthamniales - UxU in CCAC 2279/4234 - position not unambiguously alignable in other Trebouxiophyceae

Taxon/character	Evolutionary change	Characterization
ITS2 - H4 H4 3rd bp before terminal loop (126-VI/126-XI)	G-C ⇒ **G-U**	- **AUT** in Microthamniales - A-U in strains CCAP 450/4 to CCAP 450/3 - AxC in *Characium perforatum* - position not unambiguously alignable in other Trebouxiophyceae
28S - Helix C1_E1 6th bp before terminal loop	U-G ⇒ **C-G**	- **AUT** in Trebouxiophyceae - U-G in all other taxa
CCAC 5561/5547 B		
ITS2 - H1 bp 8 (8/11)	C-G ⇒ **G-C**	- **NHS** in *Microthamnion* - position not present in *Characium perforatum* - position not unambiguously alignable in other Trebouxiophyceae
ITS2 - H4 bp 3 (126/128)	G-C ⇒ **GxG**	- **NHS** in Microthamniales - G-U in CCAC 2764 B and CCAC 2804 B - position paired in most Trebouxiophyceae
28S - Spacer Helices B12'-B13 nt 3	U ⇒ **C** (AGU ⇒ AG**C**)	- **NHS** in Trebouxiophyceae - U in *Characium perforatum* - A in all other Trebouxiophyceae
28S - Helix C1_E3 bp 8	G-C ⇒ **GxG**	- **NHS** in Microthamniales - C-G in CCAC 2279/4234, *Characium perforatum* and *Leptosira terrestris* SAG 463-3 - G-C in *Eremosphaera viridis* and *Stichococcus bacillaris* SAG 379-2
rbc*L - AA 50	P ⇒ **A**	- **NHS** in Microthamniales - S in CCAC 2279/4234 - P majority of Trebouxiophyceae
CCAC 3677/3838/4161		
18S - Loop of Helix 17 nts 3 and 4	UUUC ⇒ UU**AU**	- **NHS** in Trebouxiophyceae - UUUC in majority of Trebouxiophyceae
5.8S - Spacer Helix B9-ITS2 H1 nts 1+2	UA ⇒ **AC** (UACCCC) ⇒ (**AC**CCCC)	- **NHS** in Trebouxiophyceae - UACUCC in *Characium perforatum* - UACCCC also in Trebouxiales
ITS2 - H2 bp 6 (24/37)	U-G ⇒ Ux**C**	- **NHS** in *Microthamnion* - UxC also in *Characium perforatum* - UxU in strains CCAP 450/4 to CCAC 3843 B - CxU in CCAC 2764 B and CCAC 2804 B - G-C in CCAC 4853/4854/4855 and in majority of Trebouxiophyceae
ITS2 - H2 "bp 6.1" (pyrimidine-pyrimidine-mismatch)	UxU ⇒ **CxC**	- **NHS** in Microthamniales - **CxC** also in one branch of the Chlorellales (*Parachlorella kessleri*, *Compactochlorella kochii*, *Mucidosphaerium palustre*, *Dicloster acuatus*, *Kalenjinia gelatinosa*), both *Symbiochloris symbiontica* strains and *Chloropyrula uraliensis*
ITS2 - H2 4th bp before terminal loop (29-III/29-XII)	C-G ⇒ **U-A**	- **NHS** in Microthamniales - G-C in CCAC 2279/4234, *Characium* perforatum and *Edaphochlorella mirabilis* - **U-A** also in *Didymogenes palatina*, *Meyerella planktonica*, *Dicloster acuatus*, *Kalenjinia gelatinosa*, *Gloeotila scopulina* SAG 335-8, *Stichococcus bacillaris* SAG 379-2, *Neocystis brevis*, *Coccomyxa dispar*
ITS2 - H3 5th bp before terminal loop (90/93)	U-A ⇒ **U-G**	- **NHS** in *Microthamnion* - **U-G** also in *Characium perforatum* - C-G in CCAC 2279/4234, CCAC 2916 and ACOI 1621 - position not unambiguously alignable in other Trebouxiophyceae
ITS2 - Spacer H3-H4 nts 1, 3, 4, 5	UGGAACGC ⇒ **C**G**UCC**CGC	- **NHS** in Microthamniales - GACCUGC in *Characium perforatum* - position not unambiguously alignable in other Trebouxiophyceae
28S - Helix D14 last bp	GxG ⇒ G-**U**	- **NHS** in *Microthamnion* - **U** also in *Characium perforatum* - other Trebouxiophyceae unpaired
***rbc*L - AA 270**	L ⇒ **I**	- **NHS** in Microthamniales - I also in Oocystaceae - L in all other Trebouxiophyceae

Taxon/character	Evolutionary change	Characterization
rbc*L - AA 282	H ⇒ **I**	- **NHS** in Microthamniales - **I** also in *Parachloroidium lobatum* - H in clade 4 + CCAC 4856/4857 + CCAC 5545/5520 B + CCAC 5521/5585 + CCAC 2916 + ACOI 1621 + CCAC 2279/4234 and majority of Trebouxiophyceae - N in CCAC 5561/5547 B + CCAC 4853/4854/4855 + *Characium perforatum*; also in one branch of the Oocystaceae (*Planctonemopsis distans*, *Amphikrikos nanus*, *Oocystis* sp. KMMCC 356, *Oocystidium* sp. SAG 81.80, *Oocystidium planoconvexum*) and *Botryococcus braunii* - R in *Prasiolopsis* sp. - K in *Planctonema lauterbornii* and *Tetrastrum staurogeniaeforme* - Q in *Dictyochloropsis asterochloroides* amd *Xylochloris irregularis*
CCAC 2279/CCAC 4234		
18S - Helix E10_1 2nd bp before terminal loop	G-C ⇒ **A-U**	- **NHS** in Microthamniales - G-C in majority of Trebouxiophyceae - **A-U** also in *Coccomyxa dispar* and one branch of the Watanabea-clade (*Chloroidium ellipsoideum*, *Chloroidium angustoellipsoideum*, *Chlorocloster engadinensis*) - C-G in *Symbiochloris handae* and *Symbiochloris* sp. - U-A in *Dictyochloropsis splendida* UTEX 2612
18S - Helix E23_2 bp 1	C-G ⇒ **UxU**	- **NHS** in *Microthamnion* - A-U in *Characium* clade - paired in almost all Trebouxiophyceae (C-G or U-A; A-U only in both *Symbiochloris symbiontica* strains) - UxU also in *Dictyosphaerium minutum*, Uncultured Trebouxiophyceae clone QE17, *Dictyochloropsis asterochloroide*s, *Dictyochloropsis splendida* SAG 2097, *Apatococcus* sp.
18S - Spacer Helices E10_1-11 nt 3	C ⇒ **U** (CU**C**C-CU**U**C)	- **NHS** in *Microthamnion* - **U** also in *Characium* clade and most other Trebouxiophyceae - C in one branch of *Choricystis*/*Botryococcus* clade (*Coccomyxa simplex*, *Pseudococcomyxa simplex*, *Coccomyxa galuniae*, *Coccomyxa dispar*, *Coccomyxa onubensis*, *Coccomyxa viridis*) and *Symbiochloris handae*
ITS2 - H2 4th bp before terminal loop (29-III/29-XII)	C-G ⇒ **G-C**	- **NHS** in *Microthamnion* - **G-C** also in *Characium perforatum* and *Edaphochlorella mirabilis* - U-A in CCAC 3677/3838/4161, *Didymogenes palatina*, *Meyerella planktonica*, *Dicloster acuatus*, *Kalenjinia gelatinosa*, *Gloeotila scopulina* SAG 335-8, *Stichococcus bacillaris* SAG 379-2, *Neocystis brevis*, *Coccomyxa dispar*
ITS2 - H3 bp 2 (53/122)	C-G ⇒ CxA	- **NHS** in Microthamniales - G-C in majority of Trebouxiophyceae - mismatch only in *Heveochlorella roystonensis* [U-U]
ITS2 - H3 4th bp before terminal loop (91/92)	G-C ⇒ UxU	- **NHS** in Microthamniales - U-G in ACOI 1621 - position not unambiguously alignable in other Trebouxiophyceae
28S - Helix C1_E3 **3**rd bp before terminal loop	C-G ⇒ **U-G**	- **NHS** in Microthamniales - **U-G** also in *Eremosphaera viridis* - G-C in CCAC 3677/4161 and CCAC 5561, *Leptosira terrestris* - G-U in *Stichococcus bacillaris*
28S - Helix C1_E3 bp 6	U-G ⇒ **G-U**	- **NHS** in Trebouxiophyceae - U-G in *Characium perforatum* - C-G in CCAC 5545/5520 B, CCAC 5521/5585, CCAC 2916, ACOI 1621 - G-C in *Leptosira terrestris*, *Stichococcus bacillaris* and *Eremosphaera viridis*
28S - Helix C1_E3 bp 8	G-C ⇒ **C-G**	- **NHS** in *Microthamnion* - GxG in CCAC 5561/5547 B - **C-G** also in *Characium perforatum* and *Leptosira terrestris* SAG 463-3 - G-C also in *Eremosphaera viridis* and *Stichococcus bacillaris* SAG 379-2
28S - Helix D22 bp 6	G-C ⇒ **A-U**	- **NHS** in Microthamniales
***rbc*L - AA 46**	P ⇒ **A**	- **NHS** in Microthamniales - P in *Characium perforatum* and *Ettlia pseudoalveolaris*

Taxon/character	Evolutionary change	Characterization
rbc*L - AA 50	P ⇒ **S**	- **NHS** in Microthamniales - A in CCAC 5561/5547 B - P majority of Trebouxiophyceae
***rbc*L - AA 87**	L ⇒ **I**	- **NHS** in Microthamniales - I also in majority of Trebouxiophyceae - L in *Tetrastrum staurogeniaeforme*, *Planctonema lauterbornii*, *Neocystis brevis*, *Lobosphaera incisa* CAUP H 4301 and SAG 2007, *Pseudococcomyxa simplex*
***rbc*L - AA 259**	D ⇒ **E**	- **NHS** in Microthamniales - variable in Trebouxiophyceae
CCAC 3677/3838/4161 + CCAC 2279/CCAC 4234		
ITS2 - H3 bp 5 (55-I/108-XIV)	U-G ⇒ **U-A**	- **NHS** in Microthamniales - C-G in CCAC 4853/4854/4855 - UxC in CCAP 450/2 - UxU CAUP J 1201, CCAC 8002 B, CCAC 4822 - **U-A** also in *Eremosphaera viridis*, *Chlorella vulgaris*, *Didymogenes palatina*, *Auxenochlorella protothecoides* var. *acidicola*, *Micractinium pusillum* SAG 8.93, *Auxenochlorella protothecoides* CCAP 211/11I, *Pabia signiensis*, *Koliella sempervirens*, *Neocystis brevis*, *Lunachloris lukesovae*
CCAC 4853/4854/4855		
18S - Helix E23_1 bp 4	U-A ⇒ **U-G**	- **NHS** in Microthamniales - **U-G** also in *Eremosphaera viridis*, *Oocystis solitaria*, *Tetrastrum staurogeniaeforme* and *Xylochloris irregularis* - U-A or C-G in majority of Trebouxiophyceae
ITS2 - H2 bp 6 (24/37)	U-G ⇒ **G-C**	**- NHS i**n Microthamniales - UxU in strains CCAP 450/4 to CCAC 3843 B - CxU in CCAC 2764 B and CCAC 2804 B - UxC in CCAC 3677/3838/4161 and *Characium perforatum* - **G-C** also in majority of Trebouxiophyceae
ITS2 - H3 bp 5 (55-I/108-XIV)	U-G ⇒ **C-G**	**NHS** in *Microthamnion* - UxC in CCAP 450/2 - UxU in CAUP J 1201, CCAC 8002 B, CCAC 4822 - U-A in CCAC 3677/3838/4161 and CCAC 2279/4234 - **C-G** also in *Characium perforatum*, *Franceia amphitricha*, *Oocystella oogama*, *Asterochloris phycobiontica*, *Asterochloris gaertneri*, *Coccomyxa dispar*, *Coccomyxa onubensis*, *Hemichloris antarctica*, *Parachloroidium lobatum*, *Watanabea reniformis*, *Symbiochloris handae*, *Symbiochloris* sp., *Symbiochloris symbiontica* SAG 27.81 and SAG 2070
ITS2 - H4 bp 1 (124-II/129-I)	A-U ⇒ **AxA**	- **NHS** in Microthamniales - AxC in CCAC 4819, CCAC 3546, CCAC 4818, ACOI 140 - G-C or G-U in majority of Trebouxiophyceae
ITS2 - H4 bp 2 (125/129)	G-U ⇒ **G-C**	- **NHS** in Microthamniales - majority of other Trebouxiophyceae G-U or G-C
ITS2 - H4 6th bp before terminal loop (126-III/126-XIV)	C-G ⇒ **A-U**	- **NHS** in Microthamniales - G-C in *Characium perforatum* - position not unambiguously alignable in other Trebouxiophyceae
ITS2 - H4 5th bp before term. Loop (126-IV/126-XIII)	U-A ⇒ **G-C**	- **NHS** in Microthamniales - C-G in clade 2 (except CCALA 368 [U-G] and M 2412 1 A [U-A]) and *Characium perforatum* - U-G in CCALA 368 - position not unambiguously alignable in other Trebouxiophyceae
28S - Helix B14 last bp before terminal loop	C-G ⇒ **U-A**	- **NHS** in *Microthamnion* - **U-A** also in *Characium perforatum* - C-G in all other Trebouxiophyceae
28S - Helix B17 bp 6	A-U ⇒ **G-C**	- **NHS** in Microthamniales - variable in other Trebouxiophyceae
28S - Helix B17 bp 7	G-U ⇒ **G-C**	- **NHS** in Microthamniales - variable in other Trebouxiophyceae

Taxon/character	Evolutionary change	Characterization
28S - Helix C1_E1 bp 4	GxG ⇒ **C-G**	- **NHS** in Microthamniales - C-G also in all other Trebouxiophyceae
28S - Helix C1_E1 bp 5	U-A ⇒ **C-G**	- **NHS** in Microthamniales - **C-G** also in all other Trebouxiophyceae
28S - Loop Helix C1_E1 nt 3	C ⇒ **G** (GCCA ⇒ GC**G**A)	- **NHS** in Microthamniales - **G** also in *Eremosphaera viridis* - A in *Leptosira terrestris*
28S - Helix C1_E3 bp 7	U-A ⇒ **CxU**	- **NHS** in Trebouxiophyceae - U-G in CCAC 2916 and *Characium perforatum* - GxG in CCAC 2279/4234 - G-C in CCAC 5561 - paired in Trebouxiophyceae
28S - Helix C1_E3 bp 12	C-G ⇒ **A-U**	- **NHS** in Microthamniales - **A-U** also in *Leptosira terrestris*
rbc*L - AA 230	A ⇒ **S**	- **NHS** in Microthamniales - **S** also in *Edaphochlorella mirabilis*, *Lunachloris lukesovae*, *Dictyochloropsis asterochloroides*, *Dictyochloropsis splendida*, *Chlorocloster engadinensis*, *Parachloroidium lobatum*, *Symbiochloris* sp. SAG 244.80 - A in majority of Trebouxiophyceae
Microthamnion		
18S - Helix 46 bp 6	G-U ⇒ **A-U**	- **NHS** in Microthamniales - G-U or G-C in most Trebouxiophyceae - A-U also in *Quadricoccus ellipticus*, *Franceia amphitricha*, *Micractinium pusillum* SAG 8.93, *Gloeotila scopulina* SAG 335-8, *Stichococcus bacillaris* SAG 379-2 and CCAP 379/7, *Pabia signiensis*, *Koliella sempervirens*, *Edaphochlorella mirabilis*, *Lunachloris lukesovae*, *Lobosphaera incisa* SAG 2466, *Dictyosphaerium minutum*, *Leptochlorella corticola*, Trebouxiophyceae sp. SC2-2, Uncultured Trebouxiophyceae clone QE17, *Botryococcus braunii*, *Heterochlorella luteoviridis*, *Leptosira terrestris*
28S - Helix C1_E1 2nd bp before terminal loop	G-C ⇒ **A-U**	- **NHS** in Trebouxiophyceae - G-C in *Characium perforatum* - C-G in *Eremosphaera viridis* - U-G in *Stichococcus bacillaris* SAG 379-2, *Leptosira terrestris* SAG 463-3
28S - Helix D15 3rd bp before terminal loop	G-U ⇒ **C-G**	- **NHS** in Trebouxiophyceae - G-U in *Characium perforatum* - A-U in other Trebouxiophyceae
rbcL - AA 54	G ⇒ **A**	- **A almost unique (exception: reversion to G in clade 4, dark green group)** - G in all other Trebouxiophyceae
Microthamniales		
18S - Helix 16 3rd bp before terminal loop	C-G ⇒ **A-U**	- **NHS** in Trebouxiophyceae - C-G in all other Trebouxiophyceae
18S - Spacer Helix 4'-Helix 17 last nt	UAAAUAA ⇒ AAAAUA**U**	- **NHS** in Trebouxiophyceae - AAAAUAA in *Parachloroidium lobatum*, *Dictyochloropsis asterochloroides*, *Dictyochloropsis splendida* SAG 2097 - GAAAUAA in Oocystacea (except *Eremosphaera viridis*) and one branch in the Watanabea clade (*Kalinella apyrenoidosa*, *Kalinella bambusicola*, *Heveochlorella roystonensis*, *Heterochlorella luteoviridi*s SAG 2133, SAG 2213 and SAG 211-2a, *Heveochlorella hainangensis*)
18S - Helix 17 2nd bp	A-U ⇒ **G**-U	- **NHS** in Trebouxiophyceae - U-A in *Ettlia pseudoalveolaris* and both *Symbiochloris symbiontica* strains
18S - Helix 17 bp 6	C-G ⇒ **G-C**	- **NHS** in Trebouxiophyceae - U-A in *Botryococcus braunii*
18S - Helix E23_2 bp 6	G-C ⇒ **A-U**	- **NHS** in Trebouxiophyceae - U-A in *Oocystella oogama*

Taxon/character	Evolutionary change	Characterization
18S - Helix E23_7 bp 3	G-C/U ⇒ **C-G**	- **NHS** in Trebouxiophyceae - U-A in *Quadricoccus ellipticus*, *Planctonemopsis distans*, *Amphikrikos nanus* - A-U in *Chloropyrula uraliensis* and *Apatococcus* sp. - UxU in *Prototheca wickerhamii*
18S - Spacer Helix E23_12'-Helix E23_9' 2nd nt	GAA ⇒ G**C**A	- **NHS** in Trebouxiophyceae - GAA conserved in all other Trebouxiophyceae
18S - Helix E23_13 last bp	U-A ⇒ **G**xA	- **NHS** in Trebouxiophyceae - U-A conserved in all other Trebouxiophyceae
18S - Spacer Helix 45-Helix 46 1st nt	CCUAG ⇒ **U**CUAU	- **NHS** in Trebouxiophyceae (first U unique in Microthamniales) - last position variable in Trebouxiophyceae (majority G)
18S - Spacer Helix 45'-Helix 47 nt 5	U ⇒ **C**	- **NHS** in Trebouxiophyceae - U conserved in all other Trebouxiophyceae
28S - Helix B18 1st bp	G-C ⇒ G-**U**	- **NHS** in Trebouxiophyceae
28S - Helix B18 last bp before terminal loop	C-G ⇒ **U-A**	- **NHS** in Trebouxiophyceae - U-G in *Pabia signiensis* and *Koliella sempervirens*

3.2.4. Compensatory Base Changes in the ITS2 Molecule

Due to their significance regarding species delimitations, the CBCs in the ITS2 molecule were highlighted separately (Table 12), although some of the information is redundant as it was already listed in the NHS table of the previous chapter, when a character change was unique for a certain group or branch. Homoplasious CBCs and those unique only for a subdivision of a clade however give new information as they were not listed before.

The CBCs listed in Table 12 represent those positions that were paired in all strains and had a double sided change that maintained the base pairing. Positions which offered CBCs but were not paired in every single strain, were refused.

Since none of the *Microthamnion* phylogenies was resolved in the basal part, the evolutionary direction of changes could not be estimated. Thus, all character changes were given in reference to the character state in the majority of strains. Serial numbers were applied to keep track on homoplasious CBCs: they were given the same number and can thus be retraced to all branches where they occurred. Serial numbers appearing only once consequently represent unique character changes. The branch numbers in the first column refer to the phylogeny in Figure 6.

Table 12. Compensatory Base Changes (CBCs) in the ITS2 molecule
Branch numbers refer to the phylogeny in Figure 6. ** ITS2 not sequenced in all members of the corresponding clade (confer with Supplementary Table 3, p. 92, for summary of available ITS2 sequence data and Supplementary Table 4, p. 93, for detailed information on character states at the given positions); strains listed as exceptions are representatives of certain branches in the phylogeny of Figure 6. Character changes are given in reference to the majority of strains.

Branch - Strain(s)	Compensatory Base Change (CBC) No - Helix (universal position)	Character Change
1 - Clade 1** **except**: CCAC 3843, CCAC 2942 B	10 - H4 (126-VI/126-XI)	G-C ⇒ A-U
2 - Clade 2** **except**: CCAC 2081, CALA 368, M 2412/1 A	8 - H4 (126-IV/126-XIII)	U-A ⇒ C-G
3 - CCAC 2916 + ACOI 1621	6 - H3 (90/93)	U-A ⇒ C-G
4 - ACOI 1621	5 - H3 (55-XVII/108-IV)	A-U ⇒ G-C
	8 - H4 (126-IV/126-XIII)	U-A ⇒ C-G
5 - CCAC 4547 B/5561	1 - H1 (8/11)	C-G ⇒ G-C
6 - CCAC 3677/4161*	2 - H1 (9-VII/9-X)	G-C ⇒ U-A
	3 - H2 (29-III/29-XII)	C-G ⇒ U-A
	5 - H3 (55-XVII/108-IV)	A-U ⇒ G-C
7 - CCAC 2279/4234	4 - H2 (29-III/29-XII)	C-G ⇒ G-C
	6 - H3 (90/93)	U-A ⇒ C-G
8 - CCAC 4853/4854/4855	2 - H1 (9-VII/9-X)	G-C ⇒ U-A
	7 - H4 (126-III/126-XIV)	C-G ⇒ A-U
	9 - H4 (126-IV/126-XIII)	U-A ⇒ G-C

3.2.5. Data Matrix for Alternate Tree Reconstructions

Supplementary Table 4 (p. 93) gives the data matrix used for a tree reconstruction based on molecular characters that exceeded those of a mere succession of nucleotides in sequence data. Included were 100% sequence identity for each marker separately (thus also including those regions that could not be unambiguously aligned), the position of restriction enzyme cutting sites within the 18S rRNA molecule which lead to either an additional or prevented cut, the total number of cuts with each of the above mentioned enzymes, the length of the four ITS2 helices, and all apomorphies detected in the nuclear-encoded rRNA operon data as well as the *rbc*L protein alignment (NHSs of Table 11 and CBCs in ITS2 (Table 12) were integrated there).

Apomorphies in the operon data were included with due consideration of the secondary structures of the individual markers, and were strictly confined to those positions which could be unambiguously aligned. This contained paired helix regions with single nucleotide changes as well as double-sided changes which lead to either the maintenance of a base pair (hemi CBCs and CBCs) or a mismatch. But also single-stranded spacer- and loop-motifs were incorporated as long as they were unambiguously alignable. In the *rbc*L, only those apomorphies were taken into consideration where the amino acid change resulted in different side chain properties.

3.2.6. Phylogeny Based on Phenetic Molecular Characters

From the 3800 'best' trees (i.e. all variants with a score of 315) obtained from parsimony analysis, tree number 301 was selected for further graphical revision since it was one of the first which represented the topology of the majority rule consensus, and resembled the topology of the *Microthamnion* tree with all data (Figure 6, p. 34) most. Figure 11 gives the phylogeny of said tree. The clade and grade designations with the corresponding colors introduced in Figure 6 were adopted for the phenetic phylogeny.

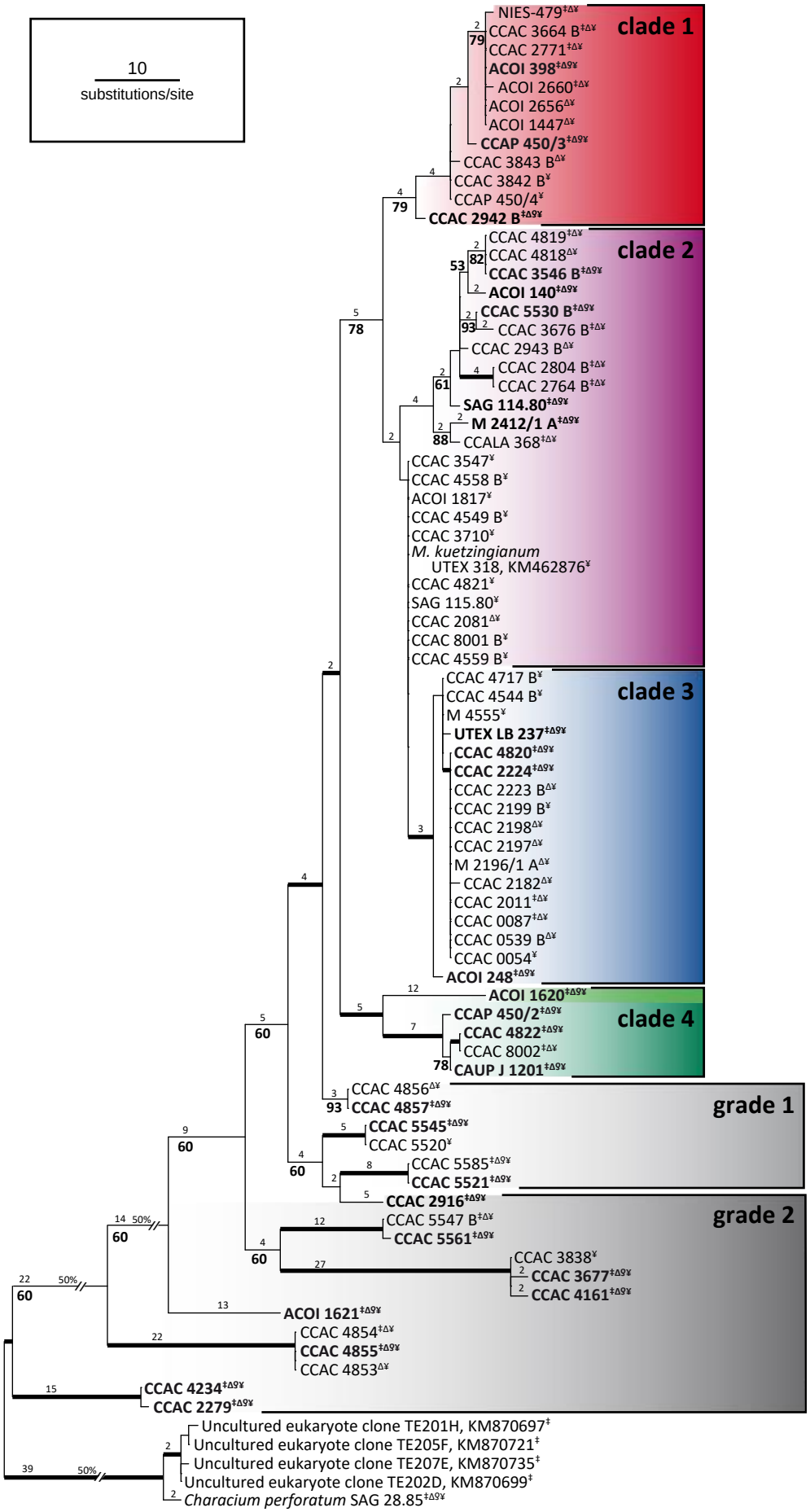
Here, clades 1, 3 and 4 were monophyletic and highly supported, whereas clade 2 was only 'defined' by the clear boundaries of clade 3, the branch establishing a sister relationship between clades 2 and 3 (which was not supported), and the clear distinction of clade 1. Clade 3 received maximum support and formed the tip of the tree, with clade 2 and clade 1 being the next descendent groups. Clade 4 was also maximally supported at the clade branch as well as the subdividing node of the dark green strains. The node corresponding to the divergence of clades 1-4 from the remaining strains was maximally supported.

The topology in the basal part of the tree differed more from the grade definitions of the *Microthamnion* multi-gene phylogeny. Here, several branches were grouped together with moderate support and thus formed monophyletic groups. One occurred in grade 2 where strains CCAC 5547 B/5561 and CCAC 3677/3838/4161 were grouped together. The other summarized strains CCAC 5545/5520 and CCAC 5585/5521 of grade 1 with strain CCAC 2916 of grade 2.

The most basal group was CCAC 2279/CCAC 4234 with the next emerging group being CCAC 4853/CCAC 4854/CCAC 4855, followed by strain ACOI 1621. All three branches only received moderate support for this topology.

Figure 11. Maximum parsimony phylogeny based on phenetic molecular characters

The tree was calculated with 112 characters, 93 of which were parsimony-informative and 19 parsimony-uninformative. From 3800 'best' trees, tree number 301 was selected for further graphic revision. Small numbers at branches: morphological character changes, large numbers in bold: MP bootstrap percentages. Branches in bold were maximally supported, very long branches were graphically reduced to 50% (50%//). Strains in bold are those used for the congruent analyses (compare figures 1-5). Sequences available for each strain are indicated by the following symbols: ‡=18S, Δ=ITS2, ϑ=28S and ¥=*rbc*L. Clade- and grade designations as well as color codes are given as introduced in Figure 6.


10
substitutions/site
clade 1
clade 2
clade 3
clade 4
grade 1
grade 2
NIES-479‡ΔΨ
CCAC 3664 B‡ΔΨ
CCAC 2771‡ΔΨ
ACOI 398‡ΔΨ
ACOI 2660‡ΔΨ
ACOI 2656ΔΨ
ACOI 1447ΔΨ
CCAP 450/3‡ΔΨ
CCAC 3843 BΔΨ
CCAC 3842 BΨ
CCAP 450/4Ψ
CCAC 2942 B‡ΔΨ
CCAC 4819‡ΔΨ
CCAC 4818ΔΨ
CCAC 3546 B‡ΔΨ
ACOI 140‡ΔΨ
CCAC 5530 B‡ΔΨ
CCAC 3676 B‡ΔΨ
CCAC 2943 BΔΨ
CCAC 2804 B‡ΔΨ
CCAC 2764 B‡ΔΨ
SAG 114.80‡ΔΨ
M 2412/1 A‡ΔΨ
CCALA 368‡ΔΨ
CCAC 3547Ψ
CCAC 4558 BΨ
ACOI 1817Ψ
CCAC 4549 BΨ
CCAC 3710Ψ
M. kuetzingianum
UTEX 318, KM462876Ψ
CCAC 4821Ψ
SAG 115.80Ψ
CCAC 2081ΔΨ
CCAC 8001 BΨ
CCAC 4559 BΨ
CCAC 4717 BΨ
CCAC 4544 BΨ
M 4555Ψ
UTEX LB 237‡ΔΨ
CCAC 4820‡ΔΨ
CCAC 2224‡ΔΨ
CCAC 2223 BΔΨ
CCAC 2199 BΨ
CCAC 2198ΔΨ
CCAC 2197ΔΨ
M 2196/1 AΔΨ
CCAC 2182ΔΨ
CCAC 2011‡ΔΨ
CCAC 0087‡ΔΨ
CCAC 0539 BΔΨ
CCAC 0054Ψ
ACOI 248‡ΔΨ
ACOI 1620‡ΔΨ
CCAP 450/2‡ΔΨ
CCAC 4822‡ΔΨ
CCAC 8002‡ΔΨ
CAUP J 1201‡ΔΨ
CCAC 4856ΔΨ
CCAC 4857‡ΔΨ
CCAC 5545‡ΔΨ
CCAC 5520Ψ
CCAC 5585‡ΔΨ
CCAC 5521‡ΔΨ
CCAC 2916‡ΔΨ
CCAC 5547 B‡ΔΨ
CCAC 5561‡ΔΨ
CCAC 3838Ψ
CCAC 3677‡ΔΨ
CCAC 4161‡ΔΨ
ACOI 1621‡ΔΨ
CCAC 4854‡ΔΨ
CCAC 4855‡ΔΨ
CCAC 4853ΔΨ
CCAC 4234‡ΔΨ
CCAC 2279‡ΔΨ
Uncultured eukaryote clone TE201H, KM870697‡
Uncultured eukaryote clone TE205F, KM870721‡
Uncultured eukaryote clone TE207E, KM870735‡
Uncultured eukaryote clone TE202D, KM870699‡
Characium perforatum SAG 28.85‡ΔΨ

4. Discussion

4.1. Phylogenetic Analyses

4.1.1. ITS2 Secondary Structure

The secondary structure of the ITS2 molecule has been extensively studied across eukaryotes (Mai & Coleman, 1997; Joseph et al., 1999; Schultz et al., 2005; Coleman, 2007), where a common clover leave like structure with four helices, connected by single-stranded spacer regions, was revealed. Detailed studies in algae were performed for the order Ulvales, where a consensus secondary structure was established based on 86 taxa (Caisová et al., 2011a). This structure was later expanded to two classes, the Ulvophyceae and Chlorophyceae, represented by the orders Ulvales, Chaetophorales, Chaetopeltidales, Oedogoniales and Sphaeropleales (Caisová et al., 2013).

The landmarks of ITS2 secondary structure prediction, specified by Caisová et al. (2013) for the Ulvophyceae and Chlorophyceae, could be reliably applied on *Microthamnion* as well (Reder, 2015). The spacer regions between helices and the first two base pairs in helices 1-3, although not completely identical, could be easily identified. Also the pyrimidine-pyrimidine mismatch in helix 2 and a GGUA motif at the 5'side of helix 3 were highly conserved and thus readily determined. With the helix boundaries defined, it was possible to fold the single helices individually and compare them pairwise which ultimately lead to the erection of a consensus secondary structure model for the complete genus. In the present study, the dataset of Reder (2015) was augmented by several new strains which also represented different genotypes. These additional sequences, having been subjected to careful folding and manual pairwise comparison with the already existing structures, offered new insights and allowed for the optimization of the previous model.

With the consensus secondary structure model established for *Microthamnion*, the transfer of this information to other trebouxiophycean taxa was smoothly pursued. Although the spacer between 5.8S and H1 showed minor length variations (one to three additional nucleotides compared to *Microthamnion*) and sequence variability, the first two base pairs of H1 were highly conserved among the Trebouxiophyceae. With the spacer between H1 and H2 being length-conserved and usually beginning with an adenine, the helix boundaries were reliably identified. The other landmarks for secondary structure prediction could also be easily determined: the pyrimidine-pyrimidine mismatch in H2 (mainly U-U), a GGUA motif on the 5' side of H3, the highly conserved spacer between H2 and H3, the first five base pairs in H2 and the first two in H3, plus the first three nucleotides on the 5' side of H4.

This approach enabled the buildup of a trebouxiophycean ITS2 alignment based on secondary structure information, which was crucial for the identification of homologous nucleotides and consequently unambiguously alignable positions

available for phylogenetic analyses (compare grey background in Figure 10, p. 42). It also facilitated the search for NHSs in these homologous regions across all trebouxiophycean taxa investigated in this study, and thus allowed to put the NHSs in *Microthamnion* in a broader context. A detailed graphic revision of a general trebouxiophycean secondary structure model as shown for *Microthamnion*, although certainly feasible, did however exceeded the purpose of this study.

4.1.2. Identification of a Promising Barcode Candidate in *Microthamnion*

The barcoding system was originally introduced as a tool for rapid species identification via a single variable DNA marker (Hebert et al., 2003). Although successfully applied in animals, where the 'correct' species boundaries can be tested against biological species (e.g. Hebert et al., 2004; Hajibabaei et al., 2006), there were also pitfalls uncovered. Incomplete lineage sorting (Wiemers & Fiedler, 2007), insufficient resolution capacities of the marker chosen (Hadi et al., 2016), or high error rates in recently diverged species (Hendrich et al., 2010) were only some of them. It has thus been proposed, to use a barcode only for a first estimation of intrageneric variability, and to corroborate species boundaries with other methods (e.g. Meyer & Paulay, 2005; Wiemers & Fiedler, 2007; Zou et al., 2016).

In the present study, a barcode was sought for the latter application rather than a means for distinct species delineation. Here, one aim was the compilation of a congruent multi-gene alignment, which should be used for several concatenated analyses with varying marker compositions, to determine the best combination for a resolved phylogeny in *Microthamnion.* As the amplification and sequencing of all markers for each of the 74 *Microthamnion* strains would have been inefficient, both time and cost wise, it was searched for a suitable barcode marker that depicted the intrageneric variability. This, in turn, facilitated the selection of reference strains for each identified group which ensured both, a balanced taxon sampling and minimized expenses in DNA amplification and sequencing. With regard to future investigations on the genus, this barcode will also be ideal to assign new *Microthamnion* isolates and environmental sequences to already identified subgroups, or quickly identify yet unknown genotypes.

From the single-gene phylogenies of all investigated markers, the *rbc*L was best suited for this task as the topology inferred from the DNA alignment revealed several well-supported groups and long branches that allowed for the selection of representatives that could then be used for more detailed comparative analyses using a congruent dataset. The *rbc*L was also the marker that was easiest to amplify as the primers worked on all strains and usually yielded a strong PCR product which made reamplification unnecessary. Besides, the sequencing reads were very clear and did hardly need manual correction. An additional advantage, being a protein coding gene, was a much easier and faster aligning process compared to the rRNA genes and especially the ITS2 molecule, as sequences only had to be aligned according to reading

frame and not to secondary structure motifs. All these facts taken into account, the *rbc*L can be considered an ideal barcode marker for *Microthamnion* which also meets the criteria proposed by Hall and collegues (2010).

4.1.3. Selection of the Best Marker Combination for a Resolved *Microthamnion* Phylogeny

The impact of the single markers on the *Microthamnion* phylogeny was exhaustively tested and compared for both, the individual markers separately as well as various combinations, all based on a congruent dataset (Figure 1 to Figure 5, pp. 28-32). This dataset comprised representatives of the groups identified via *rbc*L (DNA) single-gene analyses, thus ensuring a well-balanced taxon sampling spanning the complete *Microthamnion* diversity investigated in this study.

These analyses showed the advantages that concatenated analyses can have over phylogenies inferred from only a single marker, but also stressed the importance of a careful selection regarding the markers to be compiled (see below). When individually computed, none of the different loci was able to resolve the *Microthamnion* phylogeny, although the *rbc*L (DNA) was powerful enough to depict the intrageneric variability. The 18S and 28S rRNA genes proved to be too conserved to differentiate more than the basal lineages, a result that did not much surprise as the SSU is a marker usually better suited for higher taxonomic phylogenies (e.g. Marin et al., 2003; Guillou et al., 2004; Caisová et al., 2011b). In the LSU the variable C-domain apparently did not have enough phylogenetic signal to compensate for the remaining conserved parts.

In the combined *Microthamnion* phylogenies, the *rbc*L nucleotide data dominated the topology. The other often suggested 'species marker' ITS2 (e.g. Lewis & Flechtner, 2004; Shao et al., 2004; Young & Coleman, 2004; Walton et al., 2007; Chen et al., 2010) had far less impact on the topology, but worked out a finer distinction of the strains, especially in clades 1 and 2, and also lent support for the separation of clades 1 to 4 from the grades in the basal part, when combined with the *rbc*L nucleotide data. Surprisingly, the addition of the 18S rRNA gene, which had not much significance when computed separately, was crucial to work out the monophyly of clade 2. Interestingly, it did not matter whether the *rbc*L was included based on the nucleotide or amino acid sequence as long as all three markers were combined. As expected, the 18S also increased the support values and thus stabilized the phylogeny. The partial 28S gene sequences did not give any new insights regarding the topology, but when combined with the other markers, it significantly increased the statistical support for clade 2 and the node confirming strain CCAC 4857 to be the next relative to clades 1 to 4.

All combinations with *rbc*L nucleotide data were superior to the single-gene analyses with regard to significance and/or topology, with the exception of the tree inferred from only SSU and *rbc*L DNA data which had no altering effect. Those phylogenies without any *rbc*L data were however inferior to the *rbc*L (DNA) single-gene tree. It can thus be concluded that in *Microthamnion* the *rbc*L is the most crucial of the markers investigated and should be used in all phylogenies addressing intrageneric variability.

The best result in regard to topology and significance was achieved with all investigated markers when the *rbc*L was included with the nucleotide sequence (compare Figure 5, p. 32). The tip of the tree was well resolved with this combination, and also a closer relationship with the strains of grade 1 was worked out with at least moderate support. The basal part, i.e. the relationships among the strains of grade 2, however remained unresolved.

There were however some combinations containing ITS2 data that confirmed the most basal position of strain CCAC 4855, as the next ascending node received relatively high bootstrap percentages (70 to 80). This was interestingly only the case in combinations where *rbc*L nucleotide data was absent. Apparently, this topology was strongly favored by the ITS2, but seemed to be silenced by the dominating phylogenetic signal of the *rbc*L nucleotide dataset. In those combinations with *rbc*L protein data though, the basal position of strain CCAC 4855 was again enforced with statistical support. Without the defining *rbc*L nucleotide signal in those phylogenies however, the individual clades in the tip of the trees were either not clearly distinguished or their relationship among each other was not resolved. There was no marker combination from those investigated in this study that could provide both characteristics, a well resolved tip and statistical support for the most basal strain. As the latter finding, though interesting, did not contribute to the identification of species in the genus, the combination of all markers with the *rbc*L based on the DNA alignment was ultimately chosen for the phylogeny comprising all *Microthamnion* strains.

4.1.4. Comparison of the Congruent with the Gapped *Microthamnion* Phylogeny

The calculation of the phylogenies based on a congruent, well-balanced taxon sampling (Figure 5, p. 32) and on a gapped alignment with all available *Microthamnion* sequences (Figure 6, p. 34) served different purposes. Where the congruent dataset was required to approximate a best resolved phylogeny, the tree with all strains was needed to display the complete known *Microthamnion* diversity and allow for a phylogeny-based identification of species.

The positive effects of a well-balanced taxon sampling and an alignment without gaps (Figure 5) on the support values has been described in detail in the results part (chapter 3.1.4, p. 33). Apparently, the overrepresentation by the large clusters of highly similar genotypes in clades 1-3, as present in the gapped phylogeny (Figure 6), blurred the phylogenetic signal and resulted in generally lower support at the clade nodes and internal branches. Besides, the gaps in the alignment caused troubles for the distance methods, which often failed to resolve branches which were otherwise well supported with the other methods. The topology however was not affected at all by the gapped alignment which validates the efficient approach only to generate *rbc*L sequences for all strains and allow for missing data in (some of) the other genes when depicting the complete *Microthamnion* diversity. For the optimal statistical support of clades and

internal nodes, the phylogeny with a reduced, well-balanced taxon sampling and complete sequence data (Figure 5) can be consulted as reference.

In both phylogenies however, the relationship among the individual long branches in the basal part could not be clarified. In general, there have been two methods discussed to stabilize the topology and to increase the accuracy of phylogenetic trees. While some authors favor the increase in number of molecular markers as well as sequence length (Rosenberg & Kumar, 2001; Rokas et al., 2003), others advocate an increased taxon sampling (Graybeal, 1998; Hedtke et al., 2006). As the present phylogenies were already inferred from a combination of four markers from different loci with a wide range of conservation levels and a total length of 4605 aligned positions, which was in fact able to nicely resolve the tip of the tree, the addition of more genes seems hardly advantageous. Instead an increased taxon sampling for the basal part seems more likely to improve the phylogeny, especially, as the addition of linking taxa can also break up long branches. There were however all available strains in culture investigated and all available sequence data incorporated in this study, and also extensive BLAST searches did not reveal additional sequences. It seems, that in order to solve this problem, the collection of new samples is required, which will have to lay a special focus on habitat preferences and sampling strategies (more see below).

Another, maybe even more important, aspect for the unresolved topology in the basal part is the absence of an adequate outgroup. *Characium*, although being the next known relative, is already quite distantly related and, apart from the newly found members of the *Characium* clade, no other relatives could be determined. When an outgroup is either chosen too distantly or the next living relatives simply are to distantly related, the long outgroup branches can lead to wrong rootings as they attract long ingroup branches and thus do not give their right position (Felsenstein, 1978; Huelsenbeck, 1997; Graham et al., 2002). It is thus quite possible that with a better sampled outgroup (which includes more taxa), the position of the long branches might shift and the topology be better resolved.

4.1.5. Trebouxiophycean Phylogeny

The trebouxiophycean phylogeny calculated in this study was apparently the first to combine molecular data from the nuclear as well as the plastid genome and at the same time covering the complete known trebouxiophycean diversity. Other studies either conducted extensive studies confined to the plastid genome (Lemieux et al., 2014; Suzuki et al., 2018), were restricted to 18 rDNA (Eliáš et al., 2008; Darienko et al., 2010; Hallmann et al., 2016) or *rbc*L (Kulichová et al., 2014), or covered less diversity (Neustupa et al., 2013b; Fučíková et al., 2014).

The tree inferred from a combined 18S rDNA, ITS2, 28S rDNA and *rbc*L (DNA) analysis (gapped alignment; Figure 7, p.36), clearly proofed the monophyly of *Microthamnion*. Besides, it confirmed *Ch. perforatum* to be (one of) the next relative(s) to *Microthamnion*, but also revealed the existence of a hitherto unknown sister group to *Ch. perforatum*. This sister group was unfortunately only represented by

environmental SSU sequences and cannot be linked to any known organism or even algal cultures. Without further investigations, it can only be speculated whether these new sequences form another *Characium* species or even represent a new genus within the Microthamniales, and thus the designation '*Characium* clade' was chosen.

Whatever the taxonomic classification might be, these sequences were extremely valuable for *Microthamnion* phylogenies as they increased the available outgroup for analyses containing 18S rDNA data. There were however no other (more distantly) related taxa to *Microthamnion* revealed as further clarification of relationships among trebouxiophycean lineages was not possible. A proposed sister relationship of the Microthamniales to *Xylochloris* and *Leptosira* (Lemieux et al., 2014; Suzuki et al., 2018) could not be confirmed. Also a closer relationship with *Parietochloris,* as suggested by 18S rDNA phylogenies (Eliáš et al., 2008; Neustupa et al., 2009, 2013a,b; Darienko et al., 2010; Hallmann et al., 2016), was only indicated by topology, but lacked any statistical support except with Bayesian posterior probabilities. Besides, *Parietochloris* (and *Ettlia*) were too distantly related to benefit the outgroup of *Microthamnion*.

Due to the limited amount of available sequence data, with lots of taxa only represented by sequence of a single gene in the NCBI database, the trebouxiophycean phylogeny was calculated based on a gapped alignment. As the comparison of the *Microthamnion* congruent and gapped phylogenies already showed, the use of such a gapped alignment has its disadvantages regarding the support values. Distance methods in particular are vulnerable to gaps and it is thus not surprising, that they failed to resolve most branches. However, when compiling the trebouxiophycean alignment, the information gained by the use of several markers and a maximized coverage of the trebouxiophycean diversity was deemed to outweigh the disadvantages of a gapped alignment. As simulations have shown, the correct phylogenetic tree can be reconstructed when taxa with missing data are included as long as the overall sequence length is high (i.e. 2000 characters as opposed to 100 characters in simulations where taxa where misplaced) (Wiens, 2006). Besides, when taxa with missing data were added to a given dataset, they were shown to be equally capable of breaking up long branches as taxa with complete sequences, as long as they covered 50% of the total examined sequence length (Wiens, 2005). The gapped alignment used for the trebouxiophycean phylogeny included only those taxa, where at least the complete 18S rRNA or *rbc*L genes were covered. Taxa with only ITS2 or partial SSU sequences available were not taken into consideration, thus ensuring a high coverage of total number of characters.

Despite an extensive taxon sampling via literature and database searches, the branches separating the Microthamniales from the remaining Trebouxiophyceae and *Microthamnion* from the *Characium* clade remained very long, indicating that there either are no other living relatives, or relevant taxa are not available in databases yet. As most of the new members of the *Characium* clade were found in glacial debris (Schmidt & Darcy, 2015), and numerous other examples in the Trebouxiophyceae are

known, where sampling strategies considering more unusual habitats lead to the discovery of new lineages (e.g. Neustupa et al., 2011, 2013a,b; Gaysina et al., 2013), it is quite possible that more taxa will be revealed in the future.

4.1.6. Putative Species Inferred from the *Microthamnion* Phylogeny

The criteria based on which species are inferred from a molecular phylogenetic tree are in theory relatively simple. A species is usually defined as a monophyletic group comprising several organisms, which is clearly distinguishable from other such groups via both branch lengths and statistical support (Leliaert et al., 2014). The ideal topology would display long and well-supported inter- and short intraspecific branches, thus allowing for a clear identification of species boundaries. There are numerous examples in algae, where species have been considered to correspond to these monophyletic groups, regardless whether they were referred to as "clear-cut genotypic clusters" (Verbruggen et al., 2005), "evolutionary significant units" (Verbruggen et al., 2007; Tronholm et al., 2010) or simply "lineages" (Šlapeta et al., 2006; Belton et al., 2014) or "clades" (van der Strate et al., 2002). Boundaries are however not always that clear-cut, especially in ongoing or very recent speciation events where the variation inside a clade can be quite high. In these cases also the surrounding topology has to be taken into consideration and the phylogeny-based approach ideally be complemented with other methods of species identification.

In the tip of the *Microthamnion* phylogeny, comprising all strains with all available sequence data (Figure 6, p. 34), four monophyletic clades could be defined with a clearly distinguishable subdivision in clade 4. Whether these clades should be considered different species or where boundaries might be disputable, needs to be discussed in detail.

From the branch lengths and the support values, the dark green group of clade 4 was the most readily identified distinct species when the above mentioned criteria were applied. This group was maximally supported and had clearly discriminative intra- versus interspecific branch lengths. As the other subdivision in clade 4 (light green) only consisted of one strain, the monophyly criterion could not be applied and also support values were not available. The branch length was however so significant, that considering it also a different species was not at all far-fetched. The only reason of choosing the terminology of 'clade 4 dark green and light green' over 'clade 4 and clade 5' was, that a single strain per definition cannot be named a clade. Clade 1 was also clear cut with almost maximal support, except for the NJ methods. The latter likely being an artefact caused by the gapped alignment and an uneven taxon sampling, as in the congruent analysis (Figure 5, p. 32) the clade was resolved with all methods. In clade 3 and, especially, clade 2 the branch lengths within the clades exceeded that of the respective clade branch and also the support values were not that significant. The surrounding topology however clearly separated both clades. The strains of clade 2, although only grouped by an extremely short clade branch with low support (compare the congruent phylogeny of figure 5 for better significance), were nevertheless distinct

through the clear boundaries of clade 1 and the long and almost maximally supported node defining a sister relationship of clades 1 and 2. A more conservative possibility would, of course, have been to summarize clades 1 and 2 into a single clade and thus a single species. The clade branch in this case would have been strongly supported (maximum support in the congruent phylogeny) but the variability within this large clade would have been too high, especially, when considering the presence of a clear-cut monophyletic group (i.e. clade 1) that implied a subdivision. Clade 2 was in fact so diverse that even a division in several subclades could have been considered, as multiple monophyletic clusters were present. This thought was however refused, because clear boundaries could not be unambiguously identified. Due to the absence of statistical support at most branches, it was not evident whether only the most terminal monophyletic groups should be considered or whether boundaries lay at more internal nodes, thus grouping several clusters together. It was thus opted for the division in the two clades earlier introduced (i.e. clade 1 and 2) with clade 2 likely depicting a very recent speciation event. For clade 3 the surrounding topology was even more distinctive than for clade 2 as both branches, the one separating clades 1-3 from the remaining strains as well as the one defining a sister relationship between clades 1 and 2, were long and highly supported.

It is thus quite reasonable to acknowledge the clade boundaries with the subdivision in clade 4, and consider them 5 different species.

The basal part of the tree, i.e. grades 1 and 2, was characterized by individual long branches with high statistical support but an unresolved topology. A grouping into clades was in so far difficult as the phylogeny was clearly undersampled in this part and strains grouped together at terminal branches often originated from the same natural sample and should thus rather be considered identical clones than individual strains; which in turn made the statistical support at those nodes invalid. This referred to the branches grouping strains CCAC 4856/4857 in grade 1 and strains CCAC 5547 B/5561 and strains CCAC 4853/4854/4855 in grade 2. In those cases though, where strains from different localities were grouped, they could be termed clades albeit with only two members each (strains CCAC 3677/3838 originate from the same raw sample, thus only two members when grouped with CCAC 4161).

There were three clearly distinguishable groups/individual strains in grade 1 and six with extremely long branches in grade 2. As both, branch lengths and statistical support, when applicable, were quite compelling, these nine groups/individual strains must at least be considered different species. But as they were so distantly related to the other *Microthamnion* clades, they could even be considered different genera, especially, when taking into consideration the comparable branch lengths of the groups/individual strains of grade 2 with those between *Microthamnion* and *Characium*, which are definitely different genera. This assumption is also consistent with the assessment of Šlapeta and collegues (Šlapeta et al., 2006), who performed a multi-locus phylogenetic study on 17 *Micromonas pusilla* isolates. The unrooted

phylogeny inferred from rDNA (SSU, both ITS regions and 5.8S), *cox*I and *rbc*L data had an astonishingly similar topology with equally divergent branch lengths among some lineages. They, too, concluded that the long branched lineages might rather belong to different genera than different species, a suggestion that was later backed up when brought into a larger phylogenetic context (Marin & Melkonian, 2010).

4.2. Phenotypic Molecular Data

4.2.1. NHS and CBC Support for Species and Genus Boundaries

The concept of molecular NHSs was introduced by Marin et al. (2003), who drew parallels between these uniquely derived molecular signatures and the use of morphological characters in systematics. Due to the high degree of conservation in the ribosomal rRNA genes' secondary and tertiary structures, it was assumed that NHSs found there displayed "morphological rRNA characters" which were similarly suitable for cladistics as phenotypic characters, albeit below the resolution of electron microscopy (Marin et al., 2003). This idea was later also exerted on the plastid-encoded rRNA operon and *rbc*L genes (Marin & Melkonian, 2010; Marin, 2012). In the present study, NHSs were sought to lend support for clades identified by phylogenetic analyses and thus to reinforce species boundaries.

Unique Signatures in the *Rbc*L and Nuclear-Encoded Operon Genes

The NHSs found in the plastid-encoded *rbc*L gene were indeed in accord with the four clades defined in the *Microthamnion* phylogeny (Figure 6). Each of the individual clades gained NHS support, and in clade 4 the separation into the two subclades was acknowledged via multiple unique synapomorphies for the dark green group. Besides, unique synapomorphy residue substitutions did not occur on any internal clade branches (when clade 4 is considered two clades).

As already discussed before, from the *Microthamnion* phylogeny it can be assumed that the clades correspond to species level and that clade 4 is defined very conservatively and the subdivision into the light- and dark green group is instead preferable to mark the species boundary. This assumption is supported by the *rbc*L NHSs and leads to the conclusion that in *Microthamnion* the smallest unit identified via *rbc*L NHSs corresponds to the species level. Besides, *rbc*L NHSs lent support to the topology and thus the relationship among clades, as clades 1 to 3 were united by a unique residue substitution. Following a gap after clade 4, where neither the individual lineages of grade 1 nor their relationship among each other or that with clades 1 to 4 gained NHS support, all lineages in grade 2, except the single-stranded strains CCAC 2916 and ACOI 1621, were again defined by unique synapomorphies in the *rbc*L. These NHSs in grade 2 were however corroborated by synapomorphy support from all other markers. The largest entity defined by a unique molecular signature in the *rbc*L was *Microthamnion* on genus level, albeit with the exception of the dark green group of clade 4.

The conserved operon genes however did not hold any NHS support for the clades 1 to 3, or the dark green group of clade 4. Instead, molecular synapomorphies in the 18S, 5.8S and 28S rRNA genes were found mainly for the lineages of grade 2 and then on genus and order level. The two only exceptions were the light green group of clade 4 with a unique loop motif in 5.8S and a CBC in 18S, albeit in a variable region (V4), as well as the node summarizing strains CCAC 4856/4857 with clades 1 to 4 which was characterized by a CBC and a unique mismatch pattern in a variable helix of 28S.

The unique molecular signatures for branches grouping clades together, as well as the NHS gap between the clades and the lineages of grade 2, also lent support to the assumption that the strains investigated in this study do not only form separate *Microthamnion* species, but might in part even represent different genera. Besides, the clear hierarchy of NHSs regarding the markers in which they were found, offered new insights as to where these alternative genus boundaries might lie. The most narrow definition for *Microthamnion* based on NHS support would then encompass clades 1 to 3 as they were grouped together and were also individually characterized by NHSs in the *rbc*L, with subdivisions only supported by unique signatures in the most variable marker ITS2 (see below). The broadest genus border, on the other hand, would be all four clades plus strains CCAC 4856/4857 as this is the last node confirming relationships among groups with NHS support. Another logical and quite neat possibility would be to draw the line between the groups that were exclusively defined by NHSs in the *rbc*L or ITS2 and those with (additional) operon gene support. In that case *Microthamnion* would consist of clades 1 to 3 and the dark green group of clade 4. This arrangement would however be in conflict with the phylogeny which supports the monophyly of clade 4 and can thus not prevail.

NHS and CBC Support from the ITS2 Molecule

NHSs and CBCs in the ITS2 molecule did not draw as uniform a picture as the unique signatures from the *rbc*L and operon genes. In the ITS2 neither CBC nor NHS occurrence coincided with the clade boundaries inferred from the phylogeny, but rather clustered on opposite ends of the *Microthamnion* tree (Figure 6, p.34). In the tip, they were mainly concentrated to internal branches in clades 1 and 2, yet also identified the dark green group of clade 4 via a single unique mismatch pattern, whereas in the basal part they were constricted to lineages of grade 2.

Interestingly, all strains in clade 1 were identified via at least one NHS whereas in clade 2 only two terminal branches had NHS support. In clade 1, strain CCAC 2942 was characterized by a unique spacer motif, and the remaining strains were united by a distinct mismatch pattern in H2. Furthermore, also nested monophyletic groups obtained NHS support, one of which being the only CBC found in clade 1. In clade 2 however, NHS support was only given for two terminal branches in which one perhaps referred to only a single strain (M 2412/1 A) as for the other strain (UTEX 318) ITS2 data was not obtained, but these two nodes were even characterized by two and three unique signatures each. The other strains in clade 2 were not defined by any NHS, but

most of them were characterized by a CBC in H 4 (whose character state was shared with a CBC in ACOI 1621 and was thus not unique). This CBC did however not summarize the complete clade, but was shifted one node inwards thus leaving out strain CCAC 2081, and also strains M 2412/1 A and CCALA 368 were excluded as they either reversed to the original character state (M 2412/1 A) or displayed a hemi CBC (CCALA 368). Clade 3 was not defined by any ITS2 NHS or CBC, but in clade 4, the dark green group was identified via a mismatch position in a conserved region of H3 that was otherwise paired in the Microthamniales.

In the basal part, on the other hand, each of the lineages in grade 2 offered at least one CBC (= CCAC 4547 B/5561 and CCAC 2916; all others were supported by two or three CBCs each) and all except strain CCAC 2916 were supported by multiple NHSs. Besides, for all branches except those of the single strains CCAC 2916 and ACOI 1621, one or two CBCs each were unique and thus concurrently NHSs.

A reoccurring pattern in congruence with the *rbc*L and operon genes was however the presence of an NHS- (and CBC-) gap between the clades and the lineages of grade 2. If it were not for the unique mismatch pattern in the dark green group of H4, it would even be considerably larger than that observed in the *rbc*L. Also the origin and numbers of CBCs found for the lineages of grade 2 and the subdivisions in clade 1 and 2, followed a clear trend. While the subdivisions only offered a single CBC each, which were positioned in the extremely variable H4, the individual lineages in grade 2 were characterized by up to three CBCs each. For the single strain branches CCAC 2916 and ACOI 1621, and the lineages comprising strains CCAC 3677/3838/4161 and CCAC 2279/4234, at least one CBC was found in the conserved regions of H2 and H3. The additional CBCs for those lineages as well as all CBCs for strains CCAC 4853/4854/4855 and CCAC 4547 B/5561 were located in the variable helices 1 and 4.

According to the CBC species concept *sensu* Coleman (2000), specimen differing by at least one CBC in the conserved regions of H2 and H3 belong to different biological species, whereas CBC occurrence in the variable helices 1 and 4 was not associated with the ability/inability to cross. However, the absence of a CBC in the conserved regions does not mean that all specimen belong to the same species, and also organisms united in one CBC clade can fall into more than one species (Coleman, 2000). It has also been shown that the CBCs in H2 and H3 do not necessarily correspond to species level, but can represent genera or higher classification ranks (Caisová et al., 2011a, 2013). If one considers the lineages of grades 1 and 2 to represent different genera instead of species, in *Microthamnion*, conserved CBCs would correspond to genus level as they were only found for lineages in grade 2. Considering the fact that in the frame of the CBC species concept only a minimum number of species (or in this case genera) can be approximated, the absence of (conserved) CBCs in the strains of grade 1 or some lineages in grade 2 does not contradict the conclusions drawn from the phylogeny and the NHSs in the *rbc*L or operon genes, which suggest the separation into even more putative genera.

What the NHSs on internal branches in clades 1 and 2 mean for species boundaries can at the moment only be speculated. Strains identified by such a unique signature could, of course, form separate species, especially when taking into consideration that for most of these nodes at least one NHS was found in the conserved helices 2 and 3. There are also several strains in clade 2 where the ITS2 was not sequenced and it is thus possible that clade 2 holds even more groups identifiable via NHSs. These putative species boundaries were however not supported by any of the other molecular methods applied in this study. Without further investigation, for example on the morphology, it is for the moment more fitting to assume these variations to correspond to subspecies/varieties or ecotypes rather than species.

4.2.2. Tree Reconstructions Inferred from Phenotypic Molecular Characters

The use of morphological data for phylogenetic tree reconstructions is a generally accepted and widely used method to assess relationships among organisms. In that framework, morphological characters are translated into a discrete numerical code and are assembled in a data matrix which is then used for phylogenetic tree reconstructions, mainly by maximum parsimony analysis (e.g. Ponder & Lindberg, 1997; Mayr & Clarke, 2003; Reinert et al., 2004; Shultz, 2007). Lately, also likelihood and Bayesian approaches have been discussed (Lewis, 2001; Lee & Worthy, 2012; Wright & Hillis, 2014), but are still not as frequently applied.

The calculation of a phylogeny based on phenotypic molecular characters, as introduced in the present study, is apparently a novel approach allowing for sequence data to be interpreted in a broader context. Although the source of the data is the same, i.e. a nucleotide sequence, the type of characters inferred from that raw data is quite different. While in sequence-based phylogenies of the rRNA genes, for example, the data is aligned according to secondary structure motifs in order to identify homologous bases that can be compared, the actual secondary structure is not reflected in the analysis. When interpreting the data in the frame of a phenotypic molecular approach though, a certain paired position in the stem of a helix can be defined as a character with character states such as CBCs, hemi CBCs or mismatches, thus including secondary structural information in the analysis. Also length variabilities, which are usually discarded due to the inability to homologize bases, may be taken into consideration as helix-, loop- or complete gene lengths are discrete entities that can be integrated in the data matrix. A protein sequence can be translated into phenotypic molecular characters by taking the nature of residue substitutions into consideration, i.e. conservative changes and those with different side chain properties.

A frequently discussed difficulty of character coding in morphology is that every definition of character states in non-binary characters holds a certain level of subjectivity, as borders are often fluid and may be a matter of interpretation (Scotland et al., 2003 and references therein). With sequence data used in a phenotypical context, this problem does not arise as character states are unambiguous. There is however

always the difficulty as to how many character states should be allowed for a certain character. In the above mentioned example of a paired helix region, one possibility would be to distinguish between CBCs, hemi CBCs and mismatches. On the other hand, each nucleotide combination, paired or unpaired, could be defined as separate character states thus increasing their number and lessen the impact of the individual character on the phylogeny. In the present study the number of character states per character was not constrained in order to depict the complete variability and to avoid any interpretation biases.

The maximum parsimony tree inferred from all phenotypic molecular characters (Figure 11, p. 57) was in remarkable congruence with that of the *Microthamnion* phylogeny (Figure 6, p. 34). Clade boundaries were reinforced as strains designated to one clade in Figure 6 also fell into a single clade here. However, only clades 1, 3 and 4 were retrieved monophyletic. Clade 4 and interestingly clade 3, which gained only moderate statistical support in the phylogeny of Figure 6, were maximally supported. Clade 1, on the other hand, received only 79 points. The strains of clade 2, which already had a very short clade branch in the *Microthamnion* phylogeny based on aligned sequence data, were only distinguished through the surrounding topology. Within "clade 2" the strains with sequence data available for more than the *rbc*L however were grouped in a monophyletic cluster to the exclusion of strain CCAC 2081 in which also the ITS2 was sequenced. The strains with only *rbc*L data analyzed, grouped with that strain at the same branch.

The topology did however slightly differ from the tree calculated with aligned sequence data, as the phenotypic molecular data favored a closer relationship of the strains of clade 2 with clade 3 instead of clade 1. In the strains of grades 1 and 2 the topology was more divergent from that of Figure 6, which is not further surprising as it was not supported. The composition of individual strains grouped together by a long branch though, was consistent in both trees.

The nodes referring to putative genus boundaries were however again congruent to those in Figure 6. Interestingly, with phenotypic molecular data, the node marking the smallest possible *Microthamnion* boundary, i.e. the one summarizing clades 1 to 3, received only moderate support whereas the branch for the widest border (clades 1 to 4 plus strains CCAC 4856/4857) as well as the node summarizing clades 1 to 4 were maximally supported.

4.2.3. *In Silico* RFLP Analysis on 18S rDNA Data

Among the phenotypic molecular data collected for alternate tree reconstructions, the 18S rDNA cutting sites held a special status as the *in silico* restriction digest on *Microthamnion* SSU data mirrored the first, and to date only, molecular approach of species delimitation in *Microthamnion* (John et al., 1993), which reduced the genus to (probable) monospecificy. The *in silico* experiment in this study was performed to retrace whether the authors would have come to the same conclusion had they had the same strains available.

Three of the *Microthamnion* strains investigated by John and coworkers are still available through algal culture collections: SAG 114.80, SAG 115.80 (both clade 2) and UTEX LB 237 (clade 3). The 18S was sequenced for strains SAG 114.80 and UTEX LB 237 in the present study. SSU data for strain SAG 115.80 was not obtained because in strain CCAC 3676 the 18S was already sequenced and additional 18S data was not required for further analyses. Since the *rbc*L in both strains was completely identical it was deemed safe to assume that their sequences in the much more conserved SSU were also identical, and thus CCAC 3676 could serve as a substitute for SAG 115.80. All of the above mentioned 18S sequences showed the same RFLP pattern as the database sequence of *Microthamnion kuetzingianum* UTEX 1914 (accession number Z28974) when digested with the same enzymes as in the 1993 study. It was thus chosen as reference not only for the strains investigated by John et al. (1993), but also for the majority of the strains analyzed in the present study as they represented the same SSU genotype.

The larger amount of strains available now went along with higher genetic diversity, as variations in the RFLP patterns were found with four of the above mentioned enzymes regarding strains CCAC 2279/4234, CCAC 5561/5547 B and CCAC 3677/4161 of grade 2 (Figure 8, p. 40 and Figure 9, p. 41), which were significant enough that they would have been detected had the authors of the original study had the same strains available for their experiments. It might thus be speculated that, based on this result, a differentiation into four *Microthamnion* species would have been assumed.

But as the present study shows, all strains with varying RFLP patterns fall into grade 2 and form individual lineages there. As previously discussed, there is molecular support for the assumption that these lineages may have to be raised to genus status in the future, which would in turn prove the restriction digestion method only suitable to distinguish among (some) genera, but inept to discriminate between *Microthamnion* species as the SSU data were too conserved.

4.3. *Excursus* to Habitats and Sampling Strategies

The information available on sampling locations of the *Microthamnion* strains investigated in this study (Table 1, p. 13) was in part rather scant and did not allow for a detailed habitat discussion. Nevertheless, even from the limited information, there were some interesting findings and trends observable which will be explored in this chapter.

4.3.1. Sampling Sites and Habitats

Sampling sites were mainly distributed all over Europe, spanning a region from Scotland in the North, Slovakia in the East and the Canary Islands in the Southwest, but also included locations in the United States of America, Japan and the South Orkney

Islands in the Antarctic. When locality information was mapped on the *Microthamnion* phylogeny (see Supplementary Table 5, p. 106, for a list of strains ordered and colored according to the phylogeny of Figure 6), it became clear that strains belonging to one clade or even the exact same genotype (i.e. identical sequences in all obtained sequence data) were not restricted to a certain geographical region. For example, the strains forming the dark green group of clade 4 originated from the Antarctic, the Black Forest in Germany and the Czech Republic. The sequence-identical strains SAG 114.80, ACOI 1817 and CCAC 4821 of clade 2 were sampled in Serra da Peneda in Portugal, the Eifel in Germany and an unknown locality and strains CCAC 2011, CCAC 0539 B and CCAC 0054 of clade 3 originated from two German regions (Eifel and Harz) and Cornwall in England.

But also within a certain geographical area, the genetic variability was quite diverse. The German Eifel and Wahner Heide as well as the Austrian Waldviertel were extensively sampled and revealed a great range of diversity among the *Microthamnion* strains. When compared with the phylogeny, strains found in the Eifel, for example, were allocated to clades 1 to 3 and both grades, those from Wahner Heide clustered in clade 3 and grade 2 and those from Waldviertel belonged to clades 1 to 3.

The sampling sites within these geographical regions were however variegated in, for example, type of water body (lentic, lotic), size (e.g. lake, pond, puddle), nutrient concentration or conductivity, and also sampling strategies differed (e.g. bailed- or scratched samples, squeezed water plants). In the Eifel, for instance, samples were taken from a dam, a dry maar, a footmark, a leaf collected from a shaded pond, a gutter in a bog, an industrial area and other sites without further specification. Taking into account their diverse nature, these sampling sites may very well represent different habitats, but their exact specifications cannot be inferred from the information available.

Within one natural sample, isolates were often genetically identical in the markers investigated (e.g. CCAC 3547/4819/3546/4818/8001 B from Barranco de Azuaje, Gran Canaria, Spain or strains CCAC 4853/4854/4855 from a roadside ditch in Wahner Heide, Germany). But there were also cases, where a single sample held two genotypes which were allocated to different clades or even grades in the *Microthamnion* phylogeny. This refers to strains CCAC 2771 (clade 1) and CCAC 2764 B (clade 2) which were both isolated from a sample collected in the Austrian Waldviertel (“Fuchsteich”) and also CCAC 3710/3676 B (clade 2) and CCAC 3677/3838 (grade 2), which originated from a sample taken from a gutter in a bog among *Polytrichum commune* (Mützenicher Venn, Eifel, Germany). Interestingly, in the latter case, the two strains allocated to one clade or grade had again identical sequences among each other. Whether that means that one habitat can hold more than one species or if one sample tapped several habitats, can at the moment not be concluded as more detailed information is needed to address such a question.

But also seemingly identical habitats from different localities could hold varying genotypes. In the Austrian Waldviertel, squeezed *Utricularia* probes were taken at two individual sampling sites. One held the genetically identical strains CCAC 4559 B/ 4558 B/4549 B, which were positioned in clade 2 in the *Microthamnion* phylogeny, and the other harbored strains CCAC 4544 B/4717 B/M 4555 (also identical in sequence) from clade 3.

Although there could be no concrete pattern of habitat preferences detected, there were nevertheless tendencies observed. In general, strains from 'extremer' habitats or more creative sampling sites seemed to cluster in grades 1 and 2 whereas strains isolated from more ordinary locations were rather distributed in the clades. For instance, most of the strains originating from waste water plants were allocated to grade 1 (CCAC 5585/5521 and CCAC 5545/5520 B) and grade 2 (CCAC 5547 B/5561). But also isolates from a sample collected in a roadside ditch (CCAC 4853/4854/4855) or an industrial area (CCAC 2279) were assigned a basal position in the *Microthamnion* tree. On the other hand, strains of clade 1 were isolated mainly from rivers, canals and ponds, for most of which the sampling method (e.g. bailed or scratched sample) was not available. Clade 2 comprised a lot of strains sampled from bogs, which were sometimes obtained from squeezed water plants, and for those strains of clade 3 where information was available, also ponds or bogs were the dominant sampling sites. There were however exceptions to these trends as, for example, clade 2 also held strains sampled in a waste water plant (CCAC 5530 B), a temporary puddle (CCAC 2081) or a footmark (CCAC 4821) and grade 2 also comprised strains from bog samples (CCAC 3677/CCAC 3838) or a Scottish loch (CCAC 4161).

Some of the strains were listed with a species designation in the respective culture collections. There was however no correlation of these putative species to the phylogeny, as strains identified as *M. kuetzingianum* or *M. strictissimum* were distributed evenly over the different clades and grades.

4.3.2. Sampling Strategies and Environmental Sequences

A conspicuous finding was that BLAST searches with *Microthamnion* 18S, ITS2, 28S and *rbc*L query sequences had relatively few hits. There were only four environmental 18S sequences found, which were augmented by one 18S and one *rbc*L sequence referring to strains in culture. Considering the global distribution of *Microthamnion*, the wide range of habitats and its adaptability to extreme conditions (e.g. Hargreaves et al., 1975; Foster, 1982 and waste water), the few number of environmental sequences was rather surprising.

This might be explained by an observation made in the CCAC and the workgroup Melkonian (Prof. Dr. Michael Melkonian, University of Cologne, personal communication). There were some natural samples, i.e. those from bogs, but also those from waste water facilities, where *Microthamnion* was immediately visible and quite abundant. On the other hand, there were often samples where *Microthamnion* was not

observed directly after sampling but was found only after some time in enrichment cultures, where it then thrived. As environmental probes are usually sequenced immediately after sampling, *Microthamnion* might not always be detected as its abundance might be too low compared with other organisms. The localities where *Microthamnion* is quite abundant though (see above), are not the ones that are usually sampled in environmental studies.

All the more remarkable was the detection of the environmental clones TE204A and TE204B (accession numbers KM870711 and KM870712), which both represented new SSU sequence-types in the 18S single-gene analysis (Supplementary Figure 1, p. 97), but could unfortunately not be incorporated in the *Microthamnion* phylogeny (Figure 6). Due to the overall high degree of conservation in the SSU, strains where only 18S data was available were positioned randomly in the multi-gene phylogeny and decreased the support values tremendously (preliminary analyses, not shown); they were thus excluded.

The two above mentioned *Microthamnion* environmental clones were, together with most of the hitherto unidentified new members of the *Characium* clade, sampled and sequenced during a study of Schmidt and Darcy (2015). Although they were not highlighted in the phylogenies presented in that study (due to the restriction to the Ulotrichales), their sequences were made available through the NCBI database. The samples were taken from a terrestrial location, i.e. glacial debris on top of the Middle Fork Toklat Glacier in Alaska. This extreme environment, which was "above the highest extent of vascular and nonvascular plants but below the permanent ice line", was however assumed not to be dry during the whole year, but rather to transition between wet periods during snow melt and extended dry periods (Schmidt & Darcy, 2015). It can only be speculated, whether *Microthamnion* has mechanisms to survive such extreme conditions over a longer period of time in a vegetative state, or whether it forms resting stages that are yet unknown.

As already discussed above, most of the lineages in grades 1 and 2 originated from unusual locations with at times extremer conditions and, as the two environmental clones show, there is even more genetic diversity within *Microthamnion*, albeit not available in culture. In order to solve the problem of undersampling in that part of the phylogeny, and also to find even more *Microthamnion* variability as well as additional outgroup taxa, the focus should thus lie on exploring more unusual habitats and locations and to sample them extensively.

5. Conclusions and Outlook

The present study delineated several *Microthamnion* species with molecular methods. Concatenated analyses of the nuclear-encoded 18S and 28S rRNA genes as well as the ITS2 molecule and the plastid-encoded *rbc*L gene, allowed for the identification of four monophyletic clades (with a subdivision in clade 4) and nine distinct lineages. The boundaries for these 14 putative species were reinforced with both, unique molecular signatures (NHSs and autapomorphies) and the phylogeny inferred from phenotypic molecular characters.

The nine lineages of grades 1 and 2 were even so distantly related to the strains allocated to clades 1 to 4, that they might have to be raised to genus status in the future. Without further investigations though, especially detailed morphological studies, a diagnosis cannot be made. For the time being it should thus be refrained from arising new genus borders.

The *Microthamnion* phylogeny in which all available sequence data was incorporated, revealed a high abundance of strains allocated to clades 1 to 3, whereas the lineages in grades 1 and 2 were represented by either very few, or only single strains. In order to clarify the topology in the basal part of the tree as well as to break up the individual long branches, an increased taxon sampling is recommended. As a thorough database search did not reveal any further sequences, future studies will have to include additional field work. The emphasis should lie on exploring unusual locations and sampling sites as strains of grades 1 and 2 seemed to be associated with extremer habitats. The *rbc*L is recommended as the molecular marker to be used to quickly screen new *Microthamnion* isolates for allocations to grades 1 and 2 or to identify new genetic varieties.

The sampling strategy as described above, should also be pursued to increase the number of outgroup taxa for future *Microthamnion* tree reconstructions. Although a thorough database search and the subsequent calculation of a trebouxiophycean phylogeny revealed some additional candidates, the outgroup would still benefit from more taxa and also from more closely related organisms.

Based on the present study, strains representing putative species can be selected for detailed morphological studies on axenic cultures under identical conditions, which will be needed in order to give species and perhaps even genus diagnoses.

6. References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. **1990**. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403–410.

Bakker ME. **1995**. A deviating pattern of cell division in the green alga *Microthamnion*: ultrastructure of vegetative cell division and zoosporogenesis. *Archiv für Protistenkunde* **146**: 117–136.

Bass D, Richards TA, Matthai L, Marsh V, Cavalier-Smith T. **2007**. DNA evidence for global dispersal and probable endemicity of protozoa. *BMC Evolutionary Biology* **7**: 162.

Belton GS, Prud'homme van Reine WF, Huisman JM, Draisma SGA, Gurgel CFD. **2014**. Resolving phenotypic plasticity and species designation in the morphologically challenging *Caulerpa racemosa-peltata* complex (Caulerpaceae, Chlorophyta). *Journal of Phycology* **54**: 32–54.

Brown B, Emberson RM, Paterson AM. **1999**. Mitochondrial COI and II provide useful markers for *Wiseana* (Lepidoptera: Hepialidae) species identification. *Bulletin of Entomological Research* **89**: 287–293.

Bucklin A, Guarnieri M, Hill RS, Bentley AM, Kaartvedt S. **1999**. Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR. *Hydrobiologia* **401**: 239–254.

Caisová L, Marin B, Melkonian M. **2011a**. A close-up view on ITS2 evolution and speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). *BMC Evolutionary Biology* **11**: 262.

Caisová L, Marin B, Melkonian M. **2013**. A consensus secondary structure of ITS2 in the Chlorophyta identified by phylogenetic reconstruction. *Protist* **164**: 482–496.

Caisová L, Marin B, Sausen N, Melkonian M. **2011b**. Polyphyly of *Chaetophora* and *Stigeoclonium* within the Chaetophorales (Chlorophyceae), revealed by sequence comparisons of nuclear-encoded SSU rRNA genes. *Journal of Phycology* **47**: 164–177.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, *et al.* 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **5**: e8613.

Cho Y, Qiu Y-L, Kuhlman P, Palmer JD. **1998**. Explosive invasion of plant mitochondria by a group I intron. *PNAS* **95**: 14244–14249.

Coleman AW. **2000**. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* **151**: 1–9.

Coleman AW. **2001**. Biogeography and speciation in the *Pandorina/Volvulina* (Chlorophyta) superclade. *Journal of Phycology* **37**: 836–851.

Coleman AW. **2007**. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research* **35**: 3322–3329.

Cooke MC. **1884**. *British Fresh-water Algae. Exclusive of Desmidieae and Diatomaceae*. London: Williams & Norgate.

Côté CA, Greer CL, Peculis BA. **2002**. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* **8**: 786–797.

Darienko T, Gustavs L, Mudimu O, Menendez CR, Schumann R, Karsten U, Friedl T, Pröschold T. **2010**. *Chloroidium*, a common terrestrial coccoid green alga previously assigned to *Chlorella* (Trebouxiophyceae, Chlorophyta). *European Journal of Phycology* **45**: 79–95.

Darienko T, Pröschold T. **2015**. Genetic variability and taxonomic revision of the genus *Auxenochlorella* (Shihira et Krauss) Kalina et Puncocharova (Trebouxiophyceae, Chlorophyta). *Journal of Phycology* **51**: 394–400.

Dettman JR, Jacobson DJ, Taylor JW. **2003**. A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* **57**: 2703–2720.

Dobzhansky T. **1937**. *Genetics and the Origin of Species*. New York: Columbia University Press.

Dupuis JR, Roe AD, Sperling FAH. **2012**. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology* **21**: 4422–4436.

Eliáš M, Neustupa J, Škaloud P. **2008**. *Elliptochloris bilobata* var. *corticola* var. nov. (Trebouxiophyceae, Chlorophyta), a novel subaerial coccal green alga. *Biologia* **63**: 791–798.

Felsenstein J. **1978**. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*: 401–410.

Felsenstein J. **1985**. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Foster PL. **1982**. Species associations and metal contents of algae from rivers polluted by heavy metals. *Freshwater Biology* **12**: 17–39.

Friedl T. **1995**. Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae: A phylogenetic analysisis of 18S ribosomal RNA sequences from *Dictyochloropsis reticulata* and from members of the genus *Myrmecia* (Chlorophyta, Trebouxiophycea). *Journal of Phycology* **31**: 632–639.

Fučíková K, Lewis PO, Lewis LA. **2014**. Widespread desert affiliation of trebouxiophycean algae (Trebouxiophyceae, Chlorophyta) including discovery of three new desert genera. *Phycological Research* **62**: 294–305.

Fučíková K, Rada JC, Lukešová A, Lewis LA. **2011**. Cryptic diversity within the genus *Pseudomuriella hanagata* (Chlorophyta, Chlorophyceae, Sphaeropleales) assessed using four barcode markers. *Nova Hedwigia* **93**: 29–46.

Gaysina L, Němcová Y, Škaloud P, Ševčícová T, Eliáš M. **2013**. *Chloropyrula uraliensis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a new green coccoid alga with a unique ultrastructure, isolated from soil in South Urals. *Journal of Systematics and Evolution* **51**: 476–484.

Gonçalves H, Martínez-Solano I, Ferrand N, García-París M. **2007**. Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, *Alytes*): deep coalescence or ancestral hybridization? *Molecular Phylogenetics and Evolution* **44**: 494–500.

Gouy M, Guindon S, Gascuel O. **2010**. SeaView Version 4 : A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**: 221–224.

Graham SW, Olmstead RG, Barrett SCH. **2002**. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular Biology and Evolution* **19**: 1769–1781.

Granneman S, Petfalski E, Tollervey D. **2011**. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *The EMBO Journal* **30**: 4006–4019.

Graybeal A. **1998**. Is it better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology* **47**: 9–17.

Greger J. **1915**. Beiträge zur Kenntnis der Entwicklung und Fortpflanzung der Gattung *Microthamnion* Naeg. *Hedwigia* **56**: 374–380.

Guillou L, Eikrem W, Chrétiennot-Dinet M-J, Le Gall F, Massana R, Romari K, Pedrós-Alió C, Vaulot D. **2004**. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist* **155**: 193–214.

Guiry MD, Guiry GM. **2019**. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org; searched on 13 June 2019.

Hadi SIIA, Santana H, Brunale PPM, Gomes TG, Oliveira MD, Matthiensen A, Oliveira MEC, Silva FCP, Basil BSAF. **2016**. DNA barcoding green microalgae isolated from neotropical inland waters. *PLoS ONE* **11**: e0149284.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. **2006**. DNA barcodes distinguish species of tropical *Lepidoptera*. *PNAS* **103**: 968–971.

Hall JD, Fučíková K, Lo C, Lewis LA, Karol KG. **2010**. An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogamie, Algologie* **31**: 529–555.

Hallmann C, Hoppert M, Mudimu O, Friedl T. **2016**. Biodiversity of green algae covering artificial hard substrate surfaces in a suburban environment: a case study using molecular approaches. *Journal of Phycology* **52**: 732–744.

Hansgirg A. **1886**. Prodromus der Algenflora von Böhmen. Erster Theil. Die Rhodophyceen, Phaeophyceen und ein Theil der Chlorophyceen. *Archiv für die naturwissenschaftliche Landesdurchforschung von Böhmen* **5**: 1–96.

Hargreaves JW, Lloyd EJH, Whitton BA. **1975**. Chemistry and vegetation of highly acidic streams. *Freshwater Biology* **5**: 563–576.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR. **2003**. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**: 313–321.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. **2004**. Identification of birds through DNA barcodes. *PLoS BIOLOGY* **2**: e312.

Hedtke SM, Townsend TM, Hillis DM. **2006**. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* **55**: 522–529.

Heering W. **1914**. Chlorophyceae, III. Ulotrichales, Microsporales, Oedogoniales. In: Pascher A, ed. Die Süßwasser-Flora Deutschlands, Österreichs und der Schweiz. Jena: Fischer.

Heesch S, Broom JES, Neill KF, Farr TJ, Dalen JL, Nelson WA. **2009**. *Ulva*, *Umbraulva* and *Gemina*: genetic survey of New Zealand taxa reveals diversity and introduced species. *European Journal of Phycology* **44**: 143–154.

Hendrich L, Pons J, Ribera I, Balke M. **2010**. Mitochondrial *cox*1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS ONE* **5**: e14448.

Hennig W. **1950**. *Grundzüge einer Theorie der phylogenetischen Systematik*. Berlin: Deutscher Zentralverlag.

Hoef-Emden K. **2012**. Pitfalls of establishing DNA barcoding systems in protists: the Cryptophyceae as a test case. *PLoS ONE* **7**: e43652.

Huelsenbeck JP. **1997**. Is the Felsenstein zone a fly trap? *Systematic Biology* **46**: 69–74.

Huelsenbeck JP, Bull JJ. **1996**. A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology* **45**: 92–98.

John DM, Bhoday R, Russell SJ, Johnson LR, Gacesa P. **1993**. A molecular and morphological analysis of *Microthamnion* (Chlorophyta, Microthamniales). *Archiv für Protistenkunde* **143**: 33–39.

John DM, Johnson LR. **1987**. Observations on the development morphology, growth rate, and reproduction of *Microthamnion kuetzingianum* Naegeli (Pleurastraceae, Pleurastrales) in culture and a taxonomic assessment of the genus. *Nova Hedwigia* **44**: 25–53.

Joseph N, Krauskopf E, Vera MI, Michot B. **1999**. Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Research* **27**: 4533–4540.

Kulichová J, Škaloud P, Neustupa J. **2014**. Molecular diversity of green corticolous microalgae from two sub-mediterranean European localities. *European Journal of Phycology* **49**: 345–355.

Kützing FT. **1849**. *Species algarum*. Leipzig: F.A. Brockhaus.

Larsson A. **2014**. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**: 3276–3278.

Lee JC, Gutell RR. **2012**. A comparison of the crystal structures of eukaryotic and bacterial SSU ribosomal RNAs reveals common structural features in the hypervariable regions. *PLoS ONE* **7**: e38203.

Lee MSY, Worthy TH. **2012**. Likelihood reinstates *Archaeopteryx* as a primitive bird. *Biology Letters* **8**: 299–303.

Leliaert F, Verbruggen H, Vanormelingen P, Steen F, López-Bautista JM, Zuccarello GC, De Clerck O. **2014**. DNA-based species delimitation in algae. *European Journal of Phycology* **49**: 179–196.

Lemieux C, Otis C, Turmel M. **2014**. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evolutionary Biology* **14**: 211.

Lewis PO. **2001**. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913–925.

Lewis LA, Flechtner VR. **2004**. Cryptic species of *Scenedesmus* (Chlorophyta) from desert soil communities of western North America. *Journal of Phycology* **40**: 1127–1137.

Liu X, Zhu H, Liu B, Liu G, Hu Z. **2017**. Classification of *Planctonema*-like algae, including a new genus *Planctonemopsis* gen. nov., a new species *Planctonema gelatinosum* sp. nov. and a reinstated genus *Psephonema* (Trebouxiophyceae, Chlorophyta). *Journal of Phycology* **53**: 869–879.

Lockhart PJ, Steel MA, Hendy MD, Penny D. **1994**. Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution* **11**: 605–612.

Logares R, Rengefors K, Kremp A, Shalchian-Tabrizi K, Boltovskoy A, Tengs T, Shurtleff A, Klaveness D. **2007**. Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? *Microbial Ecology* **53**: 549–561.

Maggs CA, Douglas SE, Fenety J, Bird CJ. **1992**. A molecular and morphological analysis of the *Gymnogongrus devoniensis* (Rhodophyta) complex in the North Atlantic. *Journal of Phycology* **28**.

Mai JC, Coleman AW. **1997**. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *Journal of Molecular Evolution* **44**: 258–271.

Marin B. **2012**. Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. *Protist* **163**: 778–805.

Marin B, Melkonian M. **2010**. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* **161**: 304–336.

Marin B, Nowack ECM, Melkonian M. **2005**. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* **156**: 425–432.

Marin B, Palm A, Klingberg M, Melkonian M. **2003**. Phylogeny and taxonomic revision of plastid-containing Euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist* **154**: 99–145.

Mattox KR, Stewart KD. **1984**. Classification of the Green Algae: A concept based on comparative cytology. In: Irvine DEG, John DM, eds. Systematics of the Green Algae. London: Academic Press, 29–72.

Mayr E. **1942**. *Systematics and the Origin of Species*. New York: Columbia University Press.

Mayr G, Clarke J. **2003**. The deep divergences of neornithine birds: a phylogenetic analysis of morphological characters. *Cladistics* **19**: 527–553.

Melkonian M. **1984**. Flagellar apparatus ultrastructure in relation to green algal classification. In: Irvine DEG, John DM, eds. Systematics of the Green Algae. London: Academic Press, 73–120.

Melkonian M. **1990**. Chlorophyte orders of uncertain affinities: Order Microthamniales. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ, eds. Handbook of Protoctista. Boston: Jones and Bartlett, 652–654.

Meyer CP, Paulay G. **2005**. DNA barcoding: error rates based on comprehensive sampling. *PLoS BIOLOGY* **3**: e422.

Müller T, Philippi N, Dandekar T, Schultz J, Wolf M. **2007**. Distinguishing species. *RNA* **13**: 1469–1472.

Neustupa J, Eliáš M, Škaloud P, Němcová Y, Šejnohová L. **2011**. *Xylochloris irregularis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a novel subaerial coccoid green alga. *Phycologia* **50**: 57–66.

Neustupa J, Němcová Y, Eliáš M, Škaloud P. **2009**. *Kalinella bambusicola* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a novel coccoid *Chlorella*-like subaerial alga from Southeast Asia. *Phycological Research* **57**: 159–169.

Neustupa J, Němcová Y, Veselá J, Steinová J, Škaloud P. **2013a**. *Leptochlorella corticola* gen. et sp. nov. and *Kalinella apyrenoidosa* sp. nov.: two novel *Chlorella*-like green microalgae (Trebouxiophyceae, Chlorophyta) from subaerial habitats. *International Journal of Systematic and Evolutionary Microbiology* **63**: 377–387.

Neustupa J, Němcová Y, Veselá J, Steinová J, Škaloud P. **2013b**. *Parachloroidium* gen. nov. (Trebouxiophyceae, Chlorophyta) a novel genus of coccoid green algae from subaerial corticolous biofilms. *Phycologia* **52**: 411–421.

Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, Murakami N, Kato M. **2005**. Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella* (Saxifragaceae). *Molecular Biology and Evolution* **22**: 285–296.

Otsuka S, Suda S, Li R, Watanabe M, Oyaizu H, Matsumoto S, Watanabe MM. **1999**. Phylogenetic relationships between toxic and non-toxic strains of the genus *Microcystis* based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiology Letters* **172**: 15–21.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y, Song K. **2000**. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *PNAS* **97**: 6960–6966.

Parsons T, Maggs CA, Douglas SE. **1990**. Plastid DNA restriction analysis links the heteromorphic phases of an apomictic red algal life history. *Journal of Phycology* **26**: 495–500.

Van de Peer Y, Robbrecht E, de Hoog S, Caers A, De Rijk P, De Wachter R. **1999**. Database on the structure of small subunit ribosomal RNA. *Nucleic Acids Research* **27**: 179–183.

Ponder WF, Lindberg DR. **1997**. Towards a phylogeny of gastropod molluscs: an analysis using morphological characters. *Zoological Journal of the Linnean Society* **119**: 83–265.

Prauser H. **1957**. Untersuchungen an *Microthamnion kützingianum* Naeg. *Archiv für Mikrobiologie* **28**: 81–88.

Printz H. **1964**. Die Chaetophoralen der Binnengewässer. *Hydrobiologia* **24**: 284–287.

Pröschold T, Darienko T, Silva PC, Reisser W, Krienitz L. **2011**. The systematics of *Zoochlorella* revisited employing an integrative approach. *Environmental Microbiology* **13**: 350–364.

Pröschold T, Leliaert F. **2007**. Systematics of the green algae: conflict of classic and modern approaches. In: Brodie J, Lewis J, eds. Unravelling the algae: the past, present, and future of algal systematics. Boca Raton: CRC Press, 123–154.

Rabenhorst L. **1863**. *Kryptogamenflora von Sachsen, der Ober-Lausitz, Thüringen und Nordböhmen mit Berücksichtigung der benachbarten Länder*. Leipzig: Verlag von Eduard Kummer.

Reder T. **2015**. Species delimitation in the genus *Microthamnion* using a molecular approach. Master Thesis at the University of Cologne.

Reinert JF, Harbach RE, Kitching IANJ. **2004**. Phylogeny and classification of Aedini (Diptera: Culicidae), based on morphological characters of all life stages. *Zoological Journal of the Linnean Society* **142**: 289–368.

Reinsch PF. **1867**. *Die Algenflora des mittleren Theiles von Franken*. Nürnberg: Verlag Wilhelm Schmid.

Reinsch PF. **1878**. Contributiones ad floram Algarum aquae dulcis Promontorii Bonae Spei. *Journal of the Linnean Society of London Botany* **16**: 245–246.

Rindi F, McIvor L, Sherwood AR, Friedl T, Guiry MD, Sheath RG. **2007**. Molecular phylogeny of the green algal order Prasiolales (Trebouxiophyceae, Chlorophyta). *Journal of Phycology* **43**: 811–822.

Rokas A, Williams BL, King N, Carroll SB. **2003**. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798–804.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. **2012**. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Rosenberg MS, Kumar S. **2001**. Incomplete taxon sampling is not a problem for phylogenetic inference. *PNAS* **98**: 10751–10756.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. **1988**. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.

Sanders WB, Pérez-Ortega S, Nelsen MP, Lücking R, de los Ríos A. **2016**. *Heveochlorella* (Trebouxiophyceae): A little-known genus of unicellular green algae outside the Trebouxiales emerges unexpectedly as a major clade of lichen photobionts in foliicolous communities. *Journal of Phycology* **52**: 840–853.

Sang T, Crawford DJ, Stuessy TF. **1997**. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniacae). *American Journal of Botany* **84**: 1120–1136.

Sanger F, Nicklen S, Coulson A. **1977**. DNA sequencing with chain-terminating inhibitors. *PNAS* **74**: 5463–5467.

Saunders GW. **2008**. A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial cryptic species diversity. 1. The foliose *Dilsea-Neodilsea* complex and *Weeksia*. *Botany* **86**: 773–789.

Schillewaert S, Wacheul L, Lhomme F, Lafontaine DLJ. **2012**. The evolutionarily conserved protein LAS1 is required for pre-rRNA processing at both ends of ITS2. *Molecular and Cellular Biology* **32**: 430–444.

Schmidle W. **1899**. Einige Algen aus preussischen Hochmooren. *Hedwigia* **38**: 156–176.

Schmidt SK, Darcy JL. **2015**. Phylogeny of ulotrichalean algae from extreme high-altitude and high-latitude ecosystems. *Polar Biology* **38**: 689–697.

Schultz J, Maisel S, Gerlach D, Müller T, Wolf M. **2005**. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* **11**: 361–364.

Scotland RW, Olmstead RG, Bennett JR. **2003**. Phylogeny reconstruction: the role of morphology. *Systematic Biology* **52**: 539–548.

Shao P, Chen Y, Zhou H, Yuan J, Qu L-H, Zhao D, Lin Y-S. **2004**. Genetic variability in Gymnodiniaceae ITS regions: implications for species identification and phylogenetic analysis. *Marine Biology* **144**: 215–224.

Shultz JW. **2007**. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zoological Journal of the Linnean Society* **150**: 221–265.

Silva P. **2019**. Index Nominum Algarum. University Herbarium, University of California, Berkeley. Compiled by Paul Silva. http://ucjeps.berkeley.edu/INA.html; searched on 19 June 2019.

Škaloud P, Rindi F. **2013**. Ecological differentiation of cryptic species within an asexual protist morphospecies: a case study of filamentous green alga *Klebsormidium* (Streptophyta). *Journal of Eukaryotic Microbiology* **60**: 350–362.

Škaloud P, Steinová J, Řídká T, Vančurová L, Peksa O. **2015**. Assembling the challenging puzzle of algal biodiversity: species delimitation within the genus *Asterochloris* (Trebouxiophyceae, Chlorophyta). *Journal of Phycology* **51**: 507–527.

Šlapeta J, López-García P, Moreira D. **2006**. Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular Biology and Evolution* **23**: 23–29.

Soltis DE, Kuzoff RK. **1995**. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* **49**: 727–742.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

van der Strate HJ, Boele-Bos SA, Olsen JL, van de Zande L, Stam WT. **2002**. Phylogeographic studies in the tropical seaweed *Cladophoropsis membranacea* (Chlorophyta, Ulvophyceae) reveal a cryptic species complex. *Journal of Phycology* **38**: 572–582.

Suzuki S, Endoh R, Manabe R, Ohkuma M, Hirakawa Y. **2018**. Multiple losses of photosynthesis and convergent reductive genome evolution in the colourless green algae *Prototheca*. *Scientific Reports* **8**: 940.

Swofford DL. **2002**. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0b10*. Sinauer Associates, Sunderland, Massachusetts.

Trewick SA. **2000**. Mitochondrial DNA sequences support allozyme evidence for cryptic radiation of New Zealand *Peripatoides* (Onychophora). *Molecular Ecology* **9**: 269–281.

Tronholm A, Steen F, Tyberghein L, Leliaert F, Verbruggen H, Siguan MAR, De Clerck O. **2010**. Species delimitation, taxonomy, and biogeography of *Dictyota* in Europe (Dictyotales, Phaeophyceae). *Journal of Phycology* **46**: 1301–1321.

Tupa DD. **1974**. An investigation of certain chaetophoralean algae. *Beihefte zur Nova Hedwigia* **46**: 55–60.

Turmel M, Otis C, Lemieux C. **2002a**. 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 that led to land plants. *PNAS* **99**: 11275–11280.

Turmel M, Otis C, Lemieux C. **2002b**. The Complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Molecular Biology and Evolution* **19**: 24–38.

Turmel M, Otis C, Lemieux C. **2003**. The mitochondrial genome of *Chara vulgaris*: Insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *The Plant Cell* **15**: 1888–1903.

Ueno R, Hanagata N, Urano N, Suzuki M. **2005**. Molecular phylogeny and phenotypic variation in the heterotrophic green algal genus *Prototheca* (Trebouxiophyceae, Chlorophyta). *Journal of Phycology* **41**: 1268–1280.

Verbruggen H, De Clerck O, Kooistra WHCF, Coppejans E. **2005**. Molecular and morphometric data pinpoint species boundaries in *Halimeda* section Rhipsalis (Bryopsidales, Chlorophyta). *Journal of Phycology* **41**: 606–621.

Verbruggen H, Leliaert F, Maggs CA, Shimada S, Schils T, Provan J, Booth D, Murphy S, De Clerck O, Littler DS, *et al.* 2007. Species boundaries and phylogenetic relationships within the green algal genus *Codium* (Bryopsidales) based on plastid DNA sequences. *Molecular Phylogenetics and Evolution* **44**: 240–254.

Walton C, Somboon P, O' Loughlin SM, Zhang S, Harbach RE, Linton Y-M, Chen B, Nolan K, Duong S, Fong M-Y, *et al.* 2007. Genetic diversity and molecular identification of mosquito species in the *Anopheles maculatus* group using the ITS2 region of rDNA. *Infection, Genetics and Evolution* **7**: 93–102.

Watson MW. **1975**. Flagellar apparatus, eyespot and behaviour of *Microthamnion kuetzingianum* (Chlorophyceae) zoospores. *Journal of Phycology* **11**: 439–448.

Watson MW, Arnott HJ. **1973**. Ultrastructural morphology of *Microthamnion* zoospores. *Journal of Phycology* **9**: 15–29.

Weber A, Czygan F-C. **1972**. Chlorophylle und Carotinoide der Chaetophorineae (Chlorophyceae, Ulotrichales). *Archiv für Mikrobiologie* **84**: 243–253.

West W, West GS. **1907**. Fresh-water algae from Burma, including a few from Bengal and Madras. *Annals of the Royal Botanic Garden, Calcutta* **6**: 182–183.

Wiemers M, Fiedler K. **2007**. Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* **4:** 8.

Wiens JJ. **2005**. Can incomplete taxa rescue phylogenetic analyses from Long-Branch Attraction? *Systematic Biology* **54**: 731–742.

Wiens JJ. **2006**. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* **39**: 34–42.

Wright AM, Hillis DM. **2014**. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* **9**: e109210.

Wuyts J, Perrière G, Van de Peer Y. **2004**. The European Ribosomal RNA database. *Nucleic Acids Research* **32**: D101–D103.

Wuyts J, De Rijk P, Van de Peer Y, Pison G, Rousseeuw P, De Wachter R. **2000**. Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucleic Acids Research* **28**: 4698–4708.

Young I, Coleman AW. **2004**. The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Molecular Phylogenetics and Evolution* **30**: 236–242.

Zou S, Fei C, Wang C, Gao Z, Bao Y, He M, Wang C. **2016**. How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in *Scenedesmus* (Chlorophyceae). *Nature Publishing Group* **6**: 36822.

7. Supplementary Material

Supplementary Table 1. Bold's Basal Medium (BBM, modified)
Recipe obtained from the CCAC homepage (http://www.ccac.uni-koeln.de/sidebar/growth-media/bbm/), modified. Amount of vitamins tripled and molarity in final culture medium added.

	Components	Stock Solution (1 liter)	Addition per 1 Liter Culture Medium	Concentration in final Medium
1	$NaNO_3$	25.00 g	**10 ml**	2.94 mM
	K_2HPO_4 x 3 H_2O	9.80 g		0.43 mM
	KH_2PO_4	17.50 g		1.29 mM
	NaCl	2.50 g		0.43 mM
2	$CaCl_2$ x 2 H_2O	2.50 g	**10 ml**	0.17 mM
3	$MgSO_4$ x 7 H_2O	7.50 g	**10 ml**	0.30 mM
4	EDTA (Titriplex II)	50.00 g	**1 ml**	0.17 mM
	KOH	31.00 g		0.55 mM
5	$FeSO_4$ x 7 H_2O	4.98 g	**1 ml**	17.9 µM
	H_2SO_4 conc.	1 ml		
6	H_3BO_3	11.42 g	**1 ml**	0.18 mM
7	**Trace metals (autoclave to dissolve)**		**1 ml**	
	$ZnSO_4$ x 7 H_2O	8.82 g		30.67 µM
	MoO_3	0.71 g		4.93 µM
	$CuSO_4$ x 5 H_2O	1.57 g		6.29 µM
	$Co(NO_3)_2$ x 6 H_2O	0.49 g		1.68 µM
	$MnCl_2$ x 4 H_2O	1.44 g		7.28 µM
8	**Vitamins**		**3 ml**	
	Vitamin B_{12}	0.20 mg		0.44 nM
	Biotin	1.00 mg		12.30 nM
	Thiamine-HCl	100.00 mg		0.89 µM
	Niacinamide	0.10 mg		2.43 nM
	pH of the vitamin solution should be around pH 7.0			
	Make up to 1 liter with bidistilled or Milli-Q water (pH should be around 6.6) and autoclave.			

Supplementary Table 2. Isopropanol precipitation protocol

Step	Instruction
1.	Maintain DNA extracts on ice & perpetuate this condition
2.	Add 2/3 volume of isopropanol (pre-cooled at -20°C) and vortex
3.	Incubation at -20°C, at least 1 h
4.	Centrifuge (17,000 g, 15', 4°C) & discard the supernatant carefully
5.	Wash pellet with 1 ml 75% ethanol (pre-cooled at -20°C) and vortex
6.	Centrifuge (17,000 g, 5', 4°C) and discard the supernatant carefully
7.	Repeat the washing step
8.	Air-dry the pellet (hood) & dissolve the pellet in 20 µl sterile nuclease-free water
9.	Store at -20°C

Supplementary Table 3. Sequence data (Overview)

Sequence data available for further analysis is represented by a tick. Without brackets: generated in this study, with brackets: obtained from (Reder, 2015). Markers not sequenced are flagged with a dash. The number of sequenced nucleotides was as follows. 18S: 1661-1705, ITS2 (inclusive flanking regions of 5.8S and 28S): 872-936, 28S: 1439-1446 and rbcL: 1284.

Strain	18S	ITS2	28S	rbcL
ACOI 140	✓	✓	✓	✓
ACOI 1447	—	(✓)	—	✓
ACOI 1620	✓	(✓)	✓	✓
ACOI 1621	✓	✓	✓	✓
ACOI 1817	—	—	—	✓
ACOI 248	✓	(✓)	✓	✓
ACOI 2656	—	(✓)	—	✓
ACOI 2660	(✓)	(✓)	—	✓
ACOI 398	(✓)	(✓)	✓	✓
CAUP J 1201	(✓)	(✓)	✓	✓
CCAC 0054	—	—	—	✓
CCAC 0087	(✓)	(✓)	—	✓
CCAC 0539 B	—	(✓)	—	✓
CCAC 2011	(✓)	(✓)	—	✓
CCAC 2081	—	(✓)	—	✓
CCAC 2182	—	(✓)	—	✓
CCAC 2197 B	—	(✓)	—	✓
CCAC 2198	—	(✓)	—	✓
CCAC 2199 B	—	—	—	✓
CCAC 2223 B	—	(✓)	—	✓
CCAC 2224	(✓)	(✓)	✓	✓
CCAC 2279	(✓)	(✓)	✓	✓
CCAC 2764 B	✓	✓	—	✓
CCAC 2771	(✓)	(✓)	—	✓
CCAC 2804 B	✓	✓	—	✓
CCAC 2916	✓	✓	✓	✓
CCAC 2942 B	✓	✓	✓	✓
CCAC 2943 B	—	(✓)	—	✓
CCAC 3546	(✓)	(✓)	✓	✓
CCAC 3547	—	—	—	✓
CCAC 3664 B	(✓)	(✓)	—	✓
CCAC 3676 B	(✓)	✓	—	✓
CCAC 3677	✓	✓	✓	✓
CCAC 3710	—	—	—	✓
CCAC 3838	—	—	—	✓
CCAC 3842 B	—	—	—	✓
CCAC 3843 B	—	(✓)	—	✓
CCAC 4161	✓	✓	✓	✓
CCAC 4234	(✓)	(✓)	✓	✓
CCAC 4544 B	—	—	—	✓
CCAC 4549 B	—	—	—	✓
CCAC 4558 B	—	—	—	✓
CCAC 4559 B	—	—	—	✓
CCAC 4717 B	—	—	—	✓
CCAC 4818	—	(✓)	—	✓
CCAC 4819	(✓)	(✓)	—	✓
CCAC 4820	(✓)	(✓)	✓	✓
CCAC 4821	—	—	—	✓

Strain	18S	ITS2	28S	rbcL
CCAC 4822	✓	(✓)	✓	✓
CCAC 4853	—	✓	—	✓
CCAC 4854	✓	✓	—	✓
CCAC 4855	(✓)	✓	✓	✓
CCAC 4856	—	✓	—	✓
CCAC 4857	✓	✓	✓	✓
CCAC 5520 B	—	—	—	✓
CCAC 5521	(✓)	(✓)	✓	✓
CCAC 5530 B	(✓)	✓	✓	✓
CCAC 5545	(✓)	(✓)	✓	✓
CCAC 5547 B	(✓)	(✓)	—	✓
CCAC 5561	✓	✓	✓	✓
CCAC 5585	✓	(✓)	✓	✓
CCAC 8001 B	—	—	—	✓
CCAC 8002 B	(✓)	(✓)	—	✓
CCALA 368	(✓)	(✓)	—	✓
CCAP 450/2	(✓)	(✓)	✓	✓
CCAP 450/3	(✓)	(✓)	✓	✓
CCAP 450/4	—	—	—	✓
M 2196/1 A	—	(✓)	—	✓
M 2412/1 A	(✓)	(✓)	✓	✓
M 4555	—	—	—	✓
NIES 479	(✓)	(✓)	—	✓
SAG 114.80	✓	✓	✓	✓
SAG 115.80	—	—	—	✓
UTEX LB 237	(✓)	(✓)	✓	✓
Ch. perforatum SAG 28.85	✓	✓	✓	✓

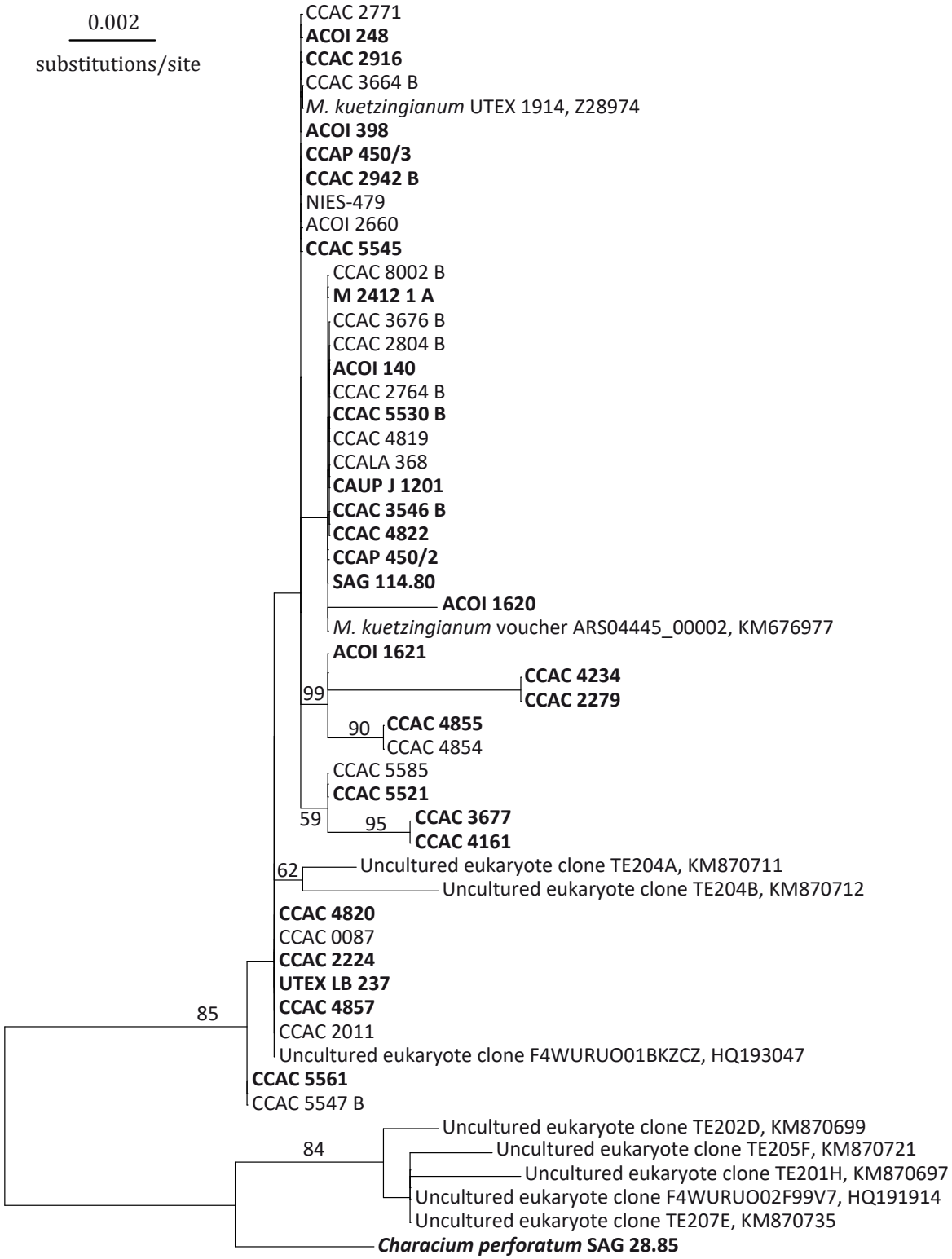
Supplementary Table 4. Data matrix of phenotypic molecular characters used for phylogenetic tree reconstructions.

Listed are all *Microthamnion* and outgroup strains that were also analyzed for Figure 6, further ENV sequences were not taken into consideration. In column "Identical sequence" all strains with identical sequences in the respective molecular marker were given the same number. The identical sequence for ITS2 is given in the matrix (grey font, marked with *) but was not analyzed in the tree reconstructions due to exceeding numbers of allowed character states. The length of ITS2 helices is given in total numbers of nucleotides for each strain. In all other columns the numbers refer to the character states given in the bottom line. A minus sign (-) is applied where no sequence data was available, a question mark (?) stands for single positions not present in the investigated strain. B.t.l. = before terminal loop.

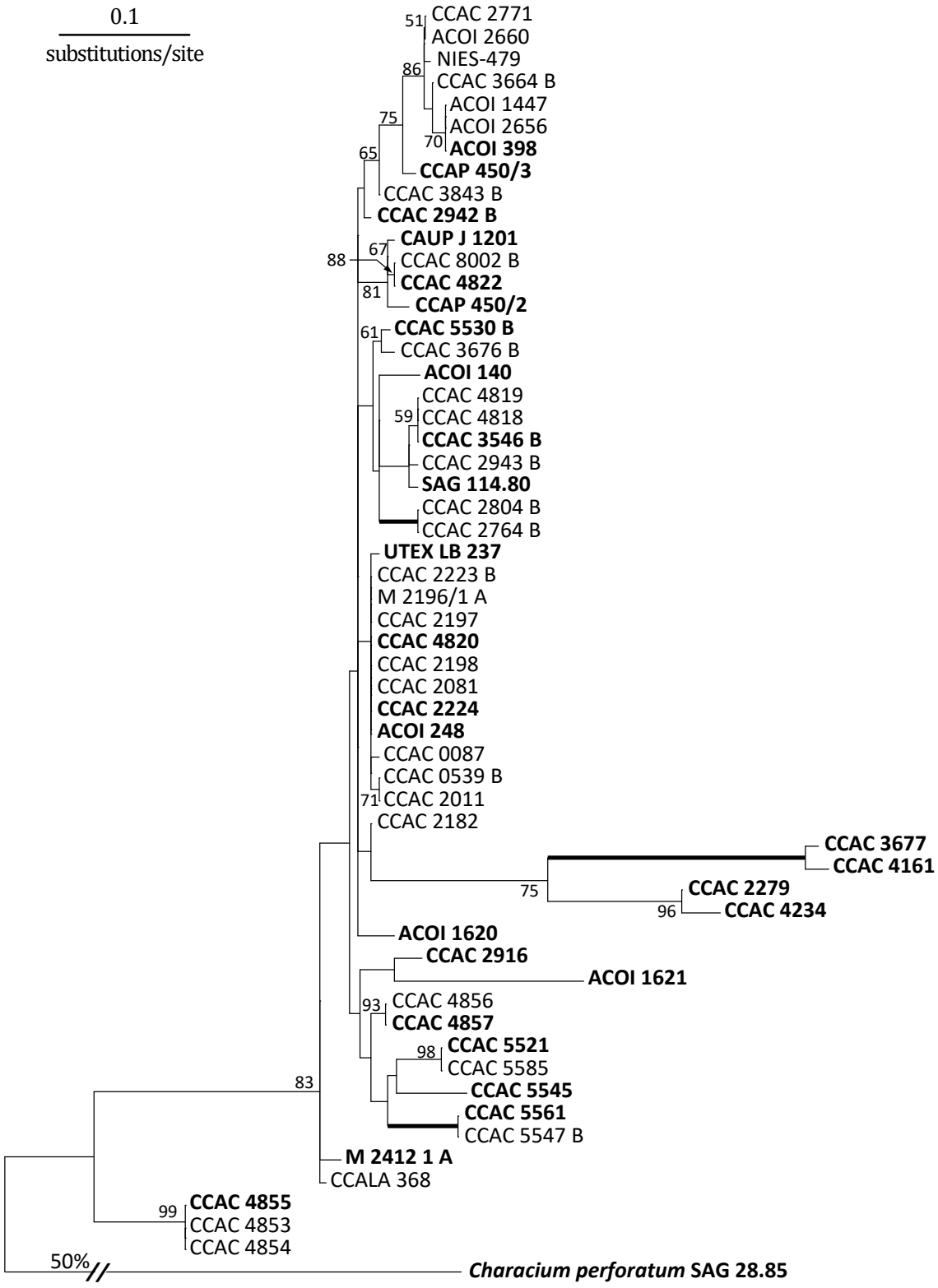
	Identical Sequence						Position of cutting sites in 18S							Total number of cuts per enzyme in 18S				Length of ITS2 Helices [no. of nts]				Apomorphies 18S																Apomorphies 5.8S		
Strain	18S	5.8S (partial)	ITS2*	28S short (89'-821')	28S long (89'-D22')	rbcL AA	H E10_1 (MspI)	H E10_1 (TaqI)	H 11 (CfoI)	H 17 (TaqI)	H E23_2 (CfoI)	H E23_2' (RsaI)	H E23_2' (CfoI)	CfoI	MspI	RsaI	TaqI	H1	H2	H3	H4	H E10_1 2nd bp b.t.l.	H E10_1 last bp b.t.l.	Spacer H E10_1 - H 11	Loop of H 17	H E23_1 bp 3	H E23_1 bp 4	H E23_2 bp 1	H E23_2 bp 4	Loop of H E23_2	H E23_7 bp 1	H E23_7 bp 2	H 46 bp 6	H 48 bp 6	H 88 bp 3	Loop H 88	Spacer H 89 - ITS2 H1
ACOI 140	6	3	28	10	14	11	2	1	2	2	2	2	2	3	2	2	3	50	44	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
ACOI 1447	-	2	14	10	-	14	-	-	-	-	-	-	-	-	-	-	-	44	45	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	1
ACOI 1620	2	1	9	8	8	4	2	1	2	2	2	2	2	3	2	2	3	57	43	89	30	1	2	1	1	1	1	1	1	2	1	1	1	1	2	2	1
ACOI 1621	10	3	3	3	3	10	2	1	2	2	2	2	2	3	2	2	3	46	38	88	53	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2
ACOI 1817	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ACOI 248	11	3	25	10	15	13	2	1	2	2	2	2	2	3	2	2	3	48	44	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ACOI 2656	-	2	14	10	-	14	-	-	-	-	-	-	-	-	-	-	-	44	45	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	1
ACOI 2660	11	2	16	10	-	14	2	1	2	2	2	2	2	3	2	2	3	44	45	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
ACOI 398	11	2	14	10	13	14	2	1	2	2	2	2	2	3	2	2	3	44	45	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
CAUP J 1201	6	3	21	10	9	3	2	1	2	2	2	2	2	3	2	2	3	44	40	89	30	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 0054	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 0087	9	3	34	10	-	13	2	1	2	2	2	2	2	3	2	2	3	50	44	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 0539 B	-	3	33	10	-	13	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 2011	9	3	33	10	-	13	2	1	2	2	2	2	2	3	2	2	3	50	44	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 2081	-	3	38	10	-	14	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 2182	-	3	36	9	-	13	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 2197 B	-	3	38	10	-	13	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 2198	-	3	38	10	-	13	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 2199 B	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 2223 B	-	3	38	10	-	13	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 2224	9	3	38	10	15	13	2	1	2	2	2	2	2	3	2	2	3	50	44	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 2279	1	2	4	5	5	1	1	2	1	2	1	2	1	1	1	2	1	31	44	88	47	2	1	2	1	1	1	2	1	1	1	2	1	1	2	1	1
CCAC 2764 B	6	3	32	10	-	14	2	1	2	2	2	2	2	3	2	2	3	50	44	89	28	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 2771	11	2	16	10	-	14	2	1	2	2	2	2	2	3	2	2	3	44	45	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
CCAC 2804 B	6	3	32	10	-	14	2	1	2	2	2	2	2	3	2	2	3	50	44	89	28	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 2916	11	3	18	10	10	5	2	1	2	2	2	2	2	3	2	2	3	48	40	89	29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 2942 B	11	2	26	10	13	14	2	1	2	2	2	2	2	3	2	2	3	46	42	89	29	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
CCAC 2943 B	-	3	30	10	-	14	-	-	-	-	-	-	-	-	-	-	-	50	45	89	27	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 3546	6	3	31	10	12	14	2	1	2	2	2	2	2	3	2	2	3	50	45	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 3547	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3664 B	11	2	17	10	-	14	2	1	2	2	2	2	2	3	2	2	3	44	45	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
CCAC 3676 B	6	3	11	10	-	14	2	1	2	2	2	2	2	3	2	2	3	53	44	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 3677	5	3	2	2	2	7	2	1	2	1	2	1	2	3	2	1	2	46	38	94	30	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	3
CCAC 3710	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3838	-	-	-	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3842 B	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3843 B	-	2	13	10	-	14	-	-	-	-	-	-	-	-	-	-	-	43	46	89	29	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	1
CCAC 4161	5	3	1	1	1	7	2	1	2	1	2	1	2	3	2	1	2	45	38	94	30	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	3
CCAC 4234	1	2	5	5	5	1	1	2	1	2	1	2	1	1	1	2	1	30	44	88	47	2	1	2	1	1	1	2	1	1	1	2	1	1	2	1	1
CCAC 4544 B	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4549 B	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4558 B	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4559 B	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4717 B	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4818	-	3	31	10	-	14	-	-	-	-	-	-	-	-	-	-	-	50	45	89	27	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 4819	6	3	31	10	-	14	2	1	2	2	2	2	2	3	2	2	3	50	45	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 4820	9	3	37	10	15	13	2	1	2	2	2	2	2	3	2	2	3	50	44	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 4821	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4822	6	3	22	10	9	3	2	1	2	2	2	2	2	3	2	2	3	44	40	89	30	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 4853	-	2	6	6	-	9	-	-	-	-	-	-	-	-	-	-	-	40	40	89	36	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	1
CCAC 4854	3	2	6	6	-	9	2	1	2	2	2	2	2	3	2	2	3	40	40	89	36	1	1	1	1	1	2	1	1	1	1	2	1	1	2	1	1
CCAC 4855	3	2	6	6	6	9	2	1	2	2	2	2	2	3	2	2	3	40	40	89	36	1	1	1	1	1	2	1	1	1	1	2	1	1	2	1	1
CCAC 4856	-	3	20	10	-	6	-	-	-	-	-	-	-	-	-	-	-	51	40	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
CCAC 4857	8	3	20	10	15	6	2	1	2	2	2	2	2	3	2	2	3	51	40	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 5520	-	-	-	-	-	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 5521	7	3	19	7	7	12	2	1	2	2	2	2	2	3	2	2	3	49	42	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 5530 B	6	3	10	10	14	14	2	1	2	2	2	2	2	3	2	2	3	52	44	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 5545	11	3	8	10	11	8	2	1	2	2	2	2	2	3	2	2	3	47	40	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 5547 B	4	3	7	4	-	2	2	1	1	2	2	2	2	2	2	2	3	46	40	88	27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 5561	4	3	7	4	4	2	2	1	1	2	2	2	2	2	2	2	3	46	40	88	27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 5585	7	3	19	7	7	12	2	1	2	2	2	2	2	3	2	2	3	49	42	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 8001 B	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 8002 B	6	3	22	10	-	3	2	1	2	2	2	2	2	3	2	2	3	44	40	89	30	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCALA 368	6	3	23	10	-	14	2	1	2	2	2	2	2	3	2	2	3	48	42	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAP 450/2	6	3	27	10	9	3	2	1	2	2	2	2	2	3	2	2	3	45	40	89	30	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAP 450/3	11	2	12	10	13	14	2	1	2	2	2	2	2	3	2	2	3	43	45	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
CCAP 450/4	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M 2196/1 A	-	3	38	10	-	13	-	-	-	-	-	-	-	-	-	-	-	50	44	89	30	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1
M 2412/1 A	6	3	24	10	9	14	2	1	2	2	2	2	2	3	2	2	3	48	42	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
M 4555	-	-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NIES 479	11	2	15	10	-	14	2	1	2	2	2	2	2	3	2	2	3	44	45	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1
SAG 114.80	6	3	29	10	14	14	2	1	2	2	2	2	2	3	2	2	3	50	48	89	27	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
SAG 115.80	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UTEX 318	-	-	-	-	-	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UTEX LB 237	9	3	35	10	15	13	2	1	2	2	2	2	2	3	2	2	3	50	44	89	30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
UEC KM870699	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	2	1	3	2	1	2	1	2	1	-	-	-
UEC KM870721	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	2	1	3	2	1	2	1	2	1	-	-	-
UEC KM870697	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	2	1	3	2	1	2	1	2	1	-	-	-
UEC KM870735	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	2	1	2	1	3	2	1	2	1	2	1	-	-	-
Ch. perforatum	12	4	39	11	16	15	-	-	-	-	-	-	-	-	-	-	-	30	46	104	60	1	1	2	1	3	1	3	1	1	2	1	2	2	1	1	4
Character states							1=no cut, 2=cut															1=G-C, 2=A-U	1=G-C, 2=A-U	1=CUCC, 2=CUUC	1=UUUC, 2=UUAU	1=G-C, 2=A-U, 3=G-U	1=U-A, 2=U-G	1=C-G, 2=U-U, 3=A-U	1=C-G, 2=U-G	1= UUU, 2= UUC	1=C-G, 2=G-U	1=G-U, 2=G-C	1=A-U, 2=G-U	1=G-C, 2=A-U	1=C-G, 2=U-A	1=CUCG, 2=UUCG	1=UACCCC, 2=UACCUC, 3=ACCCCC, 4=UACUCC

Strain	Apomorphies ITS2																														
	H1 bp 7 (7-I/11-8)	H1 bp 8 (8/11)	H1 last bp b.t.l. (9-VII/9-X)	Spacer H1-H2	H2 bp 5 (23/38)	H2 bp 6 (24/37)	H2 "bp 6.1" (pyr-pyr-missmatch)	H2 bp 9 (28/33)	H2 bp 10 (29/32)	H2 bp 11 (29-I/29-XIV)	H2 4th bp b.t.l. (29-III/29-XII)	Spacer H2-H3	H3 bp 1 (52/123)	H3 bp 2 (53/122)	H3 bp 3 (54/121)	H3 bp 4 (55/120)	H3 bp 5 (55-I/108-XIV)	H3 bp 9 (55-V/108-IX)	H3 bp 14 (55-XVII/108-IV)	H3 bp 23 (78-II/95-VII)	H3 bp 27 (86/95-III)	H3 6th bp b.t.l. (89/94)	H3 5th bp b.t.l. (90/93)	H3 4th bp b.t.l. (91/92)	Spacer H3-H4	H4 bp 1 (124-II/129-I)	H4 bp 2 (125/129)	H4 bp 3 (126/128)	H4 6th bp b.t.l. (126-III/126-XIV)	H4 5th bp b.t.l. (126-IV/126-XIII)	H4 3rd bp b.t.l. (126-VI/126-XI)	H4 2nd bp b.t.l. (126-VII/126-X)
ACOI 140	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	2	1	1	2	1	3	1
ACOI 1447	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
ACOI 1620	4	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	3	1	1	3	1	1	1	1	1	2	4	3	1
ACOI 1621	4	2	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	2	1
ACOI 1817	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ACOI 248	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
ACOI 2656	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
ACOI 2660	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
ACOI 398	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
CAUP J 1201	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	1	4	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 0054	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 0087	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 0539 B	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2011	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2081	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2182	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2197 B	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2198	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2199 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 2223 B	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2224	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2279	5	2	?	1	1	1	1	2	1	2	3	1	1	2	1	1	5	2	2	1	1	1	1	3	1	1	1	1	2	4	3	1
CCAC 2764 B	1	2	2	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	3	2	1	3	2
CCAC 2771	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
CCAC 2804 B	1	2	2	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	3	2	1	3	2
CCAC 2916	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	1	2	1	1	1	1	1	1	2	4	3	1
CCAC 2942 B	1	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 2943 B	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	1	3	1
CCAC 3546	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	2	1	1	2	1	3	1
CCAC 3547	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3664 B	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
CCAC 3676 B	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	1	3	1
CCAC 3677	1	2	1	1	1	4	2	1	1	1	2	1	2	1	3	1	5	2	1	1	1	1	2	1	2	1	1	1	2	4	3	1
CCAC 3710	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3838	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3842 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 3843 B	2	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 4161	1	2	1	1	1	4	2	1	1	1	2	1	2	1	3	2	5	2	1	1	1	1	2	1	2	1	1	1	2	4	3	1
CCAC 4234	5	2	?	1	1	1	1	2	1	2	3	1	1	2	1	1	5	2	2	1	1	1	1	3	1	1	1	1	2	4	3	1
CCAC 4544 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4549 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4558 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4559 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4717 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4818	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	2	1	1	2	1	3	1
CCAC 4819	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	2	1	1	2	1	3	1
CCAC 4820	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 4821	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 4822	1	2	2	1	1	1	1	2	1	1	1	1	2	1	1	1	4	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 4853	6	2	1	1	1	5	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	3	1	1	3	2	1	1	2	3	1
CCAC 4854	6	2	1	1	1	5	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	3	1	1	3	2	1	1	2	3	1
CCAC 4855	6	2	1	1	1	5	1	1	1	1	1	1	1	1	1	2	2	1	2	1	1	1	3	1	1	3	2	1	1	2	3	1
CCAC 4856	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 4857	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 5520	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 5521	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 5530 B	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	1	3	1
CCAC 5545	4	2	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 5547 B	5	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	2	2	4	3	1
CCAC 5561	5	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	2	2	4	3	1
CCAC 5585	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAC 8001 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CCAC 8002 B	1	2	2	1	1	1	1	2	1	1	1	1	2	1	1	1	4	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCALA 368	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	3	3	1
CCAP 450/2	1	2	2	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
CCAP 450/3	3	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	1	1
CCAP 450/4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M 2196/1 A	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
M 2412/1 A	1	2	2	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	3	1	1	1	1	1	2	4	3	1
M 4555	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NIES 479	2	2	2	1	1	2	1	1	1	1	1	1	1	1	3	1	1	1	2	2	1	1	3	1	1	1	1	1	2	4	1	1
SAG 114.80	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	1	3	1
SAG 115.80	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UTEX 318	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UTEX LB 237	1	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	2	4	3	1
UEC KM870699	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UEC KM870721	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UEC KM870697	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UEC KM870735	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ch. perforatum	6	?	3	1	2	4	1	2	2	2	3	3	1	1	1	3	2	3	2	1	1	1	2	1	3	1	2	1	3	1	4	3
Character states	1=G-C, 2=CxC, 3=UxC, 4=AxC, 5=A-U, 6=U-A	1=G-C, 2=C-G, ?=pos. not pres.	1=U-A, 2=G-C, 3=C-G, ?=pos. not pres.	1=AGC, 2=A**UC**, 3=AAC	1=C-G, 2=A-U	1=U-G, 2=UxU, 3=CxU, 4=UxC, 5= G-C	1=UxU, 2=CxC	1=U-G, 2=C-G	1=U-A, 2=U-G	1=C-G, 2=U-G	1=C-G, 2=U-A, 3=G-C	1=AAAGGCAAA 2=AAA**CU**CAAA 3=AAAUGCAAA	1=G-C, 2=G-U	1=C-G, 2=CxA	1=C-G, 2=U-G, 3=CxA	1=G-C, 2=G-U, 3=C-G	1=U-G, 2=C-G, 3=UxC, 4=UxU, 5=U-A	1=U-G, 2=UxU, 3=U-A	1=G-C, 2=A-U	1=A-U, 2=G-U, 3=AxC	1=A-U, 2=G-U	1=C-G, 2=U-G	1=C-G, 2=U-G, 3=U-A	1=G-C, 2=U-G, 3=UxU	1=UGGAACGC 2=CG**UCC**CGC 3=GACCUGC	1=A-U, 2=AxC, 3=AxA	1=G-U, 2=G-C	1=G-C, 2= GxG, 3=G-U	1=A-U, 2=C-G, 3=G-C	1=C-G, 2=G-C, 3=U-G, 4=U-A	1=A-U, 2=G-U, 3=G-C, 4= AxC	1=G-C, 2=G-U, 3=U-G

Strain	Apomorphies 28S: Spacer H B12' - H B13	H B14 3rd bp b.t.l.	H B14 last bp b.t.l.	H B17 bp 3	H B17 bp 6	H B17 bp 7	H B17 3rd bp b.t.l.	H C1_E1 bp 3	H C1_E1 bp 4	H C1_E1 bp 5	H C1_E1 bp 9	H C1_E1 bp 12	H C1_E1 bp 13	H C1_E1 bp 14	H C1_E1 6th bp b.t.l.	H C1_E1 2nd bp b.t.l.	Loop H C1_E1	H C1_E3 bp 6	H C1_E3 bp 7	H C1_E3 bp 8	H C1_E3 bp 9	H C1_E3 bp 10	H C1_E3 bp 11	H C1_E3 bp 12	H C1_E3 bp 15	H C1_E3 4th bp b.t.l.	H C1_E3 3rd bp b.t.l.	H D14 last bp b.t.l.	H D15 4th bp b.t.l.	H D15 3rd bp b.t.l.	H D18 bp 3	H D22 bp 6	Apomorphies *rbc* L AA [position in ref. to UTEX 318, KM462876]: 50	168	228	230	245	255	279	282	284	287	374	428
ACOI 140	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
ACOI 1447	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
ACOI 1620	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	2	1	1	1
ACOI 1621	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	2	1	1	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	2	2	1	2	1
ACOI 1817	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
ACOI 248	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1
ACOI 2656	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
ACOI 2660	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
ACOI 398	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CAUP J 1201	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	2	2	1	3
CCAC 0054	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	1	3	1	1	1	1
CCAC 0087	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 0539 B	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 2011	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 2081	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 2182	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 2197 B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
CCAC 2198	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 2199 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 2223 B	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 2224	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
CCAC 2279	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	3	4	2	4	2	2	1	2	1	1	1	1	1	2	2	1	2	2	1	2	2	1	2	1	2	1	1	1
CCAC 2764 B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 2771	1	1	1	1	1	1	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 2804 B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 2916	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	3	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	3	1	2	2	2	1	2	1
CCAC 2942 B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 2943 B	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 3546	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 3547	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 3664 B	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 3676 B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 3677	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	3	2	1	1	1	1	2	2	2	1	1	1	1	2	2	1	5	2	1	1	1
CCAC 3710	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 3838	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	2	1	2	2	2	5	2	1	1	1
CCAC 3842 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 3843 B	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
CCAC 4161	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	3	2	1	1	1	2	2	2	1	1	1	1	1	1	2	2	1	5	2	1	1	1
CCAC 4234	1	-	-	-	-	-	-	1	1	1	2	1	1	1	2	1	1	3	4	2	4	2	2	1	2	1	3	1	1	1	1	2	2	1	1	2	2	1	2	2	1	2	1	2	1	1
CCAC 4544 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	1	3	1	1	1	1
CCAC 4549 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 4558 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 4559 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 4717 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	1	3	1	1	1	1
CCAC 4818	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 4819	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 4820	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	3	1	1	1	1
CCAC 4821	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 4822	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	2	2	1	3
CCAC 4853	1	1	2	1	2	2	1	1	2	2	2	1	1	1	3	1	2	1	2	1	2	1	2	2	2	2	1	1	1	1	1	1	1	1	1	2	2	4	2	1	2	1
CCAC 4854	1	1	2	1	2	2	1	1	2	2	2	1	1	1	3	1	2	1	2	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	2	2	4	2	1	2	1
CCAC 4855	1	1	2	1	2	2	1	1	2	2	2	1	1	1	3	1	2	1	2	1	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	2	2	4	2	1	2	1
CCAC 4856	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	3	1	2	2	2	1	1	1
CCAC 4857	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	3	2	2	1	1	1	1	1
CCAC 5520	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	2	1	3	1	2	2	2	1	2	1
CCAC 5521	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	1	2	1	2	1
CCAC 5530 B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 5545	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	3	2	-	-	-	1	1	1	1	1	1	1	1	1	2	2	1	3	1	2	2	2	1	2	1	1
CCAC 5547 B	2	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	-	-	-	-	-	-	-	3	2	1	1	2	2	1	4	2	1	2	1	1
CCAC 5561	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	3	2	1	1	2	2	2	1	1	1	1	1	1	3	2	1	1	2	1	2	4	2	1	2	1
CCAC 5585	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	1	2	1	1	1
CCAC 8001 B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAC 8002 B	1	1	1	1	1	1	1	-	-	-	-	2	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2	1	1	1	1	1	2	2	2	1	3
CCALA 368	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
CCAP 450/2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	2	2	1	3
CCAP 450/3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CCAP 450/4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
M 2196/1 A	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	2	3	1	1	1	1
M 2412/1 A	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
M 4555	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	2	1	1	3	1	1	1	1
NIES 479	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1
SAG 114.80	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	1	1	1	1
SAG 115.80	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
UTEX 318	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	6	1	1	1	1
UTEX LB 237	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1
UEC KM870699	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UEC KM870721	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UEC KM870697	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UEC KM870735	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Ch. perforatum	1	2	2	2	1	1	2	2	1	1	2	1	1	1	1	2	1	1	5	2	1	5	2	3	1	2	1	2	2	2	2	2	1	1	2	1	1	4	2	2	2	1	2	2	2
Character states	1=AGU, 2=AGC	1=G–C, 2=C–G	1=C–G, 2=U-A	1=G–C, 2=A–U	1=A-U, 2=G–C	1=G–U, 2=C–G	1=C–G, 2=U-A	1=G–C, 2=A–U	1=GxG, 2=C–G	1=U-A, 2=C–G	1=G–U, 2=G–C	1=U-A, 2=GxA	1=A–U, 2=A-A	1=U–G, 2=C–G	1=G–U, 2=G–C, 3=U–U	1=A-U, 2=G–C	1=GCCA, 2=GC**GA**	1=U–G, 2=C–G, 3=G–U	1=U–A, 2=C–U, 3=G–C, 4=G–G, 5=U–G	1=G–C, 2=C–G, 3=G–G	1=G–G, 2=U-A, 3=C–G, 4=U–G, 5=G–U	1=C–G, 2=G–C	1=U–A, 2=C–G, 3=G–C	1=C–G, 2=A–U	1=G–C, 2=G–U	1=G–C, 2=A–U	1=G–C, 2=A–U	1=C–G, 2=G–C, 3=U–G	1=GxG, 2=G–U	1=G–C, 2=A–U	1=C–G, 2=G–U	1=C–G, 2=U-A	1=G–C, 2=A–U	1=P, 2=S, 3=A	1=G, 2=S	1=S, 2=A	1=A, 2=S	1=G, 2=A, 3=S	1=E, 2=A	1=T, 2=S	1=R, 2=H, 3=Q, 4=N, 5=I, 6=K	1=S, 2=C	1=N, 2=H	1=T, 2=V	1=T, 2=V, 3=I

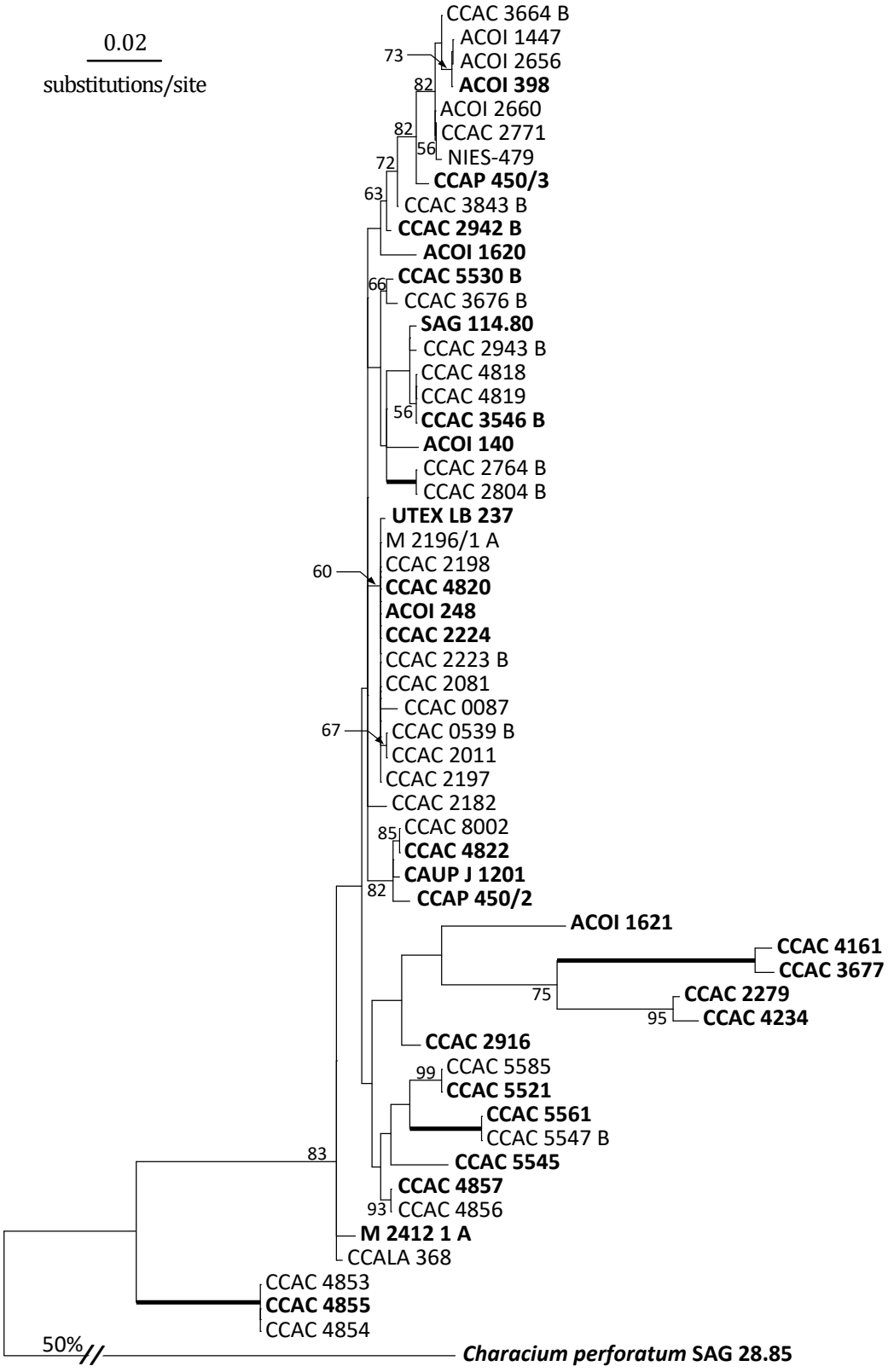


Supplementary Figure 1. 18S single-gene phylogeny
ML topology of 47 *Microthamnion* and six outgroup SSU sequences based on 1588 aligned positions. Bootstraps: ML; strains in bold were selected for a congruent dataset.



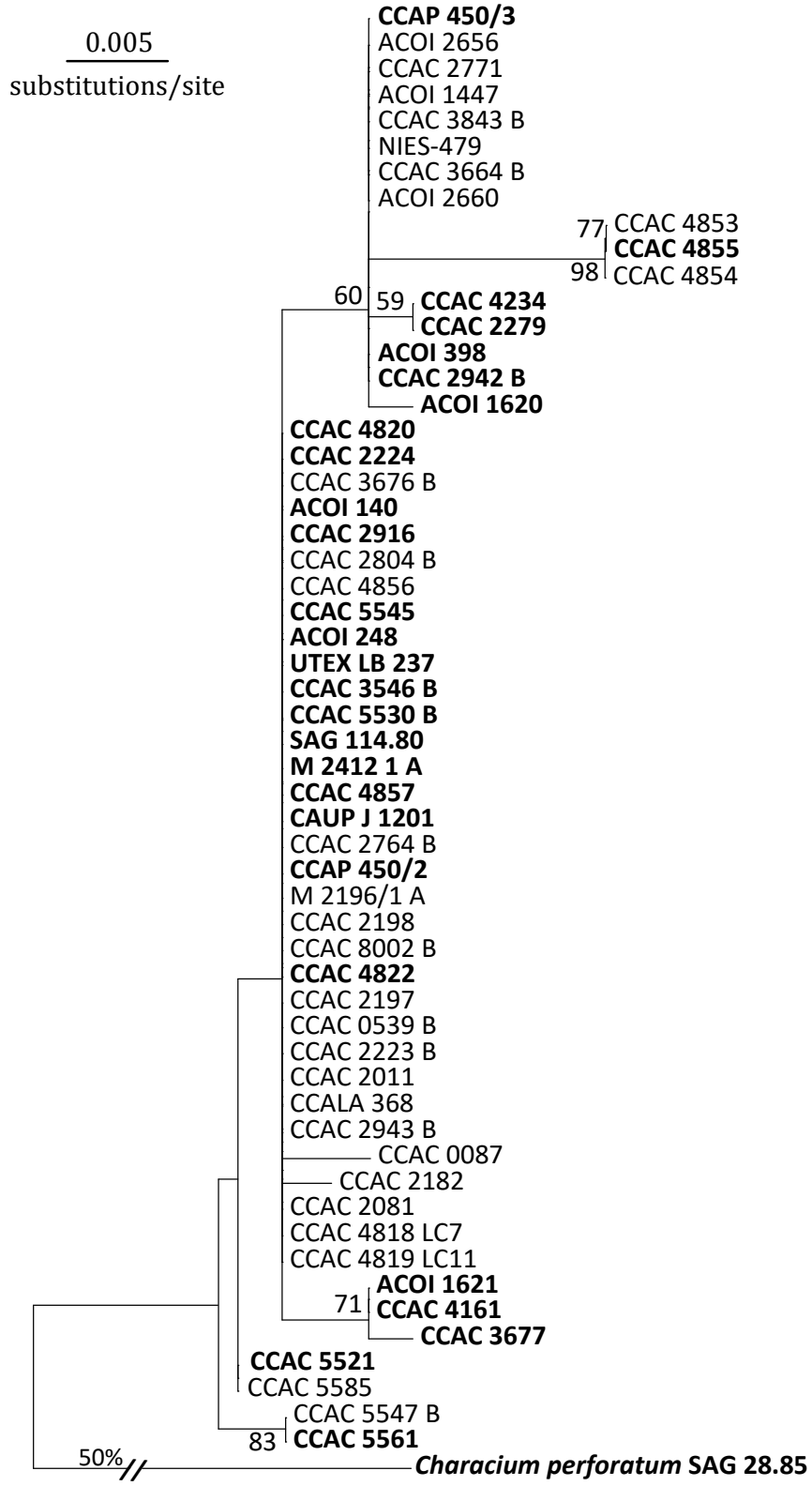
Supplementary Figure 2. ITS2 'short' single-gene phylogeny

ML topology of 56 *Microthamnion* ITS2 sequences (214 aligned positions within the processing sites C1 and C3/E) with *Characium perforatum* used as outgroup. Bootstraps: ML; branches in bold were maximally supported. A very long branch was graphically reduced to 50% (50%//). Strains in bold were selected for a congruent dataset.



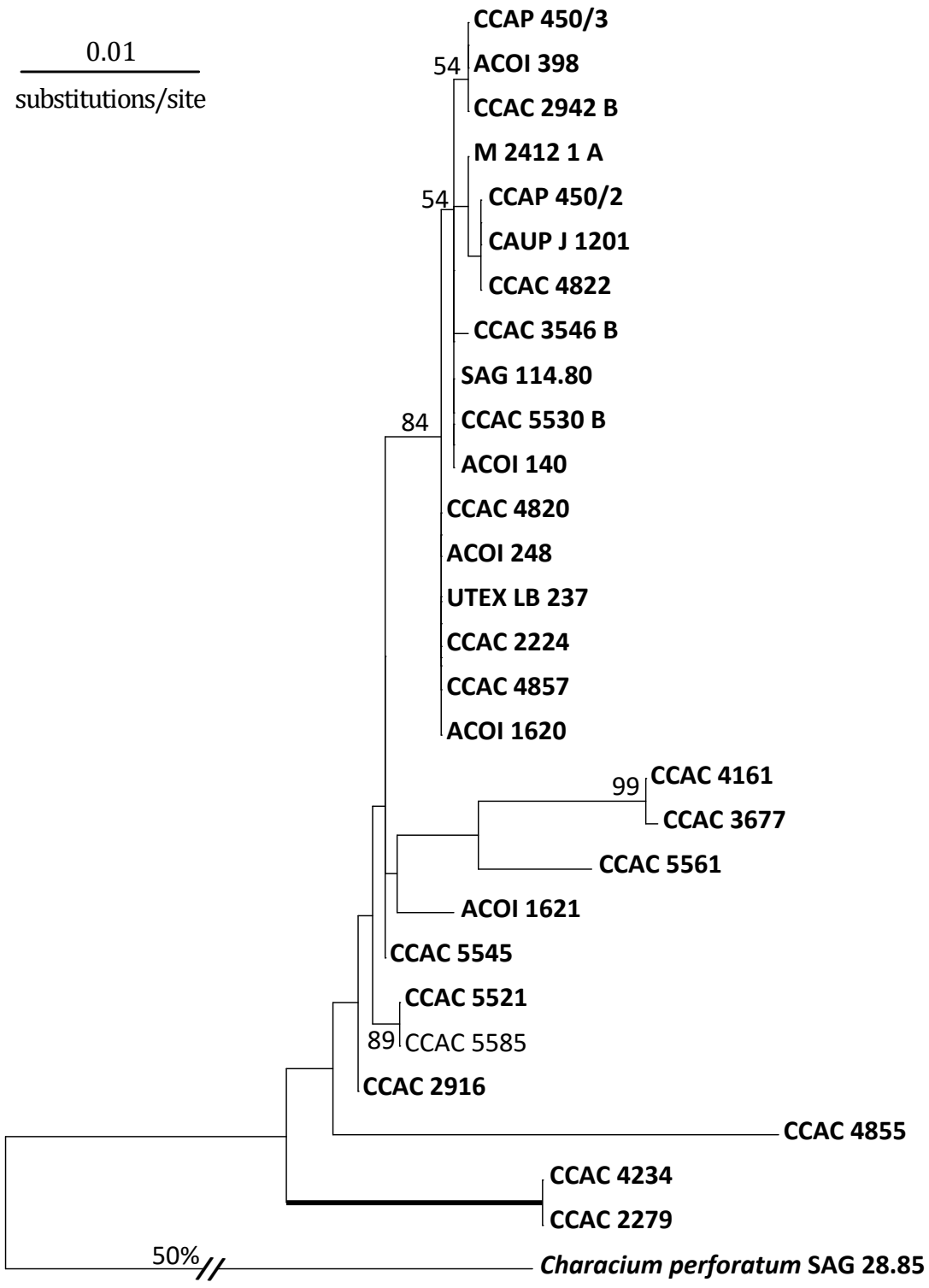
Supplementary Figure 3. ITS2 'long' single-gene phylogeny

ML topology of 56 *Microthamnion* **ITS2** sequences plus flanking regions of 5.8S and 28S (82/**214**/411 aligned positions) with *Characium perforatum* used as outgroup. Bootstraps: ML; branches in bold were maximally supported. A very long branch was graphically reduced to 50% (50%//). Strains in bold were selected for a congruent dataset.

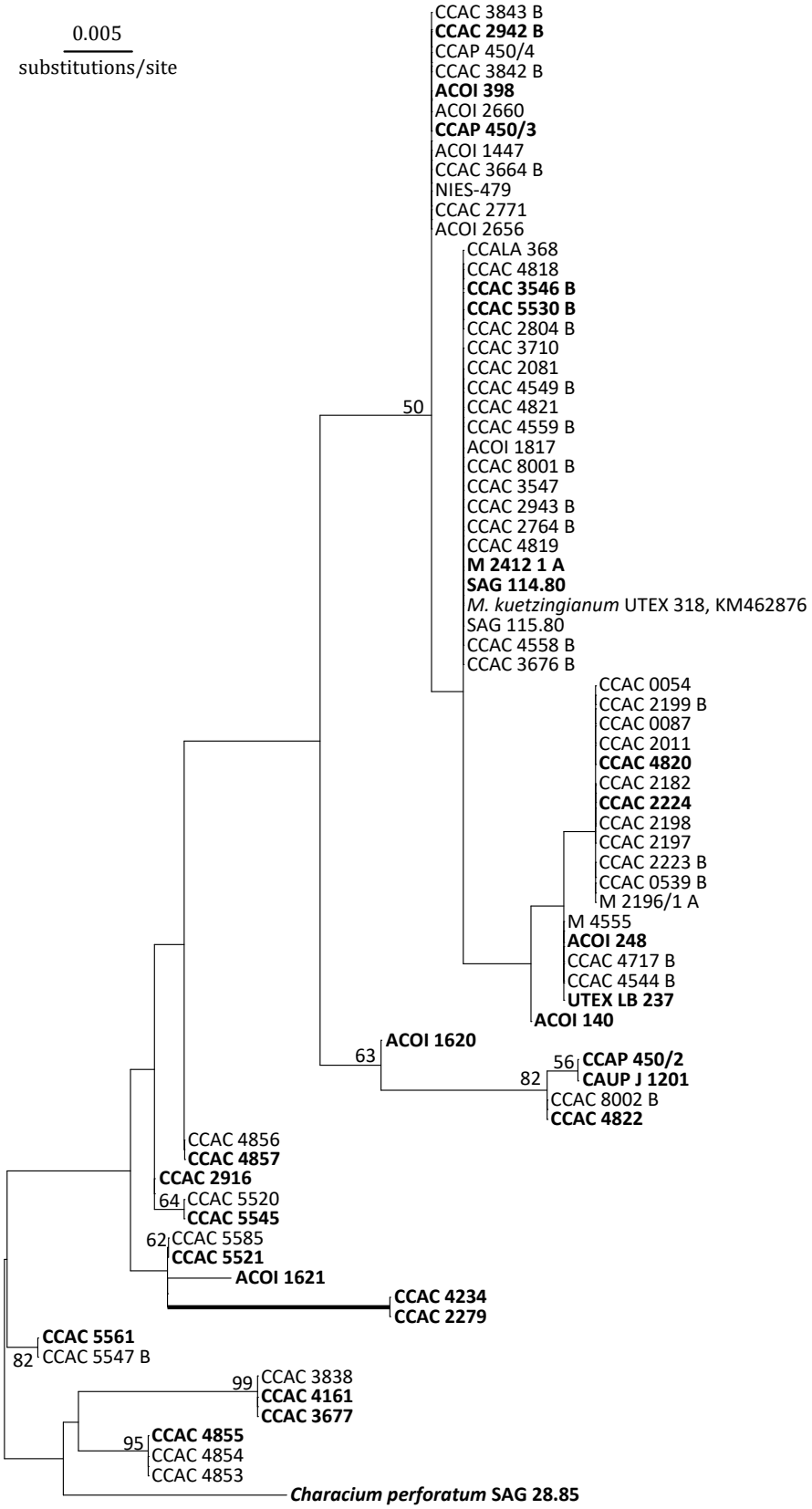


Supplementary Figure 4. 28S 'short' single-gene phylogeny

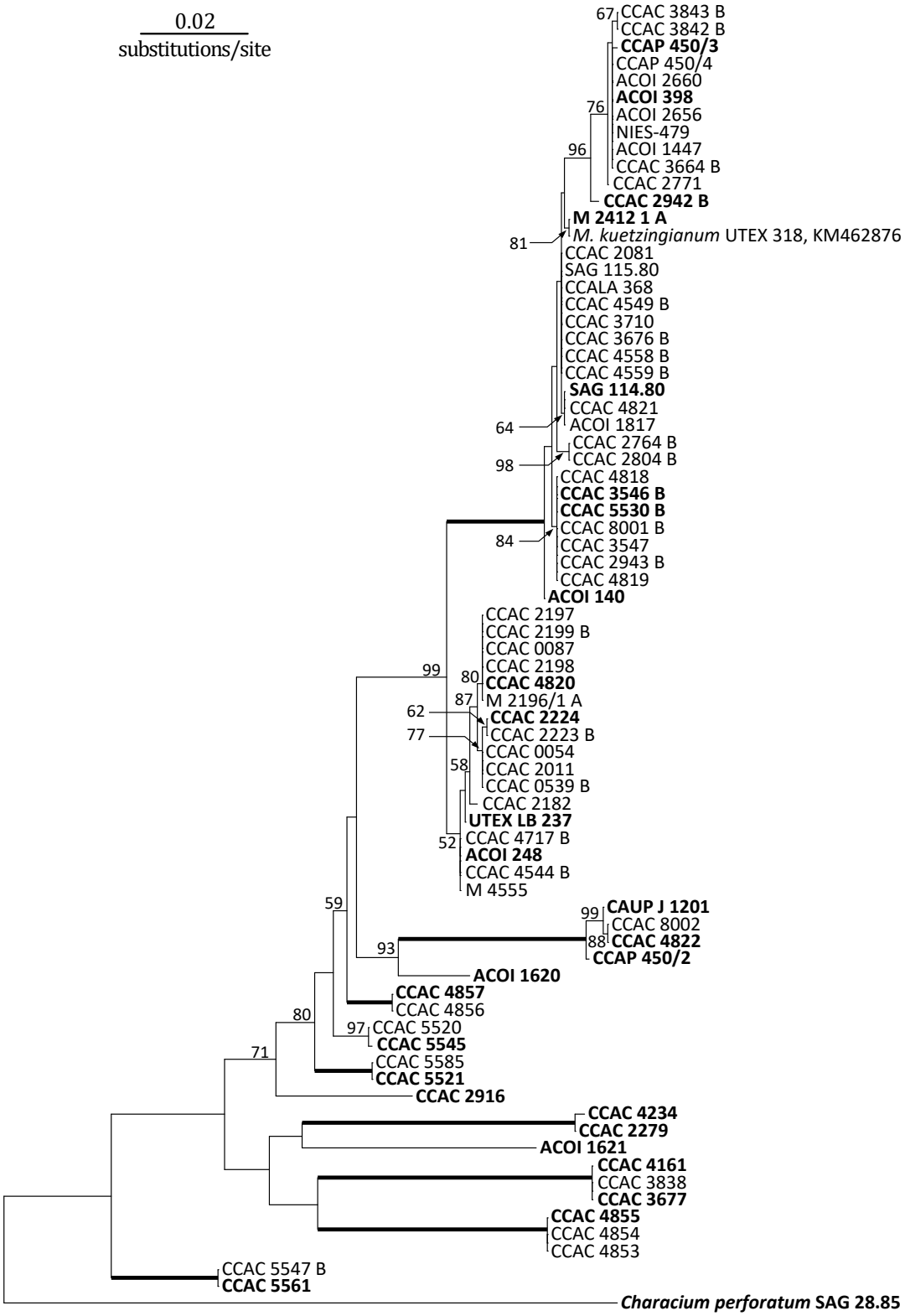
ML topology of 56 *Microthamnion* sequences covering the flanking regions of the ITS2 molecule, i.e. 82 aligned positions of 5.8S and the first 411 aligned positions of 28S. *Characium perforatum* was used as outgroup. Bootstraps: ML; a very long branch was graphically reduced to 50% (50%//). Strains in bold were selected for a congruent dataset.



Supplementary Figure 5. 28S 'long' single-gene phylogeny
ML topology of 28 *Microthamnion* sequences covering the first 1446 aligned positions of 28S. *Characium perforatum* was used as outgroup. Bootstraps: ML; branches in bold were maximally supported. A very long branch was graphically reduced to 50% (50%//). Strains in bold were selected for a congruent dataset.

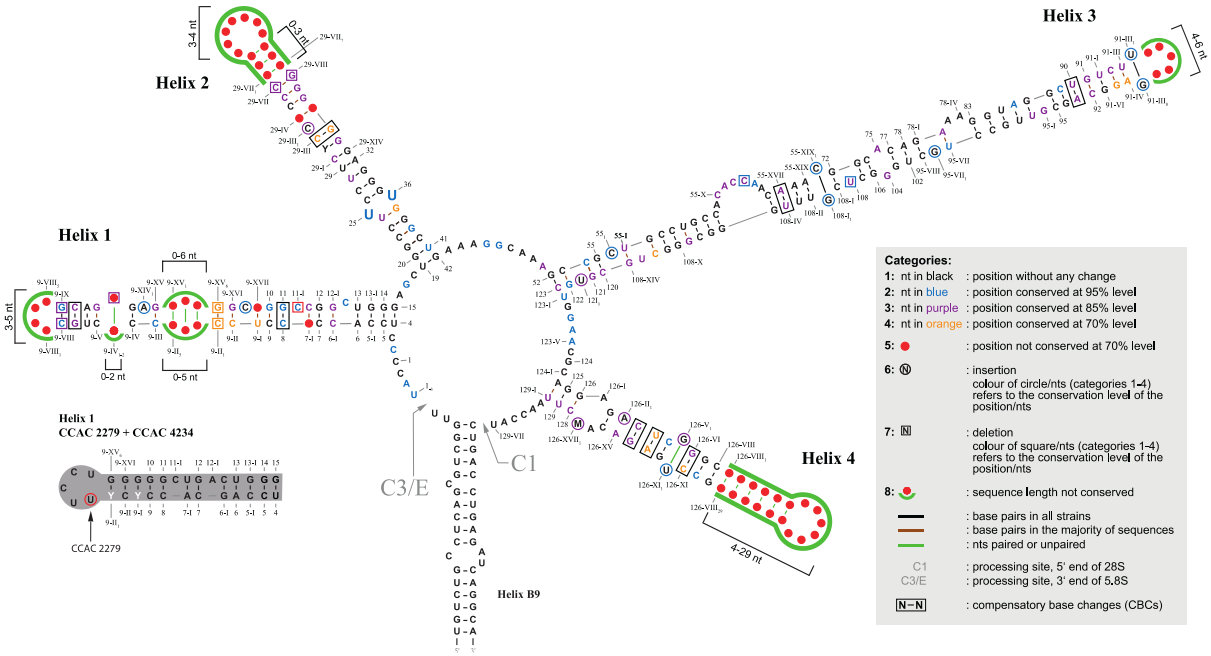


Supplementary Figure 6. *Rbc*L protein single-gene phylogeny
ML topology of 75 *Microthamnion rbc*L sequences, based on the protein alignment (425 positions). *Characium perforatum* was used as outgroup. Bootstraps: ML; branches in bold were maximally supported. Strains in bold were selected for a congruent dataset.

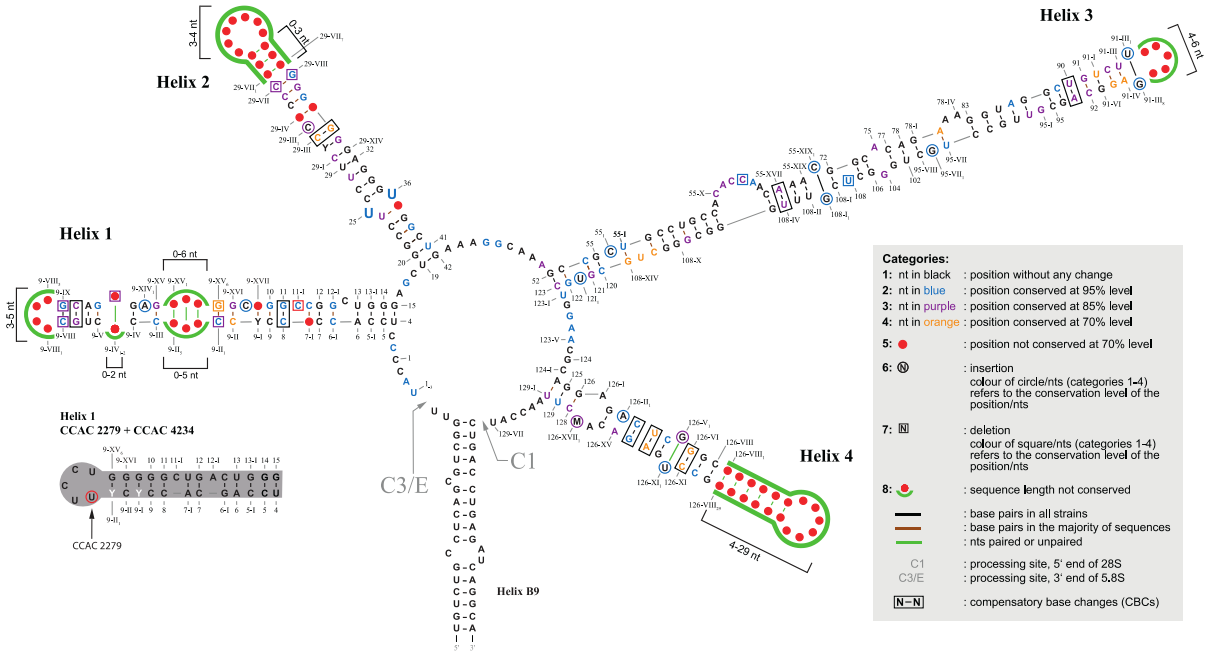


Supplementary Figure 7. *Rbc*L DNA single-gene phylogeny

ML topology of 75 *Microthamnion rbc*L sequences, based on the DNA alignment (1275 positions). *Characium perforatum* was used as outgroup. Bootstraps: ML; branches in bold were maximally supported. Strains in bold were selected for a congruent dataset.



Supplementary Figure 8. Consensus secondary structure model of all 56 ITS2 *Microthamnion* sequences
Conservation levels were adjusted to all obtained ITS2 data, including sequences from identical clones which originated from the same natural sample. The figure spans the ITS2 molecule, consisting of four paired helices (labelled Helix 1 – Helix 4) connected by single-stranded spacers and the B9 helix (site of interaction between 5.8S and 28S). The processing sites C1 and C3/E according to Côté et al (2002), Granneman et al. (2011) and Schillewaert et al. (2012) are depicted in grey. An alternate, very short Helix 1, unique for strains CCAC 2279 and CCAC 4234 is displayed separately (grey background, ambiguities in white font). Different colours of nucleotides and boxes/circles refer to levels of conservation (details are given under "Categories" in the figure). Variable parts that differed in character states and length (range in no. of nts. given for each region, indicated by black brackets) are given as red dots with a framing green line. CBC positions are highlighted by a black box.The characteristic U-U mismatch in H2 is accentuated in larger font and set apart from the helix outline. Positions of nucleotides are numbered according to Caisová et al. (2011a, 2013).



Supplementary Figure 9. Consensus secondary structure model of 45 ITS2 *Microthamnion* sequences
Conservation levels were adjusted to a reduced ITS2 dataset where identical sequences were allowed, but only when strains originated from different natural samples. The figure spans the ITS2 molecule, consisting of four paired helices (labelled Helix 1 – Helix 4) connected by single-stranded spacers and the B9 helix (site of interaction between 5.8S and 28S). The processing sites C1 and C3/E according to Côté et al (2002), Granneman et al. (2011) and Schillewaert et al. (2012) are depicted in grey. An alternate, very short Helix 1, unique for strains CCAC 2279 and CCAC 4234 is displayed separately (grey background, ambiguities in white font). Different colours of nucleotides and boxes/circles refer to levels of conservation (details are given under “Categories” in the figure). Variable parts that differed in character states and length (range in no. of nts. given for each region, indicated by black brackets) are given as red dots with a framing green line. CBC positions are highlighted by a black box.The characteristic U-U mismatch in H2 is accentuated in larger font and set apart from the helix outline. Positions of nucleotides are numbered according to Caisová et al. (2011a, 2013).

Supplementary Table 5. List of strains ordered according to the *Microthamnion* phylogeny

Strain number	cf. Species	Locality	Isolator	Year	
CCAP 450/4	*M. kuetzingianum*	England, Cornwall, by River Hayle, freshwater.	Foster	1976	Clade 1
ACOI 2656	*Microthamnion sp.*	Portugal, Serra da Estrela, pond near Lagoa Comprida.	O. Lourenço	1991	
ACOI 1447	*M. kuetzingianum*	Portugal, Mina do Vale das Gatas.	P. Ávila	2001	
ACOI 398	*M. strictissimum*	Portugal, Serra da Estrela, pond near Lagoa Comprida.	O. Lourenço	1990	
CCAC 2771	*Microthamnion* sp.	Austria, Waldviertel, Fuchsteich (near Gmünd), freshwater.	M. Melkonian	2005	
NIES 479	*M. kuetzingianum*	Japan, Hokkaido, Sapporo, Toyohira River, freshwater.	F. Kasai	1987	
ACOI 2660	*Microthamnion sp.*	Portugal, Mata Nacional de Foja, canal.	G. Carvalho	2003	
CCAC 3664 B	*Microthamnion* sp.	Germany, Eifel, Oberer Marmagener Stauteich, near shore, plankton.	M. Melkonian	2012	
CCAP 450/3	*M. kuetzingianum*	England, Cornwall, River Gannel, freshwater.	Foster	1975	
CCAC 3843 B	*Microthamnion* sp.	Germany, Stallberger Teiche, Lohmar, squeezed Sphagnum, freshwater.	M. Melkonian	2013	
CCAC 3842 B	*Microthamnion* sp.	Germany, Stallberger Teiche, Lohmar, squeezed Sphagnum, freshwater.	M. Melkonian	2013	
CCAC 2942 B	*Microthamnion* sp.	Austria, Waldviertel, Blockheide, freshwater.	M. Melkonian	2007	
CCAC 2081	*Microthamnion* sp.	Germany, Harz, Brunnenbachweg near Braunlage, temporary puddle, freshwater.	M. Melkonian	2002	Clade 2
CCAC 2764 B	*Microthamnion* sp.	Austria, Waldviertel. Fuchsteich (near Gmünd), freshwater.	M. Melkonian	2005	
CCAC 2804 B	*Microthamnion* sp.	Austria, Waldviertel, peat bog (near Heidenreichstein), freshwater.	M. Melkonian	2006	
CCALA 368	*M. kuetzingianum*	Slovakia, Orava, peat bog.	Kovacik	1983	
M 2412/1 A	*Microthamnion* sp.	Germany, Eifel, Strohner Maarchen, freshwater.	M. Melkonian	2003	
CCAC 3547	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410), freshwater.	M. Melkonian	2012	
CCAC 4819	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410, isolated from enrichment; 3N BBM), freshwater.	L. Caisová	2012	
CCAC 3546	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410), freshwater.	M. Melkonian	2012	
CCAC 4818	*Microthamnion* sp.	Spain, Gran Canaria, Barranco de Azuaje (sample 410, isolated from enrichment; SFM), freshwater	L. Caisová	2012	
CCAC 8001 B	*Microthamnion* sp.	Spain, Gran Canaria Barranco de Azuaje (sample 410, isolated from enrichment; SFM), freshwater.	L. Caisová	2012	
CCAC 2943 B	*Microthamnion* sp.	Germany, Cologne, Lindenthal, Stadtwald, freshwater.	M. Melkonian	2006	
ACOI 140	*M. kuetzingianum*	Portugal, Amieiro, pond near Arazede.	M. F. Santos	1979	
SAG 114.80	*M. kuetzingianum* (form. *M. curvatum*)	Origin unknown. Freshwater.	F. Ambard	1968	
ACOI 1817	*M. kuetzingianum*	Portugal, Serra da Peneda.	J. Paiva	2005	
CCAC 4821	*Microthamnion* sp.	Germany, Eifel, (sample 031, isolated from the enrichment; M7) Dahlemer Binz, Schlenke II (footmark), conductivity: 26,9 µS, temperature: 23,4°C, pH 4-5	L. Caisová	2012	
CCAC 5530 B	*M.* cf. *kuetzingianum*	Germany, Frechen, waste water plant, secondary settlement tank, scratch-sample, freshwater.	V. Zilz	2013	
CCAC 4559 B	*Microthamnion sp.*	Austria, Waldviertel, "Schwarzes Moos", near Brand (GPS: 48°52.33' N 14°58.9' O), squeezed Utricularia, freshwater.	M. Melkonian	2014	
CCAC 3710	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012	
SAG 115.80	*M. kuetzingianum* (formerly *M. strictissimum*)	Germany, near Hamburg, coordinates: 53.535411 /10.007172 (Lat/Long.), freshwater.	A. Weber	1969	
CCAC 3676 B	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012	
CCAC 4558 B	*Microthamnion sp.*	Austria, Waldviertel, "Schwarzes Moos", near Brand (GPS: 48°52.33' N 14°58.9' O), squeezed Utricularia, freshwater.	M. Melkonian	2014	
CCAC 4549 B	*Microthamnion sp.*	Austria, Waldviertel, "Schwarzes Moos", near Brand (GPS: 48°52.33' N 14°58.9' O), squeezed Utricularia, freshwater.	M. Melkonian	2014	
CCAC 4820	*Microthamnion* sp.	Germany, Eifel, Nettersheim (isolated from the enrichment; M7), leaf collected from a shaded pond	L. Caisová	2012	Clade 3
CCAC 0087	*M. kuetzingianum*	Germany, Lohmar near Cologne, Jexmühle, freshwater.	M. Melkonian	2002	
CCAC 2199 B	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002	
CCAC 2197 B	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002	
CCAC 2198	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002	

Strain number	cf. Species	Locality	Isolator	Year	
M 2196/1 A	*Microthamnion* sp.	Germany, Eifel, Genfbachtal near Engelgau/Ahekapelle, freshwater.	M. Melkonian	2002	Clade 3
CCAC 2223 B	*M. kuetzingianum*	Germany, near Grande, swampy area beside River Bille, freshwater.	L. Kies	1970	
CCAC 2224	*M. kuetzingianum*	Germany, near Grande, swampy area beside River Bille, freshwater.	L. Kies	1970	
CCAC 2011	*M.* cf. *strictissimum*	Germany, Cologne, Wahner Heide, location 1, freshwater.	M. Melkonian	2002	
CCAC 0539 B	*Microthamnion* sp.	Germany, Harz, freshwater.	M. Melkonian	1979	
CCAC 0054	*M. kuetzingianum*	England, Cornwall, ford near Bowithick, freshwater.	M. Melkonian	1978	
CCAC 2182	*Microthamnion* sp.	Germany, Cologne, Wahner Heide, freshwater.	M. Melkonian	2003	
UTEX LB 237	*Microthamnion* sp.	USA, Indiana, Bloomington, pond, freshwater	R. C. Starr	1953	
CCAC 4544 B	*Microthamnion sp.*	Austria, Waldviertel (sample 19a), Schremser Hochmoor, Prügelsteg (GPS: 48°47.913' N 15°6.01' O), squeezed Utricularia, freshwater.	M. Melkonian	2014	
M 4555*	*Microthamnion sp.*	Austria, Waldviertel, Schremser Hochmoor, Prügelsteg (GPS: 48°47.913' N 15°6.01' O), squeezed Utricularia, freshwater.	M. Melkonian	2014	
CCAC 4717 B	*Microthamnion sp.*	Austria, Waldviertel, Schremser Hochmoor, Prügelsteg (GPS: 48°47.913' N 15°6.01' O), squeezed Utricularia, freshwater.	M. Melkonian	2014	
ACOI 248	*M. strictissimum*	Portugal, Serra da Estrela, pond near Lagoa Comprida, plankton.	M. F. Santos	1986	
ACOI 1620	*M. kuetzingianum*	Portugal, Abrantes, Campo Militar de Sta Margarida, Barragem do Carvalhoso, canal, plankton.	M. F. Santos	2003	
CCAP 450/2	*M. kuetzingianum*	Antarctica, South Orkney Islands, Signy Island, freshwater.	Broady	1974	Clade 4
CAUP J 1201	*M. kuetzingianum*	Czech Republic, Central Bohemia, near Třtice, peat-bog "V Bahnách", soil.	Neustupa	1998	
CCAC 8002 B	*Microthamnion* sp.	Germany, Schwarzwald (sample 031, isolated from the enrichment M7) clone 1, freshwater.	L. Caisová	2012	
CCAC 4822	*Microthamnion* sp.	Germany, Schwarzwald (sample 031, isolated from the enrichment M7) clone 2, freshwater.	L. Caisová	2012	
CCAC 4856	*Microthamnion sp.*	Germany, Eifel, Dahlemer Binz, freshwater.	S. Wittek	2014	Grade 1
CCAC 4857	*Microthamnion sp.*	Germany, Eifel, Dahlemer Binz, freshwater.	S. Wittek	2014	
CCAC 5585	*M.* cf. *strictissimum*	Germany, Villau, waste water plant, aeration tank, water-body-sample, freshwater.	V. Zilz	2013	
CCAC 5521	*M.* cf. *kuetzingianum*	Germany, Frechen, waste water plant, secondary settlement tank, scratch-sample, freshwater.	V. Zilz	2013	
CCAC 5545	*M.* cf. *strictissimum*	Germany, Frechen, waste water plant, secondary settlement tank, scratch-sample, freshwater.	V. Zilz	2013	
CCAC 5520 B	*M.* cf. *kuetzingianum*	Germany, Frechen, waste water plant, Bio-P tank, water-body-sample, freshwater.	V. Zilz	2013	
CCAC 2916	*Microthamnion* sp.	Germany, Cologne, Wahner Heide, Fuchskaule,freshwater.	M. Melkonian	2006	Grade 2
ACOI 1621	*M. strictissimum*	Portugal, Abrantes, Campo Militar de Sta Margarida, lake north of Lagoa da Murta, plankton.	M. F. Santos	2003	
CCAC 5547 B	*M.* cf. *curvatum*	Germany, Glessen, waste water plant, aeration tank, scratch-sample, freshwater.	V. Zilz	2013	
CCAC 5561	*M.* cf. *kuetzingianum*	Germany, Glessen, waste water plant, aeration tank, scratch-sample, freshwater.	V. Zilz	2013	
CCAC 3677	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012	
CCAC 3838	*Microthamnion* sp.	Germany, Eifel, Mützenicher Venn, gutter in bog among Polytrichum commune, freshwater.	M. Melkonian	2012	
CCAC 4161	*Microthamnion* sp.	Scotland, Longhowe Loch, freshwater.	M. Melkonian	2013	
CCAC 2279	*Microthamnion* sp.	Germany, Eifel, Kall, industrial area 1, freshwater.	M. Melkonian	2003	
CCAC 4234	*Microthamnion* sp.	Italy, Sardinia, inland pond "Pauli Trottas", near Stagno di Cabras, pipette probe from stone and sediment, freshwater.	S. Wittek	2013	
CCAC 4854	*Microthamnion sp.*	Germany, Köln, Wahnerheide, roadside ditch, freshwater.	S. Wittek	2014	
CCAC 4855	*Microthamnion sp.*	Germany, Köln, Wahnerheide, roadside ditch, freshwater.	S. Wittek	2014	
CCAC 4853	*Microthamnion sp.*	Germany, Köln, Wahnerheide, roadside ditch, freshwater.	S. Wittek	2014	

Acknowledgements

First of all I would like to express my gratitude to Prof. Dr. Michael Melkonian for offering me the great opportunity to study in his workgroup, and for providing this interesting topic.

I would also like to thank Prof. Dr. Burkhard Becker for being the second reviewer and Prof. Dr. Siegfried Roth for agreeing to be the chairman of the examination committee.

Many thanks to Dr. Nicole Sausen, who shared her knowledge on phylogenetic analyses generously and, of course, for all the inspiring discussions and advice that helped me on the way to this thesis.

Thanks also to Dr. Lenka Caisová for fruitful discussions, helpful comments and guidance on how to establish a consensus ITS2 secondary structure model.

Many thanks to Dr. Birger Marin, whose enthusiasm sparked my interest in botany and especially in algae. He was also the one, who introduced me to the world of molecular phylogeny and remained a valuable advisor whenever I needed help with the more complicated aspects of 'reading' trees.

A big thank you to Dr. Barbara Melkonian and the CCAC staff for providing the algal cultures and the very useful information on sampling sites.

I would also like to thank the RRZK for providing access to CHEOPS which made my life much easier.

I am especially grateful that I could share my PhD experience with so many wonderful colleagues. Nicole, Doro, both Alices, Petra, Zehra, Sebastian H., Martin and Frederik – you did not only enrich my scientific life, but have become dear friends.

Finally, I would like to express my sincerest gratitude to my friends and family in the outside world. Without your love and support this project would not have been possible.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Michael Melkonian betreut worden.

Köln, den 06.12.2019

_____

Tanja Reder