universität wien

DISSERTATION

Titel der Dissertation

„Species delineation and autecology

of *Spirogyra* LINK 1820"

Verfasserin

Charlotte Maria Margarita Chen

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr. rer. nat.)

Wien, 2015

Studienkennzahl lt. Studienblatt:	A 091 444
Dissertationsgebiet lt. Studienblatt:	Ökologie
Betreuer:	ao. Univ. Prof. Mag. Dr. Michael Schagerl

Lake Egelsee, AUT, 2006

Table of Contents

Zusammenfassung

Die Grünalge *Spirogyra* (Streptophyta, Zygnemataceae) ist aufgrund ihrer typischen vegetativen Morphologie leicht identifizierbar. Sie wird charakterisiert durch unverzweigte Filamente und Zellen mit ein bis mehreren spiralig gewundenen Chloroplasten (Kadlubowska 1984; Kolkwitz & Krieger 1941; McCourt *et al.* 1986; Randhawa 1959a; Transeau 1951). Die einfache Struktur wird oft fälschlicherweise als Hinweis für eine niedrige genetische Diversität gedeutet. *Spirogyra* wird meist von phylogenetischen Analysen ausgeschlossen; unter anderem auch, weil *Spirogyra* lange Äste in Phylogenien verursacht („long branch taxon"), was durch eine hohe Evolutionsrate hervorgerufen wird. Weiters wird sie wegen der problematischen Artbestimmung von vielen ökologischen Analysen ausgenommen, obwohl die ökologische Nische der Gattung eine große Bandbreite von Süßwasser- bis Brackwasserstandorten abdeckt (Hoshaw & McCourt 1988; Rieth 1983; Simons & Van Beem 1990) und von naturbelassenen Bergseen und Moorstandorten bis hin zu hoch eutrophen Standorten vorkommt (Hainz *et al.* 2009). Die Artbestimmung in dieser Gattung ist schwierig, da nicht nur der komplette Reproduktionszyklus (Konjugation), sondern auch reife Zygosporen (Hypnozygoten) vorhanden sein müssen (Kadlubowska 1984; Kolkwitz & Krieger 1941; McCourt *et al.* 1986; Randhawa 1959a; Transeau 1951).
Studien zur genetischen Diversität sind rar. Untersuchungen zur Stellung von *Spirogyra* innerhalb der Zygnematophycea leiden unter dem geringen Stichprobenumfang; der Stammbaum ist zwar viel diskutiert, aber immer noch unklar – besonders im Bezug auf die Beziehung auf die Gattung *Sirogonium*. Um diese Frage zu beantworten und um einen bessern Einblick in die genetische Diversität der Gattung *Spirogyra* zu bekommen, haben wir folgende Untersuchungen durchgeführt.

Die verwendeten Algenproben wurden an 133 verschiedenen Standorten im Zeitraum von März 2006 bis Oktober 2007 gesammelt und kultiviert (Hainz *et al.* 2009). Ein Teil davon wurde für die genetischen Untersuchung in den folgenden Manuskripten verwendet.
Die phlyogenetischen Untersuchungen der SSU rRNA Sequenzen von 130 *Spirogyra* Kulturen zeigte Untertteilung in acht Kladen. Im ersten Teil unserer Studien entdeckten wir in etwa 60% der Sequenzen (Kladen A-D) ein 1506 Gruppe I Intron, die Sequenzen ohne Intron (Kladen E-H) bildeten zusammen eine monophyletische Gruppe, die an einem langen Ast plaziert wurde. Das in *Spirogyra* gefundene Intron gehört zu den IC Gruppe I Introns. Es liegt an derselben Stelle wie das Intron, das bei Zygnematalen gefunden wird und weist weitere Charakteristika dieses Typuses auf (typische Domänenstruktur (P1-P9), Basenkomposition, stark konservierte Regionen, die GU Paarung (auf 5' Seite) und das G (auf 3' Seite), die der Spleiß-Stelle vorangehen. Es weist sowohl Merkmale von frühen (optionale P2 Domäne) als auch von sich später abzweigenden (Variation der typsichen L5b GAAA vierfach Schleife) Desmidiaceae auf. Phylogenetische Analysen des Gruppe I Introns von *Spirogyra* zeigten einen monophyletischen Ursprung der Gattung innerhalb der Zygnematophyceae. Die Analysen zeigten weiters, daß die Mutationsrate für die SSU rRNA höher war als für das Intron selbst. Daraus schliessen wir auf einen sekundären Verlust des Introns in den Kladen E-H, der durch die hohe Mutationsrate und die lange Entwicklung verursacht wurde.
Im zweiten Teil unserer Untersuchungen lag der Fokus auf den phylogenetischen Beziehungen innerhalb *Spirogyras* und auch zu anderen Zygnematophyceen. 130 *Spirogyra* SSU rRNA Sequenzen wurden mit Hilfe komplexer Evolutionsmodelle analysiert (posterior probability, maximum likelihood, neighbor joining und maximum parsimony methods). Das Resultat waren acht Kladen innerhalb *Spirogyra*s, die die Gattung *Sirogonium* einschlossen. Die phylogenetische Beziehung zur Gattung *Spirotaenia* wurde nicht ausreichend geklärt. Die Untersuchungen zeigten, dass die 130 Sequenzen

53 Klone im Bezug auf die SSU rRNA beinhalten. Keiner dieser Klone hatte eine niedrigere Mutationsrate oder konnte an einem kürzeren Ast als die restliche Gattung plaziert werden. Weiters wurde die Monophyllie der Gattung bestätigt. Die genetischen Unterschiede innerhalb der Gattung *Spirogyra* waren größer als die innerhalb der anderen analysierten Gattungen der Zygnemataceae. Es wurde eine große Anzahl an sog. nicht-homoplasischen Synapomorphien (NHS; 114 NHS gesamt) gefunden, 41 NHS für *Spirogyra* und 73 NHS für die einzelnen Kladen. Dies betont die hohe genetische Diversität der Gattung und die Distanz zu den anderen Zygnematophyceae.

In einem dritten Schritt haben wir das Problem der Artkonzepte, Artdefinition und –zahl in der Gattung *Spirogyra* beleuchtet. Wir haben ITS2 Sekundärstrukturen und den entsprechenden Barcode miteinbezogen. Das konventionelle Artkonzept der fädigen Grünalge *Spirogyra* basiert nur auf morphologischen Merkmalen, was die phylogenetischen Beziehungen nicht wiederspiegelt. Der ITS2 Barcode wurde zur Schätzung der Artzahl unseres Datensets verwendet. Sequenzen, die sich um mindestens einen kompensatorischen Basenaustausch unterschieden, wurden als unterschiedliche Arten definiert. 51 verschiedene Arten wurden mit Hilfe von Barcoding voneinander unterschieden. Von den 120 verwendeten Sequenzen wurden 68 durch nur eine Seuqenz repräsentiert, 23 durch mehrere. Die ITS2 Sekundärstruktur wurde analysiert, aber wegen der großen Unterschiede in der Primär- und Sekundärstruktur wurde keine Konsensusstruktur gefunden. Sie zeigt trotzdem die typischen Merkmale der ITS2 Sekundärstruktur. Die üblicherweise stark konservierte Region am 5' Ende von Helix III zeigt in der Gattung *Spirogyra* einige Abwandlungen, ist aber gut identifizierbar. Einige der Sequenzen haben eine Helix IV, einige sogar einen zusätzlichen Arm an Helix III. Vegetative Morphologie, Zugehörigkeit zu Kladen und Klonen wurden auf Muster und Korrelationen hin untersucht. Es wurden einige Korrelationen gefunden, diese waren allerdings nicht signifikant genug, um Aussagen zu unterstützen, welche Morphologie bestimmten Kladen zugeordnet werden könnten und umgekehrt. Innerhalb einiger Klone wurde außerdem eine variable Morphologie beobachtet.

Mit unseren Studien haben wir versucht etwas Licht auf die phylogenetisch vernachlässigte Gattung *Spirogyra* und ihre große genetische Diversität zu werfen, die unserer Meinung nach bisher stark unterschätzt wurde. Unser Datenset deckt eine große Vielfalt an verschiedenen vegetativen Morphologien ab, die von Standorten sehr verschiedenen Typs gesammelt wurden. Wir haben bewiesen, dass vegetative Morphologie ein schlechter Indikator für genetische Diversität ist. Unsere Studien haben die außergewöhnliche Stellung der Gattung *Spirogyra* innerhalb der Zygnematopyceae und den Einschluss der Gattung *Sirogonium* bestätigt. Unsere Ergebnisse deuten darauf hin, dass eine hohe Evolutionsrate, ein abweichendes Muster in der Basen-Substituion und eine relativ rezente Ausbreitung Grund für die Platzierung an langen Ästen bei Phylogenien sein könnte, obwohl *Spirogyra* eine sich früh von den übrigen Zygnematophyten wegentwickelten Gattung ist.

Abstract

The green alga *Spirogyra* (Streptophyta, Zygnemataceae) is easily recognizable based on its vegetative morphology, which is characterized by unbranched filaments and cells that have one to several coiled chloroplasts (Kadlubowska 1984; Kolkwitz & Krieger 1941; McCourt *et al.* 1986; Randhawa 1959a; Transeau 1951). This simple structure falsely points to a low genetic diversity: *Spirogyra* is commonly excluded from phylogenetic analyses; also because the genus is known as a long-branch taxon caused by a high evolutionary rate. It is excluded from many ecological analyses because of difficulties in species determination, although its ecological niche covers a wide range from fresh to slightly brackish water bodies (Hoshaw & McCourt 1988; Rieth 1983; Simons & Van Beem 1990) and from dystrophic mountain lakes and bogs to highly eutrophicated systems (Hainz *et al.* 2009). Traditional species determination is problematic, because it requires the complete reproductive (conjugation) cycle including zygospores (hypnozygotes) (Kadlubowska 1984; Kolkwitz & Krieger 1941; McCourt *et al.* 1986; Randhawa 1959a; Transeau 1951) is necessary.
Studies elucidating genetic diversity are rare, whereas studies on the position of *Spirogyra* among the other Zygnematophyceaen genera often suffer from limited taxon sampling; the relation to these genera is much discussed but unclear – especially the relation to the genus *Sirogonium*. To resolve these questions and get a better impression of the genetic diversity of the genus *Spirogyra*, we conducted the following studies.

Algal samples used for the following studies were collected at 133 sites from March 2006 to October 2007 (Hainz *et al.* 2009) and cultivated. A selected subsample/part of them was extracted and analyzed as described in the manuscripts.
Phylogenetic analyses of SSU rDNA sequences of 130 *Spirogyra* strains revealed that these strains are subdivided into eight clades. In the first part of our studies, we discovered that approximately 60% of the assessed strains (clades A–D) contain a 1506 group I intron, whereas strains without introns form individual clades that cluster together in a long branch (E–H). The *Spirogyra* intron belongs to the IC group I introns and shares the common insertion site of the Zygnematalean intron and many common features (the typical domain structure (P1–P9), the base composition, the highly conserved regions, the GU pair preceding the 5' splice site and the G preceding the 3' splice site). It exhibits features of early desmids (optional P2 domain) as well as of later diverging desmids (variation of the typical L5b–GAAA tetraloop). Phylogenetic analyses of the *Spirogyra* group I intron showed the monophyletic origin of the genus within the Zygnematophyceae; analyses also showed a higher mutation rate in the SSU rRNA than in the intron. Therefore, we assume the secondary loss of the intron in clades E–H is caused by the high evolutionary rate of *Spirogyra* and its long evolutionary history.

The second part focused on the genetic diversity within *Spirogyra* and its phylogenetic relation to other Zygnematophyceaen taxa. The 130 *Spirogyra* SSU rDNA sequences used for phylogenetic analyses were tested using complex evolutionary models (posterior probability, maximum likelihood, neighbor joining and maximum parsimony methods). We found that the eight resulting clades formed by *Spirogyra* include the genus *Sirogonium*, but the phylogenetic relationship to the genus Spirotaenia was not resolved satisfactorily. Even though sequence comparisons revealed 53 individual clones, our results still support monophyly of the genus. Our data set did not contain a single slow-evolving taxon that would have been assigned to a shorter branch compared to the remaining sequences. The genetic distance within the genus *Spirogyra* exceeds the distances measured within any other genus of the remaining Zygnemataceae included in this study. A large

number of non-homoplasious synapomorphies (NHS; 114 NHS in total) was found for *Spirogyra* (41 NHS) and for each clade (totaling 73 NHS). This emphasizes the high genetic diversity of this genus and the distance to the remaining Zygnematophyceae.

In a third step, we addressed the issue of species definition and number in the genus *Spirogyra*. We included ITS2 secondary structure information as well as ITS2 barcode. The traditional species concept of the filamentous alga *Spirogyra* is based on morphological traits and does not reflect phylogenetic relationships. The ITS2 barcode was used to estimate the species number of our data set; sequences that differed by at least one compensatory base change were considered putative different species. Out of 120 sequences, 68 were found only once and 23 were found more than once; 51 putative species were separated by ITS2 barcoding. The ITS2 secondary structure was analyzed, but no consensus structure could be found due to the big variation in primary and secondary sequences. The overall structure shows the typical features of ITS2 secondary structure. The typically conserved region at the 5' end of helix III is less conserved in this genus, but still recognizable. Some strains exhibit a helix IV, some an extra arm in helix III. The correlation of vegetative morphology, clade grouping and information on clone grouping was analyzed statistically to find patterns. Correlations were found, but were too weak to support projections which morphology would belong to a certain clade and vice-versa. Also within single clones, some variation in morphology was found.

Our studies tried to shed light on a phylogenetically neglected genus with great genetic diversity that has been underestimated to this day. Our data set included samples from a large variety of water bodies and many different vegetative morphotypes. We proved that vegetative morphology is a poor indicator for genetic diversity. Our studies confirmed the exceptional position of the genus *Spirogyra* among the Zygnematophyceaen taxa and the inclusion of the genus *Sirogonium*. Our findings indicate that a high evolutionary rate, a different pattern in base substitution and recent radiation might be the reasons for the long-branch status of *Spirogyra*, despite it being an early diverging genus within the Zygnematophytes.

Hainz, R., C. Wöber, et al. (2009). "The relationship between *Spirogyra* (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe." Aquatic Botany 91(3): 173-180.

Hoshaw, R. W. and R. M. McCourt (1988). "The Zygnemataceae (Chlorophyta): A twenty-year update of research." Phycologia 27(4): 511-548.

Kadlubowska, J. Z. (1984). Conjugatophyceae I - Zygnemales. Süßwasserflora von Mitteleuropa, Chlorophyta VIII. H. Ettl, H. Gerloff, H. Heynig and D. Mollenhauer, Stuttgart, New York: Gustav Fischer Verlag.

Kolkwitz, R. and H. Krieger (1941). Zygnemales. Dr. L Rabenhorst's Kryptogamen-Flora von Deutschland und der Schweiz. R. Kolkwitz, Akademische Verlagsgesellschaft Becker & Erler, Leipzig. XIII, 2 Abteilung: 499 pp.

McCourt, R. M., R. W. Hoshaw, et al. (1986). "Distribution, morphological diversity and evidence for polyploidy in North American Zygnemataceae (Chlorophyta)." Journal Of Phycology 22(3): 307-315.

Randhawa, M. S. (1959). Zygnemaceae. New Delhi, Indian Council of Agricultural Research.

Rieth, A. (1983). "Eine *Spirogyra* von der Ostsee bei Zingst." Genetic Resources and Crop Evolution 31(2): 317-326.

Simons, J. and A. Van Beem (1990). "*Spirogyra* species and accompanying algae from pools and ditches in The Netherlands." Aquatic Botany 37(3): 247-269.

Transeau, E. N. (1951). The Zygnemataceae (Fresh-water conjugate algae), The Ohio State University Press.

Dissertation Outline and Contributions

This dissertation is a compilation of three manuscripts that represent the main part of my research;

1) Chen, C. and M. Schagerl (2012). "Slow evolution of 1506 group I intron in *Spirogyra* LINK 1820 (Zygnematophyceae, Streptophyta), a fast evolving lineage in the Zygnemataceae." Fottea, Olomouc, 12(1):255-272.
2) Chen, C., M. Barfuss, et al. (2012). "Hidden genetic diversity in the green alga *Spirogyra* (Zygnematophyceae, Streptophyta)." BMC Evolutionary Biology 12(1): 77.
3) Chen, C., Coleman, A,. Hainz, R., Pröschold, T. and Schagerl, M. (unpublished). "Combining vegetative morphology, genetics and environment – a polyphasic approach to gain insight into the complex relationships of *Spirogyra* (Zygnematophyceae)"

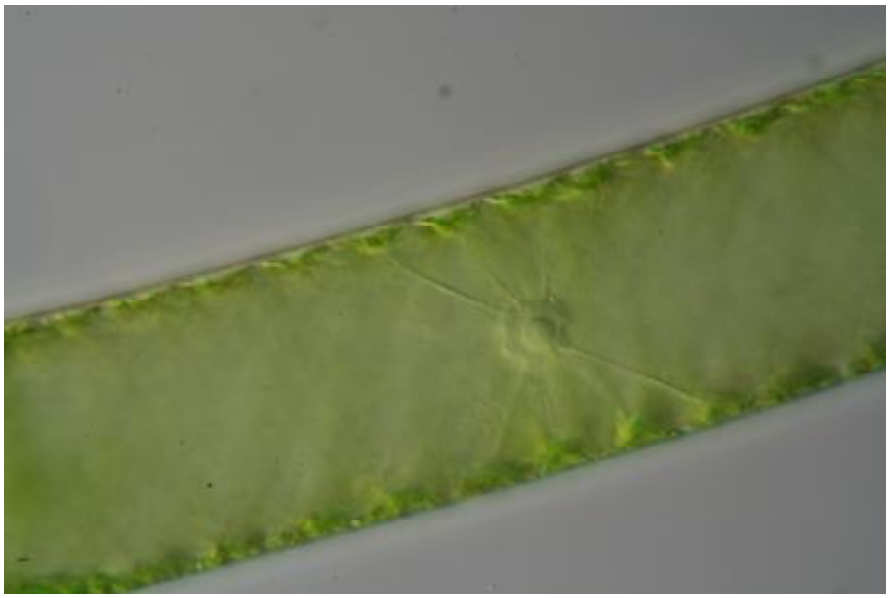
Additionally, I contributed to other work published during the project;

1) Hainz, R., C. Wöber, et al. (2009). "The relationship between *Spirogyra* (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe." Aquatic Botany 91(3): 173-180.
   My contributions: field sampling, chemical water analyses, digital photography and measurements of morphological parameters
2) Schagerl, M., R. Hainz & C. Wöber, 2009. Autecology and species delineations of the green filamentous alga *Spirogyra* LINK. Phycologia 48: 115.
   My contributions: field sampling, chemical water analyses, digital photography and measurements of morphological parameters
3) Zwirn, M. (2010). Sexuelle Fortpflanzung und taxonomische Fragestellungen bei der Süßwasseralge *Spirogyra* (Zygnemataceae, Streptopyhta). Limnology. Vienna, University of Vienna. Master: 68.
   My contributions: preparing algal material for experiments, experimental setup, supporting cultural experiments
4) Zwirn, M., Chen, C., Bohuslav, U. and M. Schagerl (2013). "Induction of sexual reproduction in *Spirogyra* clones – does an universal trigger exist?" Fottea, Olomouc, 13(1): 77-85.
   My contributions: preparing algal material for experiments, experimental setup, supporting cultural experiments

Table 1: My estimated personal input to the various publicatons and manuscripts

	Chen, C. and M. Schagerl (2012)	Chen, C., M. Barfuss, et al. (2012)	Chen, C., Coleman, A,. et al. (unpublished)	Hainz, R., C. Wöber, et al. (2009)	Schagerl, M., R. Hainz & C. Wöber, (2009)	Zwirn, M. (2010)	Zwirn, M., Chen, C., Bohuslav, U. & M. Schagerl (2013)
Field sampling	-	-	-	50%	50%	-	-
Experimental design & setup	80%	80%	80%	30%	30%	30%	30%
Preparation	100%	100%	100%	-	-	30%	30%
Laboratory work	95%	95%	95%	50%	50%	30%	30%
Data analysis	100%	95%	90%	0	0	0	0
Writing of manuscript	95%	95%	95%	0	0	0	0

The work presented here was done in the context of the project “Species delineation and autecology of *Spirogyra* LINK 1820” funded by the Austrian Science fund (P18465-B03).



Spirogyra sp.

Introduction

The genus *Spirogyra* was first described in 1820 by Link as "Thallus septatus, coniugans. Interanea in spiras torta. Fructificatio externa nulla." (von Esenbeck & Link 1820), p.5, which roughly translates to "Thallus divided by walls into connected compartments. Cell content spiral. No external fruit". This description characterizes the main features of this genus – unbranched filaments of cells connected/separated by cell walls of various structure and covered by a layer of mucilage; one to sixteen spirally coiled chloroplasts that are located within the periphery of the cylindrical wall; and a special way of reproduction typical for all Zygnematales – the so-called conjugation.

There are two typical modes of conjugation – the scalariform (ladder-like, more common) and the lateral conjugation. In the first step of the scalariform conjugation, two haploid filaments align and the differentiation of two vegetative cells into gametangia starts with the disintegration of the chloroplasts. After that, a conjugation tube is formed. During lateral conjugation on the other hand, two adjacent cells within a single filament are connected by a conjugation tube. In both types, the cytoplasm of the donating cell (male gametangium) moves through the conjugation tube into the receiving cell (female gametangium), a process which can be seen as amoeboid stage. Both gametes fuse to form a diploid hypnozygote. After meiosis and the decay of three nuclei, the hypnozygote ripens. Under favorable conditions, a haploid seedling emerges and starts forming a filament by cell divisions (Van den Hoek *et al.* 1995). For successful species identification, all stages of the reproduction cycle (vegetative filaments with vital chloroplasts, conjugating cells, ripe hypnozygotes) are essential (Kadlubowska 1984; Randhawa 1959a; Transeau 1951), which complicates taxonomy, because conjugation stages are only available in approximately 10% of findings (Hoshaw & McCourt 1988; McCourt *et al.* 1986). Additionally, sexual reproduction is not easily induced because a universal trigger remains unidentified (Allen 1958; Simons *et al.* 1982; Simons *et al.* 1984; Zwirn 2010; Zwirn *et al.* 2013).

The genus *Spirogyra* occurs worldwide – it has been recorded on all continents, even Antarctica (Hawes 1988; Hoshaw & McCourt 1988). It has a very broad ecological niche, typically favoring water bodies with low to no flow velocity (Hoshaw & McCourt 1988) ranging from oligotrophic to eutrophic levels (Hainz *et al.* 2009). *Spirogyra* is sometimes contributing to the biofilm community on hard substrata. It also forms small submerged patches and larger floating mats that are buoyant due to the oxygen-bubbles resulting from its high productivity and trapped by the dense biomass (Hoshaw & McCourt 1988). Despite its wide distribution and common occurrence, and the resulting potential for ecological studies, it is often excluded from data sets due to problems with species identification. As mentioned above, the full reproductive cycle is essential but rarely found and difficult to induce.

Important research questions are the estimation of the true species number and the application of a species concept. The most current monograph by Kadlubowska comprises 386 species (Kadlubowska 1984); in previous works by Randhawa (1959) and Transeau

(1951) the number was somewhat lower with 289 and 275 taxa, respectively. Species delineations are based on morphological traits, many of them relying on a single finding; descriptions of these traits often exhibit large overlaps and/or only differ marginally. The morphological species concept, which is applied to the genus *Spirogyra*, is not proven to represent true biological species, nor does it provide any information on the ecological or genetic diversity in a genus.

The phenomenon of polyploidy, that has been reported from the genus, presents another problem to conventional taxonomy (Allen 1958; Hoshaw *et al.* 1985; Hoshaw *et al.* 1987; McCourt & Hoshaw 1990). Filaments isolated by Allen (1958) from an unialgal culture were shown to differ in filament width, chloroplast number, chromosome number, and zygospore size, whereas zygospore ornamentation remained unchanged (Allen 1958; McCourt & Hoshaw 1990).

Different species concepts focus on different aspects of life – such as isolation, recognition, ecology, evolution and phylogeny (De Queiroz 2007; Stuessy 2009). Many of these concepts and their definitions are incompatible and lead - among other things - to different conclusions on the species number. Therefore, a single specimen can belong to different groups with a varying set of members (De Queiroz 2007; Hey 2006) – this also has been pointed out in the case of the genus *Spirogyra* (McCourt & Hoshaw 1990). Every concept claims universal practicability, however, particular criteria for identifying species are not applicable in all situations (Hey 2006). Consequences for working with an algal genus like *Spirogyra* have already been described by some authors (McCourt & Hoshaw 1990; Zwirn 2010; Zwirn *et al.* 2013).

Most phylogenetic studies including *Spirogyra* were designed to analyze relationships on the level of family and genus and did not give information on intragenic diversity or relationships (Drummond *et al.* 2005; Gontcharov *et al.* 2002; Gontcharov *et al.* 2004a; McCourt *et al.* 2000). Studies based on SSU rRNA phylogenies often suffer from a long-branch attraction (LBA) phenomenon between *Spirogyra* and the outgroup used. In some studies, even the placement of *Spirogyra* within the Zygnemophyceae was questionable (Gontcharov *et al.* 2002; Gontcharov & Melkonian 2004).

To address the problems and questions depicted above, we chose SSU rRNA and ITS2 primary and secondary structure as phylogenetic markers in our studies, described in the following manuscripts, over *rbcL* for the following reasons:

- Search of specimens with slower evolutionary rate that could be placed on a shorter branch to reduce LBA in SSU rRNA phylogenies and confirm the position of the genus *Spirogyra* within the Zygnemataceae (Gontcharov *et al.* 2002; Gontcharov *et al.* 2004a; Gontcharov & Melkonian 2004);
- The relatively poor resolution of *rbcL* phylogenies of *Spirogyra* and *Sirogonium* (Drummond *et al.* 2005; Stancheva *et al.* 2013), resulting in closely related,

geographically isolated putative species whose relationship could not be resolved by using a conserved gene such as *rbcL*;

- Survey on the presence/secondary loss of the 1506 group I intron (Gontcharov *et al.* 2002; Gontcharov *et al.* 2004a; Gontcharov & Melkonian 2004);
- ITS2 has been proven as a useful marker for barcoding questions (and hence species numbers) and hints at sexual compatibility between strains (Coleman 1999; Coleman 2000; Coleman 2003; Coleman 2007; Coleman 2009; Coleman & Mai 1997; Coleman *et al.* 1994; Coleman & Vacquier 2002; Yao *et al.* 2010; Young & Coleman 2004). The ITS2 region represents an untranslated insertion between the 5.8S RNA gene and the LSU RNA gene. Valuable information can be derived from sequence and secondary structure analyses. By comparing the universal folding pattern, compensatory base changes (CBCs) can be detected and applied to estimate whether two samples are compatible in mating type i.e. if they can reproduce sexually (Coleman 2000). The ITS2 and its secondary structure can be used to address questions that range from the level of order down to subspecies, where other markers have limited potential (Coleman 2003).

Other important subjects of our study were:

- The relationship of *Spirogyra* to the putative sub/sister-genus *Sirogonium*;
- The position of the genus within the order Zygnematales
- Can the non-correspondence of groups, as described by McCourt & Hoshaw (1990), be simplified by using additional information such as ecological niche and phylogenetic markers;
  Investigate if morphology is reflected by phylogeny, if ecology is reflected by morphology (Hainz *et al.* 2009), and if ecology is reflected by phylogeny by means of sampling specimens of different vegetative morphologies and ecological origin to cover different aspects;
- Evaluate conventional species identification characters;
- Find a method for estimating the species number within a given data set.

The algal samples used in the following manuscripts were collected at 133 sites between 2006 and 2007, together with water samples and environmental data (Hainz *et al.* 2009). Sampling areas were chosen according to geology and accessibility. . Water samples and environmental data were treated as described by Hainz (Hainz *et al.* 2009).

Algal samples were cultivated by the author at the Algensammlung Wien (ASW) and treated as described in the second manuscript (Chen *et al.* 2012).

References

Allen MA (1958) *The biology of a species complex in Spirogyra*, Indiana University.

Besendahl A, Bhattacharya D (1999) Evolutionary analyses of small-subunit rDNA coding regions and the 1506 group I introns of the Zygnematales (Charophyceae, Streptophyta). *Journal of Phycology* 35, 560-569.

Bhattacharya D (1996) Analysis of the distribution of bootstrap tree lengths using the maximum parsimony method. *Molecular Phylogenetics and Evolution* 6, 339-350.

Bhattacharya D (1998) The origin and evolution of protist group I introns. *Protist* 149, 113-122.

Bhattacharya D, Cannone J, Gutell R (2001) Group I intron lateral transfer between red and brown algal ribosomal RNA. *Current Genetics* 40, 82-90.

Bhattacharya D, Damberger S, Surek B, Melkonian M (1996a) Primary and secondary structure analyses of the rDNA group-I introns of the Zygnematales (Charophyta). *Current Genetics* 29, 282-286.

Bhattacharya D, Friedl T, Damberger S (1996b) Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion site lineages in the green algae. *Mol Biol Evol* 13, 978-989.

Bhattacharya D, Surek B, Rüsing M, Damberger S, Melkonian M (1994) Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). *Proceedings of the National Academy of Sciences of the United States of America* 91, 9916-9920.

Burke JM, Belfort M, Cech TR*, et al.* (1987) Structural conventions for group I introns. *Nucl. Acids Res.* 15, 7217-7221.

Cannone JJ, Subramanian S, Schnare MN*, et al.* (2002) The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BioMed Central Bioinformatics* 3, 15.

Cech TR (1988) Conserved sequences and structures of group I introns: building an active site for RNA catalysis - a review. *Gene* 73, 259-271.

Chen C, Barfuss M, Proschold T, Schagerl M (2012) Hidden genetic diversity in the green alga Spirogyra (Zygnematophyceae, Streptophyta). *BMC Evolutionary Biology* 12, 77.

Coleman AW (1999) Phylogenetic analysis of "*Volvocacae*" for comparative genetic studies. *Proceedings of the National Academy of Sciences of the United States of America* 96, 13892-13897.

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 (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *TRENDS in Genetics* 19, 370-375.

Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucl. Acids Res.*, gkm233.

Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* 50, 197-203.

Coleman AW, Mai JC (1997) Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relatedness. *Journal Of Molecular Evolution* 45, 168-177.

Coleman AW, Suarez A, Goff LJ (1994) Molecular delineation of species and syngens in *Volvoceaen* green algae (Chlorophyta). *Journal Of Phycology* 30, 80-90.

Coleman AW, Vacquier VD (2002) Exploring the Phylogenetic Utility of ITS Sequences for Animals: A Test Case for Abalone (Haliotis). *Journal Of Molecular Evolution* 54, 246-257.

Damberger SH, Gutell RR (1994) A comparative database of group I intron structures. *Nucl. Acids Res.* 22, 3508-3510.

De Queiroz K (2007) Species Concepts and Species Delimitation. *Systematic Biology* 56, 879-886.

Drummond CS, Hall J, Karol KG, Delwiche CF, McCourt RM (2005) Phylogeny of *Spirogyra* and *Sirogonium* (Zygnematophyceae) based on rbcL sequence data. *Journal of Phycology* 41, 1055-1064.

Dujon B (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations: a review. *Gene* 82, 91-114.

Ganesan N, Kesavan C (2009) Phylogeny of conserved adenines in linkers of group I introns. *Available from Nature Preceedings* <http://hdl.handle.net/10101/npre.2009.2909.1>.

Gilbert W (1978) Why genes in pieces? *Nature* 271, 501.

Gontcharov AA (2008) Phylogeny and classification of Zygnematophyceae (Streptophyta): Current state of affairs. *Fottea* 8, 87-104.

Gontcharov AA, Marin B, Melkonian M (2002) Molecular phylogeny of conjugating green algae (Zygnemophyceae, Streptophyta) inferred from SSU rDNA sequence comparisons. *Journal of Molecular Evolution* 56, 89-104.

Gontcharov AA, Marin B, Melkonian M (2004a) Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Mol Biol Evol* 21, 612-624.

Gontcharov AA, Marin B, Melkonian M (2004b) Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Molecular Biology and Evolution* 21, 612-624.

Gontcharov AA, Melkonian M (2004) Unusual position of the genus *Spirotaenia* (Zygnematophyceae) among streptophytes revealed by SSU rDNA and rbcL sequence comparisons. *Phycologia* 43, 105-113.

Hainz R, Wöber C, Schagerl M (2009) The relationship between Spirogyra (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe. *Aquatic Botany* 91, 173-180.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.

Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. *Trends in Genetics* 21, 111-119.

Hawes I (1988) The seasonal dynamics of Spirogyra in a shallow, maritime Antarctic lake. *Polar biology* 8, 429-437.

Hey J (2006) On the failure of modern species concepts. *Trends in Ecology & Evolution* 21, 447-450.

Hoshaw RW, McCourt RM (1988) The Zygnemataceae (Chlorophyta): A twenty-year update of research. *Phycologia* 27, 511-548.

Hoshaw RW, Wang J-C, McCourt RM, Hull HM (1985) Ploidal changes in clonal cultures of *Spirogyra communis* and implications for species definition. *American Journal of Botany* 72, 1005-1011.

Hoshaw RW, Wells CV, McCourt RM (1987) A polyploid species complex in *Spirogyra maxima* (Chlorophyta, Zygnemataceae), a species with large chromosomes. *Journal of Phycology* 23, 267-273.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754-755.

Kadlubowska JZ (1984) Conjugatophyceae I - Zygnemales. In: *Süßwasserflora von Mitteleuropa, Chlorophyta VIII* (eds. Ettl H, Gerloff H, Heynig H, Mollenhauer D). Gustav Fischer Verlag, Stuttgart, New York.

Kolkwitz R, Krieger H (1941) Zygnemales. In: *Dr. L Rabenhorst's Kryptogamen-Flora von Deutschland und der Schweiz* (ed. Kolkwitz R), p. 499 pp. Akademische Verlagsgesellschaft Becker & Erler, Leipzig.

Kusel-Fetzmann E, Schagerl M (1993) Verzeichniss der Sammlung von Algenkulturen an der Abteilung für Hydrobotanik am Institut für Pflanzenphysiologie der Universität Wien. *Phyton* 33, 209-234.

Marin B, Klingberg M, Melkonian M (1998) Phylogenetic relationships among the *Cryptophyta*: analyses of nuclear-encoded SSU rRNA sequences support the monophyly of extant plastid-containing lineages. *Protist* 149, 265-276.

Marin B, Palm A, Klingberg Max, 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.

McCourt RM, Hoshaw RW (1990) Noncorrespondence of breeding groups, morphology and monophyletic Groups in *Spirogyra* (Zygnemataceae: Chlorophyta) and the application of species concepts. *Systematic Botany* 15, 69-78.

McCourt RM, Hoshaw RW, Wang J-C (1986) Distribution, morphological diversity and evidence for polyploidy in North American Zygnemataceae (Chlorophyta). *Journal Of Phycology* 22, 307-315.

McCourt RM, Karol KG, Bell J, *et al.* (2000) Phylogeny of the conjugating green algae (Zygnemophyceae) based on rbcL sequences. *Journal of Phycology* 36, 747-758.

Michel F, Dujon B (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *The EMBO Journal* 2, 33-38.

Michel F, Westhof E (1990) Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *Journal of molecular biology* 216, 585-610.

Müller K, Borsch T, Legendre L, *et al.* (2004) Evolution of Carnivory in Lentibulariaceae and the Lamiales. *Plant Biology* 6, 477-490.

Muller KM, Cannone JJ, Gutell RR, Sheath RG (2001) A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta). *Mol Biol Evol* 18, 1654-1667.

Murphy FL, Cech TR (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *Journal of molecular biology* 236, 49-63.

Nielsen H, Johansen SD (2009) Group I introns: Moving in new directions. *RNA Biology* 6, 375-383.

Oliveira MC, Ragan MA (1994) Variant forms of a group I intron in nuclear small-subunit rRNA genes of the marine red alga *Porphyra spiralis var. amplifolia*. *Mol Biol Evol* 11, 195-207.

Philippe H (2000) Opinion: Long branch attraction and protist phylogeny. *Protist* 151, 307-316.

Posada D, Buckley T (2004) Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53, 793-808.

Posada D, Crandall K (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Pröschold T, Marin B, Schlösser UG, Melkonian M (2001) Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas Ehrenberg* and *Chloromonas Gobi*, and description of *Oogamochlamys gen. nov.* and *Lobochlamys gen. nov*. *Protist* 152, 265-300.

Randhawa MS (1959a) *Zygnemaceae* Indian Council of Agricultural Research, New Delhi.

Randhawa MS (1959b) *Zygnemaceae* Indian Council of Agricultural Research, New Delhi.

Rieth A (1983) Eine *Spirogyra* von der Ostsee bei Zingst. *Genetic Resources and Crop Evolution* 31, 317-326.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.

Russell A, Samuel R, Rupp B*, et al.* (2009) Phylogenetics and cytology of a pantropical orchid genus *Polystachya* (Polystachyinae; Vandeae; Orchidaceae); evidence from plastid DNA sequence data. *Taxon*, 16 pp.

Saldanha R, Mohr G, Belfort M, Lambowitz AM (1993) Group I and group II introns. *FASEB J.* 7, 15-24.

Simons J, Van Beem A (1990) *Spirogyra* species and accompanying algae from pools and ditches in the Netherlands. *Aquatic Botany* 37, 247-269.

Simons J, Van Beem AP, de Vries PJR (1982) Structure and chemical composition of the spore wall in Spirogyra (Zygnemataceae, Chlorophyceae). *Acta Bot. Neerl.* 31, 359-370.

Simons J, Van Beem P, De Vries PJR (1984) Induction of conjugation and spore formation in species of *Spirogyra* (Chlorophyceae, Zygnematales). *Acta Bot. Neerl.* 33, 323-334.

Stancheva R, Hall JD, McCourt RM, Sheath RG (2013) Identity and phylogenetic placement of *Spirogyra* species (Zygnematophyceae, Charophyta) from California streams and elsewhere. *Journal of Phycology* 49, 588-607.

Stuessy TF (2009) *Plant taxonomy: the systematic evaluation of comparative data* Columbia University Press.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Sinauer Associates, Sunderland, Massachusetts.

Tel-zur N, Abbo S, Myslabodski D, Mizrahi Y (1999) Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae). *Plant Molecular Biology Reporter* 17, 249-254.

Transeau EN (1951) *The Zygnemataceae (Fresh-water conjugate algae)* The Ohio State University Press.

Van den Hoek C, Mann DG, Jahns HM (1995) Algae: An introduction to phycology. Cambridge University Press, Cambridge.

von Esenbeck CGN, Link HF (1820) *Horae physicae Berolinensis: collectae ex symbolis virorum doctorum H. Linkii...; edicuravit Christianus Godof. Nees ab Esenbeck* Adolphi Marcus.

Weiser B, Noller H (2009) XRNA. University of Santa Cruz.

Yao H, Song J, Liu C*, et al.* (2010) Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. *PLoS ONE* 5, e13102.

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.

Zaug AJ, Grabowski PJ, Cech TR (1983) Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature* 301, 578-583.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* 31, 3406-3415.

Zwirn M (2010) *Sexuelle Fortpflanzung und taxonomische Fragestellungen bei der Süßwasseralge Spirogyra (Zygnemataceae, Streptopyhta)*, University of Vienna.

Zwirn M, Chen C, Uher B, Schagerl M (2013) Induction of sexual reproduction in Spirogyra clones - does an universal trigger exist? *Fottea* 13, 77-85.

Chapter 1

Slow evolution of 1506 group I intron in Spirogyra LINK 1820 (Zygnematophyceae, Streptophyta), a fast evolving lineage in the Zygnemataceae

Pond in Limsdorf, GER, 2007

Slow evolution of 1506 group I intron in *Spirogyra* LINK 1820 (Zygnematophyceae, Streptophyta), a fast evolving lineage in the Zygnemataceae

Published: Fottea 2012, 12(2):255–272.

Chen Charlotte, Schagerl Michael*

University of Vienna, Department Limnology, Althanstraße 14, A-1090 Vienna, Austria

* author of correspondence, email: michael.schagerl@univie.ac.at

Running title: 1506 group I intron in *Spirogyra* LINK 1820

Abstract

Phylogenetic analyses of SSU rDNA sequences of 130 *Spirogyra* strains have revealed that these strains were subdivided into eight clades. Approximately 60% of the assessed strains (clades A – D) contain a 1506 group I intron, whereas strains without introns form individual clades (E – H). The *Spirogyra* intron shared the common insertion site of the Zygnematalean intron (position 1506 relative to the *Escherichia coli* small-subunit rRNA). Phylogenetic analyses of the *Spirogyra* group I intron showed the monophyletic origin within the Zygnematophyceae. Therefore, we assume the secondary loss of the intron in clades E – H is caused by the high evolutionary rate of *Spirogyra* and its long evolutionary history. The *Spirogyra* intron belongs to the IC group I introns and shares many common features with the intron of other Zygnematophyceae (the typical domain structure (P1 - P9), the base composition, the highly conserved regions the U preceding the 5' splice site and the G to which it pairs, and the G preceding the 3' splice site are typical for IC group I intron). *Spirogyra* group I introns exhibit features of early desmids (optional P2 domain) as well as of later diverging desmids (variation from the typical L5b-GAAA tetraloop). The P2 domain shows an additional optional P2 sub-domain in clade B. Surprisingly, the mutation rate of the *Spirogyra* SSU rRNA exceeds the rate of the intron by far. Evolutionary rates differ significantly within the *Spirogyra* SSU rRNA accessions, but not within the respective group I intron sequences.

Keywords: *Spirogyra*, group I intron, intron subgroup IC1, optional P2 domain, secondary structure, evolutionary rate

Abbreviations: nt, nucleotide; rDNA, ribosomal DNA; SSU, small subunit; A, adenine; G, guanine; C, cytidine; U/T uridine/ thymine; MP, maximum parsimony; ML, maximum likelihood; NJ, neighbor joining; PP, posterior probability; Meso, Mesotaeniceae; Zygn, Zygnemetaceae; Desm, Desmidiaceae; NHS, Non-Homoplasious Synapomorphy; (H-)CBC, (Hemi-) Compensatory Base Change;

Introduction

The genus *Spirogyra* LINK (Zygnematophyceae, Streptophyta) is characterized by unbranched filaments, coiled chloroplasts and a special type of reproduction (conjugation) with complete absence of flagellated reproductive stages (Kadlubowska 1984; Kolkwitz & Krieger 1941; McCourt *et al.* 1986; Randhawa 1959b; Transeau 1951). The filaments are covered by a layer of mucilage, which is responsible for the slimy appearance (HOSHAW & MCCOURT 1988). The genus is distributed worldwide in fresh to slightly brackish water bodies (Hoshaw & McCourt 1988; Rieth 1983; Simons & Van Beem 1990). *Spirogyra* covers a wide ecological range from dystrophic mountain lakes and bogs to highly eutrophicated systems (HAINZ et al. 2009).

The term "intron" was introduced by Gilbert (1978) to describe intragenic regions that occur alternating with coding regions and that would subsequently be lost from the mature messenger RNA. Introns are of interest because they are used to gain insight into the evolution of protein synthesis and how primitive RNAs could have catalyzed their own replication (SALDANHA et al. 1993). Beside group I introns, group II intron and spliceosomal introns exist, which are defined by their way of splicing (self-splicing vs. RNA operated splicing) and have a distinct secondary structure (SALDANHA et al. 1993). Group I introns are an important class of RNA enzymes that also exhibit a typical secondary structure that is expressed in the sub-domains P1 – P9 and the conserved core regions P, Q, R and S (BURKE et al. 1987; CECH 1988). Five major subgroups of group I introns (IA, IB, IC, ID and IE) and 10 minor subgroups can be distinguished (CANNONE et al. 2002). The first rDNA group I Intron was discovered in the LSU rRNA of the protist *Tetrahymena thermophila* (ZAUG et al. 1983). Since then, group I introns have been found in rRNAs, mitochondrial and chloroplast genomes in a great variety of organisms from fungi (MICHEL & DUJON 1983) to many algal groups (Besendahl & Bhattacharya 1999; Bhattacharya *et al.* 2001; Damberger & Gutell 1994; Muller *et al.* 2001). So far, no function could be linked to group I introns nor do they affect the phenotype (DUJON 1989). Two theories about their manner of distribution have been formulated: vertical inheritance together with the gene they interrupt or, horizontal transfer by an unknown vector between individuals of different species, genera or families (BHATTACHARYA 1998). A common ancestry of all group I introns is likely but still remains obscure (Bhattacharya 1998; Oliveira & Ragan 1994). Group I introns are typically capable of self-splicing, but they have lost this ability within the Desmidiaceae (BESENDAHL & BHATTACHARYA 1999).

Spirogyra is a very species-rich genus within the Zygnematales with an estimated species number of 386 (KADLUBOWSKA 1984). Older monographs had somewhat lower numbers between 275 (TRANSEAU 1951) and 289 (Randhawa 1959b). Even though the current species number is questioned (HOSHAW et al. 1985), there are only few studies with adequate taxon sampling (DRUMMOND et al. 2005) that give more insight into genetic diversity in this genus. The group I intron of the genus *Spirogyra* has not yet been described in detail; Gontcharov who first discovered the *Spirogyra* group I intron, only made a brief characterization of it as a typical zygnematophycean intron, solely based on three intron sequences (GONTCHAROV et al. 2002).

Phylogenetic analyses of SSU rDNA sequences (Gontcharov *et al.* 2002; Gontcharov *et al.* 2004a; Gontcharov & Melkonian 2004) have revealed that *Spirogyra* represents a so-called "long branch" taxon sensu Philippe (2000). We wanted to investigate if the long branch attraction phenomenon (LBA) is also present in a group I intron phylogeny and if evolutionary rates of the *Spirogyra* intron is comparable to the SSU rRNA region. In this study, we provide a detailed view of the characteristics of this RNA intron in the genus *Spirogyra* and on the variability of a group I intron within a single genus. Furthermore, the phylogenetic position of the *Spirogyra* intron is evaluated and its secondary structure is compared to known group I introns of other Zygnematalean taxa to test if the intron has taxonomical value as suggested by Gontcharov, Marin et al. (2002).

Material and Methods

Origin of organisms

Spirogyra clones used in this study originated from a field survey conducted in 2006 and 2007 (Hainz et al. 2009). Single filaments were isolated by the author (CC) and incorporated into the Algenkultursammlung Wien (ASW). The non-axenic clones were maintained in 100 ml Erlenmeyer flasks with Desmids medium (Kusel-Fetzmann & Schagerl 1993) at 18°C under low light conditions at a 16:8 l:d light cycle (provided by either Philips TLD 36W/33 or Osram FQ 39W/840 LUMLUX Cool White). Because only few strains could be identified at species level, cultures were labeled with a code for the corresponding sampling site and date (Table 5). For our study, we considered 130 Spirogyra isolates from different sampling sites and with different vegetative morphologies to cover various ecological niches.

DNA extraction

Prior to extraction the cultures were transferred into a defined mineral medium (modified Woods Hole medium; (SIMONS et al. 1984)). After 4 to 6 weeks, the algae were harvested with a sterile needle and put into a sterile 2 ml microcentrifuge tube. Samples were frozen at -80°C for at least 4 h and then lyophilized for at least 48 h to improve the DNA yield. Afterwards, the samples were placed in 2 ml Eppendorf tubes containing 5 to 7 glass beads (3 mm diam.) and ground with a homogenizing mill. Total DNA was extracted following a modified CTAB protocol ((TEL-ZUR et al. 1999) modified after (RUSSELL et al. 2009)).

DNA amplification and sequencing

Primers used in this study are given in Table 1 (MARIN et al. 1998; PRÖSCHOLD et al. 2001). The PCR reaction mixture was prepared according to the manufacturer's recommendation. For each PCR reaction, a 10 µl mixture was prepared containing 9 µl ABGene Reddy Mix PCR Master Mix, 0.2 µl for each primer at 20 pM.µ l^{-1}, 0.4 µl dimethyl sulfoxide (Sigma) and 0.2 µl DNA template. When the PCR result was unsatisfactory due to low DNA concentration, up to 0.5 µl DNA template was used; when DNA template volume was increased, dimethyl sulfoxide volume was reduced to maintain the total volume of 10 µl, accepting a slight shift in the overall ratio of ingredients. The PCR reaction conditions were an initial hold at 80°C for

5 min followed by 36 cycles starting with a denaturation step at 95°C for 30 s, an annealing step at 55°C for 30 s and an extension step at 72°C for 2 min. A final extension step at 72°C for 8 min and the final hold at 4°C were performed after the 36 cycles were completed. The amplified DNA was cleaned by incubating at 37°C for 45 min, followed by denaturing at 80°C for 15 min together with the enzymes Exonuclease I and Shrimp Alkaline Phosphatase (both from Fermentas) and then subjected to a cycle sequencing reaction. The cycle sequencing reaction conditions were an initial hold at 96°C for 1 min followed by 35 cycles starting with a denaturation step at 96°C for 10 s followed by an annealing step at 50°C for 5 s and an extension step at 60°C for 4 min. The end of the cycles was followed by a final hold at 4°C. Sequencing was performed on a 16-capillary sequencer (Applied Biosystems 3130xl Genetic Analyzer) following the manufacturer's protocols. The SSU rDNA sequences were used in the phylogenetic analyses; their GenBank accession numbers are given in Table 5.

Sequence Alignment and Analysis

The group I intron sequences were manually aligned according to the secondary structures of the P, Q, R and S elements to identify homologous regions (Bhattacharya *et al.* 1994; Michel & Westhof 1990). The alignment was refined by comparison of the secondary structure of the sequences. Secondary structure was determined via the Rensselaer bioinformatics web server using mfold (ZUKER 2003). Secondary structure diagrams were generated with the software XRNA (WEISER & NOLLER 2009). Base composition was calculated from the alignment using BioEdit (HALL 1999). Only unambiguously aligned regions of the sequences were used for analyses, gap-rich regions were excluded. To compare the *Spirogyra* sequences with other Zygnematalean sequences, initial analyses of all *Spirogyra* sequences were performed to detect the variability among the *Spirogyra* strains (data not shown). Three different data sets were further analyzed: (1) For confirmation of the *Spirogyra* clade position among the other algae included, for checking the long branch status, and for demonstration of clade grouping in *Spirogyra*, an analysis of a SSU alignment comprising 12 Desmidiaceae, Mesotaeniaceae and Zygnemataceae sequences was conducted (Table 2, data set 6). Twelve *Spirogyra* sequences selected from the initial analysis were added; two sequences were taken from each clade containing the intron; one of the clades without intron. (2) For a phylogenetic understanding of the intron (Table 2, data set 7) in relation to other Zygnematophyceaen introns, an analysis of a data set comprising the same taxa as in the SSU alignment was performed. 8 *Spirogyra* sequences of the initial analysis from clades containing the intron were added. (3) The group I intron sequence of 51 *Spirogyra* strains (Table 2, data set 8). Identical sequences were excluded from analyses to reduce computing effort.

The alignment of the desmid SSU sequences consisted of 1717 unambiguously aligned bases, the respective intron alignment of 263 unambiguously aligned bases. To avoid LBA (PHILIPPE 2000) and following Gontcharov's argument, that monophyly of the Zygnematophyceae is undoubted, but that its position in the streptophytes is unclear, and therefore no suitable outgroup can be chosen, unrooted phylogenies were used for calculations in this study (Gontcharov *et al.* 2004b). The *Spirogyra* group I intron data set comprised 251

unambiguously aligned bases. To test for the best evolutionary model for the analyses, the log likelihood values of 56 models using Modeltest 3.7 were compared (Posada & Buckley 2004; Posada & Crandall 1998).

The phylogenetic trees presented were inferred by maximum likelihood settings using PAUP* 4.0b10 (SWOFFORD 2003). The best models were chosen according to the Akaike Information Criterion by Modeltest 3.7. For the desmid SSU alignment analyses the GTR+G model was chosen, for the desmid intron alignment analyses the GTR+I+G model and for the *Spirogyra* intron alignment the TrN+I+G model. Bootstrap support values were calculated by distance (neighbor joining (NJ; 1000 replicates), maximum parsimony (MP; 1000 replicates) and maximum likelihood methods (ML; 1000 replicates; using the appropriate evolutionary model for each data set). Bayesian inference was calculated using MrBayes 3.1.2 (using following settings: 1 - 3 million generations; MCMC chains = 4; all trees below the burnin value of 0,01 were excluded as burnin, the majority consensus tree was calculated using PAUP*; (HUELSENBECK & RONQUIST 2001; RONQUIST & HUELSENBECK 2003)). Details of the evolutionary models and Bayesian analyses are given in the legend of the respective figures. Relative rate tests were carried out between all accessions used for phylogeny in GRate 0.4 (MÜLLER et al. 2004).

Secondary structure analyses

To construct a putative secondary structure for *Spirogyra*, existing secondary structure models from other taxa were obtained from http://www.rna.ccbb.utexas.edu/ (CANNONE et al. 2002). Homologous regions (P, Q, R and S (Bhattacharya *et al.* 1994; Michel & Westhof 1990)) were identified, aligned, and assigned to the respective P domains. For secondary structure calculation, the individual sequence fragments were uploaded to the Rensselaer bioinformatics web server using mfold (ZUKER 2003).

Secondary structures were compared among the clades for *Spirogyra*. To find all non-homoplasious synapomorphies (NHS), the method described in Marin, Palm et al. (2003) was applied. With this method NHS, Compensatory Base Change (CBC) and Hemi-CBC (H-CBC; only one base changes, but the base pairing remains intact) can be detected. The two NHS criteria were applied: (1) absence of convergent evolution outside the clade, and (2) strict conservation within the clade.

Results

Phylogenetic analyses

Phylogenetic analyses of SSU rDNA sequences of 130 *Spirogyra* strains have revealed that these strains were subdivided into eight clades. Numerous strains of these clades (clades A – D) contain a 1506 group I intron, whereas strains without introns belong to an individual derived clade (clade E - H, Fig. 1). *Spirogyra* was analyzed together with other Zygnematalean taxa to confirm its position within the Zygnemataceae taxa, and to test whether the same position would be true for the *Spirogyra* group I intron. Seventy two out of 130 *Spirogyra* strains contained a 1506 group I intron. Additionally, base frequencies were

compared among the groups and between the SSU and the intron. Accepted secondary structure suggestions of the taxa used for the desmid alignment were compared to putative secondary structures of *Spirogyra* group I introns.

For the desmid 18s SSU phylogeny shown in Fig. 1, a data set of the SSU with a total of 1715 unambiguously aligned positions was used. The data set was analyzed to depict the clade grouping in the genus *Spirogyra* and test the group I introns as molecular marker in comparison to the well established SSU. We were also interested, if the *Spirogyra* clades established in the 18s SSU phylogeny will be reflected in the group I intron phylogeny. The *Spirogyra* group I intron alignment was analyzed separately using 263 unambiguously aligned positions (Fig. 2B). In the tree derived from intron alignment analyses (Fig. 2A), two different lineages are visible in our analyses, the Desmidiaceaen branch and the Mesotaeniaceaen/ Zygnemataceaen lineage; *Closterium littorale* could not be placed in either lineages. The only phylogenetic relation supported by bootstrap and Bayesian values outside the *Spirogyra* clade is the *Gonatozygon/Genicularia* clade with very high support, but is placed in the Mesotaeniaceaen/ Zygnemataceaen clade. All other phylogenetic relationships in these groups remain questionable. Within *Spirogyra*, clade A, B and D receive high bootstrap support. There is no clade C formed in this tree; the members of this clade are placed on individual branches basal to clade B and D. In tree derived from SSU alignment analyses (Fig. 1), more branches receive support by bootstrap and Bayesian values, still not all proposed phylogenetic relationships are reliable. Only two clades of closely related taxa are supported: the *Sphaerozosma*/*Cosmocladium* clade and the *Gonatozygon*/*Genicularia* clade. Unlike in the intron alignment, Desmidiaceae and Mesotaeniaceae/ Zygnemataceae are not separated into two lineages. The Mesotaeniaceaen/ Zygnemataceaen branch is placed within a larger clade with other taxa in basal positions. Bootstrap support within *Spirogyra* is very high – accessions that show no 1506 group I intron form a well supported clade.

Spirogyra group I introns were analyzed separately (Fig. 2B). 50 individual sequences obtained by the author and *Sirogonium sticticum* were included using 251 unambiguously aligned positions. Seven sequences had to be excluded from analyses because of sequencing problems. The tree was rooted with clade A. Clade B is not supported by all algorithms, but the taxa of clade B still form a branch together in the maximum likelihood tree. The individual branches of Clade C are well supported, but are not located in a collective branch. Clade D is well supported by Bayesian and bootstrap values.

Evolutionary rates

Evolutionary rates were inferred by pair wise comparison of unambiguously aligned positions of the Desmidiaceae/ Mesotaeniaceae/ Zygnemataceae intron (Table 2, data set 7) and the Desmidiaceae/ Mesotaeniaceae/ Zygnemataceae SSU alignment (Table 2, data set 6) in GRate (Table 3). In both data sets, a significant difference between the evolutionary rate of *Spirogyra* and all other taxa analyzed could be detected. Within the *Spirogyra* intron sequences, many accessions exhibit a rate that is not significantly different (Table 3; highlighted in grey). Whereas in the *Spirogyra* SSU sequences, more significantly different

evolutionary rates could be found. When including more taxa to the intron alignment and using not only the unambiguously aligned bases, there is no significant difference between *Spirogyra* and other genera (data not shown). However, the trend in the SSU alignment is the same – different evolutionary rate to the remaining taxa, diversity of evolutionary rates within *Spirogyra*.

Additionally, a modeltest analysis was conducted to compare substitution models for 5 data sets (Table 2), which are (1) all taxa of the SSU rRNA alignment except for *Spirogyra*, (2) *Spirogyra* accessions of the SSU rRNA alignment that contained a group I intron, (3) all *Spirogyra* accessions of the SSU rRNA alignment, (4) all taxa of the group I intron alignment except for *Spirogyra* and (5) all *Spirogyra* accessions of the intron alignment. The Desmidiaceae/ Mesotaeniaceae/ Zygnemataceae SSU alignment (data set 1) contains the largest proportion of invariant sites, the Desmidiaceae/ Mesotaeniaceae/ Zygnemataceae intron data set (data set 4) the smallest. The largest base frequency differences were found between the Desmidiaceae/ Mesotaeniaceae/ Zygnemataceae SSU (data set 1) and the *Spirogyra* intron data set (data set 5). The *Spirogyra* SSU +I data set (data set 2) has the highest base substitution rates; the *Spirogyra* intron data set (data set 5) the lowest (except for the A – U substitution rate). The *Spirogyra* intron data set exhibits the highest average G-C content (0.6108), the Desmidiaceae/ Mesotaeniaceae/ Zygnemataceae SSU the lowest (0.4628).

Base composition

Relative frequencies of the four bases were compared for the complete SSU and the intron (Fig. 3). The base composition differed considerably between the SSU rRNA and the 1506 group I intron. The SSU is relatively U – rich and C – poor; in the intron the opposite was observed. The G – content is higher than the A – content in both SSU and intron. The content of individual bases varies more in the intron than in the SSU eg., G-content 25-34% in introns, in the SSU from 26-29%; the same holds true for the other bases. These were also reflected in the Modeltest settings (see Fig. 1, 2A + B, Table 2). The relative content of U and A are very similar in the 18s SSU, whereas the content of G and C show a relatively big difference. In the intron, G and C show similar relative contents, and the percentage of A and U differ from each other.

Sequence identity of *Spirogyra*

Seventy two sequenced strains exhibited the 1506 group I intron. The total length of both the 18s SSU and the 1506 group I intron were compared individually. Out of these 72 SSU sequences 19 sequences were unique, 13 were found more than once. The group I intron exhibited 50 unique variants and only 8 types were found more than once. All *Spirogyra* strains that had identical intron sequences also had identical SSU sequences – 22 individual accessions forming 6 different groups. 58% of the strains that had identical SSU sequences, differed in intron sequences. In Fig. 2B, sequences that were obtained from isolates from the same sampling site are marked with open grey boxes, sequences from different isolation

sites with closed boxes. Not all isolates from the same site had the same intron, but only small differences (up to 5 nt) could be observed.

Secondary structures

Spirogyra group I introns showed the typical secondary structure (Fig. 4). The secondary structure of the *Spirogyra* group I intron exhibits the typical P5 extension of the IC introns (CECH 1988; MICHEL & WESTHOF 1990), which together with the P2.1 domain places them in the subgroup IC1. The overall sequence length of the intron varies considerably from 387 to 771 nt. Four taxa (all members of Clade C_1, Fig. 2B) contained an insertion in P2 (229 – 316 nt). Short open reading frames were found, but do not correspond to known genes.

Tetraloops are the most frequent forms of terminal loops, yet *Spirogyra* tetraloops do not fit the proposed consensus sequences GNRA or UNCG (MURPHY & CECH 1994) (where N = A, G, C or U and R =A or G).

The intron shows the highly conserved U preceding the 5' splice site and the G to which it pairs (Fig. 4, 1), and the G preceding the 3' splice site (CECH 1988). The most conserved regions in the *Spirogyra* 1506 group I intron are the intron core regions: P3, P7 and J7/8 and P4, P6 and J4/5.

Clade A differs in the sequence of P3 and P7 from the typical *Spirogyra* group I intron. The typical P3 sequence is CGCGACC (see Fig. 4), while the intron sequences of clade A have a mutation in the 6th nt from C to U. The typical P7 sequence is GAGGC –GCCUC. Clade A strains show a different sequence (GAGAC – GUCCU/G). The first four base pairs in P1 and P8 are highly conserved (Fig. 4, 2 & 3). They are identical to a putative consensus structure of zygnematalean taxa (BHATTACHARYA et al. 1996a). P2 has a highly conserved sub-helix that is present in all taxa (Fig. 4, 4) and shows an additional optional sub-helix that is only present in strains of clade B (Fig. 4, 5). The expected GAAA tetraloop in the P5 extension and the A-rich bulge are not present. Instead, L5b shows highly divergent sequences (Fig. 4, 6).

The remaining intron core region is highly conserved as well (P4, P6, J4/5); the peripheral extensions are more divergent (MURPHY & CECH 1994). The J4/5 contains AAA in 65 cases (Fig. 4, 7); in the other sequences the first A is changed to a G. The most frequent terminal loop in L5c (39/72) is an AACG-tetraloop (Fig. 4, 8). Domains P8 and P9 are longer in clade D (58 to 82 nt, and 80 to 105 nt, respectively). Also the P4-6 domain is longer in clade D (197 to 207 nt) than in other clades.

The search for NHS revealed 58 NHS, 4 H-CBCs and 5 CBCs (Table 4). Clade A showed the most diverging pattern from the other clades; it was characterized by 26 NHS, 3 H-CBCs (P3, P5, P7) and 2 CBCs (P3, P7). Clade B had 2 NHS and 1 CBC (P6), Clade C 18 NHS and 2 CBCs (P2.1, P6), Clade D 8 NHS and 1 H-CBC (P6) and *Sirogonium* showed 4 NHS.

Discussion

The *Spirogyra* intron is a typical Zygnematalean intron (GONTCHAROV et al. 2002) and belongs to the CI subgroup of the group I introns based on the characterization of the primary and secondary structure features by Michel & Westhof (1990) and Damberger & Gutell (1994). The introns show most features of the characteristic catalytic cores for enzymatic activity consisting of P3, P4, J4/5, P6, P7, P8 and P9 which are highly conserved (SALDANHA et al. 1993). They also exhibit typical linker regions J3/4: AAC, J6/7: UCA, J8/7: AAGAUA (GANESAN & KESAVAN 2009). As expected, the more peripheral extensions are more divergent (MURPHY & CECH 1994). Large extensions of the P5 domain are typical for such introns. However, terminal loop of P5 (L5b) does not have the expected GAAA tetraloop and also the A-rich bulge is missing, which is a characteristic of later-diverging Desmidiaceae according to Bhattacharya, Damberger et al. (1996a). All sequenced *Spirogyra* introns showed the optional P2 helix that is characteristic for saccoderm and filamentous desmids as well as for the Zygnematales (BESENDAHL & BHATTACHARYA 1999) which is highly conserved in sequence. In the strains of Clade B, an additional sub-domain in P2 could be found, but it varies in sequence and size. Four strains of Clade C displayed an insertion in P2 (229 – 316 nt), which is shorter compared to insertions of other groups such as in the Bangiales (MULLER et al. 2001), but much longer than the zygnematalean 1506 group I intron insertion found in *Mougeotia* (147 nt in P6) (BHATTACHARYA et al. 1996a).

When comparing base frequencies of *Spirogyra* SSU to the values for desmids given in Bhattacharya, Damberger et al. (1996a), a higher C-content (Spiro: 22% to 25%, desmid: ~20% to 21%), a lower U-content (Spiro: 23% to 26%, desmid: ~27%), a higher G-content (Spiro: 26% to 29%, desmid: ~26%) and a slightly lower A-content (Spiro: 23% to 26%, desmid: ~26%) was observed. A shift in relative base frequencies from U to C and from A to G is characteristic for the *Spirogyra* SSU. The percentual G+C content of the group I intron ranges from 52% to 68%. Compared to the typical desmid intron (BHATTACHARYA et al. 1996a), the *Spirogyra* intron exhibits a much higher G-C value, but for the SSU, the G+C percentage of the unambiguously aligned positions ranged from 56% to 64% and therefore agrees well with the observed range of G+C percent contents of group I intron sequences given in the ML tree in Bhattacharya, Surek et al. (1994). This C – U bias (Fig. 3) is reflected also in the modeltest results (Fig. 2A + B), which explains why *Spirogyra* is a so-called long branch taxon. The *Spirogyra* intron is different from the typical Desm/Meso/Zygn intron, but its role among this group of algae may not be exceptional. *Closterium littorale* also exhibits a very long branch in the intron phylogeny in Fig. 2A. The Desmidiaceae do not form a clade in Fig. 1 + 2A, unlike to the findings using *rbcL* phylogenetic analyses (MCCOURT et al. 2000). Only the *Gonatozygon/ Genicularia* clade and the *Sphaerozosma/ Cosmocladium* clade agrees with their *rbcL* tree and with phylogenetic trees presented by Besendahl & Bhattacharya (1999). These dissimilarities can be attributed to different taxon sampling as we did not include the same taxa to our study. However, similar to those results, *Mougeotia* and *Mesotaenium* could not be placed in a defined relationship. Because there is no bootstrap or Bayesian

support for any of the relationships among the Mesotaeniaceae/ Zygnemataceae, except for the *Spirogyra/Sirogonium* branch, the phylogenetic relationship remains unclear.

The genus *Spirogyra* shares some similarities with the genus *Mougeotia*. Besendahl did not find the 1506 group I intron in one analyzed species of *Mougeotia,* whereas other species do contain the intron (BESENDAHL & BHATTACHARYA 1999). This genus also shows both, species with and without intron. Maybe these species differences are a characteristic feature for certain Zygnematalean genera, but this hypothesis has to be solved with enhanced within-genus taxon sampling.

Although there are suggestions to use group I introns as a taxonomic marker to provide more divergent sequences for resolving evolutionary relationships (BHATTACHARYA 1998), we assume that analyses from the intron alone do not provide a reliable source for phylogenetic interpretation. As shown in our present study, introns are not necessarily present in all members of a genus, which makes it difficult to judge genetic relatedness between taxa with and without intron. Furthermore, they are more difficult to align, decisions on including or excluding characters are more crucial for analyses, and resulting longer branches and low bootstrap/Bayesian support pave the way for misinterpretation. Even within a genus, results are not sufficient for clear statements on difficult taxa. However, when intron sequences are used as a supporting marker to the conserved SSU, it may enhance bootstrap/Bayesian values. The problem with phylogenetic analyses of SSU and group I introns is that both represent rather fast evolving parts of the genome for *Spirogyra* (Gontcharov *et al.* 2004a). To level all long branch effects, it would be better to combine markers with different evolutionary rates (Gontcharov *et al.* 2004a). The use of the intron as a taxonomic marker within the genus *Spirogyra* is problematic as faster evolving taxa that experienced the loss of the intron would be excluded from analyses and interpretation.

However, it is likely that a common ancestor of the *Spirogyra* contained the group I intron that was subsequently lost in more derived clades like clade E (Fig. 1). The single origin of the Zygnematalean group I intron has already been stated and discussed by Bhattacharya (Besendahl & Bhattacharya 1999; Bhattacharya 1996; Bhattacharya *et al.* 1996b; Bhattacharya *et al.* 1994)

and Gontcharov (2008). Haugen, Simon et al. (2005) state that once an intron invades a population and becomes fixed, the homing endonuclease gene responsible for splicing activity accumulates mutations and is consequently inactivated or lost. This loss of self-splicing capacity of the 1506 intron may therefore have originated within the common ancestor of the Zygnematales (BESENDAHL & BHATTACHARYA 1999). *Spirogyra* exhibits both – presence and absence of a 1506 intron. This disposition for a secondary loss is potentially caused by accumulation of mutations because of a long evolutionary history and/or a high evolutionary rate. The loss of an optional subdomain in P2 defines the derived desmids (BESENDAHL & BHATTACHARYA 1999). The presence of this optional P2 domain – therefore a characteristic of early Desmidiaceae – supports the theory of a long evolutionary history. The absence of the GAAA tetraloop in L5b and the high divergence of the intron and the SSU rRNA among the *Spirogyra* strains and from other Desmidiaceae/ Zygnemataceae support the theory of a high evolution rate, as Gontcharov, Marin et al. (2002) already mentioned for

the Zygnematales. The phylogenies generated from the *Spirogyra* sequence data of the SSU rRNA and the 1506 group I intron are largely congruent (Fig. 1+2A), and support the hypothesis of a long term immobility and vertical inheritance of the intron for the Zygnematales (BHATTACHARYA 1998).

The variability of the intron between the individual *Spirogyra* clades is high. The intron of Clade A is characterized by 2 CBCs and 3 H-CBCs, all located in highly conserved regions of the intron core (Fig. 4), which explains the long branch formed by this clade. The remaining clades do not show as many H-CBCs or CBCs nor NHS. This reflects not only the generally high evolutionary rate in *Spirogyra* but also shows that in different clades have different evolutionary rates.

When comparing evolutionary rates among the different groups of taxa and different sequence regions used in this study, significant differences in evolutionary rate can be found among different data sets and taxa, respectively. A higher rate of base substitution can be found in the Desm/Meso/Zygn intron data set than in the Desm/Meso/Zygn SSU data set. The Desm/Meso/Zygn SSU data set also has the highest portion of invariant sites, marking it as the least variable data set. Mutations in the group I intron of the Desm/Meso/Zygn group are fixed at a higher rate than in the respective SSU rRNA sequence. However, in *Spirogyra* SSU rRNA, much higher substitution rates were found than in the group I intron sites – up to 8 times higher. *Spirogyra* not only differs significantly in evolutionary rate from the other taxa used; it also shows more individual, significantly different evolutionary rates within the SSU rRNA than within the intron. The 1506 group I introns of *Spirogyra* strains used here are more similar in evolutionary rate than the respective SSU rRNA sequences of the same *Spirogyra* strain. Selective constraints seem to be stronger in the SSU rRNA, where mutations influence the functionality of the resulting ribosomal subunit. Whereas in the group I intron with no known function (NIELSEN & JOHANSEN 2009), mutations are more likely to be silent. *Spirogyra* is subjected to the same selective constraints as other algae; still a different mutation pattern is established.

Spirogyra is an early divergent genus of the Zygnematophyceae that possesses a very high evolutionary rate concerning the intron sequences and even more in the SSU rRNA sequences. Similar to *Mougeotia*, for which a unique insertion in the P6 region was reported (BHATTACHARYA et al. 1996a), some *Spirogyra* strains have large insertions. This variability in the intron and the number of mutations in otherwise conserved regions (P7) lead to the theory that the loss of the intron is a logical and likely step in evolutionary history, and that genera that have not yet developed taxa without group I introns are evolutionary younger or slower than *Spirogyra*. Thus the genus *Spirogyra* may be a helpful tool for tracking and evaluating evolutionary processes.

Acknowledgements

We are very grateful to Prof. Annette Coleman, Brown University, for her long-lasting support and cooperativeness. This study was supported by the Austrian Science Fund, Project number P18465.

References

BESENDAHL, A. & BHATTACHARYA, D. (1999): Evolutionary analyses of small-subunit rDNA coding regions and the 1506 group I introns of the Zygnematales (Charophyceae, Streptophyta). - Journal of Phycology 35(3): 560-569.

BHATTACHARYA, D. (1996): Analysis of the distribution of bootstrap tree lengths using the maximum parsimony method. - Molecular Phylogenetics and Evolution 6(3): 339-350.

BHATTACHARYA, D. (1998): The origin and evolution of protist group I introns. - Protist 149(2): 113-122.

BHATTACHARYA, D.; CANNONE, J. & GUTELL, R. (2001): Group I intron lateral transfer between red and brown algal ribosomal RNA. - Current Genetics 40(1): 82-90.

BHATTACHARYA, D.; DAMBERGER, S.; SUREK, B. & MELKONIAN, M. (1996a): Primary and secondary structure analyses of the rDNA group-I introns of the Zygnematales (Charophyta). - Current Genetics 29(3): 282-286.

BHATTACHARYA, D.; FRIEDL, T. & DAMBERGER, S. (1996b): Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion site lineages in the green algae. - Mol Biol Evol 13(7): 978-989.

BHATTACHARYA, D.; SUREK, B.; RÜSING, M.; DAMBERGER, S. & MELKONIAN, M. (1994): Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). - Proceedings of the National Academy of Sciences of the United States of America 91(21): 9916-9920.

BURKE, J.M.; BELFORT, M.; CECH, T.R.; DAVIES, R.W.; SCHWEYEN, R.J.; SHUB, D.A.; SZOSTAK, J.W. & TABAK, H.F. (1987): Structural conventions for group I introns. - Nucl. Acids Res. 15(18): 7217-7221.

CANNONE, J.J.; SUBRAMANIAN, S.; SCHNARE, M.N.; COLLETT, J.R.; D'SOUZA, L.M.; DU, Y.; FENG, B.; LIN, N.; MADABUSI, L.V.; MUELLER, K.M.; PANDE, N.; SHANG, Z.; YU, N. & GUTELL, R.R. (2002): The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. - BioMed Central Bioinformatics 3: 15.

CECH, T.R. (1988): Conserved sequences and structures of group I introns: building an active site for RNA catalysis - a review. - Gene 73(2): 259-271.

DAMBERGER, S.H. & GUTELL, R.R. (1994): A comparative database of group I intron structures. - Nucl. Acids Res. 22(17): 3508-3510.

DRUMMOND, C.S.; HALL, J.; KAROL, K.G.; DELWICHE, C.F. & MCCOURT, R.M. (2005): Phylogeny of *Spirogyra* and *Sirogonium* (Zygnematophyceae) based on rbcL sequence data. - Journal of Phycology 41(5): 1055-1064.

DUJON, B. (1989): Group I introns as mobile genetic elements: facts and mechanistic speculations: a review. - Gene 82(1): 91-114.

GANESAN, N. & KESAVAN, C. (2009): Phylogeny of conserved adenines in linkers of group I introns. - Available from Nature Preceedings <http://hdl.handle.net/10101/npre.2009.2909.1>.

GILBERT, W. (1978): Why genes in pieces? - Nature 271(5645): 501.

GONTCHAROV, A.A. (2008): Phylogeny and classification of Zygnematophyceae (Streptophyta): Current state of affairs. - Fottea 8(2): 87-104.

GONTCHAROV, A.A.; MARIN, B. & MELKONIAN, M. (2002): Molecular phylogeny of conjugating green algae (Zygnemophyceae, Streptophyta) inferred from SSU rDNA sequence comparisons. - Journal of Molecular Evolution 56(1): 89-104.

GONTCHAROV, A.A.; MARIN, B. & MELKONIAN, M. (2004a): Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). - Molecular Biology and Evolution 21(3): 612-624.

GONTCHAROV, A.A.; MARIN, B. & MELKONIAN, M. (2004b): Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). - Mol Biol Evol 21(3): 612-624.

GONTCHAROV, A.A. & MELKONIAN, M. (2004): Unusual position of the genus *Spirotaenia* (Zygnematophyceae) among streptophytes revealed by SSU rDNA and rbcL sequence comparisons. - Phycologia 43(1): 105-113.

HAINZ, R.; WÖBER, C. & SCHAGERL, M. (2009): The relationship between *Spirogyra* (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe. - Aquatic Botany 91(3): 173-180.

HALL, T.A. (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. - Nucleic Acids Symposium Series 41: 95-98.

HAUGEN, P.; SIMON, D.M. & BHATTACHARYA, D. (2005): The natural history of group I introns. - Trends in Genetics 21(2): 111-119.

HOSHAW, R.W. & MCCOURT, R.M. (1988): The Zygnemataceae (Chlorophyta): A twenty-year update of research. - Phycologia 27(4): 511-548.

HOSHAW, R.W.; WANG, J.-C.; MCCOURT, R.M. & HULL, H.M. (1985): Ploidal changes in clonal cultures of *Spirogyra communis* and implications for species definition. - American Journal of Botany 72(7): 1005-1011.

HUELSENBECK, J.P. & RONQUIST, F. (2001): MRBAYES: Bayesian inference of phylogeny. - Bioinformatics 17: 754-755.

KADLUBOWSKA, J.Z. (1984): Conjugatophyceae I - Zygnemales. - In: ETTL, H., GERLOFF, H., HEYNIG, H. and MOLLENHAUER, D. (eds.). Süßwasserflora von Mitteleuropa, Chlorophyta VIII: Gustav Fischer Verlag.Stuttgart, New York.

KOLKWITZ, R. & KRIEGER, H. (1941): Zygnemales. - In: KOLKWITZ, R. (eds.). Dr. L Rabenhorst's Kryptogamen-Flora von Deutschland und der Schweiz XIII, 2 Abteilung: 499 pp., Akademische Verlagsgesellschaft Becker & Erler, Leipzig.

KUSEL-FETZMANN, E. & SCHAGERL, M. (1993): Verzeichniss der Sammlung von Algenkulturen an der Abteilung für Hydrobotanik am Institut für Pflanzenphysiologie der Universität Wien. - Phyton 33: 209-234.

MARIN, B.; KLINGBERG, M. & MELKONIAN, M. (1998): Phylogenetic relationships among the *Cryptophyta*: analyses of nuclear-encoded SSU rRNA sequences support the monophyly of extant plastid-containing lineages. - Protist 149(3): 265-276.

MARIN, B.; PALM, A.; KLINGBERG, M.A.X. & 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(1): 99-145.

MCCOURT, R.M.; HOSHAW, R.W. & WANG, J.-C. (1986): Distribution, morphological diversity and evidence for polyploidy in North American Zygnemataceae (Chlorophyta). - Journal of Phycology 22(3): 307-315.

MCCOURT, R.M.; KAROL, K.G.; BELL, J.; HELM-BYCHOWSKI, K.M.; GRAJEWSKA, A.; WOJCIECHOWSKI, M.F. & HOSHAW, R.W. (2000): Phylogeny of the conjugating green algae (Zygnemophyceae) based on rbcL sequences. - Journal of Phycology 36(4): 747-758.

MICHEL, F. & DUJON, B. (1983): Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. - The EMBO Journal 2(1): 33-38.

MICHEL, F. & WESTHOF, E. (1990): Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. - Journal of molecular biology 216(3): 585-610.

MÜLLER, K.; BORSCH, T.; LEGENDRE, L.; POREMBSKI, S.; THEISEN, I. & BARTHLOTT, W. (2004): Evolution of Carnivory in Lentibulariaceae and the Lamiales. - Plant Biology 6(4): 477-490.

MULLER, K.M.; CANNONE, J.J.; GUTELL, R.R. & SHEATH, R.G. (2001): A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta). - Mol Biol Evol 18(9): 1654-1667.

MURPHY, F.L. & CECH, T.R. (1994): GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. - Journal of molecular biology 236(1): 49-63.

NIELSEN, H. & JOHANSEN, S.D. (2009): Group I introns: Moving in new directions. - RNA Biology 6(4): 375-383.

OLIVEIRA, M.C. & RAGAN, M.A. (1994): Variant forms of a group I intron in nuclear small-subunit rRNA genes of the marine red alga *Porphyra spiralis var. amplifolia*. - Mol Biol Evol 11(2): 195-207.

PHILIPPE, H. (2000): Opinion: Long branch attraction and protist phylogeny. - Protist 151(4): 307-316.

POSADA, D. & BUCKLEY, T. (2004): Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. - Systematic Biology 53: 793-808.

POSADA, D. & CRANDALL, K. (1998): Modeltest: testing the model of DNA substitution. - Bioinformatics 14(9): 817-818.

PRÖSCHOLD, T.; MARIN, B.; SCHLÖSSER, U.G. & MELKONIAN, M. (2001): Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas Ehrenberg* and *Chloromonas Gobi*, and description of *Oogamochlamys gen. nov.* and *Lobochlamys gen. nov*. - Protist 152(4): 265-300.

RANDHAWA, M.S. (1959): Zygnemaceae. - 478 pp., Indian Council of Agricultural Research, New Delhi.

RIETH, A. (1983): Eine *Spirogyra* von der Ostsee bei Zingst. - Genetic Resources and Crop Evolution 31(2): 317-326.

RONQUIST, F. & HUELSENBECK, J.P. (2003): MRBAYES 3: Bayesian phylogenetic inference under mixed models. - Bioinformatics 19: 1572-1574.

RUSSELL, A.; SAMUEL, R.; RUPP, B.; BARFUSS, M.H.J.; ŠAFRAN, M.; BESENDORFER, V. & CHASE, M.W. (2009): Phylogenetics and cytology of a pantropical orchid genus *Polystachya* (Polystachyinae; Vandeae; Orchidaceae); evidence from plastid DNA sequence data. - Taxon: 16 pp.

SALDANHA, R.; MOHR, G.; BELFORT, M. & LAMBOWITZ, A.M. (1993): Group I and group II introns. - FASEB J. 7(1): 15-24.

SIMONS, J. & VAN BEEM, A. (1990): *Spirogyra* species and accompanying algae from pools and ditches in the Netherlands. - Aquatic Botany 37(3): 247-269.

SIMONS, J.; VAN BEEM, P. & DE VRIES, P.J.R. (1984): Induction of conjugation and spore formation in species of *Spirogyra* (Chlorophyceae, Zygnematales). - Acta Bot. Neerl. 33(3): 323-334.

SWOFFORD, D.L. (2003): PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Sunderland, Massachusetts, Sinauer Associates.

TEL-ZUR, N.; ABBO, S.; MYSLABODSKI, D. & MIZRAHI, Y. (1999): Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae). - Plant Molecular Biology Reporter 17(3): 249-254.

TRANSEAU, E.N. (1951): The Zygnemataceae (Fresh-water conjugate algae). - 327 pp., The Ohio State University Press.

WEISER, B. & NOLLER, H. (2009): XRNA, University of Santa Cruz.

ZAUG, A.J.; GRABOWSKI, P.J. & CECH, T.R. (1983): Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. - Nature 301(5901): 578-583.

ZUKER, M. (2003): Mfold web server for nucleic acid folding and hybridization prediction. - Nucl. Acids Res. 31(13): 3406-3415.

Table 1: List of primers used in this study.

Primer name	Sequence	Reference
EAF3	5'-TCGACAATCTGGTTGATCCTGCCAG-3'	PRÖSCHOLD et al. 2001
18sF2	5'-ACCACATCCAAGGAAGGCAGCAG-3'	This study
18sR1	5'-ACGCTATTGGAGCTGGAATTACCGC-3'	This study
18sF3	5'-AGTCCCAACCGTAAACGATGCC-3'	This study
N920R2	5'-CCCTTCCGTCAATTCCTTTAAGTTTC-3'	This study
18sR3	5'-TGTTACGACTTCTCCTTCCTCTAAACG-3'	This study
BR	5'-TTGATCCTTCTGCAGGTTCACCTAC-3'	(MARIN et al. 1998; PRÖSCHOLD et al. 2001)
1380F	5'-GCGTTGAWTACGTCCCTGCC-3'	PRÖSCHOLD et al. 2001
ITS-GF	5'-GGGATCCGTTTCCGTAGGTGAACCTGC-3'	PRÖSCHOLD et al. 2001

Table 2: Comparison of modeltest results among the individual data sets; highest and lowest value in rate matrix marked in light grey.

	incl.	incl.	pars-inf.	model			base frequencies				Substitution model/rate matrix				
	taxa	char.	char.	selected	I	G	A	C	G	U	A-C	A-G	A-U	C-G	C-U
1	11	1717	147	GTR+I+G	0.6752	0.9038	0.2627	0.2009	0.2619	0.2745	1.2538	2.0361	1.5494	0.6242	6.4562
2	9	1717	195	GTR+I+G	0.6392	0.5762	0.2569	0.2193	0.2676	0.2562	6.0200	8.3107	5.5401	3.2841	20.8136
3	13	1717	266	GTR+I+G	0.5234	0.6118	0.2438	0.2369	0.2789	0.2404	1.8712	3.1395	1.6342	1.9982	6.3806
4	11	263	124	TrN+I+G	0.2727	1.7394	0.2286	0.2655	0.2955	0.2103	1.0000	2.6770	1.0000	1.0000	6.5834
5	9	263	91	GTR+G	0.0000	0.2809	0.2392	0.2983	0.3125	0.1500	0.7473	1.4149	1.8670	0.4333	3.2024
6	24	1717	425	GTR+I+G	0.5030	0.7457	0.2499	0.2262	0.2707	0.2531	1.4387	2.3888	1.4904	1.2962	5.4133
7	20	263	168	GTR+G	0.0000	0.6668	0.2163	0.2944	0.3105	0.1788	0.8651	2.2251	1.9833	0.7013	4.6364
8	51	251	108	TrN+I+G	0.1481	0.8150	0.2381	0.2934	0.2627	0.2059	1.0000	1.4160	1.0000	1.0000	1.8766

1 - all taxa of the SSU rRNA alignment aside *Spirogyra*; 2 - *Spirogyra* accessions of the SSU rRNA alignment that contained a group I intron; 3 - all *Spirogyra* accessions of the SSU rRNA alignment; 4 - all taxa of the group I intron alignment aside *Spirogyra*; 5 - all *Spirogyra* accessions of the intron alignment; 6 – complete 18s SSU rRNA alignment; 7 – complete 1506 group I intron alignment; and 8 – *Spirogyra* 1506 group I intron alignment.

Table 3: Relative Rate Test (Müller et al., 2004); using only unambiguously aligned positions; *Gonatozygon aculteaum* as outgroup; genus *Spirogyra* marked in light grey

	taxa	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Genic|X74753	-	**	***	**	***	***	***	***	***	***	***	***	***
2	Meso|X75763	*	-	N.S.	N.S.	***	***	***	***	***	***	**	***	**
3	Zygn|X79495	N.S.	N.S.	-	N.S.	**	**	***	***	***	**	*	**	*
4	Moug|X70705	*	N.S.	*	-	**	**	***	***	***	**	**	**	**
5	WRH5	N.S.	N.S.	N.S.	N.S.	-	*	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
6	WNII3A	***	***	***	***	***	-	N.S.	N.S.	N.S.	*	*	*	N.S.
7	MRT4A	***	***	***	***	***	N.S.	-	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
8	SNG1	***	***	***	***	***	N.S.	*	-	N.S.	***	N.S.	N.S.	N.S.
9	GTH1	***	***	***	***	***	N.S.	N.S.	**	-	N.S.	N.S.	N.S.	N.S.
10	TCA2	***	***	***	***	***	N.S.	N.S.	N.S.	N.S.	-	N.S.	N.S.	N.S.
11	KRA2	***	***	***	***	***	N.S.	**	N.S.	*	*	-	N.S.	N.S.
12	TIS1	***	***	***	***	***	***	***	***	***	***	***	-	N.S.
13	Siro|AJ428076	***	***	***	***	***	***	***	***	***	***	***	N.S.	-

Intron - above diagonal, SSU - below diagonal; not significant: N.S. ($p > 0.05$; relative rates not significantly different at 5% level). Asterisks: $p = 0.05 > * > 0.01 > ** > 0.005 > ***$ (relative rates significantly different).; Genic|X74753 – *Genicularia spirotaenia*; Meso|X75763 – *Mesotaenium caldariorum*; Zygn|X79495 – *Zygnemopsis circumcarinata*; Moug|X70705 – *Mougeotia scalaris;* Siro|AJ428076 – *Sirogonium sticticum*.

Table 4: NHS, H-CBCs and CBC identified in Spirogyra 1506 group I intron; numbers highlighted in grey refer to marked bases in Fig. 4.

Number															
Clade A	1		3	4	5	6	7	8		9			11	12
Clade B														
Clade C					5				8		9	10		
Clade D		2												
Sirogonium														
Pos. in alignment	2	3	178	194	200	210	592	595	595	596	596	729	731	740
Domain	P1	P1	J1/2	P2	P2	P2	P2.1	P2.1	P2.1	P2.1	P2.1	P2.1	P2.1	P3
									CBC			CBC		CBC
NHS	A	G	U	C	A	G/G/A	A	A	C	G/G/A	C	G	A	A
original base	U	A/U	A	A	C	C	U	U	U	U	U	A	G	G

Number														
Clade A	13	14	15		17		18	19		20		21		23
Clade B														
Clade C		14				17		19	19		20			
Clade D				16	17								22	
Sirogonium														
Pos. in alignment	746	750	761	800	807	807	813	824	824	825	825	835	845	951
Domain	P3	J3/4	P4	P5	P5	P5	P5	P5	P5	P5	P5	P5b	L5b	P5c
	H-CBC						H-CBC							
NHS	U	A	G	G	G	U	U	A	U	A	G	A	G	G
original base	C	C	A	C	C	C	C	G	G	C	C	C/U	A/C/U/-	A/U

Number														
Clade A							30	31						
Clade B														
Clade C		25	26	27	28	29			31					
Clade D										31	32			
Sirogonium	24											33	34	35
Pos. in alignment	982	997	999	1001	1009	1011	1024	1031	1031	1031	1034	1075	1076	1078
Domain	L5c	P5c	P5c	P5c	P5c	P5c	P5a	P5a	P5a	P5a	P5a	P5	P5	P5
NHS	U	G/U	A	U	U	U/G	A	C	A	U	C	U	C	C
original base	A/C/G	C	C	G/K	G/-	A	G	G	G	G	A/G/-	C	A	U

Number											
Clade A		37					42	43		45	46
Clade B	36			39	40						
Clade C		37	38			41			44		
Clade D						41					
Sirogonium											
Pos. in alignment	1089	1096	1113	1114	1196	1197	1220	1222	1226	1239	1248
Domain	J4/5	P4	P6	P6	P6	P6	P7	P7	J7/3	P3	P8
			CBC	CBC	CBC	H-CBC/CBC	CBC			CBC	
NHS	G (C)	C	C	U	A (G, U)	G	A	C	C	U	A
original base	A/-/U	A	U	C	G (A)	A	G	A	G	C	G

Number												
Clade A		48	49	50	51	52				56		57
Clade B								54				
Clade C									55		56	57
Clade D	47						53					
Sirogonium												
Pos. in alignment	1331	1350	1351	1354	1357	1361	1362	1486	1487	1489	1489	1490
Domain	P8	J7/8	J7/8	P7	P7	P7	J7/9	P9	P9	P9	P9	P9
				CBC		H-CBC						
NHS	C/U	G/A	G	U	C	U/G	G	A	A/U	C/-	A	A
original base	A/G	C	C	C	U	C	A/C/U	C/G/-	G	U/G	U/G	U

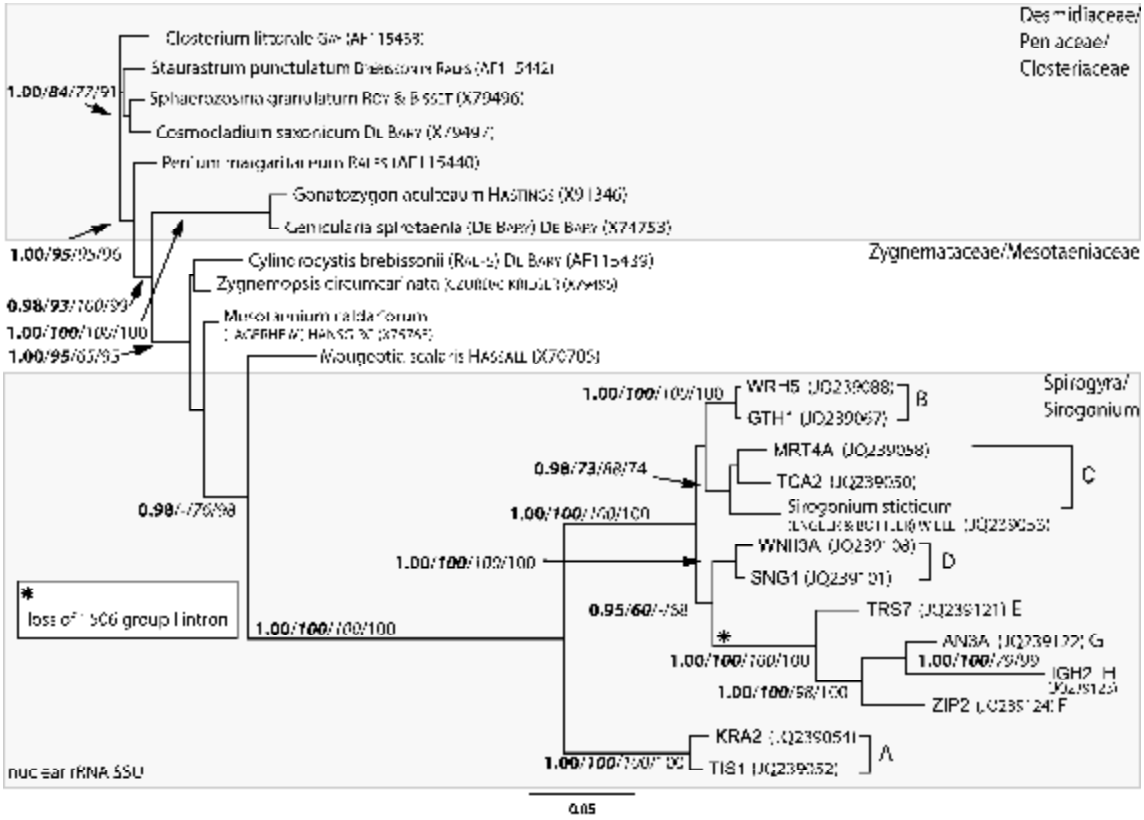


Figure 1: Desm SSU: Molecular phylogeny of Desmidiaceae, Mesotaeniaceae and Zygnemataceae based on a SSU alignment. The phylogenetic tree was inferred by maximum likelihood analyses of 1715 aligned positions of 24 taxa using PAUP* 4.0b10. As an evolutionary model, the GTR+I+G model was chosen (base frequencies: A 0.25, C 0.23, G 0.27, T 0.25; rate matrix: A-C 1. 4387, A-G 2. 3888, A-T 1. 4904, C-G 1. 2962, C-T 5.4133, G-T 1.0000) with the proportion of invariable sites (I=0. 5030) and gamma distribution parameter (G=0.7457). Bootstrap values (>50%) of the maximum likelihood (1000 replicates; bold italic), neighbor-joining (1000 replicates; italic), and maximum likelihood (1000 replicates; not bold) as well as Bayesian values (>95%;bold) were calculated by MrBayes 3.1.2 using the covariation settings (3 million generations, trees from 4100 – 30000) are given in the tree (bayesian/*ML*/*NJ*/MP).

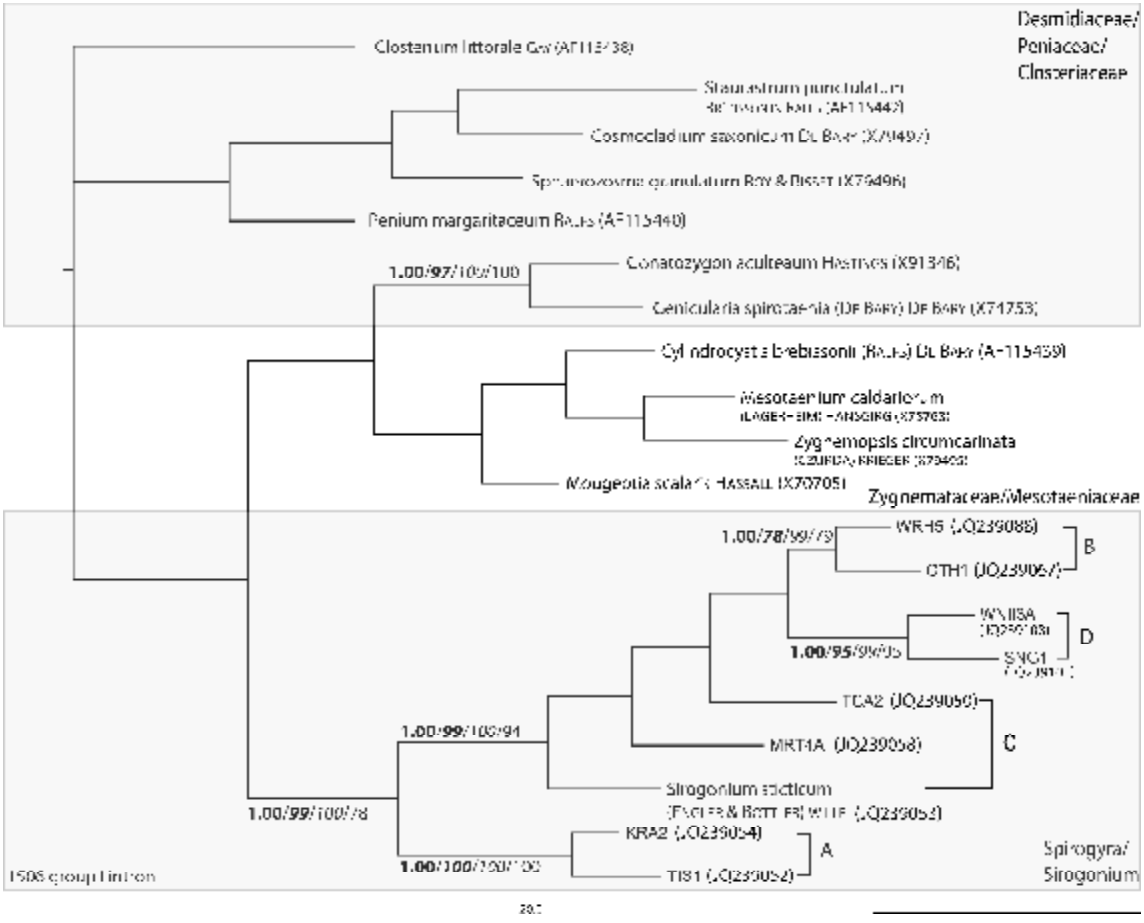


Figure 2A: Desm Intron: Molecular phylogeny of Desmidiaceae, Mesotaeniaceae and Zygnemataceae based on a group I intron alignment. The phylogenetic tree was inferred by maximum likelihood analyses of 263 aligned positions of 20 taxa using PAUP* 4.0b10. As an evolutionary model, the GTR+G model was chosen (base frequencies: A 0.22, C 0.29, G 0.31, T 0.18; rate matrix: A-C 0. 8651, A-G 2.2251, A-T 1. 9833, C-G 0. 7013, C-T 4.6364, G-T 1.0000) with the proportion of invariable sites (I=0) and gamma distribution parameter (G=0. 6668). Bootstrap values (>50%) of the maximum likelihood (1000 replicates; bold italic), neighbor-joining (1000 replicates; italic), and maximum likelihood (1000 replicates; not bold) as well as Bayesian values (>95%;bold) were calculated by MrBayes 3.1.2 using the covariation settings (1 million generations, trees from 52100 – 67300, 77000 – 100000) are given in the tree (bayesian/*ML*/*NJ*/MP).

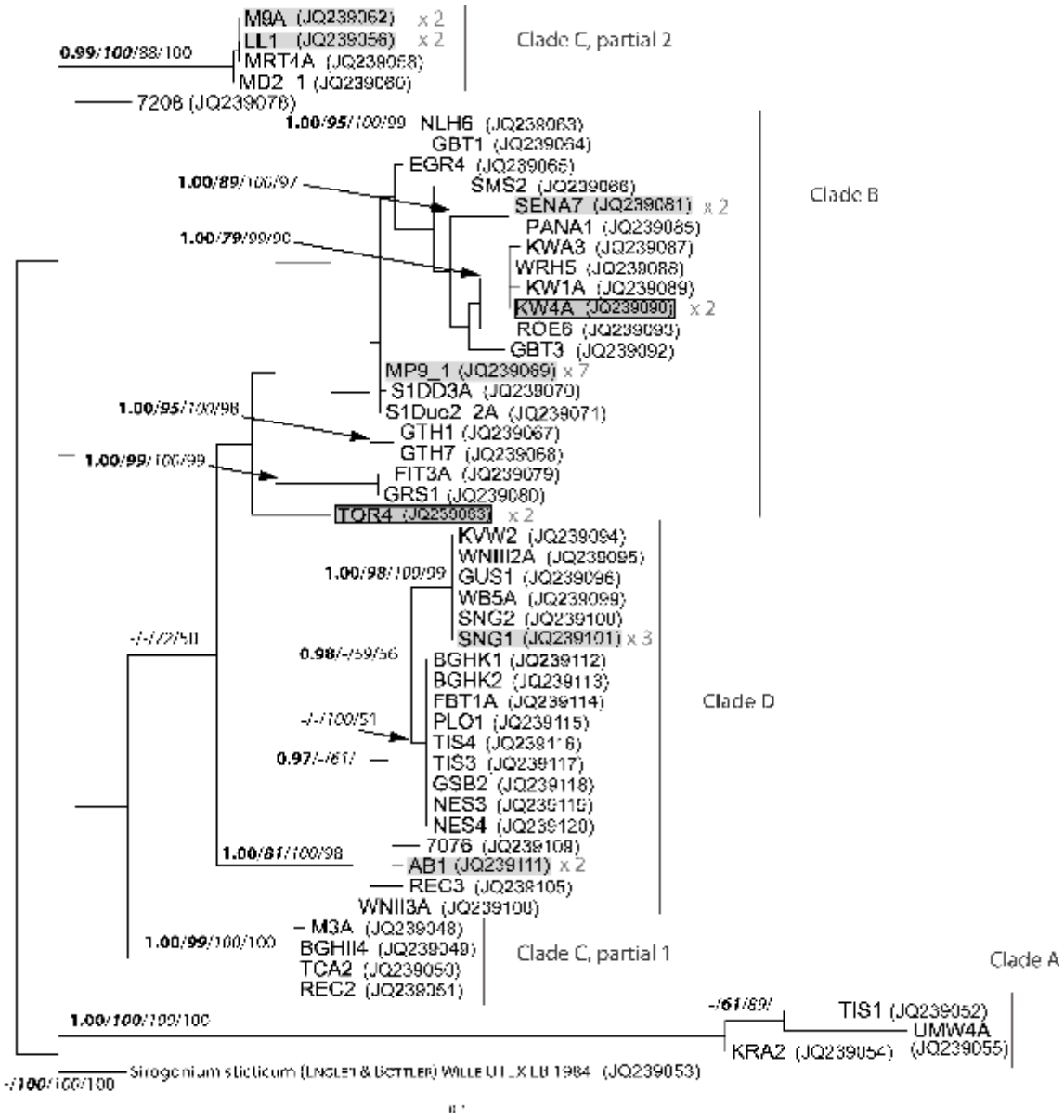


Figure 2B: *Spirogyra* Intron: Molecular phylogeny of a Spiroyra group I intron alignment. The phylogenetic tree was inferred by maximum likelihood analyses of 251 aligned positions of 52 taxa using PAUP* 4.0b10. As an evolutionary model, the TrN+I+G model was chosen (base frequencies: A 0.24, C 0.29, G 0.26, T 0.21; rate matrix: A-C 1.0000, A-G 1.4160, A-T 1.0000, C-G 1.0000, C-T 1.8766, G-T 1.0000) with the proportion of invariable sites (I=0.1481) and gamma distribution parameter (G=0.8150). Bootstrap values (>50%) of the maximum likelihood (1000 replicates; bold italic), neighbor-joining (1000 replicates; italic), and maximum likelihood (1000 replicates; not bold) as well as Bayesian values (>95%; bold) were calculated by MrBayes 3.1.2 using the covariation settings (2.5 million generations, trees from 8320 – 250000) are given in the tree (bayesian/*ML*/*NJ*/MP). Sequences marked with grey boxes were found more than once (number of clones with identical sequence given in grey after the respective sequence), for details see Results.

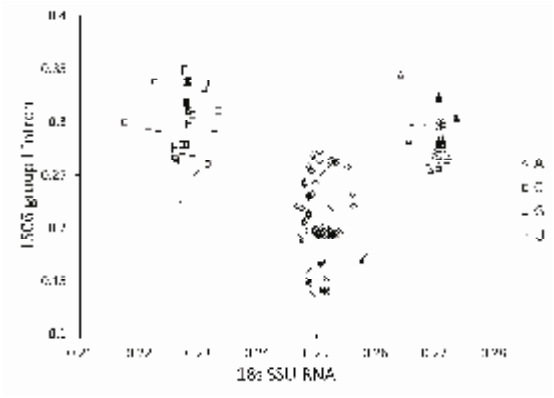


Figure 3: Comparison of base composition in *Spirogyra* SSU rRNAs (x-axis) and 1506 group I introns (y-axis); relative base frequencies are plotted on the axes; single data point may overlap and create pseudo-filled symbols.

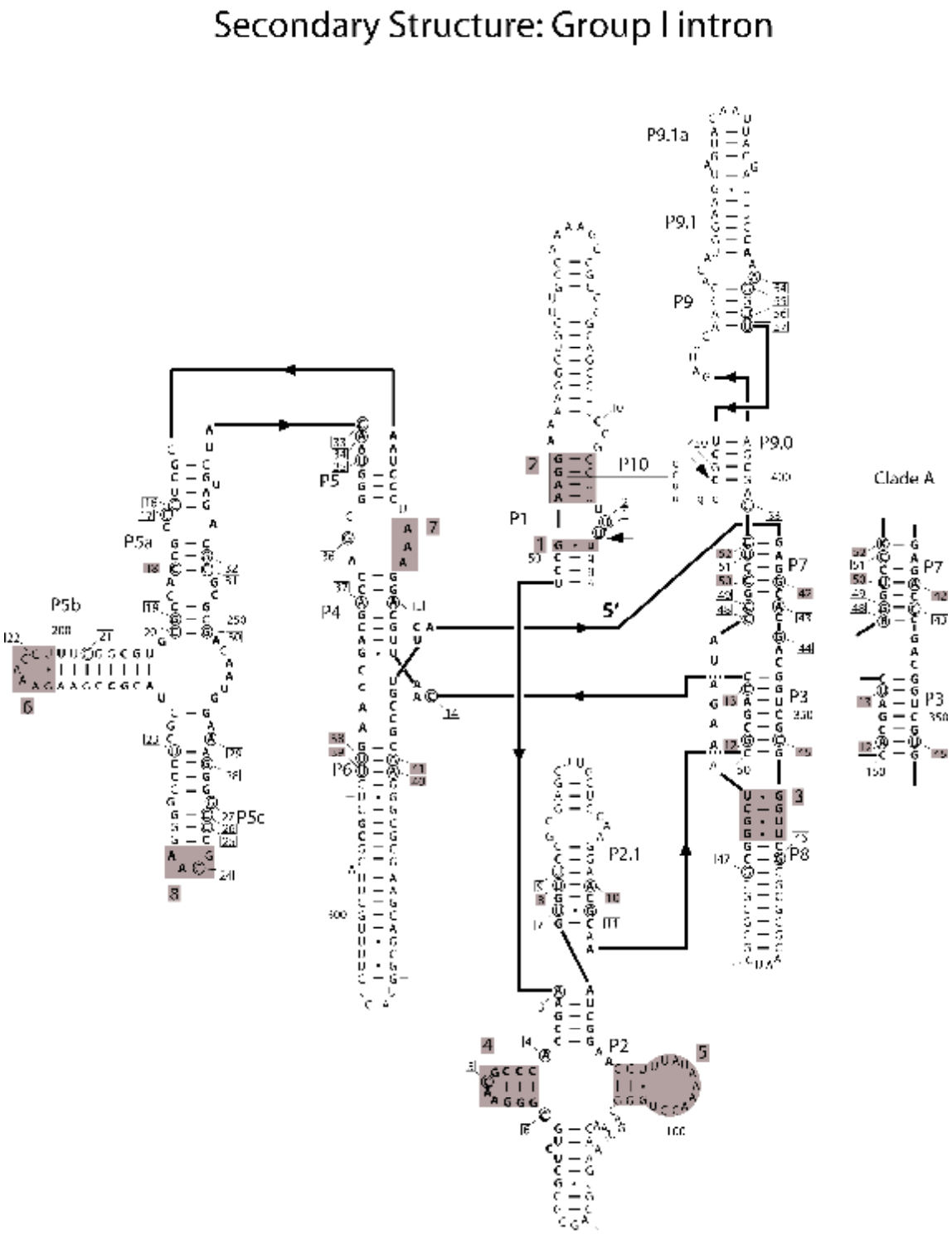


Figure 4: Putative secondary structure of the *Spirogyra* 1506 group I intron, shown by way of example of a representative sequence WRH5 (Acc.No. JQ239088; Clade B). Exon sequence in lower case, intron in upper case letters; 80% consensus sequence in bold face letters. Subdomains are numbered according to Burke, Belfort et al. 1987 and Cech 1988, nucleotide marks are set every 10 nts. NHS are numbered following Tab.1. H-CBCs and CBCs are marked in grey. Parts discussed in the results are numbered (1 – 8, see text for details) and highlighted in grey.

Table 5: Origin of isolates; Abbr. – Abbreviation; Coll. date – Collection date; repr. by – represented in SSU phylogeny by; AUT – Austria; GER – Germany; HR – Croatia; C.C. – culture collection; n.a. – not available.

Abbr.	Origin	Source	Coll. date	GPS coordinates	Accession No.	Clade	Intron	repr. by
7076	n.a.	C.C.	n.a.	n.a.	JQ239109	D	+	7076
7208	n.a.	C.C.	n.a.	n.a.	JQ239078	C	+	7208
AB1	between Ahrensdorf & Behrensdorf	GER	2007-04-17	14° 03' 32"/ 52° 10' 52"	JQ239111	D	+	AB4
AB4	between Ahrensdorf & Behrensdorf	GER	2007-04-17	14° 03' 32"/ 52° 10' 52"	JQ239110	D	+	AB4
AN3A	Alte Naufahrt	AUT	2006-05-10	16° 28' 15"/ 48° 11' 46"	JQ239122	G	-	AN3A
BEA1	Benda pond	AUT	2006-05-15	16° 20' 58"/ 48° 09' 11"	JQ239098	D	+	KVW2
BGHII4	Botanical garden Hamburg near entrance	GER	2007-04-26	09° 51' 37"/ 53° 33' 43"	JQ239049	C	+	REC2
BGHK1	Botanical garden Hamburg limestone alpine area	GER	2007-04-26	09° 51' 36"/ 53° 33' 50"	JQ239112	D	+	BGHK1
BGHK2	Botanical garden Hamburg limestone alpine area	GER	2007-04-26	09° 51' 36"/ 53° 33' 50"	JQ239113	D	+	NES3
BTSA1	Pond Suessenbrunn	AUT	2006-05-25	16° 29' 09"/ 48° 16' 36"	JQ239106	D	+	VNE2
DRS3	Dranser lake	GER	2007-04-19	12° 37' 53"/ 53° 10' 48"	JQ239059	C	+	MRT4A
EGR4	drainage trench Reddelich	GER	2007-04-23	11° 49' 52"/ 54° 04' 38"	JQ239065	B	+	EGR4
FBT1A	Fuchsbodenteich	AUT	2006-06-16	15° 51' 45"/ 48° 21' 54"	JQ239114	D	+	TIS3
FIT3A	Filmteich	AUT	2006-05-29	16° 24' 17"/ 48° 09' 01"	JQ239079	B	+	FIT3A
GBT1	gr Brunnerteich	AUT	2006-10-23	16° 18' 30"/ 48° 07' 00"	JQ239064	B	+	GBT1
GBT3	gr Brunnerteich	AUT	2006-10-23	16° 18' 30"/ 48° 07' 00"	JQ239092	B	+	GBT3
GRS1	Hamburg Großensee (pier)	GER	2007-04-25	10° 20' 51"/ 53° 36' 46"	JQ239080	B	-	GRS1
GSB2	gr. Segeberger Lake	GER	2007-04-25	10° 20' 23"/ 53° 57' 27"	JQ239118	D	+	GSB2
GTH1	garden pond Hamburg	GER	2007-04-25	10° 08' 44"/ 53° 37' 21"	JQ239067	B	+	GTH1
GTH7	garden pond Hamburg	GER	2007-04-25	10° 08' 44"/ 53° 37' 21"	JQ239068	B	+	GTH1
GUS1	Grubensee	GER	2007-04-17	13° 59' 40"/ 52° 09' 30"	JQ239096	D	+	KVW2
IGH2	Industrial area Höltingbaum	GER	2007-04-25	10° 11' 01"/ 53° 36' 44"	JQ239123	H	-	IGH2
KRA2	watercourse between Krauswitz & Schlagnitz	GER	2007-04-17	13° 52' 50"/ 52° 01' 48"	JQ239054	A	+	KRA2
KVW2	Pond near traffic circle A19 Wittstock	GER	2007-04-19	12° 27' 20"/ 53° 09' 26"	JQ239094	D	+	KVW2
KVW4	Pond near traffic circle A19 Wittstock	GER	2007-04-19	12° 27' 20"/ 53° 09' 26"	JQ239097	D	+	KVW2
KW1A	Kaiserwasser	AUT	2006-04-25	16° 25' 29"/ 48° 13' 54"	JQ239089	B	+	KWA3
KW4A	Kaiserwasser	AUT	2006-04-25	16° 25' 29"/ 48° 13' 54"	JQ239090	B	+	KWA3
KWA3	Kaiserwasser	AUT	2006-04-25	16° 25' 29"/ 48° 13' 54"	JQ239087	B	+	KWA3
LL1	local fire service pond Limsdorf	GER	2007-04-17	14° 00' 51"/ 52° 09' 34"	JQ239056	C	+	MRT4A
LL2	local fire service pond Limsdorf	GER	2007-04-17	14° 00' 51"/ 52° 09' 34"	JQ239057	C	+	MRT4A
LSB4A	Langenschoenbichl	AUT	2006-06-16	15° 59' 45"/ 48° 19' 52"	JQ239091	B	+	KWA3
M3A	St. Pölten	AUT	2006-03-20	n.a.	JQ239048	C	+	M3A
M5A	St. Pölten	AUT	2006-03-20	n.a.	JQ239075	B	+	M6A
M6A	St. Pölten	AUT	2006-03-20	n.a.	JQ239076	B	+	M6A

Table 5 continued

Abbr.	Origin	Source	Coll. date	GPS coordinates	Accession No.	Clade	Intron	repr. by
M9A	St. Pölten	AUT	2006-03-20	n.a.	JQ239062	C	+	MRT4A
M9B	St. Pölten	AUT	2006-03-20	n.a.	JQ239074	B	+	M6A
MD2-1	St. Pölten	AUT	2006-03-20	n.a.	JQ239060	C	+	MRT4A
MDA1	St Pölten	AUT	2006-03-20	n.a.	JQ239061	C	+	MRT4A
MP9-1	St Pölten	AUT	2006-04-11	n.a.	JQ239069	B	+	M6A
MP923	St. Pölten	AUT	2006-04-11	n.a.	JQ239073	B	+	M6A
MP92-6	St. Pölten	AUT	2006-04-11	n.a.	JQ239072	B	+	M6A
MRT4A	Mausrodlteich	AUT	2006-08-24	15° 02' 52"/ 47° 52' 25"	JQ239058	C	+	MRT4A
NES3	Nebelsee	GER	2007-04-19	12° 39' 58"/ 53° 14' 54"	JQ239119	D	+	NES3
NES4	Nebelsee	GER	2007-04-19	12° 39' 58"/ 53° 14' 54"	JQ239120	D	+	NES3
NLH5	watercourse between Neu-Lübbenau & Hohenbrück	GER	2007-04-19	13° 53' 38"/ 52° 04' 57"	JQ239084	B	+	NLH5
NLH6	watercourse between Neu-Lübbenau & Hohenbrück	GER	2007-04-17	13° 53' 38"/ 52° 04' 57"	JQ239063	B	+	NLH6
OSOR-IV-A1	Osor	HR	2006-05-24	14° 23' 87"/ 44° 41,732'	JQ239104	D	+	SNG1
PANA1	Panozzalacke	AUT	2006-05-10	16° 29' 16"/ 48° 10' 51"	JQ239085	B	+	PANA1
PLO1	Plötinsee Rostock	GER	2007-04-20	12° 58' 22"/ 53° 13' 14"	JQ239115	D	+	TIS3
REC2	Meadow pond Rechlin	GER	2007-04-20	12° 45' 45"/ 53° 18' 43"	JQ239051	C	+	REC2
REC3	Meadow pond Rechlin	GER	2007-04-20	12° 45' 45"/ 53° 18' 43"	JQ239105	D	+	VNE2
ROE6	Pond near Rödel	GER	2007-04-19	12° 35' 51"/ 53° 23' 06"	JQ239093	B	+	ROE6
S1DD3A	St. Pölten	AUT	2006-03-20	n.a.	JQ239070	B	+	M6A
S1DUe2-2A	St. Pölten	AUT	2006-03-20	n.a.	JQ239071	B	+	M6A
S1Due2A	St Pölten	AUT	2006-03-20	n.a.	JQ239077	B	+	M6A
SENA4	Sengsee	GER	2006-07-04	11° 18' 58"/ 47° 46' 37"	JQ239082	B	+	SENA7
SENA7	Sengsee	GER	2006-07-04	11° 18' 58"/ 47° 46' 37"	JQ239081	B	+	SENA7
SMS2	Scharmützelsee	GER	2007-04-17	14° 00' 50"/ 52° 12' 49"	JQ239066	B	+	SMS2
SNG1	Marsh north Großensee	GER	2007-04-25	10° 21' 27"/ 53° 37' 51"	JQ239101	D	+	SNG1
SNG2	Marsh north Großensee	GER	2007-04-25	10° 21' 27"/ 53° 37' 51"	JQ239100	D	+	SNG1
SNG3	Marsh north Großensee	GER	2007-04-25	10° 21' 27"/ 53° 37' 51"	JQ239103	D	+	SNG1
SNG5	Marsh north Großensee	GER	2007-04-25	10° 21' 27"/ 53° 37' 51"	JQ239102	D	+	SNG1
TCA2	Canow fish farming	GER	2007-04-20	12° 53' 23"/ 53° 11' 32"	JQ239050	C	+	REC2
TIS1	Tiefensee	GER	2007-04-17	13° 59' 22"/ 52° 09' 23"	JQ239052	A	+	TIS1
TIS3	Tiefensee	GER	2007-04-17	13° 59' 22"/ 52° 09' 23"	JQ239117	D	+	TIS3
TIS4	Tiefensee	GER	2007-04-17	13° 59' 22"/ 52° 09' 23"	JQ239116	D	+	TIS3
TOR4	Torsee	GER	2007-04-18	13° 39' 56"/ 52° 14' 24"	JQ239083	B	+	TOR4
TRS7	Tristacher Lake	AUT	2006-08-10	12° 47' 40"/ 46° 48' 29"	JQ239121	E	-	TRS7
UMW4A	Lower Muehlwasser	AUT	2006-05-09	16° 27' 24"/ 48° 12' 47"	JQ239055	A	+	UMW4A

Table 5 continued

Abbr.	Origin	Source	Coll. date	GPS coordinates	Accession No.	Clade	Intron	repr. by
UMW4A	Lower Muehlwasser	AUT	2006-05-09	16° 27' 24"/ 48° 12' 47"	JQ239055	A	+	UMW4A
UTEX LB 1984	n.a.	C.C.	n.a.	n.a.	JQ239053	C	+	UTEX LB 1984
VNE2	accretion zone Neversdorfer lake	GER	2007-04-25	10° 15' 31"/ 53° 51' 56"	JQ239107	D	+	VNE2
WB5A	Wienerberg pond2	AUT	2006-05-15	16° 20' 49"/ 48° 09' 43"	JQ239099	D	+	KVW2
WNII3A	local fire service pond Gerasdorf	AUT	2006-05-25	16° 27' 58"/ 48° 18' 00"	JQ239108	D	+	WNII3A
WNII4A	local fire service pond Gerasdorf	AUT	2006-04-25	16° 27' 58"/ 48° 18' 00"	JQ239086	B	+	PANA1
WNIII2A	Pond recreation area Seeschlacht (Vienna North III)	AUT	2006-05-25	16° 21' 49"/ 48° 17' 56"	JQ239095	D	+	KVW2
WRH5	Wandse detention reservoir	GER	2007-04-25	10° 11' 07"/ 53° 37' 17"	JQ239088	B	+	KWA3
ZIP2	Zipke	GER	2007-04-22	12° 47'02"/ 54° 20' 30"	JQ239124	F	-	ZIP2

Chapter 2

Hidden genetic diversity in the green alga Spirogyra (Zygnematophyceae, Streptophyta)

Lunz am See, AUT, 2007

Hidden genetic diversity in the green alga Spirogyra (Zygnematophyceae, Streptophyta)

Charlotte Chen[1]; Michael H. J. Barfuss[2]; Thomas Pröschold[3]; Michael Schagerl[1§]

[1]University of Vienna, Department of Limnology, Althanstraße 14, A-1090 Vienna, Austria

[2]University of Vienna, Department of Systematic and Evolutionary Botany, Rennweg 14, A-1030 Vienna, Austria

[3] University of Rostock, Department Applied Limnology, Albert-Einstein-Str. 3, D-18059, Rostock, Germany

[§]Corresponding author

Published: BMC Evolutionary Biology 2012, 12:77; doi:10.1186/1471-2148-12-77

Email addresses:

CC: charlotte.woeber@univie.ac.at

MHJB: michael.h.j.barfuss@univie.ac.at

TP: thomas.proeschold@uni-rostock.de

MS: michael.schagerl@univie.ac.at

Abstract

The unbranched filamentous green alga *Spirogyra* (Streptophyta, Zygnemataceae) is easily recognizable based on its vegetative morphology, which shows one to several spiral chloroplasts. This simple structure falsely points to a low genetic diversity: *Spirogyra* is commonly excluded from phylogenetic analyses because the genus is known as a long-branch taxon caused by a high evolutionary rate.

Results

We focused on this genetic diversity and sequenced 130 *Spirogyra* SSU rDNA strands of different origin. The resulting SSU rDNA sequences were used for phylogenetic analyses using complex evolutionary models (posterior probability, maximum likelihood, neighbor joining, and maximum parsimony methods). The sequences were between 1672 and 1779 nucleotides long. Sequence comparisons revealed 53 individual clones, but our results still support monophyly of the genus. Our data set did not contain a single slow-evolving taxon that would have been placed on a shorter branch compared to the remaining sequences. Out of 130 accessions analyzed, 72 showed a

secondary loss of the 1506 group I intron, which formed a long-branched group within the genus. The phylogenetic relationship to the genus *Spirotaenia* was not resolved satisfactorily. The genetic distance within the genus *Spirogyra* exceeded the distances measured within any other genus of the remaining Zygnemataceae included in this study.

Conclusion

Overall, we define eight distinct clades of *Spirogyra*, one of them including the genus *Sirogonium*. A large number of non-homoplasious synapomorphies (114 NHS in total) was found for *Spirogyra* (41 NHS) and for each clade (totaling 73 NHS). This emphasizes the high genetic diversity of this genus and the distance to the remaining Zygnematophyceae.

Keywords

Zygnematales, Zygnematophyceae, non-homoplasious synapomorphy, *Spirogyra*, *Sirogonium*, *Spirotaenia*, SSU rDNA, diversity

Abbreviations

bp, base pair; AU, approximately unbiased test; KH, Kishino-Hasegawa test; SH, Shimodaira-Hasegawa test; WSH, weighted SH; rbcL, ribulose-bisphosphate carboxylase large subunit gene; nt, nucleotide; rDNA, nuclear ribosomal DNA; SSU, small subunit; MP, maximum parsimony; ML, maximum likelihood; NJ, neighbor joining; PP, posterior probability; NHS, Non-Homoplasious Synapomorphy; (H-) CBC, (Hemi-) Compensatory Base Change; UD tree, user defined tree;

Background

The genus *Spirogyra* is a member of the Zygnemataceae (Zygnematophyceae, Streptophyta). It comprises unbranched, filamentous green algae that are characterized by spirally coiled chloroplasts and sexual reproduction by means of conjugation. *Spirogyra* is commonly found in stagnant or slowly flowing freshwater habitats all over the world [1, 2]. It is sometimes referred to as an alga of roadside ditches and is frequently used in introductory biology courses [3] because it often occurs in huge abundances and is easy to address at the generic level. Species definition is mainly based on hypnozygote (also known as zygospores) morphology because the simple morphology in its vegetative state does not permit species recognition. In the latest monograph of *Spirogyra* published by Kadlubowska [4], 386 species are included. They were described using morphological traits, many of them based on a single finding. Ashraf and Godward [5] suggested that the mesospore wall structure analyzed using scanning electron microscopy should be added to the species descriptions because the taxonomy of *Spirogyra* at the light microscopical level remains confusing due to overlapping morphological traits [6]. The morphological species concept, which is also applied in *Spirogyra*, is not proven to represent true biological species, nor does it provide any information on the ecological or genetic diversity in a genus. It also does not elucidate the phylogenetic relationships between taxa [7, 8]. Accordingly, the diversity of a genus remains unclear when estimates are based a single species concept. The problems arising for *Spirogyra* from findings without ripe hypnozygotes and the low success rate in inducing conjugation [9-13] call for other ways of addressing the issue of species delimitation and identification.

The Zygnematophyceae (Viridiplantae) represent the most species-rich lineage in the Streptophyta except for the embryophytic land plants [14]. Conjugating green algae including the orders Desmidiales (Desmidiaceae, Peniaceae, Closteriaceae) and Zygnematales (Mesotaeniaceae, Zygnemataceae) form a unique and distinct group. Its taxonomic and phylogenetic separation from other algae is definitive [15-18], but relationships within this group have undergone numerous

rearrangements and still remain unclear. The classification schemes within the Zygnematophyceae have been based on morphological traits such as cell size, wall structure, cellular organization or chloroplast structure [1, 14, 19-21], approaches that have been criticized in the past [22, 23]. The Zygnematales are distinguished from the Desmidiales by a smooth cell wall consisting of a single piece, lacking pores and ornamentations [19, 21], whereas Desmidiales have cell walls consisting of more than one piece with pores and ornaments [24, 25]. Filamentous forms are grouped in the family Zygnemataceae; the unicellular taxa form the family Mesotaeniaceae [4, 14, 21]. This classification, however, is artificial because polyphyly of both families has been proven by phylogenetic analyses [20, 22].

It remains unclear which growth form is primary and which derived [15]. West [26, 27] described the ancestral state as filamentous, evolving towards unicellular forms, but Yamagishi [28] stated the opposite. Since the introduction of molecular markers, efforts have been made to solve this question [20, 23, 29], but the position of the genus *Spirogyra* within the Zygnematophyceae is not fully resolved. The evolutionary rate is one possible reason for this problem: it differs considerably among Zygnemataceaen genera [30]. The uncertain position of *Spirogyra* in phylogenetic analyses is also attributed to the long-branch attraction (LBA) phenomenon [31].
Some genera of the Zygnematophyceae originally defined based on morphology have been revealed as artificial based on molecular markers (e.g., *Staurastrum;* [32]). Furthermore, SSU rDNA phylogenetic studies have often suffered from limited taxon sampling [20]. When genera are represented by just one taxon, authors are unable to address the monophyly of the phylogenetic groups, either at the generic or at higher taxonomic levels. A low number of species within a genus also hinders proving monophyly [33]. In order to assess the monophyly of the genus *Spirogyra* and to investigate whether the low diversity of its vegetative morphology is also reflected in molecular data, we sampled 130 strains of different origin. The position within the Zygnematophyceae and its long-branch position were evaluated by calculating phylogenetic trees with complex evolutionary models. Minimizing the *Spirogyra* LBA problem will also help define phylogenetic relationships among genera. Additionally, we searched for *Spirogyra* taxa with slower evolutionary rates by including isolates from a broad spatial and ecological range and different vegetative morphology. Sampling locations were chosen to cover different types of water bodies in various areas [2]. Morphologically different *Spirogyra* filaments were isolated and cultivated for later use to check if the morphological differences are also reflected in phylogenetic groups. We focused on sampling the morphological diversity of the genus. SSU rRNA was chosen over *rbcL* because other studies already demonstrated the poor resolution of this marker in the Zygnematales and Desmidiales [3, 15].

Results

Molecular phylogenetic analyses

The 130 *Spirogyra* sequences formed a monophyletic clade. This group, including *Sirogonium*, was subdivided into eight individual clades A to H. Molecular phylogenies were inferred from two data sets, one combining Zygnematophyceae and *Spirogyra* alignment and one alignment comprising only *Spirogyra* and *Sirogonium* sequences (Figures 1, 2). In the phylogeny inferred from the combined Zygnematophyceae alignment (Figure 1), very high bootstrap support was given for branches within the *Spirogyra* clade. Only few branches were without support, one indicating a possible polytomy for clade C and another one indicating the lack of support for phylogenetic resolution between clades B and C, and D to H. Only very closely related taxa received high support by bootstrap values, e.g., the *Zygnema* clade, *Sphaerozosma/Cosmocladium*, *Gonatozygon/Genicularia* and

Closterium/Cosmarium. The remaining branches lacked support of at least one method. The individual clades of *Spirotaenia* and *Spirogyra* showed very high bootstrap support from all algorithms. The Desmidiales clade was moderately supported (PP/*ML*/*NJ*/MP: -/*50*/*100*/58); the Zygnematales clade received no support at all. *Sirogonium sticticum* was placed within the *Spirogyra* clade C. None of the algal families analyzed here formed exclusive monophyletic clades.
When testing tree topologies for the combined Zygnematophyceae – *Spirogyra* alignment, the "best tree" derived from ML analysis (same as phylogeny in Figure 1) was not the overall best tree (Table 1). The original ML tree (tree 1) and the UD tree with *Spirogyra* relocated outside in ancestral position to the clades formed by the Desmidiales and the Zygnematales (tree 2) had the same likelihood and same Bayesian posterior probability values. These two trees were the only ones not rejected by the AU; the tree representing *Spirogyra* as a sister to a Zygnematales clade (including the *Spirotaenia* clade; tree 3) was rejected by all tests except SH and WSH; *Spirogyra* within the Zygnematales clade (tree 4) was also rejected by all tests except SH. All other trees were significantly worse than the best tree at $p \leq 0.05$.

In the *Spirogyra* phylogenetic tree (Figure 2), three sequences could not be placed within any clade: UTEX 1742, UTEX 1745 and 7075 share only a small portion of the identification patterns in base composition with adjacent clades. The major clades received very high bootstrap support, except for Clade E that was not supported by ML. The support for branches within clade B was poor due to high sequence similarity. Taxa with a secondary loss of the group I IC1 intron (marked in Figure 1 & 2) were clearly separated from the taxa containing the intron (clades A to D with intron; UTEX 1742, 7075, UTEX 1745, clades E to H without intron). The placement of *Sirogonium* in any other clade yielded significantly worse trees in all cases (tested with consel, Table 2). Also, the relocation of the sequences previously not included in the clades into an adjoining clade was rejected with only one exception: tree 2 (relocation of UTEX 1742 and 7075 into clade E) was not rejected by SH test.

Sequence similarities in *Spirogyra*
130 *Spirogyra* nuclear encoded SSU rRNA sequences of strains from 79 different sites were sequenced for this study and, in total, 53 different SSU rDNA types (clones) were identified. Thirty sequence types were found once; the remaining 23 were represented by up to 11 accessions. Thirty-eight clones were found only at a single site and 19 more were obtained from up to 6 different sampling sites; 47 sites were represented by just one accession, 32 with up to 10 accessions. Sixty-seven sampling sites exhibited just one clone and 12 had up to 3 different clones.

To describe the genetic variability among the discovered eight lineages (clade A-H; see above), the minimal distance and pair-wise differences were calculated in PAUP. The minimal distance (Table 3, right top) between two clades was 5.09% (clade B and D), the maximum 16.74% (clade A and H), which is 1.57 times higher than the highest value found in the remaining algal groups included in this study. This means that the within-genus difference in *Spirogyra* exceeds the differences among the remaining genera. The highest within-clade distance was observed in clade H (5.17%), followed by clade C (4.31%), whereas the other clades had comparatively low values from 0.25% to 1.23%. Clade A had the lowest distance value to the Zygnemataceae and Mesotaeniaceae used in our analyses (Zygn.: 22.47%; Desm.: 21.44%), while clade H showed the highest distances (Zygn.: 25.63%; Desm.: 25.71%). The pair-wise differences followed the same pattern (Table 3, left bottom) – the biggest difference within *Spirogyra* was recognized between clade A and H (272 nucleotides (nt) difference); the biggest difference exhibited was clade H to the remaining Zygnemataceae (417 nt difference).

Evolutionary rates in Zygnematophyceae

To test the evolutionary rates among the Zygnematophyceae, the evolutionary models of different data sets were tested by Modeltest. As shown in Table 4a, the data sets revealed major differences in base composition. Compared with other Zygnematalean taxa, the G/C content of *Spirogyra* is elevated. Additionally, high variability of sequences is indicated by a lower portion of constant nt and a different pattern in base substitution rates (Table 4a). 'Zygnemataceae + selected *Spirogyra'* and '*Spirogyra'* represent the same data sets used for phylogenetic analyses; other data sets were obtained by modifying the previous by exclusion of certain taxa (Table 4a). The biggest difference in G/C content (0.0586 units) occurred between the Zygnematophyceae data set excluding *Spirogyra* and the *Spirogyra* data set, pointing out the disparity between the two groups (for respective values see Table 4a). Interestingly, the lowest C to T substitution rate was found in the data set used for phylogenetic analyses comprising Zygnematophyceae and selected *Spirogyra* (4.5); the highest value was calculated for the Zygnematophyceae without *Spirogyra* and *Spirotaenia* (7.2). In the *Spirogyra* data set, this value is less elevated (5.3) compared to the other substitution rates.

Evolutionary rates were inferred by pair-wise comparison of unambiguously aligned positions of an rRNA SSU alignment of all sequences used in this study in GRate (Table 4b & c). The comparison was calculated among the genera (Table 4b) and among the individual sequences of the alignment (Table 4c). The genus *Spirogyra* showed significant differences to all other Zygnematophyceaen genera (Table 4b); the separate clades also revealed highly different values compared to the other Zygnematophyceae (data not shown). For the remaining genera, the picture was less clear – the evolutionary rates of some taxa such as Mesotaeniaceae, Desmidiaceae and Peniaceae did not differ from each other, but did differ from Closteriaceae. Analyses of the individual sequences (Table 4c) revealed that the evolutionary rates of all *Spirogyra* sequences differed significantly from all other Zygnematophyceaen sequences. The one exception was KRA2 from Clade A: it differed significantly from all *Spirogyra* sequences, but only from two of the Zygnematophyceaen sequences; all other differences could not be distinguished statistically. Within genera, insignificant values prevailed. The same holds true for *Spirogyra* clades – with one exception: *Spirogyra* clade G showed significant differences among its sequences. Most of the disparities among non-*Spirogyra* sequences were not statistically relevant; differences among the *Spirogyra* sequences of clades B to D were mostly not significant, whereas the differences to the clades E to H (representing accessions not containing the 1506 group I intron) were mostly significant.

Secondary structure and NHS

To discover the variable positions (compensatory base changes - CBCs/HCBCs, and non-homoplasious synapomorphies - NHSs) in the SSU rDNA of *Spirogyra*, the secondary structures of all strains were compared. No major changes were found in the overall secondary structure of the SSU rDNA (Figure 3). Variable parts (denoted by lower case letters in Figure 3) are predominantly peripheral regions such as E10, E10_1, 17, E23_1, E23_2, E23_4, E23_7, 43, 45, 46, and 49. In an alignment of all Zygnematophyceae and *Spirogyra* sequences used, 114 NHS were identified for the genus *Spirogyra* and/or the individual clades (Table 5); for the genus *Spirogyra*, 41 NHS were found (blue filled circles in Figure 3). Eight of the NHS were involved in CBCs: the first base pair (bp) in helix 29 (C-G, Nos. 62 and 71), the first C-G pair in helix 44 (Nos. 85 and 86), the fourth bp in helix 47 (C-G, Nos. 93 and 96) and the penultimate bp in helix 48 (C-G, Nos. 97 and 98). NHS for the individual clades were located in the variable parts of the SSU secondary structure, especially in E10_1, E23_1, E23_2, E23_4, E23_7 and 44 (Figure 3, green filled circles). Twenty-nine NHS were found for clade A, 5 for B, 3 for C (including *Sirogonium sticticum*), 8 for D, 3 for E, 9 for F, 4 for G and 12 for H. NHS were also found

for groups of clades (data not shown): 23 NHS for clades B to H, 2 NHS for D to H, 16 NHS for the group of taxa with the secondary loss of the group I IC1 intron (clade E to H), 7 for clades F to H and 9 for clades G and H.

Discussion

We compared 130 SSU rDNA sequences of *Spirogyra* and found a high genetic diversity that was unexpected from the phenotypes. Our phylogenetic analyses revealed that *Spirogyra* splits into eight independent lineages within Zygnematophyceae (clades A-H; Figs. 1-2). In contrast to low phenotypic and high genetic variability in *Spirogyra*, the genus *Staurastrum* – one the most species-rich genera within the Desmidiaceae comprising around 700 species – showed great variability of morphological characters such as cell shape, size or cell wall ornamentation. In contrast, it yielded no information on phylogenetic relationships or genetic distance, which was interpreted to reflect recent radiation [32]. Nonetheless, even a small number of species and their relatively uniform appearance may not guarantee generic monophyly (e.g., *Spondylosium*, *Desmidium*, *Hyalotheca*, *Netrium*, *Cylindrocystis*, *Mesotaenium*; [33]). Morphology apparently does not reflect genetic diversity in this group. Morphology is even less trustworthy in *Spirogyra*; the genus seems to be very uniform, but the species exhibit a wide overlap of character ranges [2, 6, 34, 35]. Furthermore, changes in ploidy level may occur, also affecting morphology [34, 36, 37].

The overall phylogenetic relationship of the Zygnematophyceae included in our analysis confirms other studies separating Desmidiaceae, Peniaceae and Closteriaceae from Zygnemataceae and Mesotaeniaceae, although the latter two are not resolved from each other [29, 30, 38]. Gontcharov et al. [20] already stated that Zygnematales appear to be a polyphyletic assemblage of independent clades. The families Mesotaeniaceae and Zygnemataceae are not monophyletic, proving that the cell wall traits (unornamented, unsegmented) are plesiomorphic [15].

When linking the results of our analysis to rbcL data presented by McCourt [15], the Desmidiales phylogeny is largely congruent, while the Mesotaeniaceae and Zygnemataceae (*sensu* Bold & Wynne [18]) show major differences. In McCourt's analyses, the branch comprising *Mougeotia/Mesotaenium* and *Cylindrocystis/Zygnemopsis* is placed as an ancestor to other Zygnematophyceae; the branch *Zygnema/Zygogonium* and *Spirogyra*/*Sirogonium*/*Spirotaenia* emerges at a position basal to the Desmidiales. Contrarily, in our phylogenetic tree, the Zygnematales clade is a sister to the Desmidiales clade, although little bootstrap support is given and the genera *Spirogyra*/*Sirogonium* and *Spirotaenia* form individual clades within the Zygnematales. Those clades form distinct branches basal to the remaining Zygnematales branch due to different evolutionary rates of the SSU rDNA.

One reason to choose SSU rDNA over rbcL for phylogenetic analyses is that, as in the rbcL analyses of McCourt [15], phylogenetic relationships among the Desmidiales show rather poor resolution and do not always receive bootstrap support. Furthermore, in the phylogeny in Drummond's analyses [3], members of the Zygnemataceae exhibit very long branches, whereas members of the genera *Spirogyra* and *Sirogonium* have very short branches indicating only little difference in sequence. This could lead to a misplacement of the genus *Spirogyra*, underestimate diversity in this genus and still yield the LBA problem [31].

The order Zygnematales pools taxa with differing evolutionary rates. On one hand, it is important to include these taxa into phylogenetic analyses to obtain an overview of major relationships. On the other hand, taxa with accelerated evolutionary rates often disturb the analyses. This might be the

major reason for low bootstrap support in some of the clades we analyzed. In general, the genetic diversity in the Zygnematales at the generic level has been underestimated in favor of morphological traits that proved to be uninformative at the phylogenetic level [22]. Some genera defined solely by morphological characters are probably artificial and polyphyletic [20]. Thus, the species concept in this group needs urgent revision, and the generic concept requires scrutiny. *Spirogyra* is positioned ancestral to the remaining Zygnematales (except for *Spirotaenia*) in our phylogenetic tree. By testing the UDT against the best tree derived from ML analysis, we conclude that *Spirogyra* has to be placed outside of the Zygnematales clade; this position is definitely not caused by LBA, because no other position in the phylogenetic tree was accepted by UDT testing. Such a position is also supported by analyses of combined Zygnemataceae and *Spirogyra* 1506 group I intron alignments [39].
Similar to the rbcL results of McCourt [15], our data do not support the hypothesis of monophyly for groups exhibiting a similar cell shape. In accordance to McCourt's findings, but in contrast to previous SSU rDNA analyses [29], chloroplast shape seems to be a diagnostic trait: stellate and laminate chloroplast containing taxa form two sister clades [15, 40], yet without clear indication of the ancestral chloroplast type [41].

McCourt [15] stated that derived cell and chloroplast forms of placoderm desmids are better photosynthesizers and have achieved greater evolutionary success. This is difficult to reconcile with the ecological success of some of the so-called "primitive forms" such as *Spirogyra*, which is among the most widespread and species-rich conjugating green algae [1]. High evolutionary rates seem to be more common in "primitive forms" such as *Spirogyra*, *Zygnema* and *Mougeotia*, whose evolutionary rates reportedly differ from other Zygnematales [20]. Combined, we conclude that the lack of so-called derived cell and chloroplast forms is compensated for by a high evolutionary rate. This yields a large variety of genotypes and helps cover ecological niches more quickly.

C-U ratios are generally elevated compared to the remaining substitution rates (Table 4). This is because mutations from C to U or U to C in non-coding RNA are not detrimental, as the change in base pairs from G-U or G-C does not affect secondary structure. The biggest difference within the GTR+I+G model occurs between *Spirogyra* and the remaining Zygnematophyceae without *Spirotaenia*: *Spirogyra* shows a 1.9 units lower rate, which partly explains the different mutation rate compared to other algae. Both *Spirogyra* and *Spirotaenia* show a slightly elevated A-G substitution rate (approximately 50% and 16% higher than in the Zygnematophyceae data set, respectively). *Spirogyra* displays not only a higher evolutionary rate, but also a different pattern of base substitution rates compared to the remaining Zygnematales. Evolutionary rates within *Spirogyra* are more diverse than previously expected; significant differences among clades and sequences outside the same clade prevail. Finally, clades B to D, comprising sequences with the 1506 group I intron, form a group with fewer differences compared to clades E to H, comprising taxa without the intron.

The genus *Spirogyra* is clearly monophyletic. No sequences were found that had slower evolutionary rates or that could resolve the long branch reported in previous studies [20]. The individual *Spirogyra* clades found in both of our data sets are essentially the same and well supported by bootstrap and PP values. Moreover, the phylogenetic relationship among those taxa exhibiting the secondary loss of the 1506 group I intron is identical. The *Spirogyra* clade branches are longer than the branches of most genera of Desmidiales. Two long branches within the genus *Spirogyra* are present in both trees; one separates clade A from the rest, and the other separates taxa with the secondary loss of the 1506 group I intron. *Spirogyra* taxa characterized by the absence of the 1506 group I zygnematalean intron form a distinct clade with no exception in both trees. This clearly indicates a single loss event.

This explanation is supported by the accelerated evolutionary rate of the exon region of the SSU rDNA [20]. A large number of differentiating NHS exists both within the genus *Spirogyra* (see results) and for each clade. This emphasizes the distinctness and genetic variety within the genus. In contrast, Gontcharov and Melkonian [42] found only very few NHS to circumscribe the different clades in *Cosmarium*.

Earlier hypotheses suggested the unicellular *Spirotaenia* as the ancestor of the filamentous *Spirogyra* [28, 40]. Although *Spirotaenia* shares many attributes of the genus *Spirogyra* such as chloroplast shape, absence of the 1506 group I intron [43] and an elevated evolutionary rate, molecular analyses have not proven or clarified its phylogenetic position. Our analyses revealed two well-supported branches at the base of the Zygnematalean clade, but the ancestral form remains unclear. *Spirotaenia*'s unusual position among the Zygnematales was already investigated by Gontcharov and Melkonian [43]. Their results indicate no affiliation of the two genera [43]. They therefore suggested excluding *Spirotaenia* from the Zygnematophyceae sensu stricto. Perhaps the same applies to the genus *Spirogyra,* but this remains to be proven by further genetic analyses. Sequence differences of the unambiguously aligned positions within *Spirogyra* reached 274 nt difference among the strains (a member of Clade A and one of Clade H), whereas the biggest difference found within the remaining taxa was 247 nt (*Closterium ehrenbergii* and *Spirotaenia obscura*). Also, the average, median and mode of the difference between sequences are higher in *Spirogyra* than in the remaining taxa considered. Within the clade, the range of pair-wise differences within the examined Desmidiales (Desmidiaceae, Peniaceae and Closteriaceae) ranges from 10 to 166 nt: this group includes 3 families and 7 genera. The relative distances within this group do not exceed 10.18. Within the genus *Spirogyra*, 14 individual distances between clades exhibit bigger values (Table 3). The same trend was observed in the remaining Zygnematales (Zygnemataceae and Mesotaeniaceae) – the range of pair-wise differences resulted in 27 to 173 nt, the respective distance values ranged from 1.65 to 10.63. The genetic differences within the genus *Spirogyra*, i.e. among the individual clades, exceed the differences of genera in either of the other groups. This calls for discussion and further investigation on whether *Spirogyra* still should be considered a single genus or rather be subdivided based on the clades we differentiated. This once again underlines the different evolutionary rates and reflects the great genetic variability of the genus *Spirogyra*.

Drummond *et al*. [3] found *Spirogyra* to be monophyletic, but still treated *Sirogonium* as a separate genus based on rbcL data. They were unable to discover morphological characters useful for a generic distinction, simply because the taxa are largely congruent (e.g., number of more or less loosely coiled chloroplasts, reproduction by conjugation and anisogamy of gametangial cells). The diagnostic characters are mainly based on the mode of conjugation: while *Spirogyra* develops conjugation tubes, *Sirogonium* filaments are bent towards each other and form knee-shaped bends, so-called geniculations. Drummond *et al*. [3] also considered the shape and ornamentation of the chloroplast margin as a diagnostic feature, but our observations showed this character to be variable and highly dependent on filament vitality. Other morphological characters such as chloroplast number or cell width are also known to be highly variable and could be explained by polyploidy [1, 34-37]. Other authors also found *Spirogyra* and *Sirogonium* forming a single lineage based on single gene analyses (rbcL, SSU rDNA) and combined data sets [15, 20, 22, 43]. Gontcharov stated in 2002 that the genus *Sirogonium* has to be rejected and S. *sticticum* (S. *sticticum* is the type species of the genus) has to be considered as a species within *Spirogyra* [20]. Interestingly, Czurda [10] already suggested including *Sirogonium* into *Spirogyra* as one of four subgenera. We also found *Spirogyra* to be monophyletic and inseparable from *Sirogonium*. Monophyly of the *Spirogyra*/*Sirogonium* clade

was determined in all our phylogenetic analyses, placing *Sirogonium* firmly in Clade C. All alternative tree topologies relocating *Sirogonium* outside *Spirogyra* were significantly worse than the best tree uniting *Spirogyra* and *Sirogonium*. *Sirogonium* shares NHS signatures with the genus *Spirogyra* and the clade in which it is located.

Conclusion

Spirogyra is monophyletic, incorporating the former genus *Sirogonium*. Genetic diversity and genetic distances within *Spirogyra* exceed the diversity and distances found in other Zygnematophyceaen genera. Our results suggest pursuing the question whether *Spirogyra* should be split into several genera, one of them incorporating *Sirogonium.*

In the surveyed 130 sequences, 53 individually different clones were identified – more than was expected from the simple vegetative morphology. The genus forms eight well-supported clades that differ considerably in NHS pattern – ranging from 3 to 29 NHS for a clade. The genus *Spirogyra* itself exhibits 41 NHS (4 CBCs). Characterizing those clades will require additional studies considering phylogenetic studies on ITS2 secondary structure, hypnozygote morphology, vegetative characteristics and ecology.

Methods

Origin of organisms

Spirogyra clones used in this study originated from a field survey conducted in 2006 and 2007 [2]. Single filaments were isolated by the author (CC) and incorporated into the Algenkultursammlung Wien (ASW). The non-axenic clones were maintained in 100 ml Erlenmeyer flasks with Desmids medium [44] at 18°C under low light conditions at a 16:8 l:d light cycle (provided by either Philips TLD 36W/33 or Osram FQ 39W/840 LUMLUX Cool White). Because only few strains could be identified at species level, cultures were labeled with a code for the corresponding sampling site and date (Table 6). For our study, we considered 130 *Spirogyra* isolates from different sampling sites and with different vegetative morphologies to cover various ecological niches. Additionally, we included some strains from the UTEX culture collection (UTEX 1746 *Spirogyra pratensis*; UTEX 1273 *S. crassispina*; UTEX 1683 *S. occidentalis*; UTEX1742 *S. juergensis*; UTEX 1745 *S. liana*; UTEX 2495 *S. maxima*).

DNA extraction

Prior to extraction the cultures were transferred into a defined mineral medium (modified Woods Hole medium; [9]). After 4 to 6 weeks, the algae were harvested with a sterile needle and put into a sterile 2 ml microcentrifuge tube. Samples were frozen at -80°C for at least 4 h and then lyophilized for at least 48 h to improve the DNA yield. Afterwards, the samples were placed in 2 ml Eppendorf tubes containing 5 to 7 glass beads (3 mm diam.) and ground with a homogenizing mill. Total DNA was extracted following a modified CTAB protocol ([45] modified after [46]).

DNA amplification and sequencing

Primers used in this study are given in Table 7 [47, 48]. The PCR reaction mixture was prepared according to the manufacturer's recommendation. For each PCR reaction, a 10 µl mixture was prepared containing 9 µl ABGene Reddy Mix PCR Master Mix, 0.2 µl for each primer at 20 pM.µ l^{-1}, 0.4 µl dimethyl sulfoxide (Sigma) and 0.2 µl DNA template. When the PCR result was unsatisfactory due to low DNA concentration, up to 0.5 µl DNA template was used; when DNA template volume was increased, dimethyl sulfoxide volume was reduced to maintain the total volume of 10 µl, accepting a slight shift in the overall ratio of ingredients. The PCR reaction conditions were an initial hold at 80°C

for 5 min followed by 36 cycles starting with a denaturation step at 95°C for 30 s, an annealing step at 55°C for 30 s and an extension step at 72°C for 2 min. A final extension step at 72°C for 8 min and the final hold at 4°C were performed after the 36 cycles were completed. The amplified DNA was cleaned by incubating at 37°C for 45 min, followed by denaturing at 80°C for 15 min together with the enzymes Exonuclease I and Shrimp Alkaline Phosphatase (both from Fermentas) and then subjected to a cycle sequencing reaction. The cycle sequencing reaction conditions were an initial hold at 96°C for 1 min followed by 35 cycles starting with a denaturation step at 96°C for 10 s followed by an annealing step at 50°C for 5 s and an extension step at 60°C for 4 min. The end of the cycles was followed by a final hold at 4°C. Sequencing was performed on a 16-capillary sequencer (Applied Biosystems 3130xl Genetic Analyzer) following the manufacturer's protocols. The SSU rDNA sequences were used in the phylogenetic analyses; their GenBank accession numbers are given in Table 6.

Sequence alignment and phylogenetic analysis

Sequences were aligned manually taking into account the secondary structure of the SSU rDNA [49]. The alignment was refined by comparison of the secondary structure of the sequences. Secondary structure was determined via the Rensselaer bioinformatics web server using mfold [50]. Only unambiguously aligned regions of the sequences were used for analyses; gap-rich regions were excluded. Two different data sets were analyzed: (1) the SSU alignment of 33 Zygnematophyceaen taxa (including 12 *Spirogyra* sequences representing the different clades) and (2) the SSU alignment of 55 *Spirogyra* sequences (clones with identical sequences were represented by only one sequence). The combined Zygnematophyceaen SSU dataset consisted of 1720 unambiguously aligned bases, the *Spirogyra* dataset of 1645 such bases.

The phylogenetic trees presented were inferred by maximum likelihood (ML) settings using PAUP* 4.0b10 [51], and the best models were chosen according to the Akaike Information Criterion by Modeltest 3.7 [52, 53]. To test for the best evolutionary model for the analyses, the log-likelihood values of 56 models using Modeltest 3.7 were compared. No outgroup was applied and unrooted phylogenies were used [22]. This was done to avoid LBA *sensu* Philippe [31] caused by unsuitable taxa as outgroups. This approach also follows Gontcharov's argument that monophyly of the Zygnematophyceae is undoubted but that its position within the Streptophyta is unclear and therefore no suitable outgroup can be chosen. For the Zygnematophyceaen alignment analyses, the TrN+I+G model was chosen; for the *Spirogyra* alignment the GTR+I+G was chosen by Modeltest. Individual *Spirogyra* clades were labeled with letters A to H in the sequence from basal to derived. The *Spirogyra* alignment was analyzed unrooted to avoid LBA phenomena due to different evolutionary rates [31]. Only individual sequences were used for analyses to reduce computational effort. Bayesian inference (PP) was calculated using MrBayes 3.1.2. [54, 55] using 3 million generations, sampling every 100 generations and MCMC chains = 4. All trees below the burnin value of 0.01 were discarded as burnin, the consensus tree was calculated using PAUP*. The robustness of the trees was assessed by bootstrap support values. These were calculated using the corresponding evolutionary model chosen by Modeltest by maximum likelihood (ML; 100 replicates), distance (neighbor joining (NJ); 1000 replicates), and maximum parsimony (MP; 1000 replicates) methods using the accordant settings/evolutionary model for each dataset. Insignificant values were not included in figures (PP < 0.95, ML, NJ, MP < 50%). Details of the corresponding evolutionary models and Bayesian analyses are given in the legend of the accordant figures and in Table 4a.

Additionally, a distance matrix was calculated using PAUP* to evaluate the genetic distances among the *Spirogyra* clades (Table 3). The value was obtained by dividing the number of differing bases by the number of total (=aligned) bases. Values were transformed to % values, so numbers near 0 indicate a high identity or short distance; values approximating 100 indicate low similarity and large distance. Relative rate tests were carried out among all genera and for all accessions individually used for phylogeny in GRate 0.4 [56].

Tree topology tests

User defined trees (UD-trees) were generated manually based on the "best tree" (derived from ML analysis; same topology as ML phylogeny used for Figure 1) using TreeView 1.6.6 [57]. To compare the UD-trees with the "best tree", the alignment was loaded into PAUP and site-wise log-likelihood values for each tree were calculated. The result was used as input for the program CONSEL v0.1k [58], calculating probability values according to the Kishino-Hasegawa test (KH; [59], the Shimodaira-Hasegawa test (SH; [60], both weighted=w and unweighted), and the approximately unbiased test (AU) using the multiscale bootstrap technique [61] (Table 1 and 2).

Apomorphy analysis

The secondary structure of the SSU rDNA (Figure 3) was modeled after Wuyts [49], following the same numbering pattern. To find all non-homoplasious synapomorphies (NHS), the method described by Marin et al. [62] was applied. To identify genetic characteristics for the different groups (synapomorphic signatures; [63]), the secondary structure in an alignment of Zygnematophyceae and all sequenced *Spirogyra* clones was compared, and NHS and CBCs were determined according to Marin et al. [62]. The analysis was performed with two aims: (1) to find NHS for the genus *Spirogyra* and (2) to identify NHS for each individual clade within the genus *Spirogyra*. For both, the two NHS criteria were applied: (1) absence of convergent evolution outside the clade and (2) strict conservation within the clade (Table 5).

Authors' contributions

CC co-designed the study, collected the samples, generated the sequence data, did the analyses and prepared the manuscript. MHJB helped with the lab work, obtaining and analyzing the data. TP provided conceptual guidance and supported the data analysis. MS designed the study and provided conceptual support. All authors contributed to the preparation of the manuscript, and read and approved the final version.

Acknowledgements

This study was supported by the Austrian Science Fund, project number P18465.

References

1. Hoshaw RW, McCourt RM: The Zygnemataceae (Chlorophyta): A twenty-year update of research. *Phycologia* 1988, 27(4):511-548.
2. Hainz R, Wöber C, Schagerl M: The relationship between *Spirogyra* (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe. *Aquatic Botany* 2009, 91(3):173-180.
3. Drummond CS, Hall J, Karol KG, Delwiche CF, McCourt RM: Phylogeny of *Spirogyra* and *Sirogonium* (Zygnematophyceae) based on rbcL sequence data. *Journal of Phycology* 2005, 41(5):1055-1064.
4. Kadlubowska JZ: Conjugatophyceae I - Zygnemales. In: *Süßwasserflora von Mitteleuropa, Chlorophyta VIII.* Edited by Ettl H, Gerloff H, Heynig H, Mollenhauer D: Stuttgart, New York: Gustav Fischer Verlag; 1984.
5. Ashraf M, Godward MBE: Ultrastructure and chemistry of the zygospore wall of Spirogyra. *Ann Bot* 1980, 46:485-487.
6. Kim J-H, Kim YH, Lee IKL: Morphotaxonomy of the Genus *Spirogyra* (Zygnemataceae, Chlorophyta) in Korea *Algae* 2004, 19(2):91-105.
7. Baum D: Phylogenetic species concepts. *TREE* 1992, 7(1):1-2.
8. Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG: Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution* 2003, 18(11):597-603.
9. Simons J, Van Beem P, De Vries PJR: Induction of conjugation and spore formation in species of *Spirogyra* (Chlorophyceae, Zygnematales). *Acta Bot Neerl* 1984, 33(3):323-334.
10. Czurda V: Zygnemales. Jena: Gustav Fischer; 1932.
11. Czurda V: Experimentelle Untersuchungen über die Sexualverhältnisse der Zygnemalen. *Sonderabdruck aus "Beihefte zum Bot Centralbl"* 1930, 47 Abt.I.:15 - 68.
12. Czurda V: Experimentelle Analyse der kopulationsauslösenden Bedingungen bei Mikroorganismen - 1. Untersuchungen an Algen (*Spirogyra*, *Zygnema* und *Hyalotheca*). *Sonderabdruck aus "Beihefte zum Bot Centralbl"* 1933, 51 Abt. I.:711 - 761.
13. Zwirn M: Sexuelle Fortpflanzung und taxonomische Fragestellungen bei der Süßwasseralge Spirogyra (Zygnemataceae, Streptopyhta). Vienna: University of Vienna; 2010.
14. Gerrath JF: The biology of desmids: A decade of progress, vol. 9. Bristol: Biopress; 1993.
15. McCourt RM, Karol KG, Bell J, Helm-Bychowski KM, Grajewska A, Wojciechowski MF, Hoshaw RW: Phylogeny of the conjugating green algae (Zygnemophyceae) based on rbcL sequences. *Journal of Phycology* 2000, 36(4):747-758.
16. Fott FE: Algenkunde, vol. 2nd ed. Jena: V.E.B. Fischer; 1935.
17. Round FE: The taxonimy of the Chlorophyta. II. *Br PhycolJ* 1971, 6:235-264.
18. Bold HC, Wynne MJ: Introduction to the algae, vol. 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.; 1985.
19. Prescott GW, Croasdale HT, Vinyard WC: Desmidilaes. I. Saccodermae, Mesotaeniaceae. New York: New York Botanical Garden, Bronx; 1972.
20. Gontcharov AA, Marin B, Melkonian M: Molecular phylogeny of conjugating green algae (Zygnemophyceae, Streptophyta) inferred from SSU rDNA sequence comparisons. *Journal of Molecular Evolution* 2002, 56(1):89-104.
21. Transeau EN: The Zygnemataceae (Fresh-water conjugate algae): The Ohio State University Press; 1951.
22. Gontcharov AA, Marin B, Melkonian M: Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Molecular Biology and Evolution* 2004, 21(3):612-624.
23. McCourt RM: Green algal phylogeny. *Trends in Ecology & Evolution* 1995, 10(4):159-163.

24. Mix M: Die Feinstruktur der Zellwände bei Mesotaeniaceae und Gonatozygaceae mit einer vergleichenden Betrachtung der verschiedenen Wandtypen der Conjugatophyceae und über deren systematischen Wert. *Archives of Microbiology* 1972, 81(3):197-220.
25. Mix M: Die Feinstruktur der Zellwände der Conjugaten und ihre systematische Bedeutung. *Beih Nova Hedwigia* 1975, 42:179-194.
26. West GS: On variation in the Desmidiaceae and its bearing on their classification. *JLinnSoc* 1899, 34:376-415.
27. West GS: A treatise on the British freshwater alga. Cambridge: Cambridge University Press; 1904.
28. Yamagishi T: Classification of the Zygnemataceae. *Sci Rep Tokyo Kyoiku Daigaku B* 1963, 11:191-210.
29. Besendahl A, Bhattacharya D: Evolutionary analyses of small-subunit rDNA coding regions and the 1506 group I introns of the Zygnematales (Charophyceae, Streptophyta). *Journal of Phycology* 1999, 35(3):560-569.
30. Bhattacharya D, Damberger S, Surek B, Melkonian M, Bhattacharya D: Primary and secondary structure analyses of the rDNA group-I introns of the Zygnematales (Charophyta). *Current Genetics* 1996, 29(3):282-286.
31. Philippe H: Opinion: long branch attraction and protist phylogeny. *Protist* 2000, 151(4):307-316.
32. Gontcharov AA, Melkonian M: Molecular phylogeny of *Staurastrum* Meyen ex Ralfs and related genera (Zygnematophyceae, Streptophtya) based on coding and noncoding rDNA sequence comparisons. *Journal of Phycology* 2005, 41(4):887-899.
33. Gontcharov AA, Melkonian M: A Study of Conflict between Molecular Phylogeny and Taxonomy in the *Desmidiaceae* (Streptophyta, Viridiplantae): Analyses of 291 rbcL Sequences. *Protist* 2011, 162(2):253-267.
34. Allen MA: The biology of a species complex in *Spirogyra*. Bloomington, Univ.Microfilms, Ann. Arbor., Mich., : Indiana University; 1958.
35. McCourt RM, Hoshaw RW: Noncorrespondence of breeding groups, morphology and monophyletic Groups in *Spirogyra* (Zygnemataceae: Chlorophyta) and the application of species concepts. *Systematic Botany* 1990, 15(1):69-78.
36. Hoshaw RW, Wang J-C, McCourt RM, Hull HM: Ploidal changes in clonal cultures of *Spirogyra communis* and implications for species definition. *American Journal of Botany* 1985, 72(7):1005-1011.
37. Hoshaw RW, Wells CV, McCourt RM: A polyploid species complex in *Spirogyra maxima* (Chlorophyta, Zygnemataceae), a species with large chromosomes. *Journal of Phycology* 1987, 23(s2):267-273.
38. Bhattacharya D, Surek B, Rüsing M, Damberger S, Melkonian M: Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). *Proceedings of the National Academy of Sciences of the United States of America* 1994, 91(21):9916-9920.
39. Chen C, Schagerl M: Slow evolution of 1506 group I intron in *Spirogyra* LINK 1820 (Zygnematophyceae, Streptophyta), a fast evolving lineage in the Zygnemataceae. *Fottea* 2012.
40. Randhawa MS: Zygnemaceae. New Delhi: Indian Council of Agricultural Research; 1959.
41. Palla E: Über eine neue pyrenoidlose Art und Gattung der Conjugaten. *Ber d Deutsch Bot Ges* 1894, 12:228-236.
42. Gontcharov AA, Melkonian M: In search of monophyletic taxa in the family Desmidiaceae (Zygnematophyceae, Viridiplantae): the genus *Cosmarium*. *American Journal of Botany* 2008, 95(9):1079-1095.
43. Gontcharov AA, Melkonian M: Unusual position of the genus *Spirotaenia* (Zygnematophyceae) among streptophytes revealed by SSU rDNA and rbcL sequence comparisons. *Phycologia* 2004, 43(1):105-113.

44. Kusel-Fetzmann E, Schagerl M: Verzeichniss der Sammlung von Algenkulturen an der Abteilung für Hydrobotanik am Institut für Pflanzenphysiologie der Universität Wien. *Phyton* 1993, 33:209-234.
45. Tel-zur N, Abbo S, Myslabodski D, Mizrahi Y: Modified CTAB Procedure for DNA Isolation from Epiphytic Cacti of the Genera Hylocereus and Selenicereus (Cactaceae). *Plant Molecular Biology Reporter* 1999, 17(3):249-254.
46. Russell A, Samuel R, Rupp B, Barfuss MHJ, Šafran M, Besendorfer V, Chase MW: Phylogenetics and cytology of a pantropical orchid genus *Polystachya* (Polystachyinae ; Vandeae ; Orchidaceae) ; evidence from plastid DNA sequence data. *Taxon* 2009:16 pp.
47. Marin B, Klingberg M, Melkonian M: Phylogenetic relationships among the *Cryptophyta*: analyses of nuclear-encoded SSU rRNA sequences support the monophyly of extant plastid-containing lineages. *Protist* 1998, 149(3):265-276.
48. Pröschold T, Marin B, Schlösser UG, Melkonian M: Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas Ehrenberg* and *Chloromonas Gobi*, and description of *Oogamochlamys gen. nov.* and *Lobochlamys gen. nov*. *Protist* 2001, 152(4):265-300.
49. Wuyts J, De Rijk P, Van de Peer Y, Pison G, Rousseeuw P, De Wachter R: Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucl Acids Res* 2000, 28(23):4698-4708.
50. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res* 2003, 31(13):3406-3415.
51. Swofford DL: PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). In., 4.0b10 edn. Sunderland, Massachusetts: Sinauer Associates; 2003.
52. Posada D, Buckley T: Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 2004, 53:793-808.
53. Posada D, Crandall K: Modeltest: testing the model of DNA substitution. *Bioinformatics* 1998, 14(9):817-818.
54. Huelsenbeck JP, Ronquist F: MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 2001, 17:754-755.
55. Ronquist F, Huelsenbeck JP: MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, 19:1572-1574.
56. Müller K, Borsch T, Legendre L, Porembski S, Theisen I, Barthlott W: Evolution of Carnivory in Lentibulariaceae and the Lamiales. *Plant Biology* 2004, 6(4):477-490.
57. Page RDM: TREEVIEW: An application to display phylogenetic trees on personal computers *Comput Applic Biosci* 1996, 12: 357-358.
58. Shimodaira H, Hasegawa M: CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 2001, 17(12):1246-1247.
59. Kishino H, Hasegawa M: Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* 1989, 29(2):170-179.
60. Shimodaira H, Hasegawa M: Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 1999, 16(8):1114-1116.
61. Shimodaira H: An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 2002, 51(3):492-508.
62. Marin B, Palm A, Klingberg Max, Melkonian M: Phylogeny and taxonomic revision of plastid-containing Euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist* 2003, 154(1):99-145.
63. Luo W, Pflugmacher S, Pröschold T, Walz N, Krienitz L: Genotype versus Phenotype Variability in *Chlorella* and *Micractinium* (Chlorophyta, Trebouxiophyceae). *Protist* 2006, 157(3):315-333.

Table 1: Comparison of the maximum likelihood tree (Zygnematophyceaen alignment) with user defined trees by AU (P-value of the approximately unbiased test calculated from multiscale bootstrap), PP, KH, SH and weighted SH. Trees significantly worse than the best trees at p≤ 0.05 are indicated by grey highlighting.

Tree No	AU	PP	KH	SH	WSH	Diff -lnL
1	0.927	0.368	0.497	0.995	0.995	
2	0.927	0.368	0.503	0.993	0.993	0
3	0.213	0.257	0.302	0.94	0.854	-0.359
4	0.101	0.007	0.116	0.767	0.353	-3.927
5	0.002	2.00E-10	0.003	0.305	0.010	-21.528
6	2.00E-07	2.00E-51	0	0	0	-115.716
7	2.00E-08	2.00E-162	0	0	0	-371.230
8	1.00E-08	2.00E-165	0	0	0	-378.236

Tree topologies tested: 1. Tree derived from ML analysis. 2. *Spirogyra* clade is located outside the Desmidiales and Zygnematales clades. 3. *Spirogyra* clade is a sister to the Zygnematales clade (including *Spirotaenia*). 4. *Spirogyra* clade is located in the Zygnematales sensu stricto clade. 5. *Spirogyra* is not monophyletic. 6. *Spirotaenia* is a sister to *Spirogyra*. 7. *Sirogonium* is located outside the *Spirogyra* clade. 8. *Sirogonium* is located within the Zygnematales clade.

Changes in tree topology of trees 4, 5, 7 and 8 are indicated in Figure 1 – ATT4 marks the insertion point for alternative placement of the *Spirogyra* clade for tree topology 4; ATT5 (3 times) mark the insertion points for Spirogyra clade A, clade B to D, and clade E to H, respectively, for tree topology 5; ATT7 marks insertion point for *Sirogonium* for tree topology 7; and ATT8 marks insertion point for *Sirogonium* for tree topology 8.

Table 2: Comparison of the maximum likelihood tree (*Spirogyra* alignment) with user defined trees by AU, KH, SH and weighted SH. Trees significantly worse than the best trees at p≤ 0.05 are indicated by grey highlighting.

Tree No	AU	PP	KH	SH	WSH	Diff -lnL
1	0.989	1	0.968	1	1	0
2	0.012	0.000007	0.032	0.818	0.131	-11.8495
3	0.00007	2e-23	0	0.126	0	-52.3597
4	0.0002	5e-25	0.0001	0.125	0.0002	-55.9536
5	0.0000003	2e-28	0	0.084	0	-63.9885
6	0.002	3e-118	0	0	0	-270.5954
7	4e-63	5e-133	0	0	0	-304.6919
8	4e-66	2e-146	0	0	0	-335.5584

Tree topologies tested: 1. Tree derived from ML analysis. 2. UTEX 1742 and 7075 are located in clade E. 3. UTEX 1745 is located in clade E. 4. *Sirogonium* is located in clade B. 5. UTEX 1745 belongs to clade F. 6. UTEX 1742 and 7075 belong to clade D. 7. UTEX 1742 trough clade H are non-monophyletic. 8. *Sirogonium* is located in clade A.

Table 3: Table of distances between *Spirogyra* clades; distance measure in the upper right part, pair wise differences in the lower left part

	Clade A	Clade B	Clade C	Clade D	Clade E	Clade F	Clade G	Clade H	Zygn	Desm
Clade A	-	11,43	13,46	11,31	12,81	13,79	13,16	16,74	22,47	21,44
Clade B	186	-	7,60	5,09	9,21	9,52	9,21	13,51	23,05	22,22
Clade C	219	124	-	8,59	11,00	12,04	11,55	15,30	23,23	22,09
Clade D	184	83	140	-	8,36	8,91	8,36	13,09	22,76	22,15
Clade E	208	150	179	136	-	7,30	6,92	10,78	24,26	23,48
Clade F	224	155	196	145	119	-	5,46	9,76	25,31	24,66
Clade G	214	150	188	136	113	89	-	8,70	24,86	23,73
Clade H	272	220	249	213	176	159	142	-	25,63	25,71
Zygnematales	366	375	378	370	395	411	405	417	-	19,49
Desmidiales	349	361	359	360	382	400	386	418	319	-
within clade range distance	1,04-1.29	0.06-1.29	0.25-4.56	0.06-1.17	1.17-2.15	0.18-0.31	0.00-0.43	0.06-5.23	1.65-10.63	0.61-10.18
Pair wise differences	17-21	1-20	4-61	1-19	19-35	3-5	0-7	0-51	27-173	10-166

Table 4a: Summary of evolutionary models (chosen by Modeltest) and character states for all individual and combined data sets.

	1.	2.	3.	4.	5.	6.
No. Taxa included	33	20	55	13	17	3
Model	TrN+I+G	GTR+I+G	GTR+I+G	GTR+I+G	GTR+I+G	TrN+I+G
-lnL	10368.3486	6458.2769	7311.2354	5725.9985	5393.957	3020.6091
I	0.4608	0.5560	0.6009	0.5228	0.5774	0.8131
G	0.6376	0.7144	0.6856	0.5791	0.6728	-
Base frequencies						
A	0.2545	0.2598	0.2338	0.2436	0.2595	equal rates
C	0.2280	0.2067	0.2439	0.2354	0.2077	
G	0.2665	0.2657	0.2871	0.2808	0.2651	
T	0.2510	0.2678	0.2351	0.2403	0.2677	
G-C	0.4945	0.4724	0.5310	0.5162	0.4728	
Rate matrix						
[A<->C]	1.0000	1.0169	1.4341	1.7371	1.2473	1.0000
[A<->G]	1.8721	1.9209	2.6641	2.7392	1.8624	2.1626
[A<->T]	1.0000	1.2931	1.2357	1.3361	1.5619	1.0000
[C<->G]	1.0000	0.7339	1.6993	1.8149	0.6646	1.0000
[C<->T]	4.5252	5.9738	5.2526	5.9522	7.1560	5.2068
[G<->T]	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Character status						
aligned nt	1720	1720	1645	1720	1720	1720
constant nt	1136	1320	1258	1351	1393	1607
MP-informative	479	256	330	269	177	0
MP-uninformative	105	144	57	100	150	113

Data sets used: 1. Zygnematophyceae and 12 *Spirogyra* sequences (same as in combined SSU alignment used in Figure 1). 2. Zygnematophyceae. 3. all *Spirogyra* sequences (same as *Spirogyra* alignment used in Figure 2). 4. 12 *Spirogyra* sequences (same *Spirogyra* sequences as data set 1.). 5. Zygnematophyceae without *Spirotaenia*. 6. *Spirotaenia*

Table 4b: Results of the Relative Rate Test carried out in GRate (Müller et al., 2004); using only unambiguously aligned positions; not significant: N.S. ($p > 0.05$; relative rates not significantly different at 5% level). Asterisks: $p = 0.05 > * > 0.01 > ** > 0.005 > ***$ (relative rates significantly different).

taxa	Peniaceae	Closteriaceae	Mesotaeniaceae	Zygnemataceae	Spirogyra
Desmidiaceae	***	***	N.S.	N.S.	***
Peniaceae		*	N.S.	*	***
Closteriaceae			*	***	***
Mesotaeniaceae				N.S.	***
Zygnemataceae					***

Table 7: List of primers used in this study

Primer name	Sequence	Reference
EAF3	5'-TCGACAATCTGGTTGATCCTGCCAG-3'	[42]
18sF2	5'-ACCACATCCAAGGAAGGCAGCAG-3'	This study
18sR1	5'-ACGCTATTGGAGCTGGAATTACCGC-3'	This study
18sF3	5'-AGTCCCAACCGTAAACGATGCC-3'	This study
N920R2	5'-CCCTTCCGTCAATTCCTTTAAGTTTC-3'	This study
18sR3	5'-TGTTACGACTTCTCCTTCCTCTAAACG-3'	This study
BR	5'-TTGATCCTTCTGCAGGTTCACCTAC-3'	[41]

Table 4c: Results of the Relative Rate Test carried out in GRate (Müller et al., 2004); using only unambiguously aligned positions of all sequences used in this study; not significant: N.S. ($p > 0.05$; relative rates not significantly different at 5% level). Asterisks: $p = 0.05 > * > 0.01 > ** > 0.005 > ***$ (relative rates significantly different).

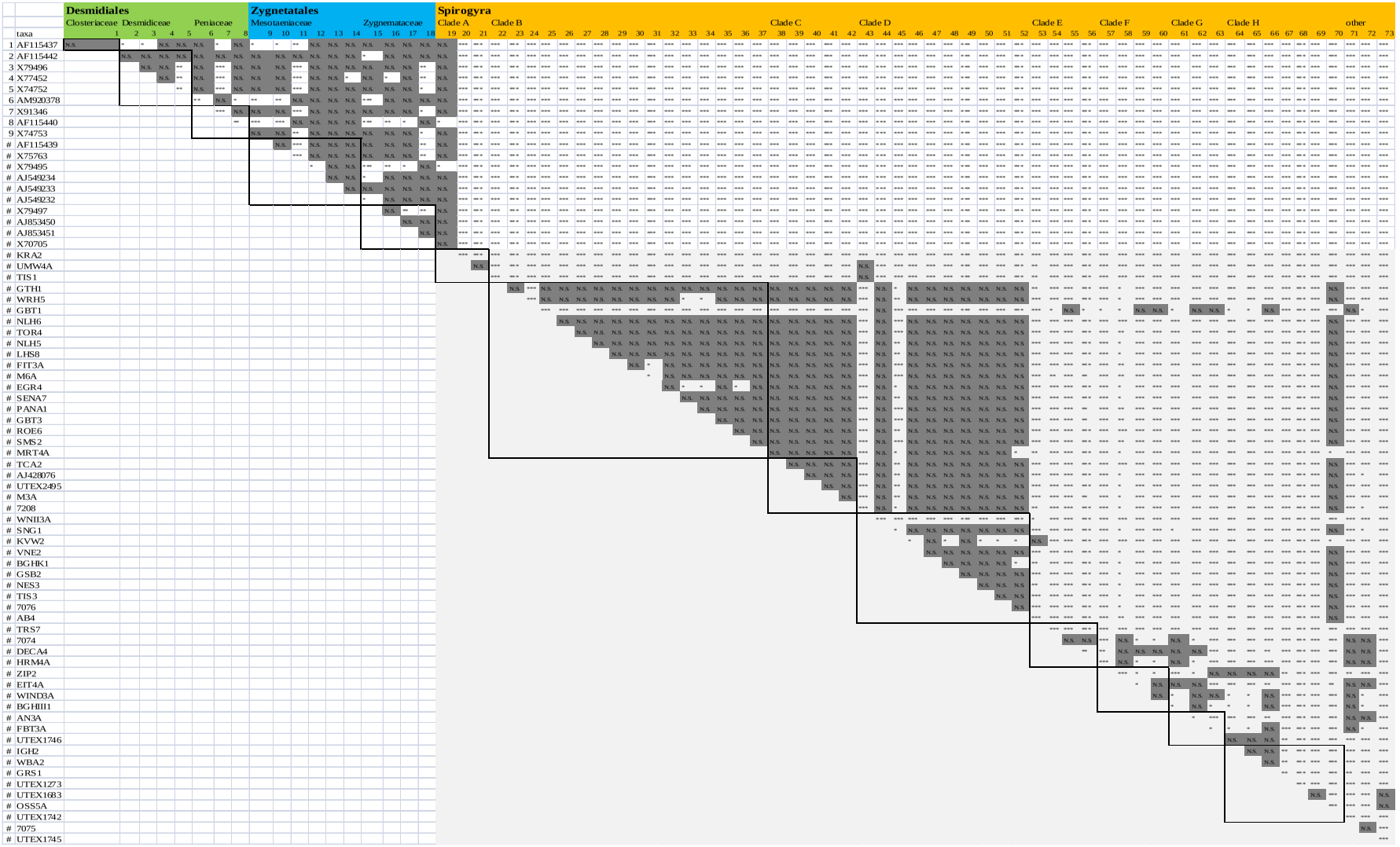


Table 5: List of NHS found for the genus *Spirogyra* and clades

Number *Spirogyra*	1	2	3						9					14				18													
Clade A							7	8				12				16			19							27			29	30	31
Clade B																															
Clade C																															
Clade D													13				17			20						26		28			
Clade E																							24								
Clade F						6				10											21	22	23								
Clade G											11																				
Clade H				4	5										15						21				25						
Position in	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alignment	0	1	1	1	2	2	2	2	2	2	2	2	2	2	2	3	3	3	4	5	5	5	5	5	5	5	5	5	6	6	
	9	4	7	9	0	1	1	2	3	4	4	5	6	8	8	1	5	6	5	2	2	3	3	3	5	5	6	7	7	4	7
	1	0	0	1	2	4	9	8	9	2	5	9	3	2	3	4	1	5	2	1	7	0	1	3	0	1	3	5	7	0	7
Helix	6	J8-9	L9	J9-	L5	J5-	J5-	BE10_1	E10_1	E10_1	E10_1	E10_1	BE10_1	E10_1	JE10_1-	11	B12	L12	B19	B17	L17	17	17	17	L18	L18	L18	J18-	J18-	J3-	E23_1
number				10		E10_1	E10_1								11													19	19	22	
without : helix, J: link, L: endloop, B: buldge																															
D13324	U	A	C	A	C	C	U	A	C	G	U	A	~	A	A	G	C	A	A	U	A	U	C	U	A	C	C	U	A	G	U
X74000	U	A	C	A	C	C	U	A	C	G	U	A	~	A	A	G	C	A	A	U	A	U	C	U	A	C	C	U	A	G	U
AF115438	U	A	C	A	C	U	U	A	C	A	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	U	C	G	U
AF115437	C	G	C	U	C	U	U	C	G	G	C	G	U	A	A	G	C	A	A	U	G	U	C	U	C	G	C	U	A	G	G
AF115442	U	A	C	A	C	C	U	A	C	A	U	G	U	A	A	G	C	A	A	C	G	U	U	U	A	C	C	U	C	G	U
X91346	U	A	C	A	U	U	U	A	C	G	U	G	U	A	A	G	C	A	A	U	G	U	C	C	A	C	C	U	C	G	U
AF115440	U	A	C	A	U	C	U	A	C	A	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	U	C	G	U
X79496	U	A	C	A	C	C	U	A	C	A	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	U	C	G	U
X77452	U	A	C	A	C	U	U	A	C	A	U	G	U	A	A	G	C	A	A	U	G	U	U	U	A	C	C	U	C	G	U
X74752	U	A	C	A	C	C	U	A	C	A	U	G	U	A	A	G	C	A	A	U	G	U	U	U	A	C	C	U	C	G	U
AF115439	U	A	C	A	C	C	U	A	G	G	U	G	U	A	A	G	C	A	A	U	G	U	C	A	A	C	C	C	A	G	U
AM920378	U	A	C	A	C	C	U	A	C	G	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	U	C	G	U
X75763	U	A	C	A	U	C	U	A	G	G	U	G	U	A	A	G	~	A	A	U	G	U	C	U	A	C	C	C	C	G	C
X74753	U	A	C	A	U	U	U	A	C	G	U	G	U	A	A	G	C	A	A	U	G	U	C	C	A	C	C	U	C	G	U
X79495	U	A	C	A	U	C	U	A	G	G	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	C	C	G	U
X79497	U	A	C	A	C	C	U	A	C	A	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	U	C	G	U
AJ853450	U	A	C	A	U	C	U	A	G	G	U	G	U	A	A	G	C	A	A	U	G	U	C	~	A	C	C	C	C	G	U
AJ853451	U	A	C	A	C	C	U	A	G	G	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	C	C	G	U
AJ549234	N	A	C	A	U	U	U	A	G	G	C	G	C	A	A	G	C	A	A	U	G	U	C	U	A	C	C	U	C	G	U
AJ549233	U	A	C	A	U	U	U	A	G	G	U	G	C	A	A	G	C	A	A	U	G	U	C	A	A	C	C	U	C	G	U
AJ549232	U	A	C	A	C	U	U	A	G	G	U	G	U	A	A	G	C	A	A	U	G	U	C	U	A	C	C	~	A	G	U
X70705	U	A	C	U	A	U	U	A	G	G	U	G	U	A	A	G	~	A	A	U	G	U	C	U	A	C	C	U	C	G	C
KRA2_A	G	U	A	A	A	U	G	U	U	G	U	U	U	G	~	U	C	U	U	U	G	A	U	U	G	G	U	U	U	A	G
TIS1_A	G	U	A	A	A	U	G	U	U	G	U	U	U	G	~	U	C	U	U	U	G	A	U	U	G	G	U	U	U	A	G
WRH5_B	G	U	A	A	C	C	U	A	U	G	U	G	C	G	U	G	C	U	A	U	G	A	C	U	G	G	C	U	C	G	A
GTH1_B	G	U	A	A	C	C	U	A	U	G	U	G	C	G	U	G	C	U	A	U	G	A	C	U	G	G	C	U	C	G	A
MRT4A_C	G	U	A	A	C	C	U	A	U	G	U	G	C	G	U	G	C	U	A	U	G	A	C	U	G	G	C	U	C	G	A
TCA2_C	G	U	A	A	C	C	U	A	U	G	U	A	C	G	U	G	C	U	A	U	G	A	C	U	G	G	C	U	C	G	A
AJ428076	G	U	A	A	C	U	U	A	U	G	U	A	U	G	U	G	C	U	A	U	G	A	C	U	G	G	C	U	C	G	A
WNII3A_D	G	U	A	A	C	C	U	A	U	G	C	G	A	G	U	G	U	U	A	A	G	A	C	U	G	U	C	A	C	G	A
SNG1_D	G	U	A	A	C	C	U	A	U	G	U	G	A	G	U	G	U	U	A	A	G	A	U	U	G	U	C	A	C	G	A
TRS7_E	G	U	A	C	C	C	U	A	U	G	C	A	C	G	U	G	C	U	A	U	G	A	C	G	G	G	C	U	C	G	A
ZIP2_F	G	U	A	C	C	A	U	A	U	C	~	A	C	G	U	G	C	U	A	U	U	G	A	U	G	G	C	U	C	G	A
AN3A_G	G	U	A	C	C	C	U	C	U	G	G	A	C	G	U	G	C	U	A	U	A	C	G	U	G	G	C	U	C	G	C
IGH2_H	G	U	A	G	G	U	U	C	U	G	U	A	C	G	C	G	C	U	A	U	C	C	G	U	U	G	C	U	C	G	A

Table 5: List of NHS found for the genus *Spirogyra* and clades (continued)

Number *Spirogyra*																										
Clade A	32						38	39	40	41			44	45				49	50	51	52	53	54		56
Clade B																								55	
Clade C													44												
Clade D												43													
Clade E					36			39																	
Clade F									40							47									
Clade G			34	35																					
Clade H		33				37				41	42				46		48								
Position in Alignment	0 6 7 8	0 6 8 6	0 6 9 1	0 6 9 3	0 7 0 1	0 7 0 6	0 7 0 7	0 7 0 9	0 7 1 0	0 7 4 3	0 7 4 6	0 7 5 0	0 7 5 1	0 7 5 6	0 7 6 1	0 7 6 5	0 7 8 0	0 7 8 7	0 8 1 5	0 8 3 3	0 8 4 5	0 8 5 1	0 8 7 8	0 8 9 2	0 9 7 9
Helix number	BE23_1	E23_2	BE23_2	BE23_2	LE23_2	E23_2	E23_2	E23_2	E23_2	JE23_1-E23_4	E23_4	BE23_4	E23_4	E23_4	E23_4	E23_4	E23_4	BE23_4	E23_10	LE23_11	JE23_12-E23_9	JE23_12-E23_9	E23_14	E23_13	25
without : helix, J: link, L: endloop, B: buldge																									
D13324	G	U	C	G	C	U	G	A	C	U	U	G	G	G	C	G	G	A	G	U	A	U	U	U	U
X74000	G	U	C	G	C	U	G	A	C	U	U	G	G	G	C	G	G	A	G	U	A	U	U	U	U
AF115438	U	U	C	G	~	G	A	A	C	U	U	G	G	G	C	A	G	A	G	U	A	U	U	G	U
AF115437	U	U	C	G	C	G	A	A	C	U	U	G	G	G	C	G	G	G	G	~	A	U	U	G	U
AF115442	U	U	C	G	C	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
X91346	U	U	C	G	U	G	A	A	C	U	U	G	G	A	U	A	A	A	G	U	A	U	U	G	U
AF115440	U	U	C	G	~	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
X79496	U	U	C	G	C	G	A	A	C	U	U	G	G	G	C	G	G	A	G	U	A	U	U	G	U
X77452	U	U	C	G	C	G	A	A	C	U	U	G	G	G	U	A	A	A	G	U	A	U	U	G	U
X74752	U	U	C	G	C	G	A	A	C	U	U	G	G	G	C	G	G	A	G	U	A	U	U	G	U
AF115439	U	U	C	G	U	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
AM920378	U	U	C	G	~	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
X75763	G	U	C	G	C	G	A	A	C	U	U	G	G	A	U	G	G	A	G	U	A	U	U	G	U
X74753	U	U	C	G	C	G	A	A	C	U	U	G	G	G	U	G	A	A	G	U	A	U	U	G	U
X79495	U	U	C	G	C	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
X79497	U	U	C	G	G	G	A	A	C	U	U	G	G	G	C	G	G	A	G	U	A	U	U	G	U
AJ853450	U	U	C	G	U	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
AJ853451	U	U	C	G	U	G	A	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	U	G	U
AJ549234	G	C	U	C	U	A	G	G	C	U	U	G	G	G	C	G	G	A	G	U	A	U	G	U	U
AJ549233	G	U	A	U	C	A	G	A	C	U	U	G	G	G	C	A	G	A	G	U	A	U	G	U	U
AJ549232	G	A	U	U	C	G	U	A	G	U	U	G	G	G	C	A	G	A	G	U	A	U	G	U	U
X70705	G	U	C	G	C	G	~	A	C	U	U	G	G	G	U	G	G	A	G	U	A	U	G	G	U
KRA2_A	A	U	C	U	U	G	C	U	A	C	U	G	A	U	C	G	G	U	U	C	C	C	C	C	C
TIS1_A	A	U	C	U	U	G	C	U	A	C	U	G	A	U	C	G	G	U	U	C	C	C	C	C	C
WRH5_B	C	U	C	U	U	G	A	A	C	A	U	U	C	G	C	G	G	A	U	U	A	U	C	A	C
GTH1_B	C	U	C	U	C	G	A	A	C	A	U	U	C	G	C	G	G	A	U	U	A	U	C	A	C
MRT4A_C	C	U	C	U	U	G	A	A	C	A	U	U	U	G	C	G	G	A	U	A	A	U	C	G	C
TCA2_C	G	U	C	U	C	G	A	A	C	A	U	U	U	G	C	G	G	A	U	U	A	U	C	G	C
AJ428076	U	U	C	U	U	G	A	A	C	A	U	U	U	G	C	G	G	A	U	U	A	U	C	G	C
WNII3A_D	C	U	C	U	U	G	A	A	C	A	U	A	C	G	U	G	G	A	U	U	A	U	C	C	C
SNG1_D	C	U	C	U	U	G	A	A	C	A	U	A	C	G	C	G	G	A	U	U	A	U	C	C	C
TRS7_E	G	U	C	C	A	G	G	C	G	A	U	U	C	G	C	G	G	C	U	G	A	U	C	U	C
ZIP2_F	G	C	C	C	C	G	A	A	U	A	A	U	C	G	C	U	G	C	U	~	A	U	C	U	C
AN3A_G	G	U	G	A	C	U	U	A	C	A	A	U	C	G	C	G	G	C	U	G	A	U	C	U	C
IGH2_H	G	G	U	G	C	C	G	A	C	G	C	U	C	C	G	G	C	C	U	~	A	U	C	U	C

Table 5: List of NHS found for the genus *Spirogyra* and clades (continued)

Number *Spirogyra*		58	59	60		62	63	64	65				69	70	71	72	73	74	75			78	79	80	81				85	86		88			
Clade A					61					66	67	68								76													89	
Clade B																										83	84						90	
Clade C																																		
Clade D	57																														87			
Clade E																																		
Clade F																									82									91
Clade G																																		
Clade H																					77													
Position in	0	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
Alignment	9	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	4	4	4	4	4	4	5	5	5	5
	8	5	7	8	9	0	0	0	0	1	1	4	4	4	4	5	3	6	6	8	9	9	0	2	3	1	5	6	6	7	7	4	5	6	8
	1	1	9	5	2	0	1	2	4	6	8	0	1	4	5	7	9	1	9	9	1	3	4	3	3	9	4	0	2	4	7	2	1	3	4
Helix number	25	B27	27	J27-28	J27-28	J28-29	29	29	B29	29	29	B29	29	B29	29	J30-28	34	L35	J35-36	B37	B37	B37	L37	37	38	43	43	J43-44	44	44	J44-38	J34-45	B45	J45-46	46
without : helix, J: link, L: endloop, B: buldge																																		
D13324	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	G	G	U	U	G	C	G	G	U	A	U	C	U	U	G
X74000	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	G	G	U	U	G	C	G	G	U	A	U	C	U	U	G
AF115438	A	U	U	A	A	U	A	G	G	~	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
AF115437	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	G	U	C	U	C	A
AF115442	A	U	U	A	A	U	A	G	G	~	~	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
X91346	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	C	A
AF115440	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
X79496	A	U	U	A	A	U	A	G	G	~	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
X77452	A	U	U	A	A	U	A	G	G	~	~	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
X74752	A	U	U	A	A	U	A	G	G	~	~	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
AF115439	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
AM920378	A	U	U	A	A	U	A	G	G	~	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
X75763	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
X74753	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	~	A
X79495	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
X79497	A	U	U	A	A	U	A	G	G	~	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	A
AJ853450	A	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
AJ853451	A	U	U	A	A	U	A	G	G	G	U	C	A	U	C	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
AJ549234	G	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	G	U	C	U	U	G
AJ549233	G	U	U	A	A	U	A	G	G	G	G	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
AJ549232	G	U	U	A	A	U	A	G	G	G	U	C	A	U	U	A	C	A	A	G	A	G	U	U	G	C	G	A	C	G	U	C	U	U	G
X70705	A	U	U	A	A	U	A	G	G	G	U	C	A	U	C	A	C	A	A	G	A	G	U	U	G	C	G	A	U	A	U	C	U	U	G
KRA2_A	A	G	C	G	U	C	C	C	A	A	C	A	G	C	G	G	G	G	U	U	A	C	C	C	A	C	G	A	G	C	U	A	C	U	G
TIS1_A	A	G	C	G	U	C	C	C	A	A	C	A	G	C	G	G	G	G	U	U	A	C	C	C	A	C	G	A	G	C	U	A	C	U	G
WRH5_B	A	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	C	A	U	G	C	U	A	U	A	G
GTH1_B	A	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	C	A	U	G	C	U	A	U	A	G
MRT4A_C	A	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	G	G	G	G	C	U	A	U	C	G
TCA2_C	C	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	G	G	G	G	C	U	A	U	C	G
AJ428076	A	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	G	G	G	G	C	U	A	U	C	G
WNII3A_D	U	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	C	G	G	G	C	C	A	U	U	G
SNG1_D	U	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	C	G	G	G	C	C	A	U	U	G
TRS7_E	U	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	C	G	G	G	C	U	A	U	C	G
ZIP2_F	C	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	U	G	G	G	C	U	A	G	C	U
AN3A_G	C	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	G	C	C	C	A	C	G	G	G	C	U	A	G	C	G
IGH2_H	C	G	C	G	A	C	C	C	A	G	A	C	G	C	G	G	G	G	U	G	C	C	C	C	A	C	G	G	G	C	U	A	G	C	G

Table 5: List of NHS found for the genus *Spirogyra* and clades (continued)

Number *Spirogyra*		93	94		96	97	98	99	100	101			104			107		109
Clade A	92											103			106			
Clade B											102							
Clade C														105			108	
Clade D																		
Clade E																		
Clade F																		
Clade G				95														
Clade H																		
Position in Alignment	1619	1620	1622	1633	1635	1665	1675	1680	1684	1729	1738	1759	1760	1766	1772	1778	1789	1806
Helix number	47	47	L47	L47	47	48	48	48	J48-32	B49	B49	B49	49	B49	B49	L49	B49	49
without : helix, J: link, L: endloop, B: buldge																		
D13324	A	U	A	U	A	A	U	U	U	U	C	~	G	G	C	G	C	C
X74000	A	U	A	U	A	A	U	U	U	U	C	~	G	G	C	~	C	C
AF115438	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	U	C
AF115437	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	C	C	C
AF115442	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	U	C
X91346	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	C	A
AF115440	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	C	C
X79496	A	U	A	U	A	A	U	U	U	U	C	C	U	G	C	G	U	C
X77452	A	U	A	U	A	A	U	U	U	U	C	C	G	~	C	G	U	C
X74752	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	U	C
AF115439	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	U	C
AM920378	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	U	C
X75763	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	U	C
X74753	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	C	A
X79495	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	U	C
X79497	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	U	C
AJ853450	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	U	C
AJ853451	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	U	C
AJ549234	A	U	A	U	A	A	U	U	U	U	C	C	G	G	C	G	U	C
AJ549233	A	U	A	U	A	A	U	U	U	U	C	C	A	G	C	G	U	C
AJ549232	A	U	A	U	A	A	U	U	U	U	C	C	A	G	C	G	U	C
X70705	A	U	A	U	A	A	U	U	U	U	C	U	G	G	C	G	C	C
KRA2_A	C	C	C	U	G	C	G	C	C	A	C	C	C	G	A	A	G	U
TIS1_A	C	C	C	U	G	C	G	C	C	A	C	C	C	G	A	A	G	U
WRH5_B	A	C	C	U	G	C	G	C	C	A	U	C	C	G	C	A	G	U
GTH1_B	A	C	C	U	G	C	G	C	C	A	U	C	C	G	C	A	G	U
MRT4A_C	A	C	C	U	G	C	G	C	C	A	C	C	C	A	C	A	A	U
TCA2_C	A	C	C	U	G	C	G	C	C	A	C	C	C	A	C	A	A	U
AJ428076	A	C	C	U	G	C	G	~	C	A	C	C	C	G	C	A	A	U
WNII3A_D	A	C	C	U	G	C	G	C	C	A	C	C	C	G	C	A	G	U
SNG1_D	A	C	C	U	G	C	G	C	C	A	C	C	C	G	C	A	G	U
TRS7_E	A	C	C	U	G	C	G	C	C	A	C	C	C	G	C	A	G	U
ZIP2_F	A	C	C	C	G	C	G	C	C	A	C	C	C	G	C	A	G	U
AN3A_G	A	C	C	G	G	C	G	C	C	A	C	C	C	G	C	A	G	U
IGH2_H	A	C	C	C	G	C	G	C	C	A	C	C	C	G	C	A	C	U

Table 6: Origin of isolates

Abbreviation	Origin	Source	Accession No.	Clade	Intron	represented by	GPS Coordinates	
7074		ASW	JQ290232	E	-	7074	N.A.	N.A.
7075		ASW	JQ290233	D-E2	-	7075	N.A.	N.A.
7076		ASW	JQ239109	D	+	7076	N.A.	N.A.
7208		ASW	JQ239078	C	+	7208	N.A.	N.A.
071015-1e	Fish pond Franzen	AUT	JQ290230	F	-	BGHIII1	15° 23' 49"	48° 37' 07"
1020A1	Pond 1020, Vienna	AUT	JQ290231	D	+	TIS3	16° 21' 09"	48° 09' 21"
AB1	between Ahrensdorf & Behrensdorf	GER	JQ239111	D	+	AB4	14° 03' 32"	52° 10' 52"
AB4	between Ahrensdorf & Behrensdorf	GER	JQ239110	D	+	AB4	14° 03' 32"	52° 10' 52"
ADA1	Alte Donau	AUT	JQ290234	D-E2	-	7075	16° 24' 29"	48° 14' 58"
ADA2	Alte Donau	AUT	JQ290235	D-E2	-	7075	16° 24' 29"	48° 14' 58"
ADA6	Alte Donau	AUT	JQ290236	D-E2	-	7075	16° 24' 29"	48° 14' 58"
AN3A	Alte Naufahrt	AUT	JQ239122	G	-	AN3A	16° 28' 15"	48° 11' 46"
AN4B	Alte Naufahrt	AUT	JQ290237	H	-	IGH2	16° 28' 15"	48° 11' 46"
BEA1	Benda pond	AUT	JQ239098	D	+	KVW2	16° 20' 58"	48° 09' 11"
BGHC1	botanical garden Hamburg chinese pond	GER	JQ290238	H	-	GRS1	09° 51' 40"	53° 33' 52"
BGHII4	Botanical garden Hamburg near entrance	GER	JQ239049	C	+	REC2	09° 51' 37"	53° 33' 43"
BGHIII1	botanical garden Hamburg pond for halophilic plants	GER	JQ290239	F	-	BGHIII1	09° 51' 36"	53° 33' 47"
BGHIII2	botanical garden Hamburg pond for halophilic plants	GER	JQ290240	F	-	BGHIII1	09° 51' 36"	53° 33' 47"
BGHK1	Botanical garden Hamburg limestone alpine area	GER	JQ239112	D	+	BGHK1	09° 51' 36"	53° 33' 50"
BGHK2	Botanical garden Hamburg limestone alpine area	GER	JQ239113	D	+	NES3	09° 51' 36"	53° 33' 50"
BTSA1	Pond Suessenbrunn	AUT	JQ239106	D	+	VNE2	16° 29' 09"	48° 16' 36"
BTSA3	Pond Suessenbrunn	AUT	JQ290241	H	-	SOT1A	16° 29' 09"	48° 16' 36"
DECA3	Dechantlacke	AUT	JQ290242	D-E2	-	7075	16° 28' 34"	48° 11' 33"
DECA4	Dechantlacke	AUT	JQ290243	E	-	DECA4	16° 28' 34"	48° 11' 33"
DRS1	Dranser lake	GER	JQ290244	F	-	BGHIII1	12° 37' 53"	53° 10' 48"
DRS2	Dranser lake	GER	JQ290245	F	-	BGHIII1	12° 37' 53"	53° 10' 48"
DRS3	Dranser lake	GER	JQ239059	C	+	MRT4A	12° 37' 53"	53° 10' 48"
EGR4	drainage trench Reddelich	GER	JQ239065	B	+	EGR4	11° 49' 52"	54° 04' 38"
EIT4A	Eitzenberger pond	GER	JQ290246	F	-	EIT4A	11° 21' 50"	47° 47' 11"
FBT1A	Fuchsbodenteich	AUT	JQ239114	D	+	TIS3	15° 51' 45"	48° 21' 54"
FBT3A	Fuchsbodenteich	AUT	JQ290247	G	-	FBT3A	15° 51' 45"	48° 21' 54"
FIT3A	Filmteich	AUT	JQ239079	B	+	FIT3A	16° 24' 17"	48° 09' 01"
GBT1	gr Brunnerteich	AUT	JQ239064	B	+	GBT1	16° 18' 30"	48° 07' 00"
GBT3	gr Brunnerteich	AUT	JQ239092	B	+	GBT3	16° 18' 30"	48° 07' 00"
GRS1	Hamburg Großensee (pier)	GER	JQ239080	H	-	GRS1	10° 20' 51"	53° 36' 46"
GRS2	Hamburg Großensee (pier)	GER	JQ290248	H	-	GRS1	10° 20' 51"	53° 36' 46"
GSB2	gr. Segeberger Lake	AUT	JQ239118	D	+	GSB2	10° 20' 23"	53° 57' 27"
GTH1	garden pond Hamburg	GER	JQ239067	B	+	GTH1	10° 08' 44"	53° 37' 21"
GTH7	garden pond Hamburg	GER	JQ239068	B	+	GTH1	10° 08' 44"	53° 37' 21"
GUS1	Grubensee	AUT	JQ239096	D	+	KVW2	13° 59' 40"	52° 09' 30"
HRM4A	Rotmoos Lunz	AUT	JQ290249	E	-	HRM4A	N.A.	N.A.
IGH2	Industrial area Höltingbaum	GER	JQ239123	H	-	IGH2	10° 11' 01"	53° 36' 44"
IGH3	Industrial area Höltingbaum	GER	JQ290250	H	-	IGH2	10° 11' 01"	53° 36' 44"
IGH4	Industrial area Höltingbaum	GER	JQ290251	H	-	IGH2	10° 11' 01"	53° 36' 44"
KAT1A	Kastanienalleeteich	AUT	JQ290252	H	-	SOT1A	16° 20' 49"	48° 09' 26"
KRA2	watercourse between Krauswitz & Schlagnitz	GER	JQ239054	A	+	KRA2	13° 52' 50"	52° 01' 48"
KVW2	Pond near traffic circle A19 Wittstock	GER	JQ239094	D	+	KVW2	12° 27' 20"	53° 09' 26"
KVW4	Pond near traffic circle A19 Wittstock	GER	JQ239097	D	+	KVW2	12° 27' 20"	53° 09' 26"
KW1A	Kaiserwasser	AUT	JQ239089	B	+	KWA3	16° 25' 29"	48° 13' 54"
KW4A	Kaiserwasser	AUT	JQ239090	B	+	KWA3	16° 25' 29"	48° 13' 54"
KWA3	Kaiserwasser	AUT	JQ239087	B	+	KWA3	16° 25' 29"	48° 13' 54"
LHS4	Langhagensee	GER	JQ290253	F	-	WIND3A	12° 40' 31"	53° 14' 53"
LHS5	Langhagensee	GER	JQ290254	F	-	WIND3A	12° 40' 31"	53° 14' 53"

Table 6: Origin of isolates (continued)

Abbreviation	Origin	Source	Accession No.	Clade	Intron	represented by	GPS Coordinates	
LHS6	Langhagensee	GER	JQ290255	F	-	WIND3A	12° 40' 31"	53° 14' 53"
LHS8	Langhagensee	GER	JQ290256	B	+	LHS8	12° 40' 31"	53° 14' 53"
LHS9	Langhagensee	GER	JQ290257	F	-	WIND3A	12° 40' 31"	53° 14' 53"
LL1	local fire service pond Limsdorf	GER	JQ239056	C	+	MRT4A	14° 00' 51"	52° 09' 34"
LL2	local fire service pond Limsdorf	GER	JQ239057	C	+	MRT4A	14° 00' 51"	52° 09' 34"
LSB4A	Langenschoenbichl	AUT	JQ239091	B	+	KWA3	15° 59' 45"	48° 19' 52"
LUA1	Lunzer Untersee outflow	AUT	JQ290258	F	-	WIND3A	15° 02' 25"	47° 51' 12"
LUBS1	Lunzer lake, stone near boathouse	AUT	JQ290259	E	-	DECA4	15° 03' 41"	47° 51' 16"
M3A	St. Pölten	AUT	JQ239048	C	+	M3A	N.A.	N.A.
M5A	St. Pölten	AUT	JQ239075	B	+	M6A	N.A.	N.A.
M6A	St. Pölten	AUT	JQ239076	B	+	M6A	N.A.	N.A.
M9A	St. Pölten	AUT	JQ239062	C	+	MRT4A	N.A.	N.A.
M9B	St. Pölten	AUT	JQ239074	B	+	M6A	N.A.	N.A.
MD2-1	St. Pölten	AUT	JQ239060	C	+	MRT4A	N.A.	N.A.
MDA1	St. Pölten	AUT	JQ239061	C	+	MRT4A	N.A.	N.A.
MIL10A	Millstaetter Lake	AUT	JQ290260	H	-	GRS1	13° 31' 10"	46° 48' 51"
MP9-1	St. Pölten	AUT	JQ239069	B	+	M6A	N.A.	N.A.
MP92-3	St. Pölten	AUT	JQ239073	B	+	M6A	N.A.	N.A.
MP92-6	St. Pölten	AUT	JQ239072	B	+	M6A	N.A.	N.A.
MRT4A	Mausrodlteich	AUT	JQ239058	C	+	MRT4A	15° 02' 52"	47° 52' 25"
NES3	Nebelsee	GER	JQ239119	D	+	NES3	12° 39' 58"	53° 14' 54"
NES4	Nebelsee	GER	JQ239120	D	+	NES3	12° 39' 58"	53° 14' 54"
NLH5	watercourse between Neu-Lübbenau & Hohenbrück	GER	JQ239084	B	+	NLH5	13° 53' 38"	52° 04' 57"
NLH6	watercourse between Neu-Lübbenau & Hohenbrück	GER	JQ239063	B	+	NLH6	13° 53' 38"	52° 04' 57"
OROSIVA1	Osor	HR	JQ239104	D	+	SNG1	14° 23' 87"	44° 41,732'
OSS5A	Ossiacher lake	AUT	JQ290283	H	-	OSS5A	13° 58' 51"	46° 40' 37"
PANA1	Panozzalacke	AUT	JQ239085	B	+	PANA1	16° 29' 16"	48° 10' 51"
PLO1	Plötinsee, Rostock	GER	JQ239115	D	+	TIS3	12° 58' 22"	53° 13' 14"
REC2	Meadow pond Rechlin	GER	JQ239051	C	+	REC2	12° 45' 45"	53° 18' 43"
REC3	Meadow pond Rechlin	GER	JQ239105	D	+	VNE2	12° 45' 45"	53° 18' 43"
ROE2	Pond near Rödel	GER	JQ290261	F	-	EIT4A	12° 35' 51"	53° 23' 06"
ROE6	Pond near Rödel	GER	JQ239093	B	+	ROE6	12° 35' 51"	53° 23' 06"
S1DD3A	St. Pölten	AUT	JQ239070	B	+	M6A	N.A.	N.A.
S1Due2-2A	St. Pölten	AUT	JQ239071	B	+	M6A	N.A.	N.A.
S1Due2A	St. Pölten	AUT	JQ239077	B	+	M6A	N.A.	N.A.
SENA4	Sengsee	GER	JQ239082	B	+	SENA7	11° 18' 58"	47° 46' 37"
SENA7	Sengsee	GER	JQ239081	B	+	SENA7	11° 18' 58"	47° 46' 37"
SHS6A	Schilfhuettensee	GER	JQ290262	F	-	WIND3A	11° 19' 02"	47° 46' 26"
SKA1	Kritzendorf	AUT	JQ290263	F	-	WIND3A	16° 18' 48"	48° 20' 06"
SKA3	Kritzendorf	AUT	JQ290264	F	-	WIND3A	16° 18' 48"	48° 20' 06"
SKR1	pond near Kröpelin, Rostock	GER	JQ290265	F	-	WIND3A	11° 48' 45"	54° 04' 25"
SKR2	pond near Kröpelin, Rostock	GER	JQ290266	F	-	WIND3A	11° 48' 45"	54° 04' 25"
SMS2	Scharmützelsee	GER	JQ239066	B	+	SMS2	14° 00' 50"	52° 12' 49"
SNG1	Marsh north Großensee	GER	JQ239101	D	+	SNG1	10° 21' 27"	53° 37' 51"
SNG2	Marsh north Großensee	GER	JQ239100	D	+	SNG1	10° 21' 27"	53° 37' 51"
SNG3	Marsh north Großensee	GER	JQ239103	D	+	SNG1	10° 21' 27"	53° 37' 51"
SNG5	Marsh north Großensee	GER	JQ239102	D	+	SNG1	10° 21' 27"	53° 37' 51"
SOT1A	Stierofenteich	AUT	JQ290267	H	-	SOT1A	16° 20' 57"	48° 09' 23"
SOTA1	Stierofenteich	AUT	JQ290268	H	-	SOT1A	16° 20' 57"	48° 09' 23"
SPS3	Springsee	GER	JQ290269	D-E2	-	7075	13° 59' 19"	52° 10' 20"
TCA2	Canow fish farming	GER	JQ239050	C	+	REC2	12° 53' 23"	53° 11' 32"
TIS1	Tiefensee	GER	JQ239052	A	+	TIS1	13° 59' 22"	52° 09' 23"
TIS3	Tiefensee	GER	JQ239117	D	+	TIS3	13° 59' 22"	52° 09' 23"

Table 6: Origin of isolates (continued)

Abbreviation	Origin	Source	Accession No.	Clade	Intron	represented by	GPS Coordinates	
TIS4	Tiefensee	GER	JQ239116	D	+	TIS3	13° 59' 22"	52° 09' 23"
TOR4	Torsee	GER	JQ239083	B	+	TOR4	13° 39' 56"	52° 14' 24"
TRS7	Tristacher Lake	AUT	JQ239121	E	-	TRS7	12° 47' 40"	46° 48' 29"
UMW4A	Lower Muehlwasser	AUT	JQ239055	A	+	UMW4A	16° 27' 20"	48° 12' 48"
UTEX LB 1984		UTEX	JQ239053	C	+	UTEX1984	N.A.	N.A.
UTEX1273 Sp. crassispina		UTEX	JQ290270	H	-	UTEX1273	N.A.	N.A.
UTEX1683 Sp. occidentalis		UTEX	JQ290271	H	-	UTEX1683	N.A.	N.A.
UTEX1742 Sp. juergensis		UTEX	JQ290272	D-E1	-	UTEX1742	N.A.	N.A.
UTEX1745 Sp. liana		UTEX	JQ290273	E-F	-	UTEX1745	N.A.	N.A.
UTEX1746 Sp. pratensis		UTEX	JQ290275	G	-	UTEX1746	N.A.	N.A.
UTEX2495 Sp. maxima		UTEX	JQ290274	C	+	UTEX2495	N.A.	N.A.
VNE2	accretion zone Neversdorfer lake	GER	JQ239107	D	+	VNE2	10° 15' 31"	53° 51' 56"
W4A	bog-drain, Windtal	I	JQ290276	F	-	WIND3A	N.A.	N.A.
WB5A	Wienerberg pond2	AUT	JQ239099	D	+	KVW2	16° 20' 49"	48° 09' 43"
WBA2	rivulet, Wienerberg	AUT	JQ290277	H	-	WBA2	16° 21' 06"	48° 09' 29"
WBA3	rivulet, Wienerberg	AUT	JQ290278	H	-	IGH2	16° 21' 06"	48° 09' 29"
WBT2A	Wienerberg pond2	AUT	JQ290279	H	-	GRS1	16° 20' 49"	48° 09' 43"
WIND3A	bog-drain, Windtal	I	JQ290280	F	-	WIND3A	N.A.	N.A.
WNII3A	local fire service pond Gerasdorf	AUT	JQ239108	D	+	WNII3A	16° 27' 58"	48° 18' 00"
WNII4A	local fire service pond Gerasdorf	AUT	JQ239086	B	+	PANA1	16° 27' 58"	48° 18' 00"
WNIII2A	Pond recreation area Seeschlacht	AUT	JQ239095	D	+	KVW2	16° 21' 49"	48° 17' 56"
WOE4A	western pond near oilharbor	AUT	JQ290281	D-E2	-	7075	16° 31' 07"	48° 10' 00"
WRH5	Wandse detention reservoir	GER	JQ239088	B	+	KWA3	10° 11' 07"	53° 37' 17"
ZIP1	Zipke	GER	JQ290282	F	-	ZIP2	12° 47'02"	54° 20' 30"
ZIP2	Zipke	GER	JQ239124	F	-	ZIP2	12° 47'02"	54° 20' 30"

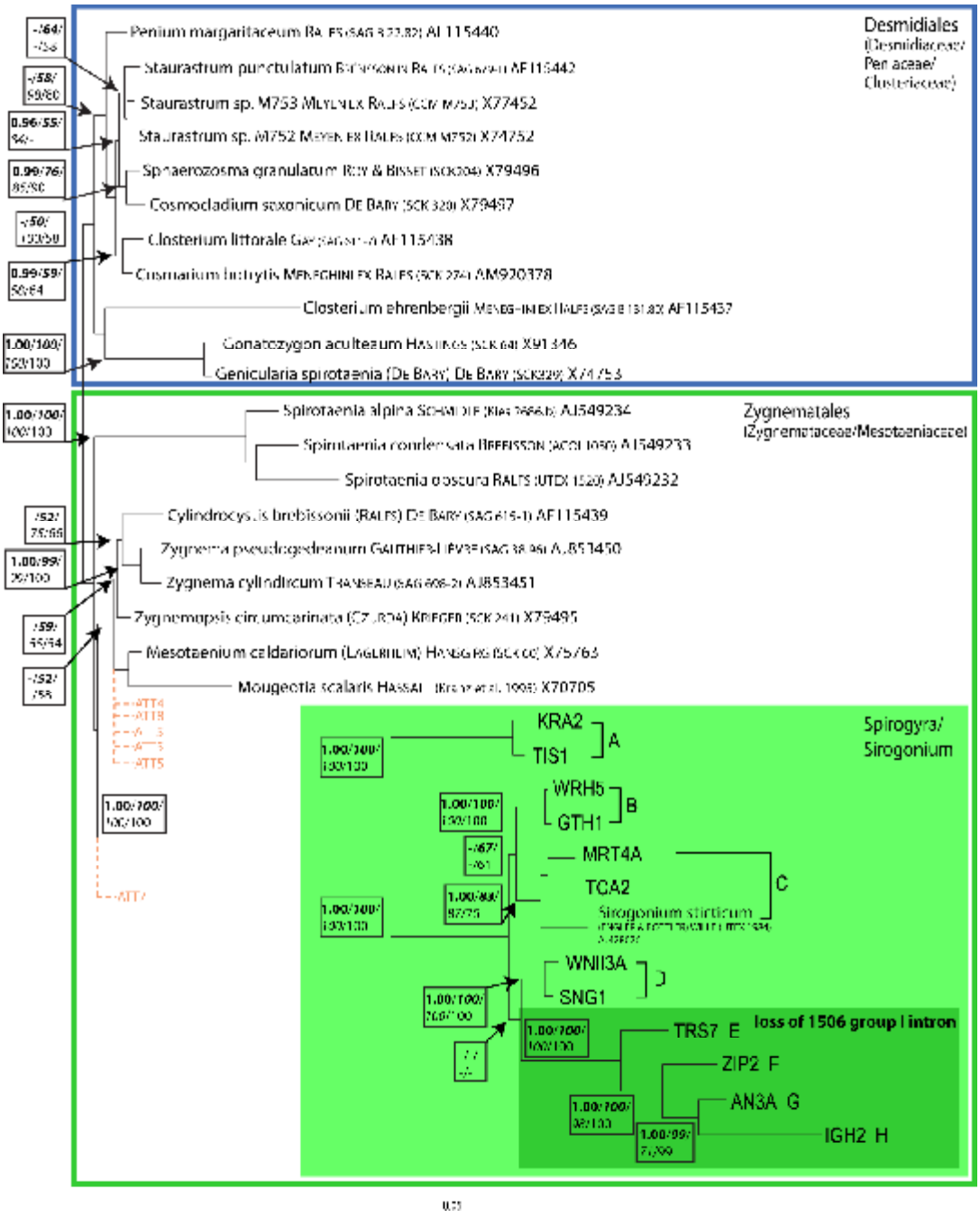


Figure 1: Combined Zygnematophyceae and *Spirogyra* SSU rRNA phylogeny: Molecular phylogeny of Desmidiaceae, Peniaceae, Closteriaceae, Mesotaeniaceae and Zygnemataceae based on SSU alignment. The phylogenetic tree was inferred by maximum likelihood analyses of 1720 aligned positions of 33 taxa using PAUP* 4.0b10. TrN+G+I was chosen as best evolutionary model (base frequencies: A 0.25, C 0.23, G 0.27, T 0.25; rate matrix: A-C 1.0000, A-G 1.8721, A-T 1.0000, C-G 1.0000, C-T 4.5252, G-T 1.0000) with the proportion of invariable sites (I= 0.4608) and gamma distribution parameter (G= 0.6376). Posterior Probabilities (>95%; bold; calculated by MrBayes 3.1.2 using the covariation settings (3 million generations, trees from 4100 – 30000)) as well as bootstrap values (>50%) of the maximum likelihood (100 replicates; bold italic), neighbor-joining (1000 replicates; italic), and maximum parsimony (1000 replicates; not bold) are given in the tree (PP/*ML*/*NJ*/MP). No outgroup was used. ATT4, 5, 7 and 8 refer to alternative tree topologies tested with consel – please refer to Table 1 for details.

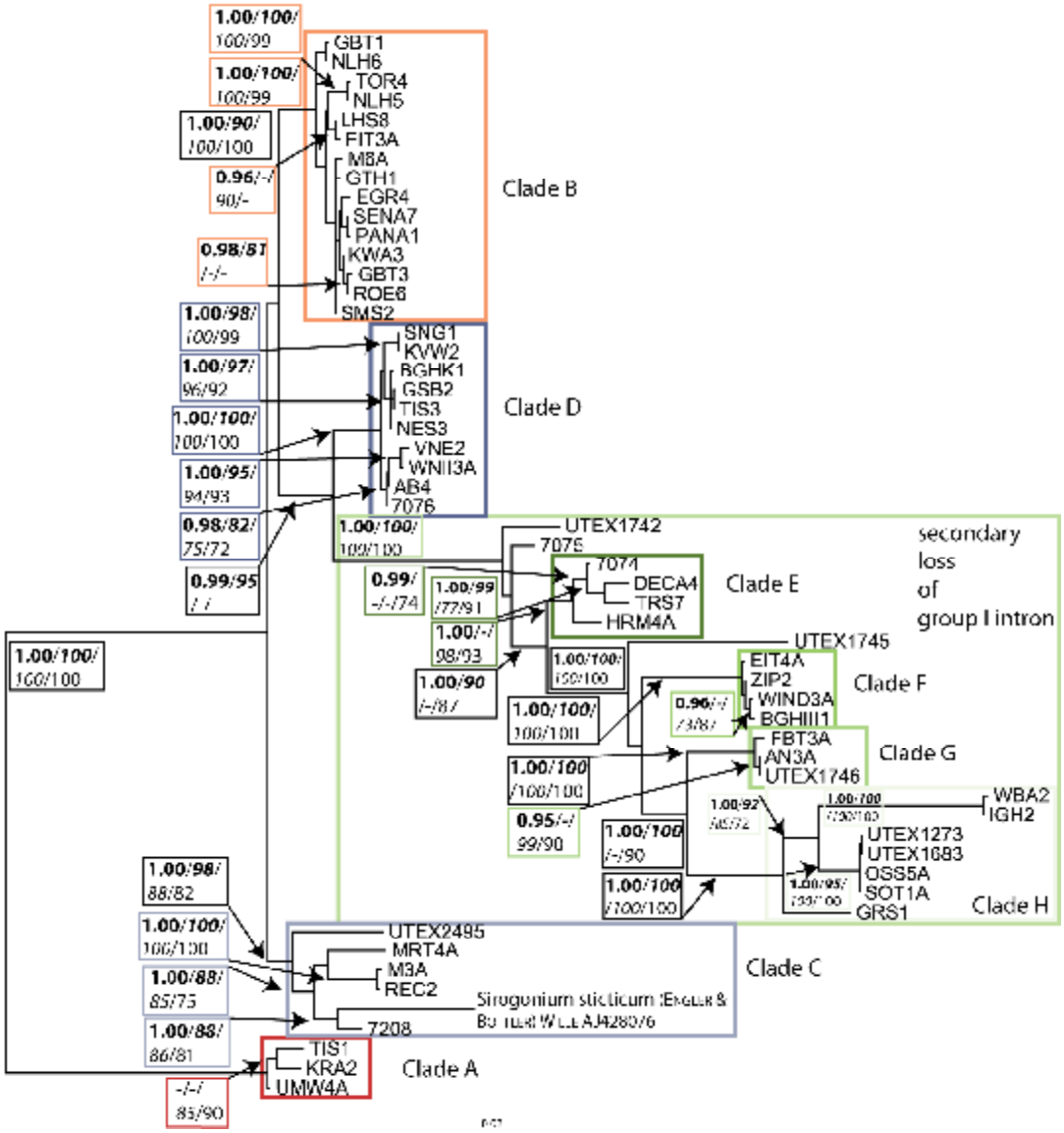


Figure 2: Spirogyra SSU rRNA phylogeny: Unrooted molecular phylogeny of Spirogyra based on SSU alignment. Individual clades are highlighted in white boxes; group with secondary loss of group I IC1 intron highlighted by light green-box. The phylogenetic tree was inferred by maximum likelihood analyses of 1645 aligned positions of 55 taxa using PAUP* 4.0b10. GTR+G+I was chosen as best evolutionary model (base frequencies: A 0.23, C 0.24, G 0.29, T 0.24; rate matrix: A-C 1.4341, A-G 2.6641, A-T 1.2357, C-G 1.6993, C-T 5.2526, G-T 1.0000) with the proportion of invariable sites (I= 0.6009) and gamma distribution parameter (G= 0.6856). Posterior Probabilities (>95%; bold; calculated by MrBayes 3.1.2 using the covariation settings (2 million generations, trees from 11070 – 20000)) as well as bootstrap values (>50%) of the maximum likelihood (100 replicates; bold italic), neighbor-joining (1000 replicates; italic), and maximum parsimony (1000 replicates; not bold) are given in the tree (PP/ML/NJ/MP). No outgroup was used, tree was rooted using clade A.

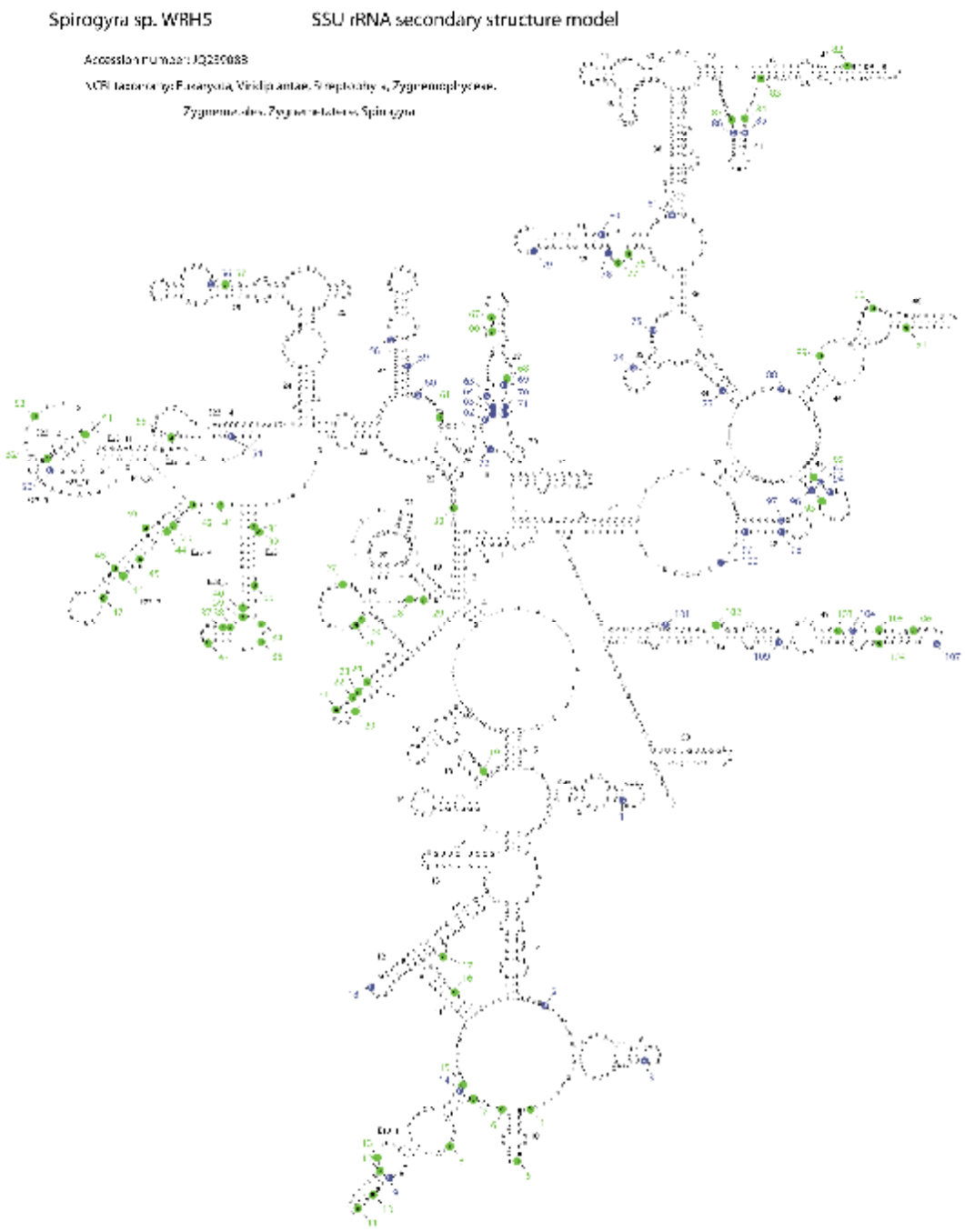


Figure 3: Putative secondary structure model of the nuclear encoded SSU rRNA of Spirogyra sp.; Strain WRH5 (Acc.No. JQ 239088) from Clade B was chosen as representative sequence; 100% consensus bases for the genus Spirogyra are given in upper case letters, variable positions in lower case letters; NHS for Spirogyra are marked with blue circles, NHS for one of the clades are marked by green circles, the assigned number indicates the number in the table; see Table 5 for details.

Chapter 3

Combining vegetative morphology, genetics and environment – a polyphasic approach on the complex relationships of *Spirogyra* (Zygnematophyceae, Streptophyta)

Spirogyra sp.

Combining vegetative morphology, genetics and environment – a polyphasic approach on the complex relationships of *Spirogyra* (Zygnematophyceae, Streptophyta)

Chen Charlotte[1], Coleman Annette[2], Hainz Roland[3], Pröschold Thomas[1], Schagerl Michael[1]

[1] Department of Limnology & Oceanography, University of Vienna, Althanstraße 14, Vienna A-1090, Austria
[2] Division of Biology and Medicine, Brown University, Providence, RI 02912, USA
[3] DWS Hydro-Ökologie GmbH, Zentagasse 47/5, A-1050 Vienna, Austria

Correspondence: Michael Schagerl, E-mail: michael.schagerl@univie.ac.at

Running title: Combining vegetative morphology, genetics and environment

Abstract

For species identification of the filamentous green alga *Spirogyra*, the complete reproductive cycle including ripe hypnozygotes are essential, which are found only sporadically. Moreover, its traditional species concept is based on morphological traits, obviously not reflecting phylogenetic relationships. These problems call for alternative approaches to elucidate systematics of this taxon. We combined SSU and ITS2 sequences and its secondary structure to uncover phylogenetic relationships within *Spirogyra.* The genus was split into eight clades, each receiving very high support by all algorithms applied. Out of 112 sequences considered, 68 were found only once and 23 were found at least twice. ITS2 barcoding was used to estimate the species number of our data set; 51 sequences that differed by at least one compensatory base change were considered putative different species. The ITS2 secondary structure was analyzed and showed the typical hallmarks, but no consensus structure could be found due to large variation of sequences. The usually conserved region on the 5' end of helix III is less conserved in this genus, but still recognizable. Some strains had a helix IV, some an extra arm in helix III.
Vegetative morphology, clade grouping and information on clone grouping, respectively, were analyzed statistically to find patterns. Although with only weak support, some correlations were found which also point at the hidden diversity of *Spirogyra*.

Keywords
Conjugatophyceae, Zygnematales, barcode, ITS2, SSU, CBC, secondary structure

Introduction

The species concept traditionally applied to the genus *Spirogyra* is based on morphology, which neither reflects phylogenetic relationships nor gives insight to the genus diversity (Chen *et al.* 2012; Chen & Schagerl 2012). As testing the interbreeding capability of *Spirogyra* strains resulted in only low success (Czurda 1930; Czurda 1933; Simons *et al.* 1982; Simons *et al.* 1984; Zwirn 2010), other paths have to be taken to get insight into the hidden diversity of this genus.
Many organisms only rarely reproduce sexually, so that they are effectively asexual or clonal (Birky & Barraclough 2009). Furthermore, not all criteria for identifying species are applicable in all situations and applying multiple species concepts will give conflicting results (i.e. concerning species number) (Hey 2006). Also within *Spirogyra,* we struggle with harmonizing the given biological settings and

conventional taxonomy; finding a species concept, which covers the issue of what a species is and how to identify a species is therefore one of the key questions for this genus.

Spirogyra is widely distributed in almost all freshwater habitats, often produce a high biomass and play therefore an important role in the ecology of those water bodies. The identification of Spirogyra at generic level is very easy because of spiral chloroplasts. However, the morphological features such as the formation and ornamentation of zygospores do not often occur in natural samples. The traditional species concept in Spirogyra is solely based on those features. More than 350 species have been described within this genus (Kadlubowska 1984). To identify the organism of this genus at species level require therefore an alternative methods such as molecular phylogeny and DNA Barcoding.

Spirogyra was shown to have a fast mutation rate (Chen & Schagerl 2012; Gontcharov *et al.* 2002; Gontcharov *et al.* 2004), speciation on molecular level is presumably faster than on morphological level. Furthermore, many morphological traits are likely to have evolved in parallel in different lineages within the genus (Stancheva *et al.* 2013), which makes it difficult to define and to identify species correctly in *Spirogyra*. According to De Queiroz (2007), relying on single character differences to detect the presence of another species is problematic, because the presence of a single property does not guarantee that a set of populations possessing that property represents separate species but only that the presence of a single property constitutes evidence supporting that hypothesis.

To find the common denominator of an applicable species concept and to combine the morphological and phylogenetic approach, we added barcoding as a tool. A possible pitfall to the term barcode is the ambiguous useage of the term “barcoding” and the dual use of several techniques in both identification and classification (Ross *et al.* 2008). The term “barcoding” has been applied widely to include the identification, delimitation, and description of species by the use of genetic distance comparisons, hierarchical clustering, distance thresholds, and other methods based on single gene comparisons (Ross *et al.* 2008). The prerequisites for a useful barcode can be outlined as follows: it possesses genetic interspecific distances exceeding intraspecific distances (monophyletic clusters of representative barcodes for a species); it consists of a sequence length obtainable in a single amplification and it contains conserved flanking fragments so facilitate the design of universal primers (Coleman 2009; Moniz & Kaczmarska 2010; Yao *et al.* 2010).

The internal transcribed spacer 2 (ITS2) has been widely used for phylogenetic analyses at species level (Mueller *et al.* 2007; Selig *et al.* 2008), but serves also as a target molecule for barcoding (Demchenko *et al.* 2010). This region can be used for barcoding without cloning because the sequences appear to homogenize sufficiently rapidly to behave as a single copy gene (Fabry *et al.* 1999). Despite variability in ITS2 sequence and length, its rRNA secondary transcript structure contains a highly conserved core (Coleman, 2003; Schultz et al., 2005). Because ITS2 is a noncoding fragment, aligning proves to be more difficult based on sequences only, so more recent works suggest the inclusion of the secondary structure (Coleman 2000; Coleman 2003; Coleman 2007; Coleman & Mai 1997; Mueller *et al.* 2007; Schultz *et al.* 2006). The use of both the ITS2 secondary structure and the barcoding region permits an independent estimate of the phylogenetic relationships of close and distantly related organisms in one study while keeping branching very similar to phylogenies obtained from other markers (Coleman & Vacquier 2002).

The correlation between ITS2-derived subclades and mating compatibilities was assessed by many authors (Coleman & Vacquier 2002; Fabry *et al.* 1999); compensatory base changes (CBC) were used as species delimiter. CBCs are mutations that occur in both nucleotides of a paired structural position while retaining the paired nucleotide bond (Coleman & Vacquier 2002; Gutell *et al.* 1994; Ruhl *et al.* 2010). The main argument for using CBCs as species delimiter is that once a CBC has appeared in the ITS2 of a group of related organisms, their mating genes will also have diverged to the point where two subgroups no longer interact at the gamete level (Coleman 2000). Many studies revealed that a CBC in the conserved parts of the ITS2 secondary structure between two organisms correlated with sexual incompatibility (Mueller *et al.* 2007). Coleman also stated that if strains differ by even a one-sided CBC among the relatively conserved pairing positions, their gametes will probably not interact at all (Coleman 2000). A hemi-CBC (hCBC) is a mutation of a single nucleotide in a paired structural position while maintaining the nucleotide bond. To date, there is no example known of an exception

to this rule, while identity for the entire ITS2 correlates with significant interbreeding potential (Coleman 2009). A CBC is rather a measure of elapsed evolutionary time, indicating that sufficient time has passed to make a speciation event very likely and is not an indispensable criterion (Coleman & Vacquier 2002; Mueller *et al.* 2007), which was tested on the DNA sequences for two gene products that interact at a major step in gamete recognition at fertilization of the gastropod *Haliotis* (Swanson & Vacquier 1998).

Our study is inspired by a common scientific wish for an ideal taxonomic world, where one morphology equals one interbreeding group equals one species (Coleman 2009). To address the issue of species recognition in *Spirogyra*, we utilized the ITS2 barcoding region alongside with the accordant ITS2 secondary structure and SSU. We wanted to find an ITS2 secondary structure model representing *Spirogyra*, or for every clade, respectively. With the aid of barcodes, we analyzed the clades for species numbers. We tried to gain insights if vegetative morphology is reflected in phylogeny and if these morphological traits would be of equal importance. Also the correlations between environmental parameters and phylogeny were investigated to reveal possible patterns.

Material and Methods

Algal Material, environmental data, DNA extraction, amplification and sequencing

Water samples and environmental data were sampled and analyzed as described by Hainz et al (Hainz *et al.* 2009). Sampling and isolation of algal material, DNA extraction, amplification and sequencing were carried out as described in Chen et al. (2012).

Primers used in this study were: 1380F (GCGTTGAWTACGTCCCTGCC), ITS-GF (GGGATCCGTTTCCGTAGGTGAACCTGC), ITS-GR (GGGATCCATATGCTTAAGTTCAGCGGGT), ITS055R (CTCCTTGGTCCGTGTTTCAAGACGGG), 5.8sF (CGATGGATATCTTGGYTCTCGCAACG) and 5.8sR (GGACGCTCAGCCAGGCATACTCCGC. Because only few strains could be identified, cultures were labeled with a code for the accordant sampling site and date (see Table S2). Additional strains were obtained from the UTEX culture collection.

Sequence Alignment and Phylogenetic Analysis

Two different data sets were analyzed: 1) a concatenated SSU and barcode alignment of 120 strains and 2) complete ITS2 of 112 strains. Identical sequences were removed before the analyses in order to reduce computational effort. Sequences were tested for the 5.8S–28S rRNA gene interaction to determine if the complete ITS2 was present. Only sequences with a complete ITS2 were included in this study. For 1) the concatenated data set, a barcode was extracted from an ITS2 alignment, using the basal 6 base pairs of helix I, the basal 12 bp of helix II (including the mismatch) and all paired bases of helix III (Demchenko *et al.* 2010; Luo *et al.* 2010; Yao *et al.* 2010). Sequences were aligned manually with consideration of the SSU rDNA secondary structure. The alignment was refined by comparison of the secondary structure of the sequences. Secondary structure was determined using the program mfold (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form ;(Zuker 2003)). SSU sequences were used from Chen et al. (2012); accession numbers for accordant sequences are given there. Only unambiguously aligned regions of the sequences were used for analyses. Gap-rich regions were excluded. The concatenated SSU/barcode dataset consisted of 1747 unambiguously aligned bases. The phylogenetic tree presented in Figure 1 was inferred by maximum likelihood settings using PAUP* 4.0b10 (Swofford 2003), the best models was chosen according to the Akaike Information Criterion by Modeltest. To test for the best evolutionary model for the analyses, the log.likelihood values of 56 models using Modeltest 3.7 (Posada & Buckley 2004; Posada & Crandall 1998) were compared. For the concatenated SSU/barcode alignment analyses the TrN+I+G model was chosen. The robustness of the trees was assessed by bootstrap support values calculated by neighbor joining (NJ; 1000 replicates), maximum parsimony (MP; 1000 replicates) and maximum likelihood (ML; 100 replicates) methods using the accordant evolutionary model for each dataset. Bayesian inference (PP) was calculated (using following settings: 2.5 million generations; sampling every 100 generations; MCMC chains = 4; all trees below the burnin value of 0.01 were discarded as burnin, the consensus tree was calculated using PAUP*) using MrBayes 3.1.2. (Huelsenbeck &

Ronquist 2001; Ronquist & Huelsenbeck 2003). Not significant values were not included in figures (PP less than 0.95, NJ, ML, MP less than 50%).
The ITS2 alignment was treated as suggested in Merget & Wolf (2010) and Schultz & Wolf (Merget & Wolf 2010; 2009). In a first step, secondary structures were determined as described before and transcribed to an extended .fasta (in bracket dot bracket format). This was used as input for 4SALE 1.7 (Seibel *et al.* 2008; Seibel *et al.* 2006). Aligning was done automatically using the integrated ClustalW (Thompson *et al.* 1994). The CBC Matrix feature in 4SALE was used to calculate a CBC table. The resulting alignment was translated to a one-letter encoded alignment, translating the 4 nucleotides in three different structural stages (unpaired, paired left, paired right) to a 12 letter alphabet. The alignment output file was imported to ProfDist 0.9.9 (Qt-Version) (Friedrich *et al.* 2005) and processed as described in Ruhl et al. (2010). Based on the one letter encoded alignment a distance tree was calculated using an ITS2 specific General Time Reversible Substitution Model (GTR).

Number of CBCs/pairwise differences of accessions (Tab. 2) were obtained by analyzing the ITS2 alignment used for Fig. 2. For better readability, identical ITS2 sequences are represented by just one sequence. CBCs between sequences in the upper right part of Table 2 were calculated using 4SALE; the pairwise distances in lower left part were calculated using PAUP. Clades were marked with a black outline. A value of "0" in both the CBC and the pairwise distances part of the table indicated identical ITS2 sequences; the SSU sequence of the assessed strain could however still be unequal.

Morphological and environmental data

Morphological data was obtained from healthy growing cultures as described in Hainz et al. (2009). To detect correlations between phylogenetic data and vegetative morphology, a mantel test was carried out in R (Team 2008). Two matrices were used: a) the concatenated SSU& barcode data set (same as for phylogenetic analyses) and b) a set of vegetative morphology parameters (end wall structure; cell width average and median; cell length average; minimum, maximum and median number of chloroplasts; chloroplast periodicity, length:width ratio of the cells).
Statistics linking the phylogenetic tree with morphology data and environmental variables were performed with the statistical package Primer V6.1.15 (Clarke & Gorley 2006). For comparison of community similarities, non-metric multidimensional-scaling (NMDS) was based on the distance matrix obtained from PAUP. NMDS resulted in a 2-dimensional solution of 0.04 minimum stress (Kruskal stress value; 1000 iterations). The taxon pattern along artificial before the to calculations. LINKTREE routine was applied to seek for distinct vegetative morphological features of groups, which are responsible for dichotomous group separation (999 permutations, minimum group size = 1, minimum split size = 4; minimum split r = 0.2 p = 0.05). We did not search for correlations between vegetative morphology and environmental data, because the morphology used in this study was derived from cultivated material that could have experienced morphological changes due to culture conditions that differ from the natural habitat. The mantel test in R focused on the quality of the correlations between phylogeny and vegetative morphology; the analyses done in PRIMER aimed to reveal correlations between environmental data and phylogeny, and find morphological characters that could be used for separating clades.

Results

Phylogenetic analyses

The analyses of the concatenated SSU& barcode data set (Fig.1) showed a division of the 120 strains into eight clades A-H. Clades were very diverse – they comprised unequal numbers of strains – some were small (clades A, E, G, and H), some very large (clades B, and D). Also, the branch lengths within clades differed – clade A, C and H exhibited the longest branches, whereas clades D and F had very short branches, indicating closely related strains. The bootstrap and Bayesian support for each clade branching was very high for all clades with the exception of those of clade C that received slightly lower support (PP/ML/NJ/MP – 0.97/91/80/78). Clade C was positioned as an early branching clade

along with clade A; they showed relatively large distances to the remaining clades which indicates big genetic differences. Strains without the 1506 group I intron are grouped together in clades E-H, including the sequences between clade D and E.
Only very few strains could be determined at species level. The strain WRH5, positioned in clade B, was identified as *Spirogyra velata* (Nordstedt) and used to depict secondary structure of ITS2 (Fig. 6). Clade C contained *Sirogonium sticticum* ((Smith) Kützing; UTEX 1985) and *S. maxima* ((Hassall) Wittrock; UTEX 2495). *S. maxima* was positioned on an early diverging branch in relation to the remaining members of this clade. Strain WOE4 was identified as *S. subsalsa* (Kützing) and placed together with *S. juergensii* (Kützing; UTEX 1742) basal to the group showing the secondary loss of the group I intron. *S. liana* (Transeau; UTEX 1745) was positioned between clade E and the remaining clades. Strain ZIP2 was identified as *S. polymorpha* (Kirchner) and placed in clade F. *S. pratensis* (Transeau; UTEX 1746) was located in one of the most derived branches in clade G. Due to sequencing problems in the SSU region, *S. varians* ((Hassall) Kützing; UTEX 479), *S. gracilis* (Kützing; UTEX 1743) and *S. communis* ((Hassall) Kützing; UTEX 2466) had to be excluded from the data set described above. These strains were however included in the analyses of the ITS alignment (see below) and were placed in clade D.

The ITS2 alignment analyzed with ProfDist (Fig. 2) showed a division into several clades – clade A was the same as in the concatenated data set, clade B, D, F and G were also completely recovered. Clade C of the concatenated data set was split into two clades – one with high support located near clade A (containing two individual clones, one of them *S. maxima*), the other one with slightly lower support near clade B (containing 6 strains including *Sirogonium sticticum*). Strains forming clades E and H in the concatenated data set (see above) were regrouped. Another important difference between the two phylogenies is that the strains without group I intron form a distinct clade in the SSU & Barcode alignment, whereas they do not form a branch together in the ITS2 alignment.

Sequence identity
120 sequences were used in this study; 68 sequences were individual both for the SSU and ITS2, 23 were found more than once.
Each clade holds a number of barcodes (Tab. 1), representing a putative number of species. Consdering these numbers, we estimated a total number of 51 species covered with 116 sequences included in the ITS2 alignment (including the UTEX strains). The most species rich clade was clade B with 16 individual barcodes/sequences. Table 1 was split into two parts – (1) – SSU & barcode: the concatenated dataset used for Fig. 1; (2) – ITS2: the dataset used for Fig. 2. For (1) 112 accessions were used, some differed in the SSU region, but not in the barcode; a few accessions had to be excluded due to sequencing problems in the SSU; in (2) 116 accessions were used, not all clades from SSU & barcode phylogeny were recovered, members of clade H were merged with members of clade B.

Morphological data
of the morphology of the members belonging to the different clades are shown in Fig 3. Boxplots were used to describe morphological traits of each clade; the tree structure on the left hand side is based on SSU & barcode phylogeny. Some clades could be easily distinguished from the others – clade A contained strains with high chloroplast numbers (3 – 9); clade C comprised strains with very large cell widths (27 – 88µm) and a medium number of chloroplasts (1 – 6). Chloroplast periodicity of clade C (i.e. the longitudinal distance that is needed by a chloroplast to make a single turn in the cell) was comparable to values of clade A and H (data not shown) although *Sirogonium sticticum* and *S. maxima* were placed in this clade; clade H contained members with 2 – 5 chloroplasts and a median cell width of 34 – 44µm. Clade B comprised strains with both replicate and plane cross walls; it also had cells with a high cell length:width ratio (i.e. cells were usually longer). The remaining clades D, E, DE, F and G) could not be distinguished from each other based on their vegetative morphology. The parameters showed large overlaps; also the presence or absence of rhizoids followed no general pattern. Descriptions should however be treated with caution as some clades contain considerably

less samples than others (e.g. clade A vs. clade B). Small clades therefore usually showed less variation due to the smaller amount of data sets included. Interestingly, clade A was an exception, as 3 very differing strains were grouped together covering a wide scale of cell width (26 – 78µm) and median chloroplast number (3 – 9).

Morphological data of clones

17 clones were identified in our data set, meaning that the respective sequences were represented by more than one accession. Clones were named arbitrarily with letters and compared concerning their vegetative morphology. 2 clones showed only replicate cross walls (ay and o; both n=2), other accessions with replicate cross walls were single sequences. The remaining clones had only plane cross walls.

Cell length median varied considerably within clones, as it is dependent on cell growth and division stages; the highest difference between longest and shortest was 253.2µm (clone o), the shortest 4.4µm (clone ai); cell widths were not as variable, only 3 clones had large min-max differences: clone x 60,7µm; clone aa 36.1µm; clone a 15.2µm. Most clones contained 1 – 2 chloroplasts, clone aa had up to 8 chloroplasts, clone x up to 5 and clone w had 3 to 4 chloroplasts. We found rhizoids in five clones (aa, ae, ag, e, f); all of them had one accession in clade D, although accessions with rhizoids were also present in other clades. Chloroplast periodicity and length/width ratio varied considerably within clones, as both are influenced by cell growth.

Secondary structure analyses

No consensus structure valid for the whole genus *Spirogyra* could be found; secondary structure was highly variable, even within the clades. The putative secondary structure of *S. velata* WRH5 (clade B) is given in Figure 6. Bases used for barcoding are highlighted in grey; flanking regions 5.8S rRNA and LSU rRNA are marked by brackets; the pyrimidine mismatch is indicated with arrows; the conserved part of the 5' end of helix III is marked in blue, which in this case was located between two bulges and was also interrupted by a bulge.

The ITS2 sequence length of *Spirogyra* was 247 – 374 nt, the GC content ranged from 49% to 79 %. Helix I was unbranched and 12 – 67 nt long; helix II was unbranched and 22 – 92 nt long; helix III was branched in some sequences and 91 – 192 nt long; helix IV was not found in all sequences and was 7 – 61 nt long. A 50% consensus representation of the helices of each clade can be found in Figures 7 to 11. Clade G was excluded from analyses, because it had too few sequences.

Helix II always displayed the pyrimidine-pyrimidine mismatch within the first four nt. 4 sequences had only 3 helices, 2 sequences had an extra helix V. The presence/absence of helix IV was not reflected in clade grouping. Helix III was always the longest and most variable one. 18 sequences exhibited a side-loop in helix III which made it more difficult to align them and their barcodes, respectively. Most sequences with the extra helix III arm were found in clade C (12 out of 13 sequences), few in clade D (3 out of 32 sequences) and only one in clade E and G, respectively. The terminal tip/side on the 5' side showed a more conserved region in helix III, although it was not always found near the tip as expected. The CGGAUA/G pattern (Fig. 6 & 10) could be detected when looking at paired bases only, because it was interrupted with bulges and therefore not easily recognized. Also, it was not conserved across all sequences – some sequences didn't exhibit this sequence (clade A) or only a derived version of it (few sequences of clade C). The last A was substituted by a G in clade F. Also the barcode varied in length due to the length variation of helix III – it ranges from 91 to 180 nt.

Statistical analyses

Vegetative morphology and information on clade grouping from the concatenated SSU & barcode phylogeny were analyzed combined in a mantel test (Tab. 3). When using the whole alignment, a weak but highly significant correlation was found. When testing each clade separately, only clade B and H were positively correlated to their respective matrices, the others not. When distinguishing

clades with intron present from clades without intron, again a highly significant correlation was obtained. Among the morphological characters considered, only cell width related characters showed significant correlations (data not shown).
When analyzing each clade/with intron/without intron grouping against each morphological character, two clades (E & G) had to be excluded because they had too few objects in them; also cell wall (cw) structure and minimum number of chloroplasts (chl) had to be excluded from some analyses, because only one state was present in the comparisons. Correlations were found for clade B and cell width (cwi), average cwi (cwi-a) and cell with to cell length (cl) ratio; clade C and cwi, average cwi, minimum and maximum number of chloroplasts; clade H and cwi-cl ratio; for the group without 1506 group I intron and cross wall structure, cwi, cl, average cwi, cwi-cl ratio, minimum, maximum and average number of chloroplasts; for the group containing the 1506 group I intron cwi and average cwi.
When using vegetative morphology parameters individually against the same phylogenetic sub-/groups as before a similar picture was revealed. All correlations are significant, but only weak indicating a correlation between clade-grouping and vegetative morphology.
Also, the analyses of the vegetative morphology and clade grouping data (Fig. 5) with the program PRIMER did not reveal a clear pattern of vegetative morphology depending on phylogenetic relationships. Except clade A, no clade was separated clearly from the others based on the morphological data considered. A large bulk of 85 strains remained due to similarities in vegetative morphology. Taxonomically important traits as transverse cross wall structure were of no importance.
Correlations between environmental parameters and phylogeny remained unclear; the NMDS plot based on the distance matrix calculated with PAUP reflects phylogeny, but no information on environmental preferences of single clades/ groups could be derived (Fig. 4).

Discussion

With using only barcoding for phylogenetic analyses, many nt changes remain unaccounted for (Ruhl *et al.* 2010) and thus may not represent the true phylogeny of the studied group. We therefore refrained from analyses based on solely the barcodes. RbcL was shown to be insufficient for species separation as it represents a conserved region among closely related species (Drummond *et al.* 2005; Hall *et al.* 2008; McCourt *et al.* 2000; Stancheva *et al.* 2013).
SSU rDNA proved to be variable enough to establish a valid phylogeny consisting of 8 clades (Chen *et al.* 2012; Chen & Schagerl 2012). In this study, we additionally sequenced ITS1, but we were not able to use this region due to its extremely high variability (data not shown). Phylogenetic analyses based on ITS2 were also difficult to conduct, because high variability in primary and secondary structures allowed only limited application of methods commonly used. Differences among clades and even within clades were too big for standard comparisons and standard CBC analyses. No general ITS2 secondary structure for the genus *Spirogyra* could be identified; even for clades, it did not seem reasonable to state an overall consensus structure. Helices had to be analyzed individually to make reliable predictions.
We combined the barcode derived from ITS2 to enhance phylogenetic resolution of the SSU rDNA phylogenies and to predict species numbers within the phylogenetic clades. Although high conservation rates within taxonomic groups were reported from other studies (Coleman & Mai 1997; Merget & Wolf 2010), *Spirogyra* ITS2 sequences were altogether not easy to align – even when considering its secondary structure. So-called hallmarks of ITS2 secondary structure are observed in most organisms. Typically, the ITS2 has four helices – helix I and IV being the most variable regions, helix III the longest one and helix II with a conserved pyrimidine mismatch within the basal seven nts. Some groups show derivations from the basic ITS2 theme: e.g. angiosperms are known to form three to six helices (Coleman & Mai 1997). However, helix II and III are recognizable and common to essentially all eukaryotes (Coleman 2007). Helix III contains a very conserved motif at the 5' side of the apex (mostly involving a GUU triplet) and a pyrimidine-pyrimidine mismatch within the first 7 base pairs of helix II (Coleman 2003; Coleman 2007; Coleman & Mai 1997; Coleman & Vacquier 2002; Schultz *et al.* 2005; Young & Coleman 2004). The ITS2 secondary structure of *Spirogyra* fits this

description – four to five helices, conserved pyrimidine mismatch and a long helix III. Helix III usually displays the most conserved region on its 5' side near the terminus. In *Spirogyra* this site is by far not as conserved as in other taxa. Helix III is always by far the longest helix, but differs in length and structure between the strains – it can be multiple branched or single stranded.
Helix IV was not observed in some *Spirogyra* sequences, nor did it follow any pattern i.e. there was no group that could be differentiated from the rest by the presence or absence of helix IV. There seems to be less selective pressure to preserve this helix as it is often not present in organisms (Leliaert *et al.* 2009). The chlorophyte *Boodlea* (Cladophorales) also shows severe deviations from the usual form of ITS2 secondary structure (Leliaert *et al.* 2009).
Secondary structure was included in our phylogenetic analyses, as accuracy and robustness of NJ trees were shown to increase (Keller *et al.* 2010). The same is assumed for other methods such as ML, bayesian inference or MP. The inclusion of a higher number of nts only enhances tree robustness but not accuracy, because the secondary structure of non-coding RNA has more impact on the cell function than the mere sequence. By including secondary structure, phylogenetic analyses can be used at high and low levels of taxonomy. Usually, mutations are located on the distal parts of the helices, leaving the central core conserved (Coleman & Mai 1997). For the exceptional cases, CBCs are equally distributed over all four helices (Mueller *et al.* 2007). Helices are generally more conserved in the basal region, except for helix III that has its most conserved region distal on its 5' side of the apex (Coleman 2007). As accessions of the same clade differed in up to 11 CBCs and accessions of individual clades differed in up to 13 CBCs, the location of CBCs was not mapped in this study.
As mentioned by Koetschan et al. (2009), the CBC criterion only works in one direction, which means that the absence of a CBC does not imply that two organisms belong to the same species. Mueller et al. calculated a conditional probability of 93 % that two species can be distinguished, given that at least one CBC is observed (Mueller *et al.* 2007). The authors also observed that the lack of a CBC between two taxa resulted in a probability of only 77% that they belong to the same species. Following the indication of Coleman and Mai (1997) that only organisms can mate that do not differ by at most one CBC in the conserved region, the sampled *Spirogyra* sequences represent 51 individual putative species. Some species were represented by only one accession; others were represented by clones that consisted of 10 accessions.
Even if problems arise as described by Leliaert, such as conflicts between species concepts or forms of identical morphology (Leliaert *et al.* 2009), combining ITS2 and the barcode thereof is a valuable tool for resolving phylogenetic questions in organisms in which morphological discrimination is difficult like algae, mosses, fungi and other organisms (Leliaert *et al.* 2009; Merget & Wolf 2010). Although infection of samples with fungi leading to wrong species numbers have been observed (Yao *et al.* 2010), the ITS2 is proven to be a good species indicator. A unique advantage of the ITS2 is that the resulting alignment contains information related to the level of the biological species (Coleman 2007).
To date, no satisfying and accurate overall species concept for algae is available, because of different modes of reproduction, adaptation to the environment (i.e. polyploidy, phenotypic variability), examples given in Stuessy (2009) and Zwirn (2010). The term algae is moreover functional and does not represent a monophyletic group. By using barcoding we tried to simplify and shorten analyses/computational time, which was however unsuccessful due to the large number of hemi-CBCs and CBCs in the barcoding region. Using the ITS2 barcode as estimator for species number and the ITS2 secondary structure as support for phylogenetic analyses brought however additional insight to the genetic diversity of *Spirogyra*.
Statistical analyses combining vegetative morphology and phylogenetic information showed a significant, but weak correlation. Most important morphology traits were cell width related parameters. Transverse cell wall structure did not result in a clear pattern. Also chloroplast data and cell length yielded no valuable information on relationships, as both parameters are indicators of cell growth and the physiological state. Some groups (groups with intron; groups without intron; clades B, C, H) show a weak correlation with vegetative morphology, but the pattern is not sufficient to make predictions of clade/group memberships. Summarizing up, in our dataset vegetative

morphology is not clearly reflected in phylogeny (neither based on the concatenated SSU & barcode nor on the ITS2).

We found that geographical distribution, genetic diversity and phenotype variability/morphology do not coincide sufficiently in *Spirogyra*. Distribution records based on solely morphology - whether on vegetative or reproductive features or any combination - should therefore be interpreted with caution. We agree with Stancheva et al (2013) that structural similar species/strains from distant regions might be genetically dissimilar and that *Spirogyra* does contain a large number of cryptic species (Chen *et al.* 2012).

Since morphological features of the vegetative state alone cannot be used for species identification, additional characters have to be included for species delineation. Other traits remain to be tested – such as hypnozygote morphology, chromosome number and DNA content. In the Volvoceaen group, chromosome numbers and nutrient demands show only little overall correlation with family structure (Coleman 1999). Similar to the assumption of Van Hannen et al., who worked on *Scenedesmus,* many ecomorphs might exist mimicking large species numbers. For *Scenedesmus,* 1300 taxa were described which probably represent no more than 30 true species (Van Hannen *et al.* 2000).

The classification of Kadlubowska (1984) into three sections *Conjugata* (Vaucher) Hansg., *Colligata* Kadlub. and *Salmacis* (Bory) Hansg. was not retrieved in our data set. It may be argued that this can also be attributed to limited sampling, as no samples with external rings of wall material at the cell junctions (section *Colligata*) were included in our data set. But even the other division by plane (section *Conjugata*) or replicate cross walls (section *Salmacis*) was not reflected in our study. Features such as cell width covered a wide range (60.7µm in clone x) supporting the idea of ecotypes. Because hypnozygote size is strongly dependent on cell size, we suppose that the trend for clones with wide ranges of measurements is continued in this feature as well.

When comparing our phylogeny resulting from the SSU & barcode alignment to the phylogeny based on rbcL presented in Stancheva (2013), following similarities could be noted: SSU-clade C of our phylogeny and rbcL-clade II of the rbcL both include the UTEX strains *Sirogonium sticticum* and *S. maxima*; the cell width range (27-88µm; 32-150µm) and number of chloroplasts (1-6; 3-8) show a relatively good match. SSU-clade D and rbcL-clade V also include the same UTEX strains *S. varians, S. communis* and *S. gracilis*; cell width range (17-42µm; 25-64µm) and number of chloroplasts (1-2;1-2) also show a good match here. The other UTEX strains included in both studies were not grouped together. SSU-clade A and rbcL-clade VII show some similarity in chloroplast number (3-9; 4-10), but not in cell width (26-78µm, 80-108µm); the relative position to the remaining clades is however similar. SSU-clade B and rbcL-clade I could also be corresponding clades as both contain all the samples with replicate transverse cell walls. So there are some parallels in both phylogenies that have to be tested in future work.

When defining morphotypes as already done by Hainz et al. (2009), cross wall structure was the most important character followed by the number of chloroplasts and cell width. These traits were however not reflected phylogenetic analyses performed in this study.

According to our results and to reports from other studies (Gontcharov *et al.* 2002; Gontcharov *et al.* 2004; Hall *et al.* 2008; Stancheva *et al.* 2013), *Sirogonium* should be merged with *Spirogyra*. On the other hand, a split of the genus *Spirogyra* into several genera also seems reasonable because of its large hidden diversity. Before this step, we highly recommend a detailed study including various phylogenetic markers from different regions (ITS2 including secondary structure and barcode, SSU, rbcL), morphological characters (vegetative and reproductive features) as well as at least some ecological information, covering as much of the variability as possible. Because only few reliable morphological parameters are at hand, the big genetic variability of this genus is definitely underestimated. Having to rely solely on morphology for species identification has also to be scrutinized because of polyploid species complexes (Drummond *et al.* 2005; Hoshaw *et al.* 1985; Hoshaw *et al.* 1987; McCourt & Hoshaw 1990; McCourt *et al.* 1986; McCourt *et al.* 2000; Wang *et al.* 1986).

A few things have to be kept in mind when interpreting phylogenetic results - molecular methods are heuristic, which only allow to make probabilistic statements of species identity (Ross *et al.* 2008).

Even with this criticism, it can be said that even in our limited data set of roughly 120 *Spirogyra* accessions, at least 51 species were separated from each other by at least 1 CBC.
Questions that need to be addressed before eventually solving the species problem based on sequence data were formulated by Silva (2008): how much within species variation is allowable before we need to consider a second species in the genus; how much between-species difference is allowable before we need to consider placement of the species in two different genera? In *Spirogyra*, the phylogenetic variation among clades exceeds by far the variability of vegetative morphology. Even the relatively conserved ITS2 secondary structure varied considerably among clades; use of the ITS2 barcode helped getting a step further towards solving the species question. Species are not easily defined and recognized by morphology in the genus *Spirogyra*; it is more likely an evolutionary nightmare revealed by ITS 2 rDNA sequences.

Acknowledgements
We are very grateful to the team of Prof. Tod Stuessy and Prof. Rose Samuel at the Institute of Botany, Rennweg, University of Vienna, for the support and cooperativeness. This study was supported by the Austrian Science Fund, Project number P18465.

References

Birky CWJ, Barraclough TG (2009) Asexual Speciation. In: *Lost Sex* (eds. Schön I, Martens K, van Dijk P), pp. 201-216. Springer, Dordrecht Heidelberg London New York.

Chen C, Barfuss M, Proschold T, Schagerl M (2012) Hidden genetic diversity in the green alga Spirogyra (Zygnematophyceae, Streptophyta). *BMC Evolutionary Biology* 12, 77.

Chen C, Schagerl M (2012) Slow evolution of 1506 group I intron in *Spirogyra* LINK 1820 (Zygnematophyceae, Streptophyta), a fast evolving lineage in the Zygnemataceae. *Fottea*.

Clarke K, Gorley R (2006) PRIMER v6: User Manual/Tutorial. In: *PRIMER-E*, Plymouth.

Coleman AW (1999) Phylogenetic analysis of "*Volvocacae*" for comparative genetic studies. *Proceedings of the National Academy of Sciences of the United States of America* 96, 13892-13897.

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 (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *TRENDS in Genetics* 19, 370-375.

Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucl. Acids Res.*, gkm233.

Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* 50, 197-203.

Coleman AW, Mai JC (1997) Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relatedness. *Journal Of Molecular Evolution* 45, 168-177.

Coleman AW, Vacquier VD (2002) Exploring the Phylogenetic Utility of ITS Sequences for Animals: A Test Case for Abalone (Haliotis). *Journal Of Molecular Evolution* 54, 246-257.

Czurda V (1930) Experimentelle Untersuchungen über die Sexualverhältnisse der Zygnemalen. *Sonderabdruck aus "Beihefte zum Bot. Centralbl."* 47 Abt.I., 15 - 68.

Czurda V (1933) Experimentelle Analyse der kopulationsauslösenden Bedingungen bei Mikroorganismen - 1. Untersuchungen an Algen (*Spirogyra*, *Zygnema* und *Hyalotheca*). *Sonderabdruck aus "Beihefte zum Bot. Centralbl."* 51 Abt. I., 711 - 761.

De Queiroz K (2007) Species Concepts and Species Delimitation. *Systematic Biology* 56, 879-886.

Demchenko E, Mikhailyuk T, Coleman A, Pröschold T (2010) Generic and species concepts in Microglena (previously the Chlamydomonas monadina group) revised using an integrative approach. . *European Journal of Phycology* 47, 264-290.

Drummond CS, Hall J, Karol KG, Delwiche CF, McCourt RM (2005) Phylogeny of *Spirogyra* and *Sirogonium* (Zygnematophyceae) based on rbcL sequence data. *Journal of Phycology* 41, 1055-1064.

Fabry S, Köhler A, Coleman AW (1999) Intraspecies Analysis: Comparison of ITS Sequence Data and Gene Intron Sequence Data with Breeding Data for a Worldwide Collection of *Gonium pectorale*. *Journal Of Molecular Evolution* 48, 94-101.

Friedrich J, Dandekar T, Wolf M, Müller T (2005) ProfDist: a tool for the construction of large phylogenetic trees based on profile distances. *Bioinformatics* 21, 2108-2109.

Gontcharov AA, Marin B, Melkonian M (2002) Molecular phylogeny of conjugating green algae (Zygnemophyceae, Streptophyta) inferred from SSU rDNA sequence comparisons. *Journal of Molecular Evolution* 56, 89-104.

Gontcharov AA, Marin B, Melkonian M (2004) Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Mol Biol Evol* 21, 612-624.

Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Mol. Biol. Rev.* 58, 10-26.

Hainz R, Wöber C, Schagerl M (2009) The relationship between Spirogyra (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe. *Aquatic Botany* 91, 173-180.

Hall JD, Karol KG, McCourt RM, Delwiche CF (2008) Phylogeny of the Conjugating Green Algae Based on Cloroplast and Mitochondrial Nucleotide Sequence Data. *Journal Of Phycology* 44, 467-477.

Hey J (2006) On the failure of modern species concepts. *Trends in Ecology & Evolution* 21, 447-450.

Hoshaw RW, Wang J-C, McCourt RM, Hull HM (1985) Ploidal changes in clonal cultures of *Spirogyra communis* and implications for species definition. *American Journal of Botany* 72, 1005-1011.

Hoshaw RW, Wells CV, McCourt RM (1987) A polyploid species complex in *Spirogyra maxima* (Chlorophyta, Zygnemataceae), a species with large chromosomes. *Journal of Phycology* 23, 267-273.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754-755.

Kadlubowska JZ (1984) Conjugatophyceae I - Zygnemales. In: *Süßwasserflora von Mitteleuropa, Chlorophyta VIII* (eds. Ettl H, Gerloff H, Heynig H, Mollenhauer D). Stuttgart, New York: Gustav Fischer Verlag.

Keller A, Forster F, Muller T*, et al.* (2010) Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biology Direct* 5, 4.

Koetschan C, Förster F, Keller A*, et al.* (2009) The ITS2 Database III - sequences and structures for phylogeny. *Nucleic Acids Research* 38, D275-D279.

Leliaert F, Verbruggen H, Wysor B, Clerck OD (2009) DNA taxonomy in morphologically plastic taxa: Algorithmic species delimitation in the Boodlea complex (Chlorophyta: Cladophorales). *Molecular Phylogenetics and Evolution* 53, 122-133.

Luo W, Pröschold T, Bock C, Krienitz L (2010) Generic concept in Chlorella-related coccoid green algae (Chlorophyta, Trebouxiophyceae). *Plant Biology* 12, 545-553.

McCourt RM, Hoshaw RW (1990) Noncorrespondence of breeding groups, morphology and monophyletic groups in *Spirogyra* (Zygnemataceae: Chlorophyta) and the application of species concepts. *Systematic Botany* 15, 69-78.

McCourt RM, Hoshaw RW, Wang J-C (1986) Distribution, morphological diversity and evidence for polyploidy in North American Zygnemataceae (Chlorophyta). *Journal Of Phycology* 22, 307-315.

McCourt RM, Karol KG, Bell J*, et al.* (2000) Phylogeny of the conjugationg green algae (Zygnemophyceae) based on rbcL sequences. *Journal Of Phycology* 36, 747-758.

Merget B, Wolf M (2010) A molecular phylogeny of Hypnales (Bryophyta) inferred from ITS2 sequence-structure data. *BMC Research Notes* 3, 320.

Moniz MBJ, Kaczmarska I (2010) Barcoding of Diatoms: Nuclear Encoded ITS Revisited. *Protist* 161, 7-34.

Mueller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species. *RNA* 13, 1469-1472.

Posada D, Buckley T (2004) Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology* 53, 793-808.

Posada D, Crandall K (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.

Ross HA, Murugan S, Sibon Li WL (2008) Testing the Reliability of Genetic Methods of Species Identification via Simulation. *Systematic Biology* 57, 216-230.

Ruhl MW, Wolf M, Jenkins TM (2010) Compensatory base changes illuminate morphologically difficult taxonomy. *Molecular Phylogenetics and Evolution* 54, 664-669.

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

Schultz J, Mueller T, Achtziger M*, et al.* (2006) The internal transcribed spacer 2 database-a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res*, W704 - 707.

Schultz J, Wolf M (2009) ITS2 sequence-structure analysis in phylogenetics: A how-to manual for molecular systematics. *Molecular Phylogenetics and Evolution* 52, 520-523.

Seibel P, Muller T, Dandekar T, Wolf M (2008) Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE %U http://www.biomedcentral.com/1756-0500/1/91. *BMC Research Notes* 1, 91.

Seibel PN, Mueller T, Dandekar T, Schultz J, Wolf M (2006) 4SALE- a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics %M doi:10.1186/1471-2105-7-498* 7, 498.

Selig C, Wolf M, MÃ¼ller T, Dandekar T, Schultz Jr (2008) The ITS2 Database II: homology modelling RNA structure for molecular systematics. *Nucleic Acids Research* 36, D377-D380.

Silva PC (2008) Historical Review of Attempts to Decrease Subjectivity in Species Identification, with Particular Regard to Algae. *Protist* 159, 153-161.

Simons J, Van Beem AP, de Vries PJR (1982) Structure and chemical composition of the spore wall in Spirogyra (Zygnemataceae, Chlorophyceae). *Acta Bot. Neerl.* 31, 359-370.

Simons J, Van Beem P, De Vries PJR (1984) Induction of conjugation and spore formation in species of *Spirogyra* (Chlorophyceae, Zygnematales). *Acta Bot. Neerl.* 33, 323-334.

Stancheva R, Hall JD, McCourt RM, Sheath RG (2013) Identity and phylogenetic placement of Spirogyra species (Zygnematophyceae, Charophyta) from California streams and elsewhere1. *Journal Of Phycology* 49, 588-607.

Stuessy TF (2009) *Plant taxonomy: the systematic evaluation of comparative data* Columbia University Press.

Swanson WJ, Vacquier VD (1998) Concerted Evolution in an Egg Receptor for a Rapidly Evolving Abalone Sperm Protein. *Science* 281, 710-712.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Team RDC (2008) R: A Language and Environment for Statistical Computing Version 2.6.2 (ed. Computing RFfS), Vienna, Austria.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice

Nucl. Acids Res. 22, 4673-4680.

Van Hannen EJ, Lürling M, Van Donk E (2000) Sequence analysis of the ITS-2 region: A tool to identiify strains of *Scenedesmus* (Chlororphyceae). *Journal Of Phycology* 36, 605-607.

Wang J-C, Hoshaw RW, McCourt RM (1986) A Polyploid Species Complex of *Spirogyra communis* (Chlorophyta) Occuring in Nature. *Journal Of Phycology* 22, 102-107.

Yao H, Song J, Liu C*, et al.* (2010) Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. *PLoS ONE* 5, e13102.

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.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* 31, 3406-3415.

Zwirn M (2010) *Sexuelle Fortpflanzung und taxonomische Fragestellungen bei der Süßwasseralge Spirogyra (Zygnemataceae, Streptopyhta)*, University of Vienna.

1 - SSU&barcode Clade	total sequences	unique sequences	barcodes
A	3	3	3
B	31	18	16
C	14	7	7
D	28	16	10
E	3	3	3
F	20	10	5
G	2	1	1
H	3	3	3
DE	8	3	3
total	112	64	51

2 - ITS2 Clade			
A	3	3	3
B	35	18	17
C	14	7	7
D	31	15	13
E	3	3	3
F	20	10	5
G	2	2	2
DE	8	3	3
total	116	61	53

Table 2: Comparison of number of barcodes in different groups based on the data sets used for phylogeny (SSU&barcode alignment; ITS2 alignment); total number of sequences indicates number of accessions included in analyses; unique sequences indicates number without identical clones (identical in SSU and ITS2); barcode indicates number of individual different barcodes per clade/group;

Table 2: Upper right part contains the number of CBCs between sequences (calculated using 4Sale), lower left part contains pair wise distances between sequences (calculated using PAUP); both based on an ITS2 alignment (same as used for phylogeny) of all sequences.

Accessions are named: clade_culturename, clades are marked with thick black outline in matrix; 0 CBCs between accessions and distance of 0 is marked in green – only if both values are 0, ITS2 sequences are identical.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	A_KRA2		1	11	6	4	7	4	5	5	4	5	6	6	5	6	7	6	8	8	4	5	5	4	0	5	4	3	3	3	3
2	A_UMW4A	0,42		3	5	4	4	5	4	3	5	4	3	5	4	3	4	4	5	5	2	4	4	3	0	4	3	3	4	4	2
3	A_TIS1	0,40	0,43		6	4	5	3	5	4	4	5	4	5	4	4	5	5	6	6	4	7	7	3	0	7	3	5	3	3	2
4	B_M5A=MP926=MP923=M6A=S1Due2A=S1DD3A=M7_2=WBT2A=S1Due22A	0,61	0,63	0,60		6	4	5	9	4	2	2	3	2	6	3	3	4	8	8	7	6	6	4	4	6	5	9	12	12	7
5	B_KW4A=KW1A=KWA3=LSB4A	0,60	0,61	0,60	0,35		7	7	3	8	7	5	5	6	6	5	6	6	9	9	6	4	4	7	6	4	5	7	4	4	5
6	B_WRH5	0,60	0,61	0,60	0,35	0		3	9	7	4	4	7	4	5	7	4	4	7	7	7	6	6	5	3	6	5	9	8	8	5
7	B_SENA2	0,60	0,60	0,62	0,38	0,38	0,39		10	7	2	3	4	4	3	4	2	2	4	4	5	3	3	3	4	3	2	4	2	2	3
8	B_ROE6	0,63	0,63	0,57	0,41	0,38	0,39	0,31		10	10	8	7	8	7	7	8	8	11	11	6	3	3	7	6	3	6	6	5	5	6
9	B_SMS2	0,60	0,62	0,52	0,36	0,40	0,41	0,31	0,33		7	6	5	5	7	5	5	7	9	9	9	7	7	5	6	7	7	7	7	7	6
10	B_SENA7	0,58	0,56	0,53	0,41	0,42	0,43	0,37	0,37	0,37		4	5	5	3	5	2	2	3	3	6	5	5	4	4	5	4	5	4	4	4
11	B_SENA4	0,58	0,56	0,55	0,42	0,42	0,43	0,39	0,39	0,40	0,02		4	2	6	4	4	5	9	9	8	7	7	5	6	7	6	9	10	10	6
12	B_PANA1=WNII4A	0,59	0,56	0,56	0,37	0,44	0,44	0,37	0,38	0,38	0,18	0,19		3	8	0	6	6	8	8	10	8	8	5	8	8	6	10	8	8	6
13	B_M9B=7074	0,58	0,60	0,59	0,33	0,43	0,43	0,37	0,44	0,43	0,34	0,35	0,25		6	3	3	4	9	9	7	5	5	3	4	5	4	8	10	10	5
14	B_EGR4	0,63	0,64	0,51	0,38	0,46	0,47	0,40	0,45	0,38	0,41	0,42	0,39	0,32		7	3	2	1	1	6	4	4	3	2	4	3	5	4	4	3
15	B_NLH6	0,65	0,65	0,59	0,47	0,46	0,47	0,50	0,46	0,45	0,47	0,45	0,44	0,42	0,47		6	6	8	8	10	8	8	5	8	8	6	10	8	8	6
16	B_GBT1	0,63	0,64	0,56	0,42	0,38	0,38	0,38	0,34	0,31	0,34	0,36	0,38	0,34	0,45	0,27		1	3	3	9	7	7	5	5	7	5	10	4	4	3
17	B_TOR4=NLH5	0,61	0,61	0,56	0,44	0,51	0,51	0,49	0,44	0,48	0,46	0,48	0,43	0,39	0,45	0,51	0,48		3	3	6	5	5	5	4	5	5	7	3	3	3
18	B_LHS8	0,61	0,60	0,65	0,43	0,46	0,47	0,42	0,32	0,46	0,39	0,41	0,35	0,39	0,47	0,52	0,40	0,28		0	8	7	7	5	5	7	6	10	10	10	6
19	B_GBT3	0,64	0,59	0,52	0,48	0,48	0,49	0,44	0,37	0,38	0,48	0,50	0,45	0,52	0,48	0,50	0,45	0,48	0,46		8	7	7	5	5	7	6	10	10	10	6
20	C_M3A	0,60	0,55	0,58	0,62	0,62	0,63	0,51	0,58	0,55	0,53	0,55	0,57	0,57	0,59	0,63	0,61	0,56	0,54	0,60		3	3	2	3	3	3	7	3	3	1
21	C_TCA2=BGHII4	0,60	0,56	0,58	0,63	0,61	0,62	0,52	0,58	0,55	0,53	0,55	0,56	0,57	0,59	0,63	0,61	0,57	0,58	0,60	0,06		0	3	1	0	1	6	2	2	1
22	C_LL1=MDA1=LL2=M9A=MRT4A=DRS3=MD21	0,63	0,59	0,51	0,57	0,67	0,67	0,58	0,54	0,53	0,59	0,61	0,55	0,55	0,59	0,57	0,55	0,56	0,52	0,51	0,60	0,60		3	1	0	1	6	2	2	1
23	C_UTEX2495	0,56	0,57	0,51	0,56	0,55	0,56	0,53	0,52	0,48	0,55	0,55	0,56	0,58	0,50	0,52	0,52	0,55	0,61	0,56	0,60	0,60	0,65		6	4	3	5	2	2	2
24	C_Sirogonium_UTEX1985	0,64	0,61	0,59	0,45	0,52	0,52	0,55	0,53	0,53	0,50	0,52	0,52	0,54	0,54	0,58	0,50	0,54	0,57	0,46	0,65	0,66	0,55	0,63		1	4	4	5	5	3
25	C_WOE4A=ADA6=7075=ADA1=ADA2=DECA3	0,69	0,59	0,59	0,68	0,51	0,52	0,64	0,65	0,62	0,65	0,65	0,67	0,70	0,66	0,64	0,64	0,62	0,66	0,67	0,66	0,66	0,67	0,56	0,68		1	6	2	2	1
26	C_REC2	0,60	0,56	0,58	0,63	0,61	0,62	0,52	0,58	0,55	0,53	0,55	0,56	0,57	0,59	0,63	0,61	0,57	0,58	0,60	0,06	0	0,60	0,60	0,66	0,66		7	5	5	3
27	C_7208	0,61	0,52	0,52	0,51	0,52	0,52	0,51	0,43	0,52	0,45	0,48	0,46	0,51	0,51	0,52	0,49	0,45	0,53	0,51	0,52	0,49	0,57	0,59	0,54	0,62	0,49		5	5	3
28	D_7075	0,59	0,62	0,53	0,61	0,60	0,60	0,55	0,48	0,56	0,52	0,54	0,54	0,60	0,55	0,65	0,54	0,52	0,52	0,58	0,64	0,65	0,60	0,50	0,58	0,56	0,65	0,63		0	0
29	D_REC3=VNE2=BTSA1	0,63	0,58	0,58	0,54	0,54	0,55	0,53	0,57	0,54	0,58	0,59	0,55	0,56	0,59	0,54	0,51	0,50	0,54	0,54	0,55	0,56	0,58	0,61	0,55	0,60	0,56	0,52	0,53		0
30	D_WNII3A	0,65	0,60	0,64	0,54	0,50	0,51	0,52	0,53	0,51	0,55	0,56	0,54	0,48	0,48	0,54	0,48	0,45	0,50	0,46	0,56	0,56	0,57	0,61	0,51	0,58	0,56	0,50	0,52	0,15	
31	D_UTEX479	0,63	0,58	0,58	0,54	0,54	0,55	0,53	0,57	0,54	0,58	0,59	0,55	0,56	0,59	0,54	0,51	0,50	0,54	0,54	0,55	0,56	0,58	0,61	0,55	0,60	0,56	0,52	0,53	0	0,15
32	D_AB1	0,64	0,62	0,59	0,55	0,50	0,50	0,48	0,41	0,50	0,49	0,51	0,53	0,55	0,56	0,51	0,46	0,52	0,49	0,51	0,58	0,58	0,59	0,56	0,52	0,59	0,58	0,48	0,45	0,33	0,28
33	D_NES3=NES4=TIS3=TIS4=GSB2=PLO1	0,65	0,56	0,59	0,49	0,51	0,52	0,52	0,47	0,49	0,52	0,54	0,51	0,54	0,54	0,57	0,46	0,50	0,50	0,52	0,52	0,53	0,61	0,60	0,50	0,57	0,53	0,46	0,56	0,31	0,26
34	D_BGHK1=BGHK2	0,66	0,55	0,58	0,49	0,49	0,51	0,50	0,46	0,49	0,51	0,53	0,52	0,56	0,54	0,57	0,46	0,50	0,51	0,50	0,54	0,54	0,61	0,60	0,49	0,56	0,54	0,46	0,54	0,31	0,27
35	D_FBT1A=1020A1	0,65	0,56	0,59	0,49	0,51	0,52	0,52	0,47	0,49	0,52	0,54	0,51	0,54	0,54	0,57	0,46	0,50	0,50	0,52	0,52	0,53	0,61	0,60	0,50	0,57	0,53	0,46	0,56	0,31	0,26
36	D_SNG1	0,60	0,63	0,59	0,56	0,50	0,51	0,47	0,38	0,44	0,55	0,57	0,55	0,56	0,55	0,54	0,50	0,49	0,49	0,46	0,53	0,54	0,61	0,58	0,56	0,60	0,54	0,46	0,51	0,33	0,27
37	D_SNG2=SNG3=SNG5=OSORIVA1=BEA1=WB5A=WNIII2A	0,60	0,63	0,59	0,56	0,50	0,51	0,47	0,38	0,44	0,55	0,57	0,55	0,56	0,55	0,54	0,50	0,49	0,49	0,46	0,53	0,54	0,61	0,58	0,56	0,60	0,54	0,46	0,51	0,33	0,27
38	D_KVW2=KVW4	0,60	0,63	0,59	0,56	0,50	0,51	0,47	0,38	0,44	0,55	0,57	0,55	0,56	0,55	0,54	0,50	0,49	0,49	0,46	0,53	0,54	0,61	0,58	0,56	0,60	0,54	0,46	0,51	0,33	0,27
39	D_GUS1	0,60	0,63	0,59	0,56	0,50	0,51	0,47	0,38	0,44	0,55	0,57	0,55	0,56	0,55	0,54	0,50	0,49	0,49	0,46	0,53	0,54	0,61	0,58	0,56	0,60	0,54	0,46	0,51	0,33	0,27
40	D_7076	0,59	0,61	0,53	0,52	0,51	0,51	0,49	0,38	0,47	0,49	0,51	0,46	0,50	0,51	0,49	0,44	0,50	0,47	0,49	0,57	0,57	0,60	0,50	0,53	0,62	0,57	0,41	0,46	0,36	0,33
41	D_UTEX1743	0,65	0,62	0,57	0,54	0,51	0,50	0,50	0,40	0,52	0,47	0,49	0,51	0,55	0,55	0,51	0,47	0,50	0,49	0,50	0,62	0,63	0,58	0,54	0,54	0,60	0,63	0,48	0,45	0,40	0,33
42	D_UTEX2466	0,59	0,61	0,54	0,53	0,54	0,54	0,56	0,51	0,50	0,52	0,53	0,57	0,58	0,56	0,53	0,47	0,54	0,59	0,44	0,56	0,57	0,57	0,56	0,51	0,56	0,57	0,52	0,49	0,41	0,42
43	D_AB4	0,65	0,62	0,57	0,54	0,51	0,50	0,50	0,40	0,52	0,47	0,49	0,51	0,55	0,55	0,51	0,47	0,50	0,49	0,50	0,62	0,63	0,58	0,54	0,54	0,60	0,63	0,48	0,45	0,40	0,33
44	E_HRM4A	0,54	0,60	0,50	0,52	0,56	0,56	0,53	0,57	0,51	0,54	0,56	0,53	0,58	0,54	0,53	0,51	0,56	0,49	0,59	0,52	0,56	0,63	0,50	0,64	0,59	0,56	0,58	0,46	0,55	0,56
45	E_TRS7	0,62	0,56	0,44	0,52	0,59	0,60	0,60	0,56	0,55	0,54	0,56	0,56	0,65	0,57	0,67	0,60	0,66	0,65	0,54	0,65	0,65	0,62	0,57	0,57	0,60	0,65	0,61	0,41	0,49	0,55
46	F_LHS5=LHS4=LHS6=SHS6A=LHS9	0,64	0,59	0,57	0,59	0,46	0,47	0,52	0,49	0,54	0,55	0,53	0,49	0,58	0,65	0,58	0,53	0,53	0,58	0,56	0,54	0,53	0,63	0,60	0,65	0,49	0,53	0,54	0,50	0,51	0,52
47	F_SKR1=SKR2	0,62	0,57	0,54	0,60	0,48	0,49	0,52	0,49	0,54	0,55	0,57	0,49	0,58	0,61	0,61	0,53	0,53	0,58	0,56	0,54	0,54	0,63	0,58	0,64	0,47	0,54	0,55	0,47	0,52	0,53
48	F_SKA1	0,62	0,57	0,54	0,60	0,48	0,49	0,52	0,49	0,54	0,55	0,57	0,49	0,58	0,61	0,61	0,53	0,53	0,58	0,56	0,54	0,54	0,63	0,58	0,64	0,47	0,54	0,55	0,47	0,52	0,53
49	F_SKA3	0,62	0,57	0,54	0,60	0,48	0,49	0,52	0,49	0,54	0,55	0,57	0,49	0,58	0,61	0,61	0,53	0,53	0,58	0,56	0,54	0,54	0,63	0,58	0,64	0,47	0,54	0,55	0,47	0,52	0,53
50	F_LUA1	0,62	0,56	0,56	0,58	0,50	0,51	0,50	0,49	0,55	0,56	0,58	0,55	0,56	0,57	0,61	0,51	0,52	0,59	0,54	0,54	0,53	0,66	0,57	0,65	0,48	0,53	0,52	0,50	0,53	0,50
51	F_BGHIII1=BGHIII2=0710151e	0,61	0,58	0,46	0,53	0,49	0,50	0,57	0,52	0,54	0,53	0,51	0,56	0,64	0,60	0,54	0,54	0,56	0,64	0,54	0,67	0,66	0,64	0,53	0,52	0,53	0,66	0,46	0,54	0,56	0,57
52	F_DRS1=DRS2	0,63	0,59	0,46	0,53	0,50	0,51	0,57	0,52	0,54	0,53	0,51	0,56	0,64	0,60	0,55	0,54	0,56	0,64	0,54	0,67	0,67	0,64	0,53	0,51	0,53	0,67	0,47	0,54	0,56	0,58
53	F_W4A=WIND3A	0,65	0,62	0,55	0,56	0,50	0,51	0,53	0,54	0,52	0,48	0,46	0,52	0,56	0,58	0,50	0,53	0,56	0,61	0,56	0,60	0,59	0,62	0,58	0,60	0,50	0,59	0,52	0,54	0,54	0,52
54	F_ROE2=EIT4A	0,61	0,61	0,54	0,52	0,50	0,51	0,52	0,54	0,49	0,57	0,59	0,50	0,57	0,58	0,62	0,57	0,50	0,55	0,55	0,55	0,55	0,67	0,56	0,57	0,48	0,55	0,54	0,48	0,58	0,56
55	F_ZIP2	0,61	0,62	0,54	0,53	0,48	0,49	0,54	0,53	0,52	0,57	0,55	0,52	0,57	0,59	0,61	0,57	0,51	0,54	0,55	0,55	0,55	0,66	0,54	0,58	0,51	0,55	0,56	0,48	0,56	0,55
56	G_UTEX1746	0,64	0,70	0,56	0,57	0,56	0,57	0,58	0,53	0,56	0,53	0,55	0,58	0,60	0,59	0,63	0,55	0,55	0,58	0,61	0,61	0,61	0,67	0,63	0,58	0,68	0,61	0,56	0,53	0,58	0,55
57	H_FIT3A=GRS1	0,69	0,64	0,57	0,49	0,46	0,46	0,47	0,40	0,53	0,54	0,54	0,51	0,47	0,50	0,45	0,43	0,49	0,47	0,49	0,67	0,67	0,60	0,59	0,58	0,62	0,67	0,45	0,64	0,58	0,55
58	H_GTH7=GTH1=KAT1A=GRS2	0,66	0,60	0,64	0,36	0,38	0,38	0,39	0,38	0,44	0,34	0,35	0,40	0,44	0,49	0,50	0,44	0,49	0,42	0,53	0,47	0,47	0,60	0,55	0,54	0,59	0,47	0,49	0,55	0,51	0,50
59	UTEX_1742_Sp_juergensii	0,63	0,62	0,51	0,44	0,55	0,55	0,47	0,49	0,47	0,54	0,56	0,57	0,52	0,54	0,50	0,45	0,51	0,53	0,59	0,54	0,55	0,59	0,58	0,66	0,66	0,55	0,55	0,58	0,50	0,55
60	UTEX1745_Sp_liana	0,61	0,53	0,48	0,60	0,55	0,56	0,53	0,53	0,49	0,47	0,45	0,52	0,55	0,57	0,52	0,45	0,56	0,50	0,57	0,49	0,49	0,60	0,50	0,61	0,59	0,49	0,57	0,43	0,56	0,60

		31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
1	A_KRA2	3	3	3	3	3	7	3	3	4	3	4	3	3	3	2	3	3	2	3	3	3	2	3	2	3	1	7	5	5	4
2	A_UMW4A	2	2	4	4	2	6	2	2	3	2	3	2	2	2	2	4	4	3	4	4	3	3	4	3	3	4	5	4	6	3
3	A_TIS1	2	3	3	3	3	6	3	4	4	3	4	2	3	2	2	4	4	3	4	4	3	3	4	3	3	2	6	5	5	3
4	B_M5A=MP926=MP923=M6A=S1Due2A=S1DD3A=M7_2=WBT2A=S1Due22A	7	9	12	12	10	7	10	12	8	6	9	7	10	9	7	13	13	11	13	13	12	11	12	11	12	5	2	11	11	5
5	B_KW4A=KW1A=KWA3=LSB4A	5	7	4	4	8	7	8	6	7	5	8	5	8	9	7	6	6	4	6	6	5	4	5	4	5	2	3	3	8	5
6	B_WRH5	5	8	8	8	8	6	8	9	9	7	9	5	8	7	6	12	12	10	12	12	10	10	10	9	10	7	7	9	10	3
7	B_SENA2	3	4	2	2	4	4	4	6	5	4	6	3	4	5	7	7	7	5	7	7	6	5	6	5	6	4	3	7	8	4
8	B_ROE6	6	7	5	5	8	7	8	7	6	4	7	6	8	9	8	7	7	5	7	7	5	5	6	5	5	3	5	6	10	5
9	B_SMS2	6	9	7	7	9	9	9	11	9	8	10	6	9	7	8	9	9	8	9	9	8	8	9	8	8	2	5	12	8	6
10	B_SENA7	4	5	4	4	5	5	5	6	5	4	6	4	5	9	7	8	8	6	8	8	7	6	7	6	7	4	4	6	9	5
11	B_SENA4	6	9	10	10	10	8	10	14	9	7	10	6	10	10	9	13	13	11	13	13	12	11	12	11	12	6	4	9	10	4
12	B_PANA1=WNII4A	6	11	8	8	11	9	11	12	11	9	12	6	11	11	9	11	11	9	11	11	9	9	10	9	9	3	3	8	11	5
13	B_M9B=7074	5	8	10	10	9	7	9	12	8	6	9	5	9	8	7	12	12	10	12	12	11	10	11	10	11	5	2	12	10	3
14	B_EGR4	3	4	4	4	3	3	3	5	5	2	6	3	3	5	5	5	5	3	5	5	5	3	4	3	5	3	6	10	5	4
15	B_NLH6	6	11	8	8	11	9	11	12	11	9	12	6	11	11	9	11	11	9	11	11	9	9	10	9	9	3	3	8	11	5
16	B_GBT1	3	6	4	4	6	5	6	11	7	6	8	3	6	7	8	8	8	6	8	8	8	6	7	6	8	4	6	9	8	4
17	B_TOR4=NLH5	3	6	3	3	6	5	6	9	7	6	8	3	6	6	7	9	9	7	9	9	9	7	8	7	9	4	5	7	7	4
18	B_LHS8	6	10	10	10	10	5	10	12	10	7	11	6	10	9	9	12	12	10	12	12	11	10	11	10	11	6	8	11	11	5
19	B_GBT3	6	10	10	10	10	5	10	12	10	7	11	6	10	9	9	12	12	10	12	12	11	10	11	10	11	6	8	11	11	5
20	C_M3A	1	3	3	3	3	2	3	7	3	3	3	1	3	2	4	5	5	6	5	5	7	5	5	6	7	2	8	7	3	2
21	C_TCA2=BGHII4	1	2	2	2	2	0	2	4	2	2	2	1	2	2	3	4	4	3	4	4	3	2	3	3	3	0	7	6	3	1
22	C_LL1=MDA1=LL2=M9A=MRT4A=DRS3=MD21	1	2	2	2	2	0	2	4	2	2	2	1	2	2	3	4	4	3	4	4	3	2	3	3	3	0	7	6	3	1
23	C_UTEX2495	2	3	2	2	3	3	3	4	3	3	4	2	3	4	5	5	5	4	5	5	5	4	4	4	5	3	4	10	4	2
24	C_Sirogonium_UTEX1985	3	5	5	5	5	5	5	3	5	5	4	3	5	2	4	3	3	3	3	3	2	2	3	3	2	3	8	7	3	0
25	C_WOE4A=ADA6=7075=ADA1=ADA2=DECA3	1	2	2	2	2	0	2	4	2	2	2	1	2	2	3	4	4	3	4	4	3	2	3	3	3	0	7	6	3	1
26	C_REC2	3	5	5	5	5	2	5	4	5	5	4	3	5	4	5	5	5	4	5	5	5	3	4	4	5	1	7	9	6	3
27	C_7208	3	4	5	5	4	4	4	5	5	4	5	3	4	2	4	4	4	3	4	4	3	3	4	3	3	1	12	9	3	3
28	D_7075	0	1	0	0	1	1	1	8	0	0	1	0	1	6	6	6	6	5	6	6	5	5	5	5	5	5	9	9	6	1
29	D_REC3=VNE2=BTSA1	0	1	0	0	1	1	1	8	0	0	1	0	1	6	6	6	6	5	6	6	5	5	5	5	5	5	9	9	6	1
30	D_WNII3A	0	2	0	0	3	2	3	6	0	0	1	0	3	5	10	4	4	3	4	4	4	3	3	3	4	5	8	8	3	1
31	D_UTEX479		2	0	0	3	2	3	6	0	0	1	0	3	5	10	4	4	3	4	4	4	3	3	3	4	5	8	8	3	1
32	D_AB1	0,33		2	2	2	2	2	7	2	0	3	3	2	5	7	5	5	4	5	5	5	4	4	4	5	4	12	11	4	1
33	D_NES3=NES4=TIS3=TIS4=GSB2=PLO1	0,31	0,32		0	1	1	1	8	0	0	1	0	1	6	6	6	6	5	6	6	5	5	5	5	5	5	9	9	6	1
34	D_BGHK1=BGHK2	0,31	0,30	0,03		1	1	1	8	0	0	1	0	1	6	6	6	6	5	6	6	5	5	5	5	5	5	9	9	6	1
35	D_FBT1A=1020A1	0,31	0,32	0	0,03		2	0	6	0	0	1	3	0	6	8	6	6	5	6	6	6	5	5	5	6	4	11	10	6	1
36	D_SNG1	0,33	0,23	0,30	0,32	0,30		2	1	2	1	2	2	2	3	2	1	1	0	1	1	1	0	0	0	1	0	8	9	2	2
37	D_SNG2=SNG3=SNG5=OSORIVA1=BEA1=WB5A=WNIII2A	0,33	0,23	0,30	0,32	0,30	0		6	0	0	1	3	0	6	8	6	6	5	6	6	6	5	5	5	6	4	11	10	6	1
38	D_KVW2=KVW4	0,33	0,23	0,30	0,32	0,30	0	0		6	5	6	5	6	2	4	6	6	6	6	6	7	6	4	4	7	8	8	7	6	0
39	D_GUS1	0,33	0,23	0,30	0,32	0,30	0	0	0		0	1	0	0	5	7	5	5	4	5	5	5	4	4	4	5	4	11	9	4	1
40	D_7076	0,36	0,13	0,32	0,30	0,32	0,24	0,24	0,24	0,24		1	0	0	3	4	2	2	1	2	2	2	1	1	1	2	1	11	7	1	1
41	D_UTEX1743	0,40	0,06	0,37	0,35	0,37	0,31	0,31	0,31	0,31	0,18		2	2	5	8	6	6	5	6	6	5	4	5	5	5	4	11	10	5	1
42	D_UTEX2466	0,41	0,45	0,47	0,47	0,47	0,45	0,45	0,45	0,45	0,47	0,47		3	5	10	4	4	3	4	4	4	3	3	3	4	5	8	8	3	1
43	D_AB4	0,40	0,06	0,37	0,35	0,37	0,31	0,31	0,31	0,31	0,18	0	0,47		6	8	6	6	5	6	6	6	5	5	5	6	4	11	10	6	1
44	E_HRM4A	0,55	0,57	0,57	0,56	0,57	0,58	0,58	0,58	0,58	0,53	0,61	0,49	0,61		2	4	4	4	4	4	4	4	3	3	4	6	10	10	5	2
45	E_TRS7	0,49	0,58	0,58	0,57	0,58	0,60	0,60	0,60	0,60	0,57	0,53	0,48	0,53	0,40		3	3	3	3	3	4	3	2	2	4	4	6	7	2	0
46	F_LHS5=LHS4=LHS6=SHS6A=LHS9	0,51	0,50	0,55	0,51	0,55	0,51	0,51	0,51	0,51	0,52	0,47	0,46	0,47	0,53	0,51		0	0	0	0	0	0	2	1	0	7	10	8	5	2
47	F_SKR1=SKR2	0,52	0,50	0,55	0,51	0,55	0,51	0,51	0,51	0,51	0,52	0,47	0,47	0,47	0,51	0,48	0,03		0	0	0	0	0	2	1	0	7	10	8	5	2
48	F_SKA1	0,52	0,50	0,55	0,51	0,55	0,51	0,51	0,51	0,51	0,52	0,47	0,47	0,47	0,51	0,48	0,03	0		0	0	0	0	2	1	0	7	8	6	5	1
49	F_SKA3	0,52	0,50	0,55	0,51	0,55	0,51	0,51	0,51	0,51	0,52	0,47	0,47	0,47	0,51	0,48	0,03	0	0		0	0	0	2	1	0	7	10	8	5	2
50	F_LUA1	0,53	0,50	0,54	0,53	0,54	0,51	0,51	0,51	0,51	0,54	0,49	0,47	0,49	0,54	0,49	0,13	0,10	0,10	0,10		0	0	2	1	0	7	10	8	5	2
51	F_BGHIII1=BGHIII2=0710151e	0,56	0,54	0,57	0,56	0,57	0,51	0,51	0,51	0,51	0,52	0,52	0,48	0,52	0,55	0,45	0,25	0,28	0,28	0,28	0,26		1	0	0	0	4	8	8	6	2
52	F_DRS1=DRS2	0,56	0,54	0,57	0,56	0,57	0,52	0,52	0,52	0,52	0,52	0,52	0,48	0,52	0,55	0,45	0,25	0,28	0,28	0,28	0,26	0,01		2	1	1	6	8	6	6	1
53	F_W4A=WIND3A	0,54	0,58	0,58	0,54	0,58	0,54	0,54	0,54	0,54	0,54	0,54	0,48	0,54	0,54	0,49	0,17	0,20	0,20	0,20	0,20	0,18	0,18		0	0	6	10	8	6	2
54	F_ROE2=EIT4A	0,58	0,54	0,53	0,53	0,53	0,51	0,51	0,51	0,51	0,53	0,54	0,52	0,54	0,50	0,49	0,23	0,20	0,20	0,20	0,24	0,22	0,22	0,30		0	5	8	6	5	1
55	F_ZIP2	0,56	0,53	0,54	0,51	0,54	0,53	0,53	0,53	0,53	0,52	0,53	0,48	0,53	0,54	0,50	0,22	0,25	0,25	0,25	0,27	0,21	0,21	0,27	0,07		4	8	8	6	2
56	G_UTEX1746	0,58	0,56	0,56	0,56	0,56	0,60	0,60	0,60	0,60	0,58	0,52	0,52	0,52	0,61	0,61	0,52	0,53	0,53	0,53	0,56	0,58	0,58	0,56	0,61	0,57		2	4	7	1
57	H_FIT3A=GRS1	0,58	0,57	0,57	0,56	0,57	0,59	0,59	0,59	0,59	0,54	0,56	0,58	0,56	0,58	0,60	0,54	0,56	0,56	0,56	0,55	0,50	0,50	0,53	0,57	0,56	0,64		9	11	5
58	H_GTH7=GTH1=KAT1A=GRS2	0,51	0,52	0,51	0,51	0,51	0,58	0,58	0,58	0,58	0,52	0,53	0,52	0,53	0,50	0,61	0,50	0,50	0,50	0,50	0,49	0,55	0,55	0,51	0,50	0,49	0,53	0,49		10	5
59	UTEX_1742_Sp_juergensii	0,50	0,53	0,51	0,52	0,51	0,51	0,51	0,51	0,51	0,51	0,56	0,53	0,56	0,49	0,53	0,55	0,56	0,56	0,56	0,53	0,55	0,54	0,55	0,53	0,54	0,56	0,60	0,47		2
60	UTEX1745_Sp_liana	0,56	0,55	0,57	0,56	0,57	0,60	0,60	0,60	0,60	0,51	0,56	0,47	0,56	0,46	0,48	0,45	0,49	0,49	0,49	0,50	0,46	0,46	0,47	0,50	0,45	0,52	0,61	0,47	0,53	

Table 3: Results from a mantel test carried out using morphological data and clade grouping; clades E and G had to be excluded, they included too few sequences; not significant: n.s. ($p > 0.05$; relative rates not significantly different at 5% level). Asterisks: $p = 0.05 > * > 0.01 > ** > 0.005 > ***$ (relative rates significantly different); cw: cell end wall structure; cwi: cell width; cl: cell length; chl min: minimum number of chloroplasts; chl max: maximum number of chloroplasts; chl-p: chloroplast periodicity; chl-a: average number of chloroplasts; cwi-a: average cwi; cwi-cl: ratio cwi-cl.

	total	r^2	cw	r^2	cwi	r^2	cl	r^2	chl min	r^2	chl max	r^2	chl-p	r^2	chl-a	r^2	cwi-a	r^2	cwi-cl	r^2
total	***	0,052	- n.s.	0,013	*	0,030	- n.s.	0,001	n.s.	0,167	n.s.	0,124	- n.s.	0,002	n.s.	0,152	*	0,029	- n.s.	0,004
clade A	- n.s.	0,466	n.a.	n.a.	- n.s.	0,919	n.s.	0,002	- n.s.	0,571	- n.s.	0,481	- n.s.	0,395	- n.s.	0,519	- n.s.	0,942	- n.s.	0,670
clade B	*	0,094	n.s.	0,181	*	0,043	n.s.	0,017	- n.s.	0,004	- n.s.	0,000	- n.s.	0,001	- n.s.	0,001	*	0,051	*	0,126
clade C	n.s.	0,174	n.a.	n.a.	***	0,170	n.s.	0,011	*	0,076	**	0,154	n.a.	n.a.	***	0,150	***	0,175	n.s.	0,010
clade D	n.s.	0,005	n.a.	n.a.	n.s.	0,193	- n.s.	0,017	n.a.	n.a.	- n.s.	0,002	n.s.	0,015	- n.s.	0,002	n.s.	0,196	n.s.	0,000
clade F	- n.s.	0,086	- n.s.	0,012	- n.s.	0,003	- n.s.	0,050	n.a.	n.a.	- n.s.	0,016	- n.s.	0,043	- n.s.	0,016	- n.s.	0,002	- n.s.	0,037
clade H	*	0,489	n.s.	0,944	*	0,190	n.s.	0,002	n.s.	0,199	n.s.	0,282	n.s.	0,000	n.s.	0,449	n.s.	0,152	*	0,430
intron +	***	0,146	***	0,206	***	0,255	***	0,148	***	0,191	***	0,155	- n.s.	0,006	***	0,202	***	0,259	***	0,068
intron -	n.s.	0,496	- n.s.	0,010	***	0,135	- n.s.	0,003	n.s.	0,420	n.s.	0,371	- n.s.	0,001	n.s.	0,413	***	0,133	- n.s.	0,004

Table 4: Comparison of vegetative morphology of clones; clones named with letters arbitrarily – missing letters were single accession clones; phylogenetic position of clones can be found in Fig. 1; pl – plane, repl – replicate; length and width are given in µm; n.r. – no rhizoids present, rhiz – rhizoids present; n – number of accessions.

Clone	Clade in SSU& barcode phylogeny	End wall	Cell Length	C.L. diff	Cell Width	C.W. diff	Chloroplast no.	Chloroplast periodicity	Rhizoids	Length/Width ratio	n
a	B	pl	48,9 - 299,1	250,2	14,3 - 29,5	15,2	1 - 2	15,9 - 49,1	n.r.	1,8 - 11	10
aa	CDE	pl	92,7 - 157,4	64,7	26,1 - 62,2	36,1	1 - 8	24,5 - 90,5	rhiz	2,5 - 4,1	6
ae	D	pl	60,2 - 81,9	21,7	23,3 - 25,6	2,3	1 - 1	16,6 -32,1	rhiz	2,4 - 3,5	3
ag	D	pl	78,7 - 92,4	13,7	21 - 28,8	7,8	1 - 2	25 - 39,2	rhiz	3,1 - 4,2	3
ai	D	pl	106,8 - 111,2	4,4	19,3 - 21,4	2,1	1 - 2	24,9 - 33,5	n.r.	5,2 - 5,5	2
aj	D	pl	56,9 - 78,8	21,9	17,5 - 19,3	1,8	1 - 2	21 - 24,5	n.r.	3 - 4,5	3
as	F	pl	133,8 - 214,5	80,7	24 - 24,3	0,3	1 - 2	34,1 - 38,7	n.r.	5,5 - 8,9	2
at	F	pl	76,8 - 84,1	7,3	21,9 - 23,1	1,2	1 - 2	20,4 - 23,1	n.r.	3,5 - 3,6	2
ay	B	repl	251 - 262,8	11,8	29,5 - 30,5	1	1 - 3	47,5 - 87	n.r.	8,2 - 8,9	2
b	B	pl	91,3 - 207,9	116,6	30,1 - 34,5	4,4	1 - 2	36,2 - 72,1	n.r.	4,3 - 6,7	5
c	F	pl, repl	60,6 - 259,9	199,3	23,1 - 28,6	5,5	1 - 3	15,4 - 47,7	n.r.	2,5 - 9,9	10
d	F	pl	121,2 - 143,2	22	21,1 - 26,3	5,2	1 - 2	21,8 - 25,9	n.r.	5,3 - 6,5	5
e	D	pl	64,1 - 99,4	35,3	18,5 - 23,9	5,4	1 - 2	17,3 - 25,1	rhiz	2,9 - 4,9	4
f	D	pl	68,4 - 158,9	90,5	23,2 - 30	6,8	1 - 2	17 - 40,5	rhiz	2,9 - 5,6	8
o	B	repl	158,7 - 411,9	253,2	25,2 - 25,6	0,4	1 - 2	29,7 - 39,2	n.r.	6,3 - 16,1	2
w	C	pl	99,8 - 159,8	60	52,9 - 54,2	1,3	3 - 4	61 - 80	n.r.	1,9 - 2,9	2
x	C	pl	137,7 - 292,5	154,8	26,8 - 87,5	60,7	1 - 5	31,1 - 97,3	n.r.	2,4 - 5,1	8

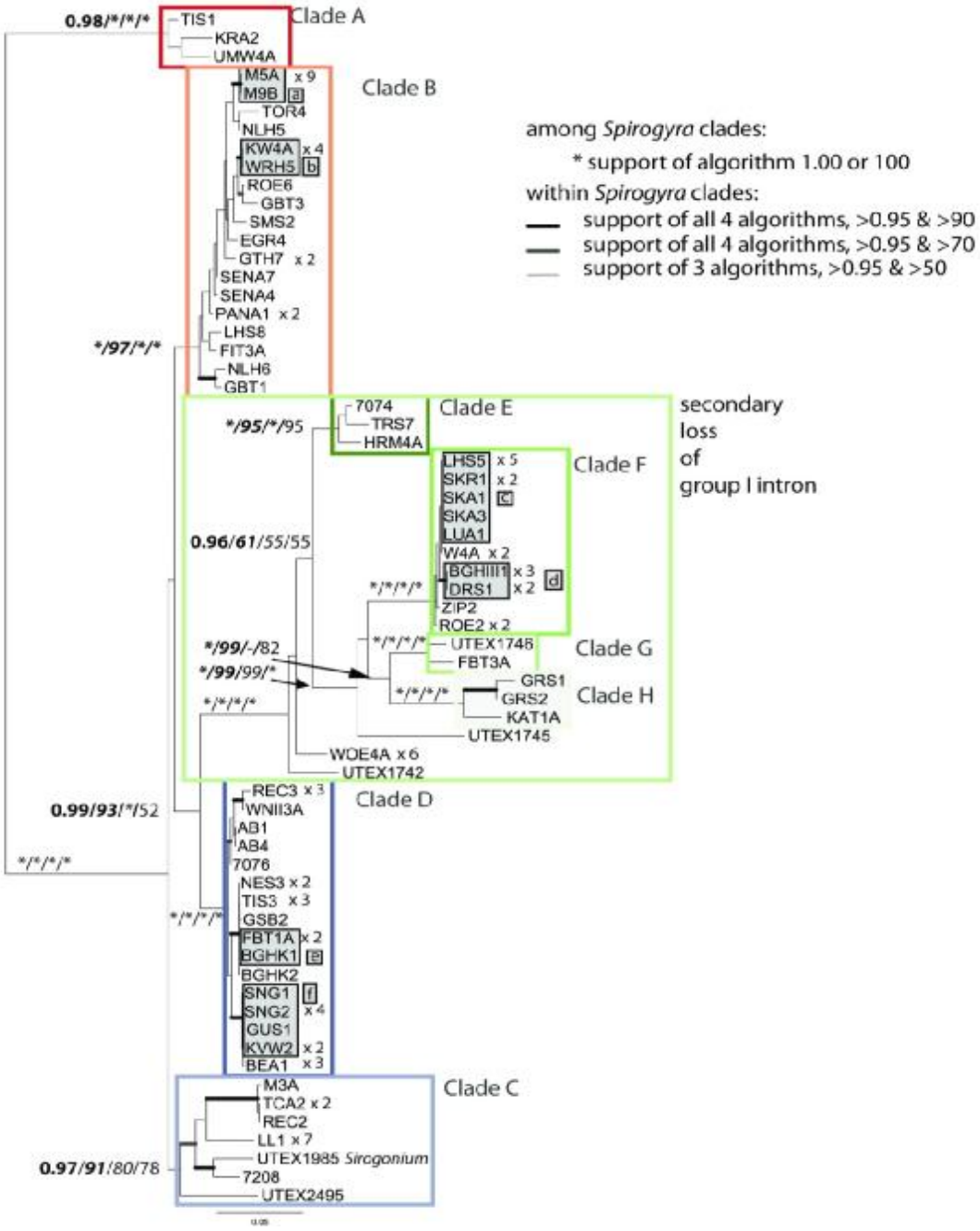


Figure 1: Phylogeny based on SSU&barcode alignment calculated using maximum likelihood algorithm in PAUP; numbers on branches indicate support from Bayesian inference, maximum likelihood, neighbor joining and maximum parsimony algorithm; only significant values are displayed (B.I. > 0.95, > 50 for others); for reading convenience, values of 1.00 and 100 were replaced by *, high support of algorithms within clades is indicated by line width (see caption in phylogeny for details); boxes labeled a – f indicate accessions that exhibit identical barcodes, but differing SSU; "x 2" indicates accession that represents 2 clones with identical SSU&barcode;

The phylogenetic tree was inferred by maximum likelihood analyses of 1747 aligned positions of 112 taxa using PAUP* 4.0b10. TrN+G+I was chosen as best evolutionary model (base frequencies: A 0.23, C 0.26, G 0.29, T 0.22; rate matrix: A-C 1.0000, A-G 1.7365, A-T 1.0000, C-G 1.0000, C-T 3.7417, G-T 1.0000) with the proportion of invariable sites (I= 0.6133) and gamma distribution parameter (G= 0.6145). Posterior Probabilities (>95%; bold; calculated by MrBayes 3.1.2 using the covariation settings (4 million generations, trees from 12580 – 40000)) as well as bootstrap values (>50%) of the maximum likelihood (100 replicates; bold italic), neighbor-joining (1000 replicates; italic), and maximum parsimony (1000 replicates; not bold) are given in the tree (PP/ML/NJ/MP). No outgroup was used.

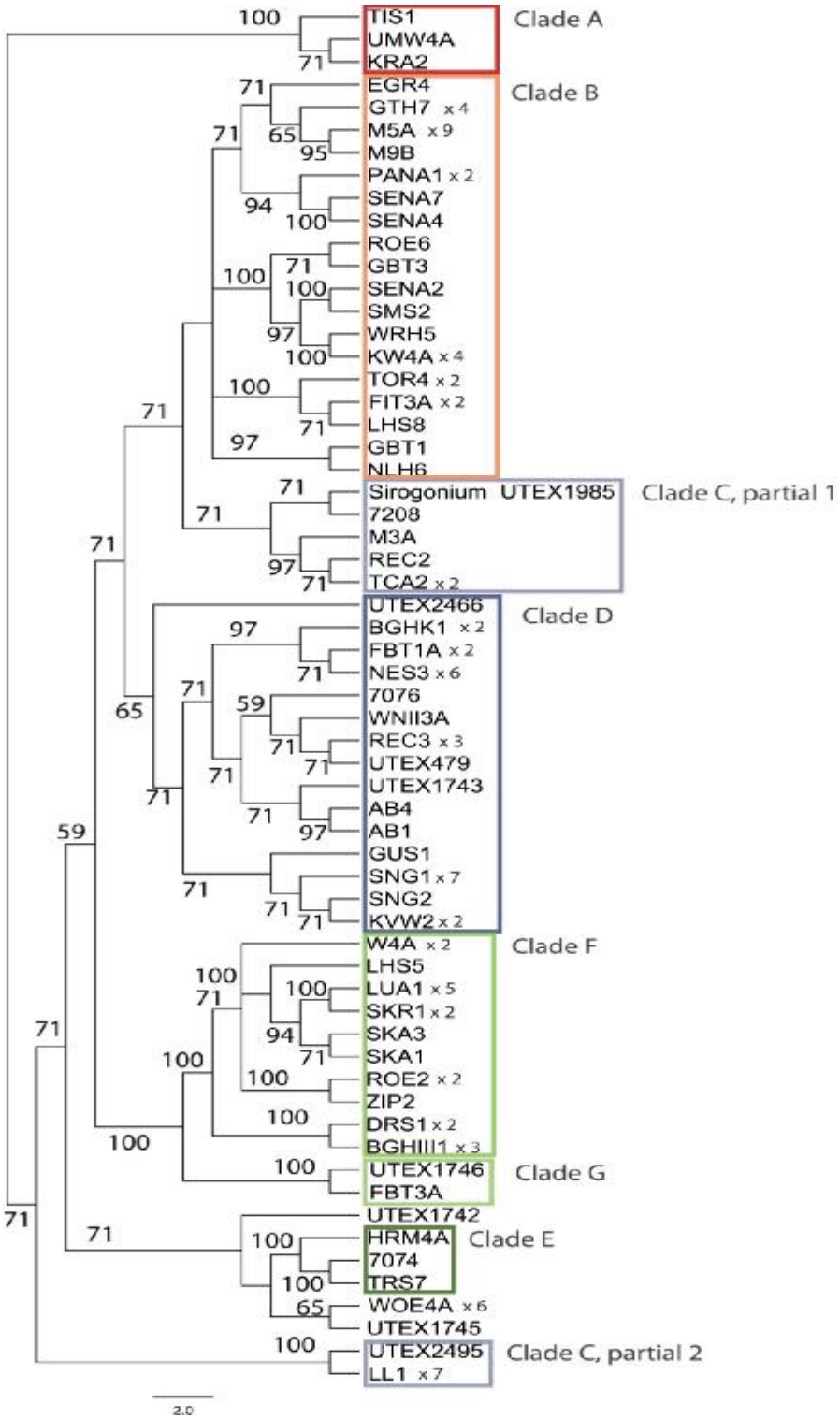


Figure 2: Phylogeny based on complete ITS2 alignment, calculated using Profdist; numbers on branches indicate support for branching; only significant values (>50%) are displayed; "x 2" indicates that accession represents 2 identical clones; clades are marked with colored boxes; we tried to find same grouping of accessions as in SSU&barcode based phylogeny; only clade H was not recovered; clade grouping of clades pooling accessions not containing the 1506 group I intron (clade E, F, G &H) also not recovered;

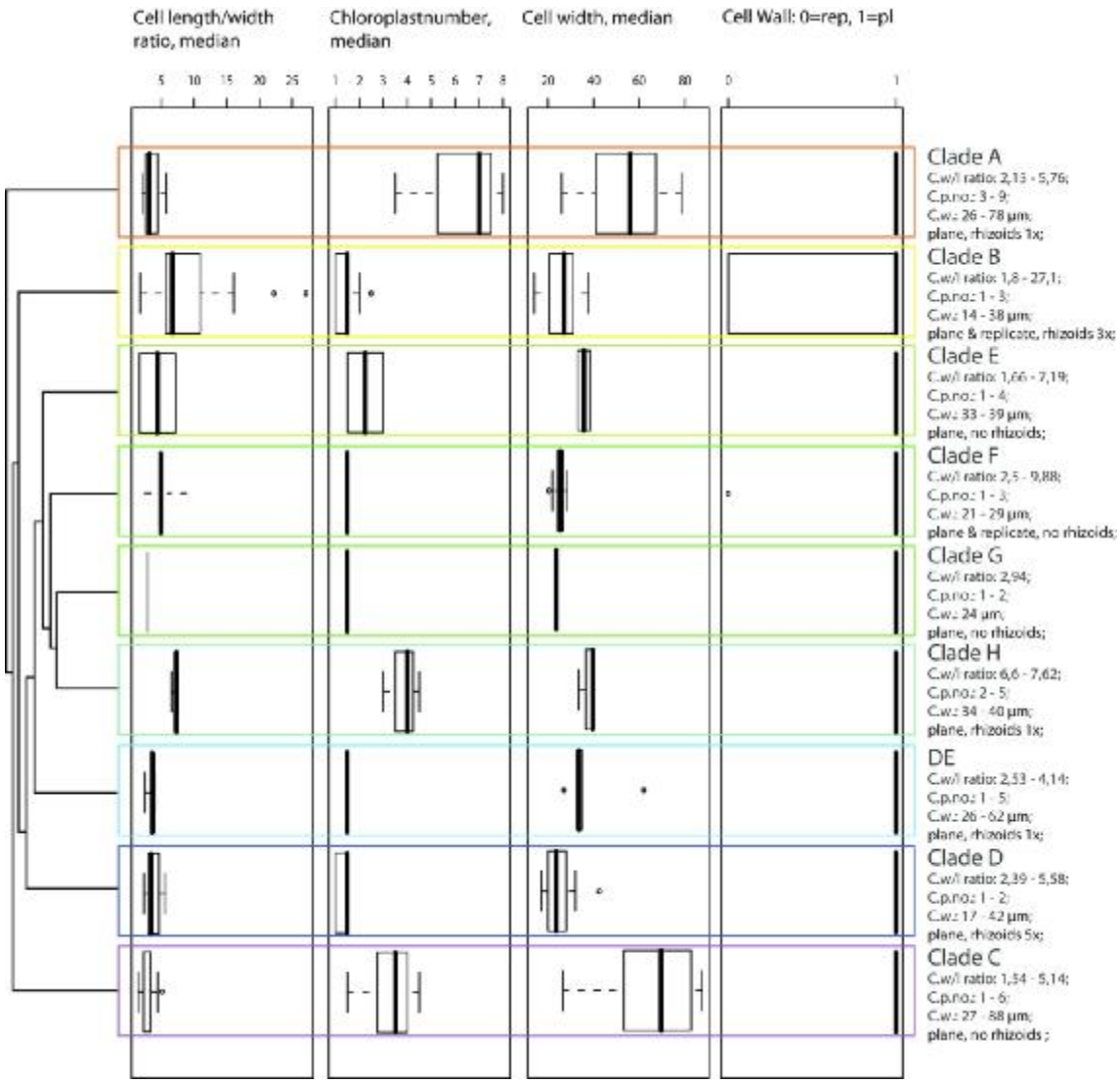


Figure 3: Tree structure on the left hand side is based on SSU&barcode phylogeny; box plot calculated using R; Cell length/width ratio, Chloroplast number and cell width in box plots are median values; unit of cell width is µm; cell end wall structure is either 1 = plane or 0 = replicate; values given with clades on the right hand side are actual minima and maxima of the respective characters; box plots are displayed with median, 95% range whiskers and outliers (circles); an overlap of boxes indicates no significant difference between clades;

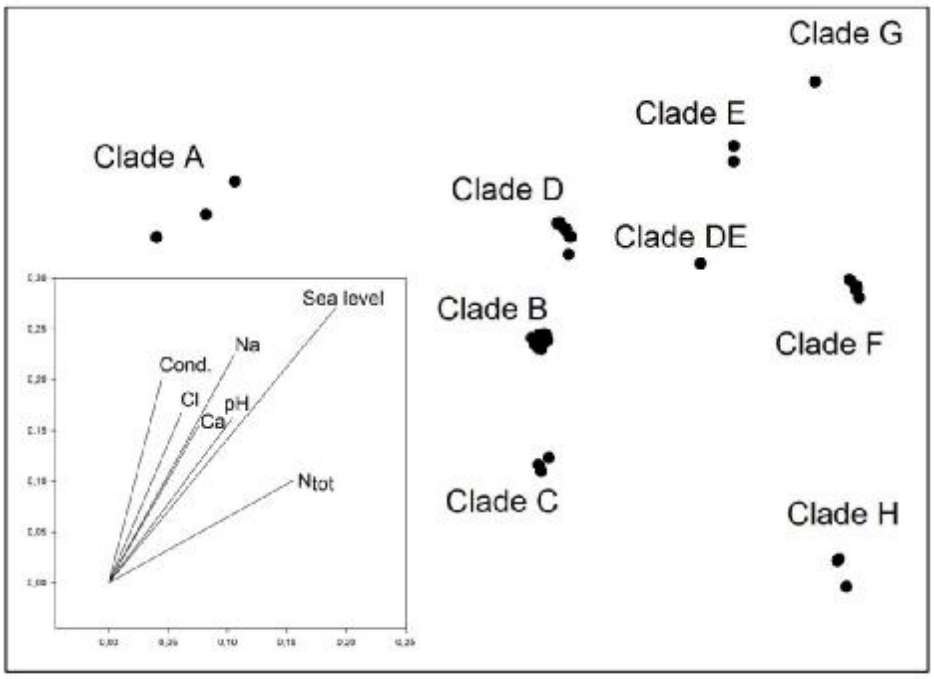


Figure 4: NMDS-plot based on the distance matrix calculated in PAUP. Correlations (>0,15; Pearson) of environmental data is shown in the small box on the lower left of the plot; environmental correlations of following parameters present: Cond.: conductivity; Cl: chlorine; Na: sodium; Ca: calcium; Ntot: level of total sodium; pH; sea level.

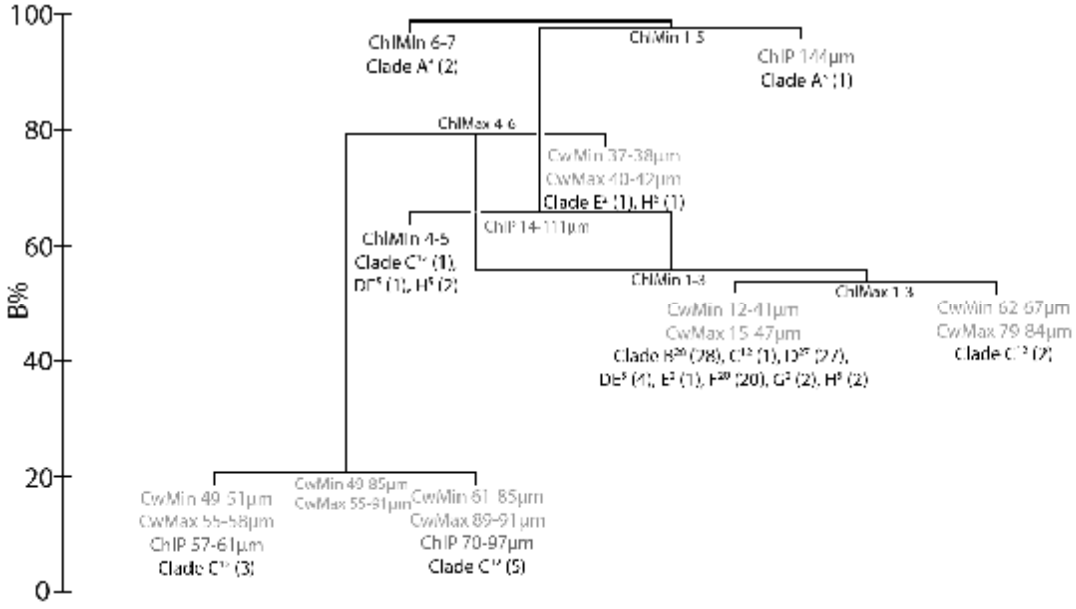


Figure 5: LINKTREE dendrogram showing significant morphological traits for pruning groups; number of strains involved in group given in brackets; total number of strains in clades given in superscript; in dark gray: ChlMin: minimum number of chloroplasts; ChlMax: maximum number of chloroplasts; in medium gray: ChlP: chloroplast periodicity; in light gray: CwMin: minimum cell width; CwMax: maximum cell width.

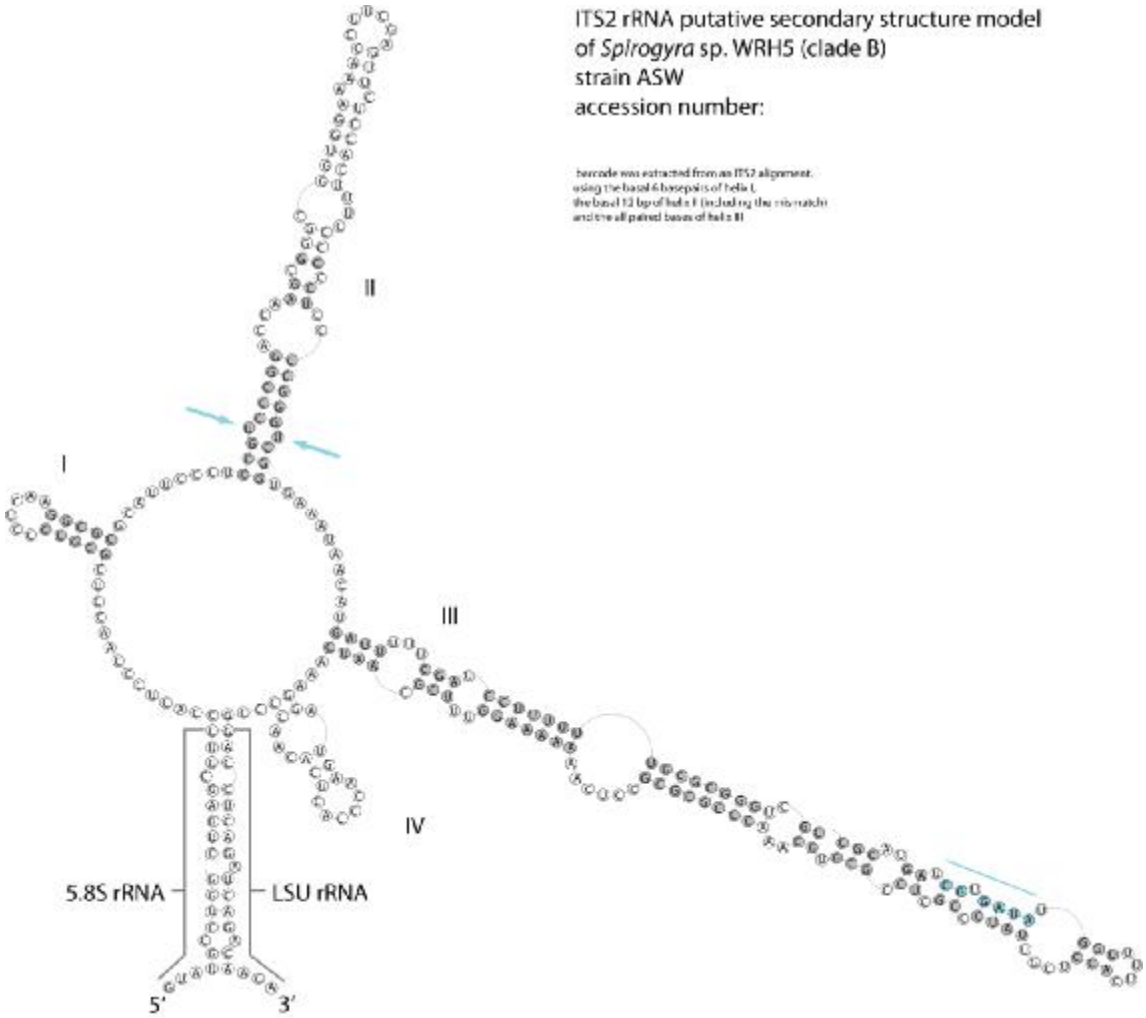


Figure 6: ITS2 rDNA putative secondary structure model of Spirogyra sp. WRH5 (clade B); bases included in barcode are marked in dark grey; flanking regions 5.8S rRNA and LSU rRNA are marked by brackets; pyrimidine mismatch (here: U-U) marked with blue arrows; conserved sequence at 5' end of helix III marked in blue.

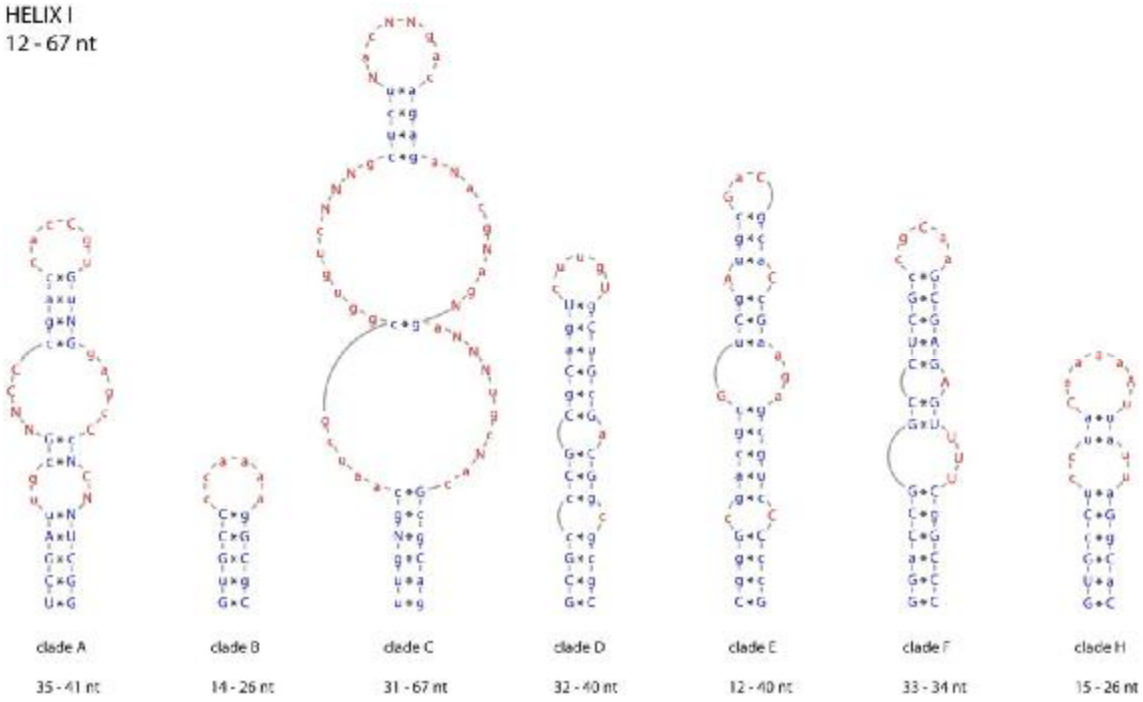


Figure 7: 50% consensus sequence & structure for helix I of clades based on complete ITS2 alignment; alignment and consensus calculated using MARNA – a capital letter indicates the full conservation of a nucleotide. A lowercase letter indicates a nucleotide conservation in at least 50% of all sequences. The consensus structures reflect the conservation of base –pairs in per cent occurring in all structures. Blue letters indicate base pairs; red letters bulges and loops.

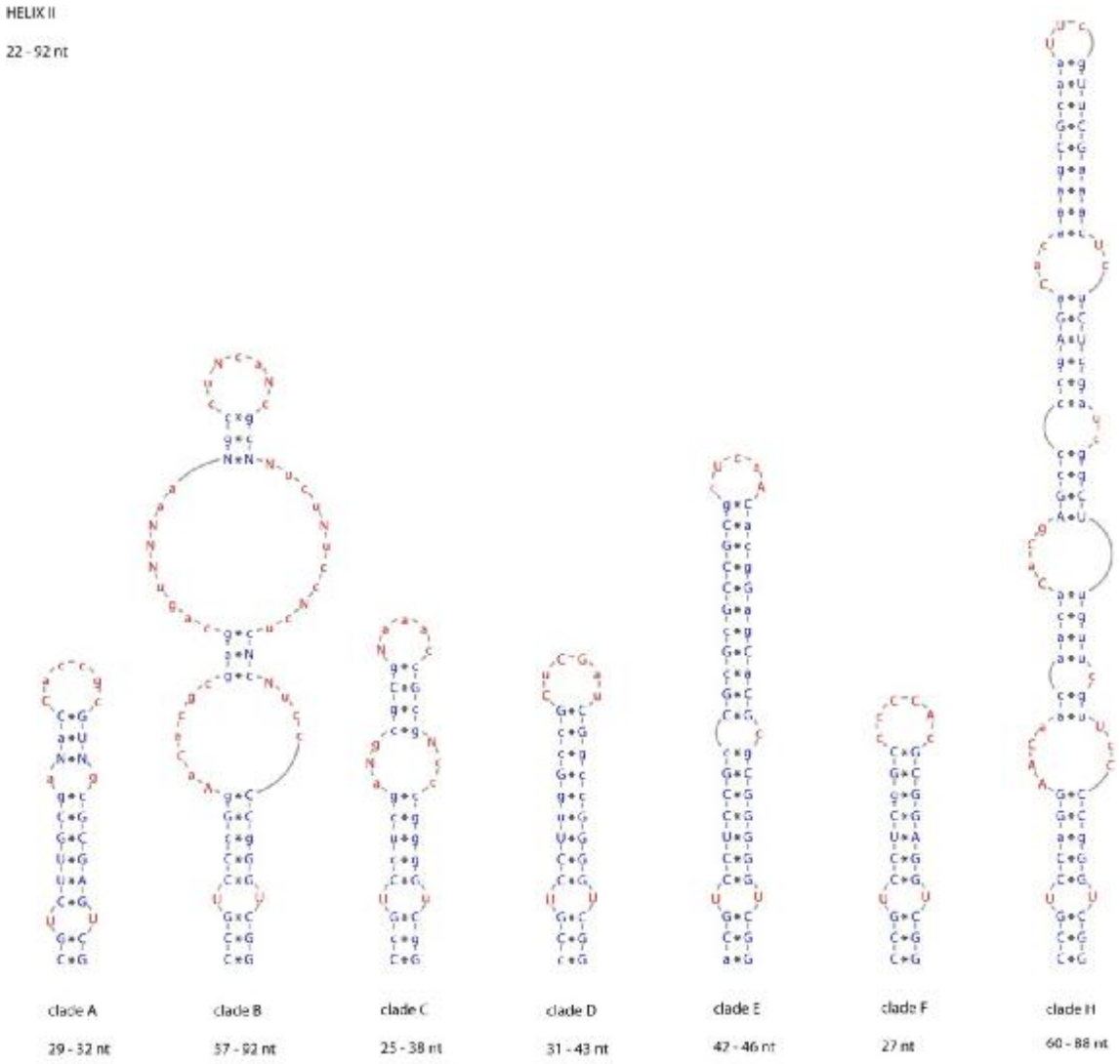


Figure 8: 50% consensus sequence & structure for helix II of clades based on complete ITS2 alignment; alignment and consensus calculated using MARNA – a capital letter indicates the full conservation of a nucleotide. A lowercase letter indicates a nucleotide conservation in at least 50% of all sequences. The consensus structures reflect the conservation of base –pairs in per cent occurring in all structures. Blue letters indicate base pairs; red letters bulges and loops.

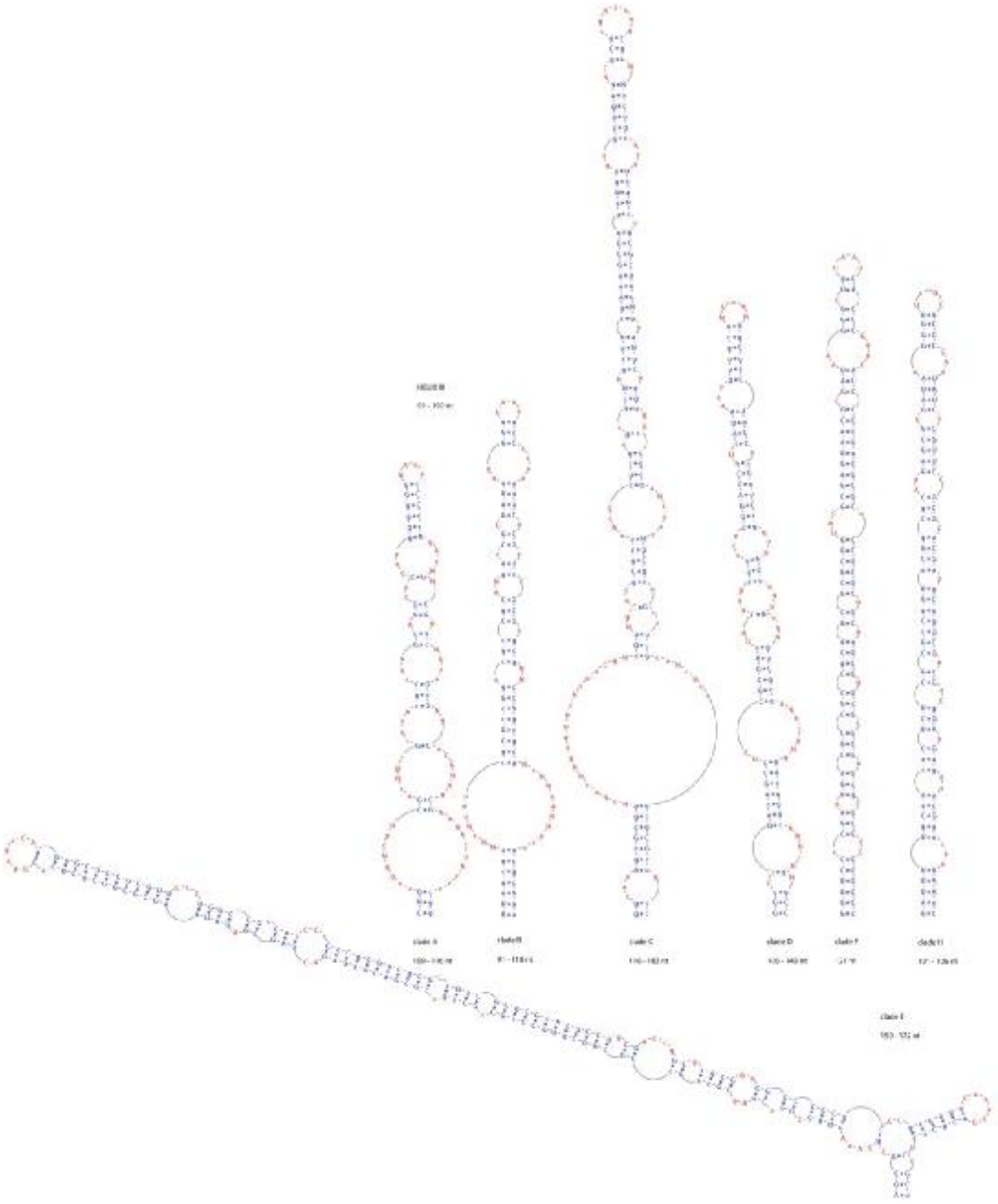


Figure 9: 50% consensus sequence & structure for complete helix III of clades based on complete ITS2 alignment; alignment and consensus calculated using MARNA – a capital letter indicates the full conservation of a nucleotide. A lowercase letter indicates a nucleotide conservation in at least 50% of all sequences. The consensus structures reflect the conservation of base –pairs in per cent occurring in all structures. Blue letters indicate base pairs; red letters bulges and loops.

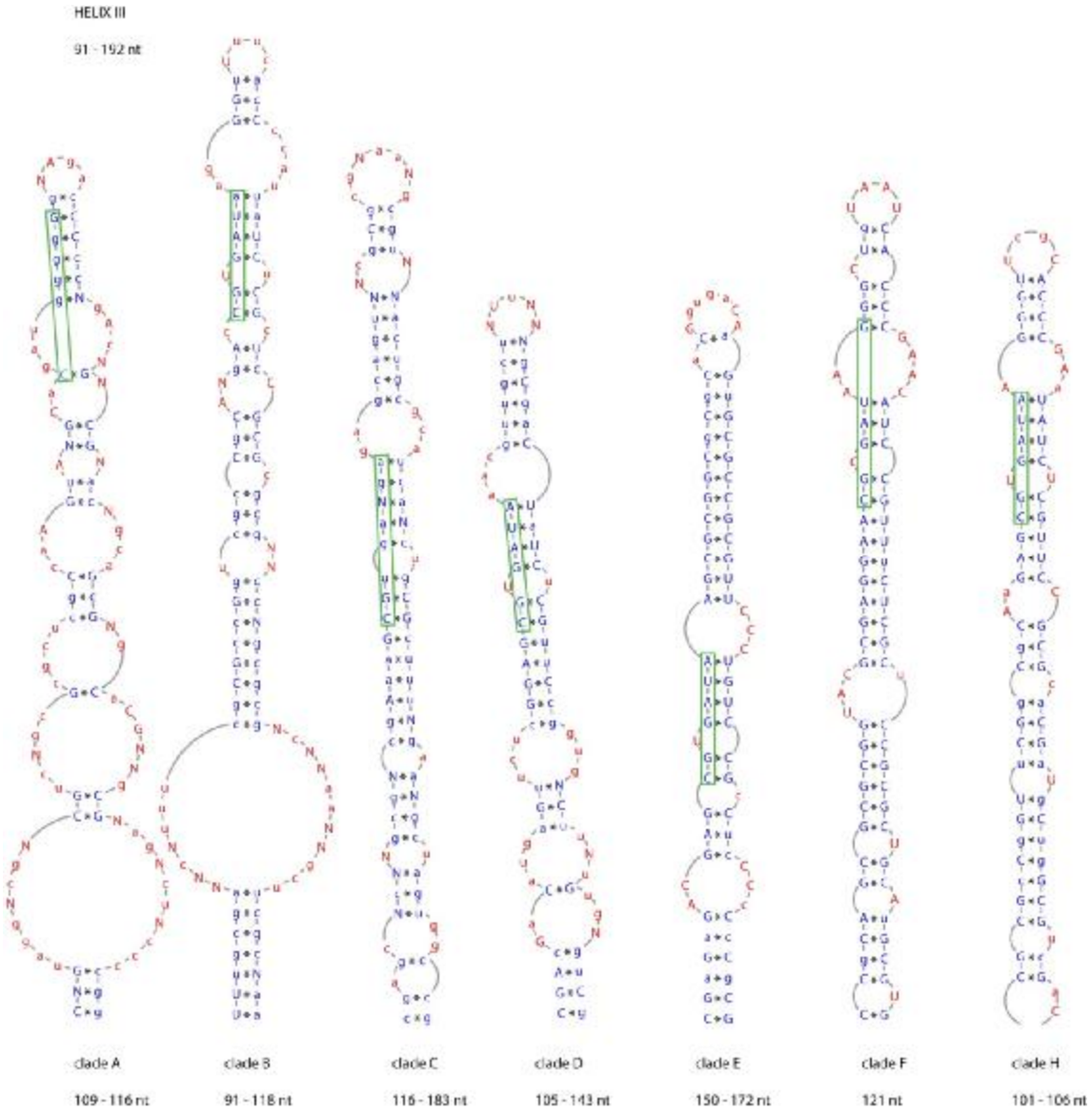


Figure 10: 50% consensus sequence & structure for 5' end of helix III of clades based on complete ITS2 alignment; alignment and consensus calculated using MARNA – a capital letter indicates the full conservation of a nucleotide. A lowercase letter indicates nucleotide conservation in at least 50% of all sequences. The consensus structures reflect the conservation of base –pairs in per cent occurring in all structures. Blue letters indicate base pairs; red letters bulges and loops. Conserved region is marked with green box.

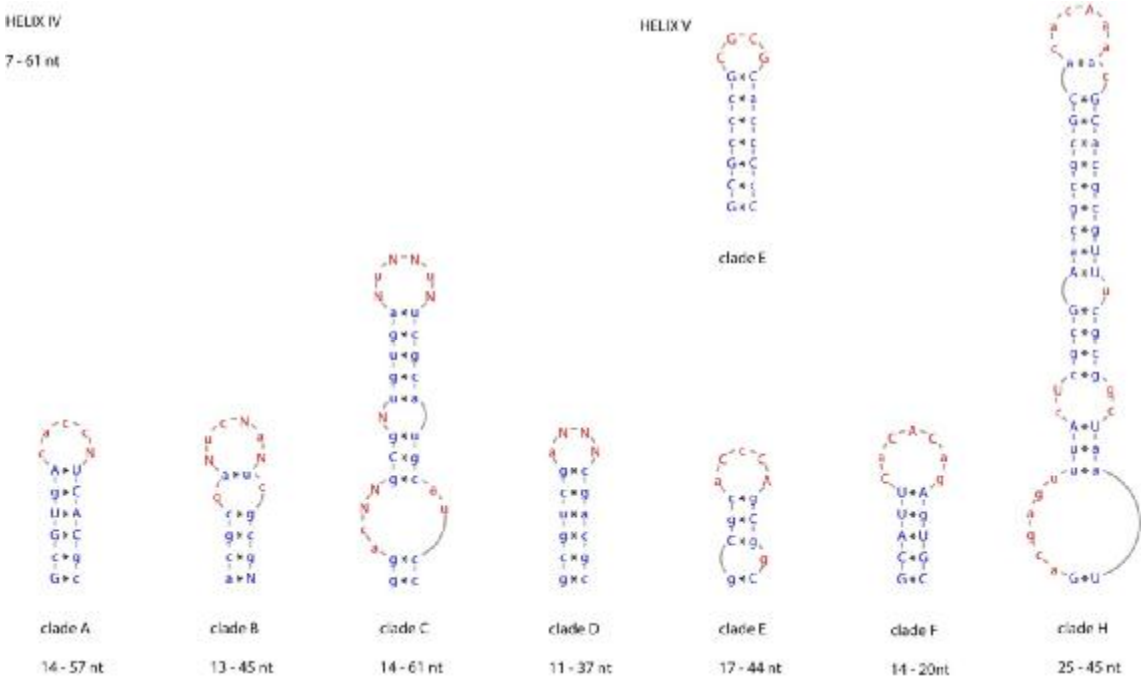


Figure 11: 50% consensus sequence & structure for helix IV and helix V of clades based on complete ITS2 alignment; alignment and consensus calculated using MARNA – a capital letter indicates the full conservation of a nucleotide. A lowercase letter indicates a nucleotide conservation in at least 50% of all sequences. The consensus structures reflect the conservation of base –pairs in per cent occurring in all structures. Blue letters indicate base pairs; red letters bulges and loops.

S1:

Contains 1-8 coded (see below) barcode of all sequences used.
Structure:
Clade_sequence
(first 6 base pairs of helix I)-(first 12 base pairs including mismatch of helix II)-(all pairs bases of helix III)

Pairing -> number/color
A-U -> 1
U-A -> 2
G-C -> 3
C-G -> 4
G-U -> 5
U-G -> 6
mismatch -> 7
deletion -> 8

A_KRA2

A_UMW4A

A_TIS1

B_M5A_MP926_MP923_M6A_S1Due2A_S1DD3A_M7_2_WBT2A_S1Due22A

B_KW4A_KW1A_KWA3_LSB4A

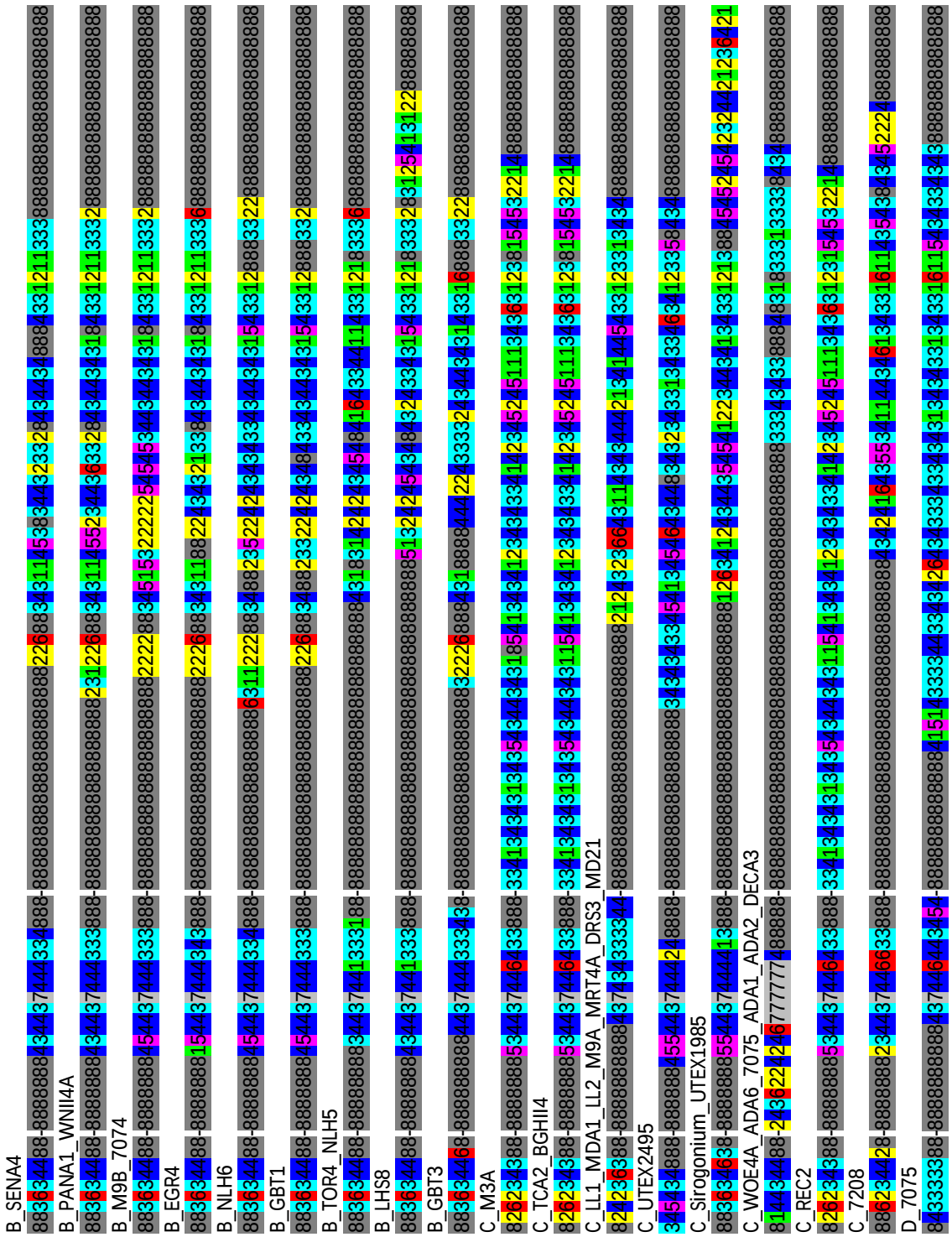
B_WRH5

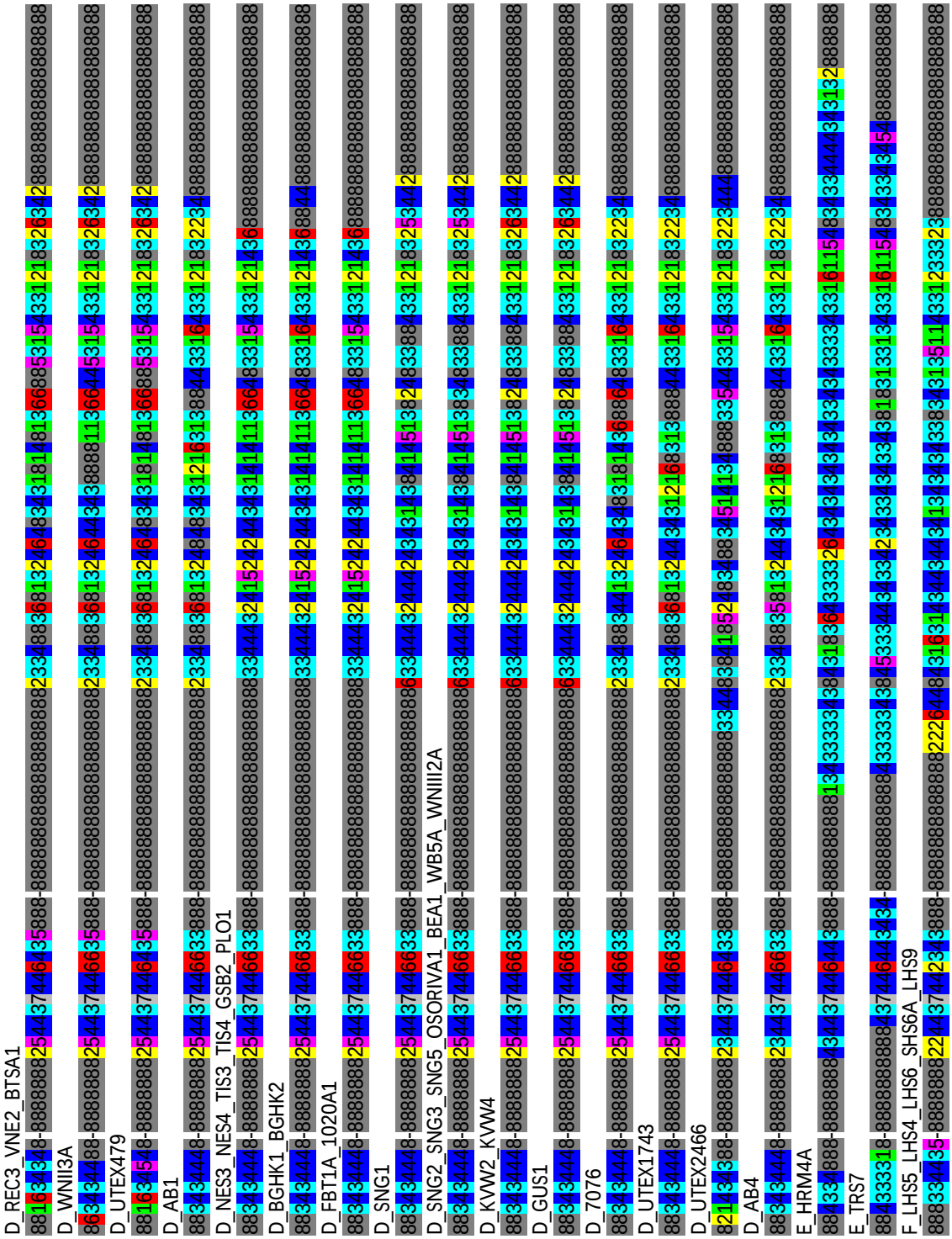
B_SENA2

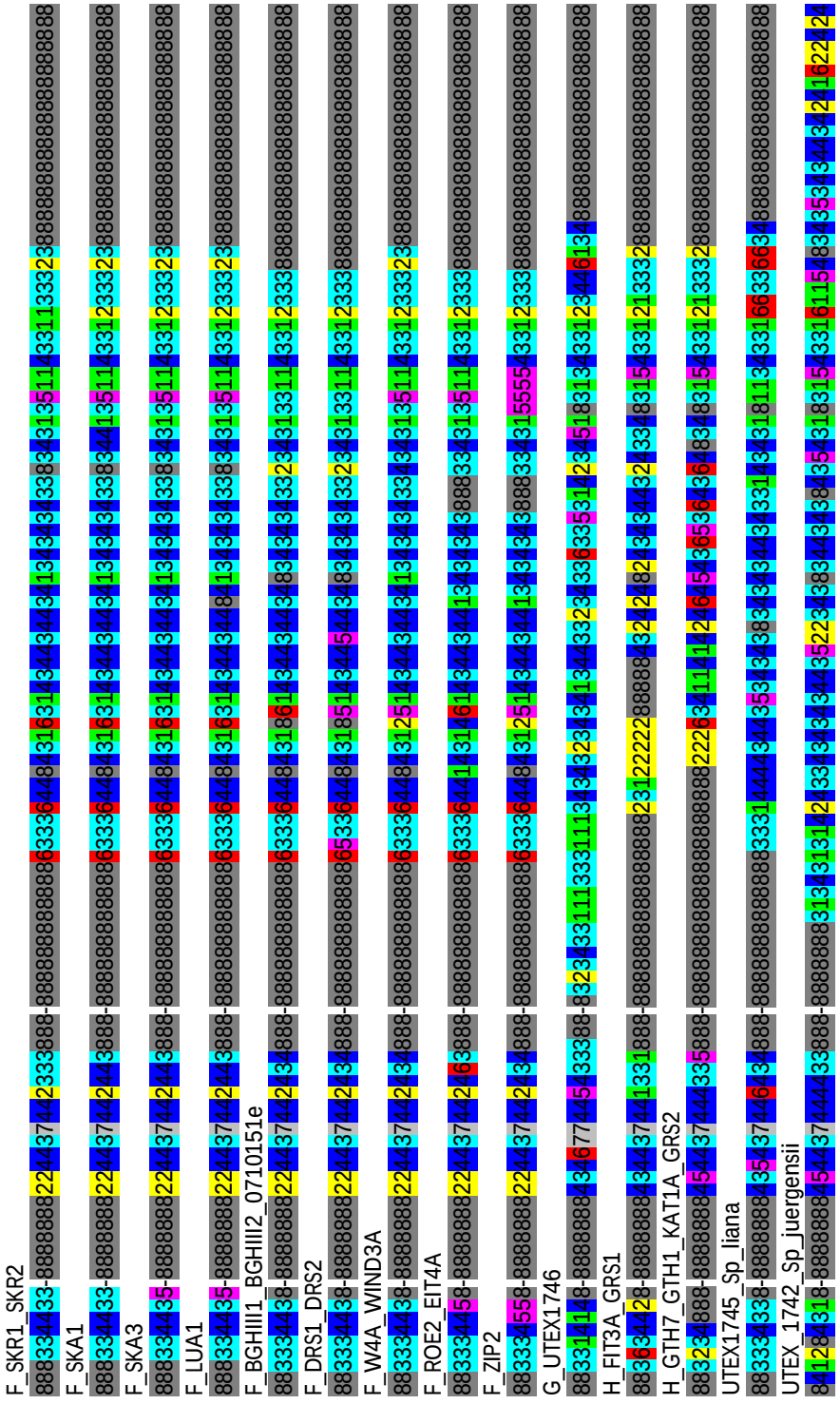
B_ROE6

B_SMS2

B_SENA7


B_SENA4
B_PANA1_WNII4A
B_M9B_7074
B_EGR4
B_NLH6
B_GBT1
B_TOR4_NLH5
B_LHS8
B_GBT3
C_M3A
C_TCA2_BGHII4
C_LL1_MDA1_LL2_M9A_MRT4A_DRS3_MD21
C_UTEX2495
C_Sirogonium_UTEX1985
C_WOE4A_ADA6_7075_ADA1_ADA2_DECA3
C_REC2
C_7208
D_7075


D_REC3_VNE2_BTSA1
D_WNII3A
D_UTEX479
D_AB1
D_NES3_NE54_TIS3_TIS4_GSB2_PLO1
D_BGHK1_BGHK2
D_FBT1A_1020A1
D_SNG1
D_SNG2_SNG3_SNG5_OSORIVA1_BEA1_WB5A_WNIII2A
D_KVW2_KVW4
D_GUS1
D_7076
D_UTEX1743
D_UTEX2466
D_AB4
E_HRM4A
E_TRS7
F_LHS5_LHS4_LHS6_SHS6A_LHS9


F_SKR1_SKR2
F_SKA1
F_SKA3
F_LUA1
F_BGHIII1_BGHIII2_07101051e
F_DRS1_DRS2
F_W4A_WIND3A
F_ROE2_EIT4A
F_ZIP2
G_UTEX1746
H_FIT3A_GRS1
H_GTH7_GTH1_KAT1A_GRS2
UTEX1745_Sp_liana
UTEX_1742_Sp_juergensii

Conclusions

Although the filamentous green alga *Spirogyra* is easily recognized and often used for teaching, it has largely been disregarded in research. Genetic studies in particular have been neglected. The work presented here revealed an unexpected picture of the hidden diversity of this genus. We combined environmental, morphological and genetic information retrieved from our field studies, cultivation work and sequencing. The focus was on genetic aspects using different phylogenetic methods and markers.

- We could confirm that *Spirogyra* is a monophyletic genus. Moreover, the genus *Sirogonium* is inseparable from *Spirogyra* (shared NHS pattern; alternative tree topologies tested; ITS2 secondary structure).

- We discovered an intron which exhibited slow evolution and characteristics of a 1506 group I intron in half of the clades (A – D - representing roughly 60% of the strains sampled); clades E – H do not have this intron and form a long branch within the genus, indicating a secondary loss in a common ancestor of these clades. The base composition differs considerably between 18s SSU rRNA and 1506 group I intron. The described intron is of typical structure with some deviations (such as A-rich bulge and P5 extension missing) that point to an overall high mutation rate and long evolutionary history (Chapter 1).

- Although exhibiting a very uniform morphology, *Spirogyra* shows a hidden genetic diversity. The genetic distances within this genus are much larger than expected, especially considering the low variability of vegetative morphology or the lack of distinctive characters of vegetative stages. Even dividing the genus into several genera could be argued based on our findings. The genus is sub-divided into eight clades (A – H) with strong support for each clade by all algorithms applied. Clades vary considerably from each other concerning number of members, branch length and genetics within clade variability and also exhibit unequal evolutionary rates (Chapter 2).

- In a polyphasic approach, the barcode extracted from the ITS2 region and the CBC criterion were used as delimiter to estimate 51 putative species in our data set of roughly 120 sequenced samples. Subsequently, we analyzed correlations among phylogeny, vegetative morphology and related environmental data. Morphology and phylogeny were found to coincide; correlations however were too weak to make reliable projections. No correlations were found between environmental data and phylogenetic information (Chapter 3).

Curriculum Vitae

Personal data

Name	Charlotte Chen (né Wöber)
Date of birth	03 May 1976
Contact	charlotte.chen@gmx.net; charlotte.woeber@univie.ac.at
Gender	female

Academic qualifications

2001 – 2004	Study at University of Vienna: Biology, specialization in ecology Core area: Vegetation and landscape ecology, Geographic information systems Master thesis: "Vegetation of the Göller, a mountain in the limestone foothills of the Alps, test case for objective methods in GIS" Supervisor: Ao. Univ. Prof. Dr. Gert Michael Steiner and Ass. Prof. Dr. Karl Reiter
1994 – 2001	Study at University of Vienna: Biology
1986 – 1994	BG/BRG Tulln

Research involvement and teaching

2006 – 2010	University of Vienna, Department of Limnology, Work group Phycology, scientific and research work in the context of the project "Species delineation and autecology of Spirogyra LINK 1820" funded by the Austrian Science fund (P18465-B03).
2003 & 2004	Tutor for the course „Application of geographic information systems".
2003 – 2006	University of Vienna, Institute of Ecology and Nature Conservation, Department of Hydrobotany, scientific and research work for: midcc (FFG7-Project, „multifunctional integrated study danube corridor and catchment"); Tisza-River (FFG7-Project) PI: Ao. Univ.-Prof. Dr. Georg Janauer, technical coordination: D.I. Norbert Exler.

September 2003: Survey and mapping of macrophytes, Breg&Brigach, Germany.

February 2005: MIDCC Project meeting, Mosonmagyarovar, Hungary.

July 2006: research trip, Iffeldorf, Germany

April 2007: research trip, Rostock, Hamburg, Germany

March 2009: research trip, Brown University, Rhode Island, USA

Conferences and seminars

C. Wöber, B. Schmidt, N. Exler, G. Janauer. Results from field investigation Breg &Brigach. MIDCC Project workshop, Mosonmagyarovar, Hungary.

C. Wöber, M. Schagerl. Artabgrenzung und Autökologie von Spirogyra LINK 1820. Abteilungsseminar Department für Meeresbiologie

C. Wöber, M. Schagerl. Species delineation and autecology of the genus Spirogyra LINK 1820. Seminar des Biodiversitätszentrums Rennweg

C. Wöber, R. Hainz, M. Schagerl. Autökologie der Gattung Spirogyra LINK 1820. 12. Tagung der DBG Sektion Phykologie 2008, Wittenberg, Deutschland.

C. Wöber, R. Hainz, M. Zwirn, G. Abo El Soad, M. Schagerl. Species delineation and autecology of Spirogyra – a project outline. SIL Austria Meeting 2008, Lunz am See.

C. Chen, M. Schagerl. Group I introns in the genus Spirogyra. Seminar der Arbeitsgruppe Phykologie

C. Chen, M. Schagerl. A comparison of ITS2 secondary structures in Spirogyra LINK 1820, Work in progress. Seminar der Arbeitsgruppe Phykologie

C. Chen, M. Schagerl. Latest results from the Spirogyra project, 18s SSU rDNA vs. Morphology, 1st DELI Workshop, Department für Limnologie, Wien.

C. Chen, M. Schagerl. Spirogyra LINK 1820, SSU Genotypes vs. Morphotypes. 13. Tagung der DBG Sektion Phykologie 2010, Insel Reichenau, Deutschland.

C. Chen, M. Schagerl. Species delineation and autecology of the genus Spirogyra LINK 1820. NOBIS AUSTRIA Young Researchers' Day, 2014 Wien.

Fields of interest

Geographical information systems (spatial analyses); Phycology (Algal isolation and culture, algal ecology, algal morphology); Phylogeny (general aspects of genetic relatedness); Aquatic ecology

Publication records

C.Chen,M.HJ. Barfuss, T. Pröschold, M. Schagerl. Hidden genetic diversity in the green alga Spirogyra (Zygnematophyceae, Streptophyta). BMC Evolutionary Biology 2012, 12:77. [ISI Impact factor 3.29]

Zwirn, M., Chen, C.,Wagih Moustafa Abo El-Soud, G., Schagerl, M. 2013. Induction of sexual reproduction in the genus Spirogyra (Zygnematophyceae, Streptophyta) and related problems. Fottea 13: 77-85. [ISI Impact factor 1.37]

Chen, C. & Schagerl, M. 2012. Slow evolution of 1506 group I intron in Spirogyra LINK 1820 (Zygnematophyceae, Streptophyta), a fast evolving lineage in the Zygnemataceae. Fottea 12: 255-272. [ISI Impact factor 1.37]

Schagerl, M., R. Hainz & C. Wöber, 2009. Autecology and species delineations of the green filamentous alga Spyrogyra LINK. Phycologia 48: 115. [ISI Impact factor 2.00]

Hainz, R., C. Wöber & M. Schagerl, 2009. The relationship between Spirogyra (Zygnematophyceae, Streptophyta) filament type groups and environmental conditions in Central Europe. Aquatic Botany 91: 173-180. [ISI Impact factor 1.516]

Acknowledgements

The authors thank the Austrian Science fund for financing the project “Species delineation and autecology of Spirogyra LINK 1820” (P18465-B0). We gratefully acknowledge Hubert Kraill (University of Vienna) for the ion analysis. Uta Raeder, Viktoria Tscherne (Limnologische Station Iffeldorf – TU München), Ulf Karsten, Henning Baudler, Jana Wölfel (Insitut für Biowissenschaften – University of Rostock), Dieter Hanelt, Ludwig Kies (Fachbereich Biologie - University of Hamburg) kindly supported the collections in Germany. Further thanks go to Tod F. Stuessy, Rose Samuel, Michael Barfuss, Elfriede Grasserbauer, Verena Klejna and Gudrun Kohl (Department of Botany and Biodiversity Research, University of Vienna) for their valuable help with the sequencing lab work.
I personally want to thank my supervisor Michael Schagerl for giving me the opportunity to participate in this research project and take part in illuminating many fascinating scientific questions. His patience, questions and motivation encouraged me throughout the process. Thank you Michi, I know it was not always easy. I am very grateful for the long lasting support, friendship, friendly cooperativeness and guidance from Annette Coleman (Brown University of Providence, Rhode Island, USA) and Thomas Pröschold (University of Vienna). Thank you for all the discussions and scientific input. Special thanks go to my colleagues and friends of the Phycology team at the University of Vienna, who not only participated in the collection trips and helped with lab work, but without whom I would never have completed this work and who have endured much with me – Roland Hainz, Melanie Zwirn and Martin Gruber. Thank you making field trips fun. I want to thank all people whom I meet in this time that I may not have mentioned namely.
I am also gratefull for my family, my husband, and two lovely daughters who went trough all ups and downs of this work with me. I want to dedicate this work to late mother Margarita Wöber who died on July 21th 2014.