

RNA consensus structures for inferring green algal phylogeny: A three–taxon analysis for *Golenkinia/Jenufa*, Sphaeropleales and Volvocales (Chlorophyta, Chlorophyceae)

Verena CZECH¹ & Matthias WOLF^{1*}

¹Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany;

*Corresponding author e–mail: matthias.wolf@biozentrum.uni-wuerzburg.de

Abstract: Background and Aims: “Order please!” was a recently published, very nice and apt title for a persistent problem in green algae phylogenetics. The green algal class Chlorophyceae comprises two clades (SV and OCC) made up of five orders (Sphaeropleales, Volvocales, Oedogoniales, Chaetopeltidales and Chaetophorales). A variety of data and methods have shown that one further group (*Golenkinia+Jenufa*) cannot be unambiguously placed among the five orders. In addition, concerning Sphaeropleaceae and/or Treubarinia the monophyly of Sphaeropleales and Volvocales remain unresolved.

Methods: IUPAC (International Union of Pure and Applied Chemistry) encoded consensus sequences for different 18S rDNA data sets as well as corresponding 18S rDNA consensus structures were assembled for Sphaeropleales, Volvocales (with or without Treubarinia) and a *Golenkinia+Jenufa* assemblage to independently infer phylogenetic relationships in a three–taxon analysis.

Using a known template structure, individual 18S rDNA secondary structures were predicted by homology modelling. Sequences and their individual secondary structures, automatically encoded by a 12–letter alphabet (each nucleotide with its three structural states, paired left, paired right, or unpaired), were simultaneously aligned; consensus structures and IUPAC encoded consensus sequences were read out from the different sequence–structure alignments.

Key Results: In contrast to previous studies using 18S rDNA data, results of this study corroborate chloroplast data and strongly support a sister group relationship between *Golenkinia+Jenufa* and the Sphaeropleales. The *Golenkinia+Jenufa* assemblage shows 330 matches to the Sphaeropleales (sequence–structure consensus) but only 214 matches to the Volvocales (sequence–structure consensus). Phylogenetically informative nucleotides are highlighted and visualized in their structural context taking into account structural domains (I–IV) and hypervariable regions (V2–V9). Rooting the three–taxon scenario remains difficult because the extremely long branches of *Golenkinia* and *Jenufa* are attracted to a chosen outgroup, reducing bootstrap support in any conceivable four–taxon tree obtained by parsimony or profile–neighbor–joining analysis.

Conclusions: This algal case study of 18S rDNA consensus data (IUPAC encoded sequences and consensus structures) coupled with profile distances between groups of sequences, demonstrated that a phylogenetic problem can be reduced to a three–taxon analysis.

Key words: 18S, algae, Chlamydomonadales, consensus, phylogeny, profile distances, RNA, secondary structure, Treubarinia, V4 region

INTRODUCTION

The Chlorophyceae are one of the classes of green algae, distinguished mainly on the basis of ultrastructural morphology. Despite a large number of unifying features, a considerable amount of variation exists among members of the class (LELIAERT et al. 2012). Chlorophyceae consist of two clades: one clade contains Sphaeropleales and Volvocales (or Chlamydomonadales) (SV–clade) and one clade comprises Oedogoniales, Chaetopeltidales

and Chaetophorales (OCC–clade) (e.g. TURMEL et al. 2008). The monophyly of each order was well supported by molecular (nuclear and chloroplast) evidence (BUCHHEIM et al. 2001, 2012, 2017; WOLF et al. 2002, 2003; KELLER et al. 2008; TURMEL et al. 2008; LELIAERT et al. 2012; TIPPERY et al. 2012; HODAČ 2012, 2016; FUČÍKOVÁ et al. 2014, 2016, 2019; LEMIEUX et al. 2015; PROCHÁZKOVÁ et al. 2015; SONG et al. 2017; FANG et al. 2018). Nonetheless, phylogenetic relationships within and between these higher–ranked taxa are not fully

understood despite the intensive research of the last decade. In particular, *Jenufa* Němcová, M. Eliáš, Škaloud et Neustupa (NĚMCOVÁ et al. 2011) and/or *Golenkinia* Chodat (CHODAT 1894), have been assigned to either Sphaeropleales (LEMIEUX et al. 2015; BUCHHEIM et al. 2017; FANG et al. 2018; HE et al. 2018; FUČÍKOVÁ et al. 2019) or Volvocales (WOLF et al. 2003; HODAČ 2012, 2016; PROCHÁZKOVÁ et al. 2015; BUCHHEIM et al. 2017; SONG et al. 2017; FUČÍKOVÁ et al. 2019) – depending on the marker and with varying support. A few investigations left the position of *Jenufa*+*Golenkinia* completely unresolved (cf. NĚMCOVÁ et al. 2011; WOLF & BUCHHEIM 2018; FUČÍKOVÁ et al. 2019), a situation also known for Treubarinia, in most cases related to Volvocales, but with low support (cf. BUCHHEIM et al. 2017).

The latest update comes from FUČÍKOVÁ et al. (2019). Fuciková et al. sequenced chloroplast genomes to compile a data set of 58 protein-coding genes of 68 chlorophycean taxa. Concatenated analyses yielded at least two well-supported phylogenies: nucleotide data supported the traditional classification with the inclusion of the enigmatic Treubarinia into Sphaeropleales *sensu lato*. However, amino acid data yielded equally strong support for Sphaeropleaceae as sister to Volvocales, with the rest of the taxa traditionally classified in Sphaeropleales in a separate clade, and Treubarinia as sister to all of the above. Single-gene and other supplementary analyses indicated that the data have low phylogenetic signal at these critical nodes. In other words, while sequence and gene order data support the deep split between the SV and OCC lineages, multiple phylogenetic hypotheses are possible for Sphaeropleales *sensu lato*. To sum up, FUČÍKOVÁ et al. (2019) identified two strongly and unequivocally supported clades: Treubarinia and

Scenedesminia (core Sphaeropleales), as well as other smaller groups that could serve a practical purpose as named clades (e.g. Sphaeropleaceae, Microsporaceae and a *Golenkinia/Jenufa*-clade).

In almost all phylogenetic studies (not only in green algal phylogenetics) one of the most frequently used markers has been the 18S rDNA. Sequence alignments were sometimes guided by secondary structure information or sequence–structure information (encoded into a new alphabet) was used simultaneously to infer real sequence–structure alignments. The structures were predicted either by folding individual sequences, by homology modeling (using a known template) or consensus structures were created during the alignment process. Phylogenetic trees were obtained based on the primary sequence alignment, on structure–guided sequence alignments, or on encoded sequence–structure data. In a few rare cases, profile–distances were determined to reconstruct trees directly for higher ranked taxa using the various alignment methods. Elements of the different approaches are reviewed in KELLER et al. (2010) and WOLF et al. (2014). This study focuses on the phylogenetic position of *Jenufa aeroterrestica* Procházková et Neustupa (PROCHÁZKOVÁ & NEUSTUPA 2015) and *Golenkinia brevispicula* Hegewald et Schnepf (HEGEWALD & SCHNEPF 1984). For the first time, 18S rDNA consensus sequences and 18S rDNA consensus secondary structures — for different sets of Volvocales (with or without Treubarinia), for Sphaeropleales (including Sphaeropleaceae) and for a *Golenkinia*+*Jenufa* assemblage (swinging between both clades, Volvocales and Sphaeropleales) — were obtained and compared to infer rooted, three–taxa phylogenies for the clades in question. In a first for green algal research, the 18S

Table 1. 18S rDNA variability map with special reference to the V4 region. Three sequence–structure pairs (the consensus from Sphaeropleales, the consensus from Volvocales and the consensus from a *Golenkinia*+*Jenufa* assemblage) mapped to each other from the point of view of the *Golenkinia*+*Jenufa* sequence–structure consensus (Volvocales were plotted with or without Treubarinia and *Hafniomonas*).

	<i>Golenkinia</i> + <i>Jenufa</i> sequence–structure consensus		<i>Golenkinia</i> + <i>Jenufa</i> sequence–structure consensus (Volvocales without Treubarinia/ <i>Hafniomonas</i>)	
	18S	V4	18S	V4
Identical nucleotides in all three sequence–structure pairs	1083	107	1120	115
identical with/tendency towards Volvocales	214(100/114)*	30(9/21)*	237(107/130)*	38(11/27)*
identical with/tendency towards Sphae- ropleales	330(98/232)*	82(14/68)*	272(79/193)*	69(16/53)*
unique characters	48	7	48	8
no allocation possible (due to IUPAC)	116	32	103	27
alignment length	1791	258	1780	257

*(paired/unpaired nucleotides)

rDNA variability map we present highlights phylogenetically informative nucleotides in their structural context, taking into account structural domains (cf. STIEGLER et al. 1981) and hypervariable regions (cf. KI 2012) of the 18S rRNA gene. All consensus sequences and consensus structures are obtained from highly supported monophyletic entities derived from maximum likelihood based 18S rDNA sequence–structure phylogenies (BUCHHEIM et al. 2017). However, concerning *Golenkinia* and *Jenufa* our study is not about an exact phylogenetic position, but about an affinity towards either Sphaeropleales or Volvocales that eventually could be deduced from 18S rDNA sequence–structure data.

MATERIAL AND METHODS

Taxon Sampling, Secondary Structure Prediction, Sequence–Structure Alignment, Consensus Structures, Structure Visualization, Consensus sequences, Clade Selection and Phylogenetic Tree Reconstruction.

We have used the only available 18S rDNA sequence–structure data set for green algae from BUCHHEIM et al. (2017) and added the 18S rDNA sequence–structure information from *J. aeroterrestica* (supplementary Table S1, Data S1). The data set from BUCHHEIM et al. (2017) contains 68 representative sequence–structure pairs for the Chlorophyceae, originally obtained from THESES db, the algae 18S rDNA sequence–structure database for inferring phylogenies (MARIN RODRIGUES et al. 2017). Like secondary structures in THESES db, the secondary structure of the 18S rDNA from *J. aeroterrestica* was obtained by homology modeling (cf. WOLF et al. 2005) using *Heterochlorella luteoviridis* (Chodat) Neustupa, Němcová, M. Eliáš et Škaloud (NEUSTUPA et al. 2009) as template. The template secondary structure (without pseudoknots) was obtained from the Comparative RNA Web (CRW) site (CANNONE et al. 2002). For alignment, the four nucleotides multiplied by three states (unpaired, paired left and paired right) are encoded by a 12 letter alphabet (reviewed in WOLF et al. 2014). Using a 12×12 sequence–structure specific scoring–matrix (SEIBEL et al. 2006), global multiple sequence–structure alignments

were automatically generated in ClustalW2 1.83 (LARKIN et al. 2007) as implemented in 4SALE 1.7.1 (SEIBEL et al. 2006, 2008). Manually transcribed 100% consensus structures (an explanatory illustration can be found in the supplementary material, Fig. S1) were plotted using PseudoViewer (BYUN & HAN 2006) and further edited in CorelDRAW (www.corel.com). Phylogenetically informative nucleotides are highlighted and visualized in their structural context, taking into account structural domains (STIEGLER et al. 1981) and hypervariable regions (KI 2012). International Union of Pure and Applied Chemistry (IUPAC) encoded consensus sequences were read out using the software ALIGN (HEPPERLE 2004). A full length sequence–structure alignment or an alignment for the V4 region of the 18S rRNA gene (the latter one is typically used in biodiversity studies) were prepared. The IUPAC encoded consensus sequences, as well as corresponding consensus structures, were assembled (1) for the outgroup (two representatives from the OCC clade), (2) for Sphaeropleales, including Sphaeropleaceae (39 sequences, highly supported as monophyletic in BUCHHEIM et al. 2017), (3) for Volvocales (26 sequences, due to low, respectively high bootstrap support in BUCHHEIM et al. (2017), with or without Treubarinia and *Hafniomonas* (five sequences) and (4) for a *Golenkinia+Jenufa* assemblage (two sequences), in order to independently infer rooted phylogenetic relationships for the three clades in question. To circumvent long branch attraction phenomena (in particular concerning the outgroup) reconstructed four–taxa trees were further compared to midpoint rooted three–taxa trees. Phylogenetic trees for IUPAC encoded consensus sequences were inferred by seqboot (100 bootstrap replicates) and dnapsars (maximum parsimony, MP) as implemented in PHYLIP (FELSENSTEIN 1985; FELSENSTEIN 1989). The ambiguity inherent in the IUPAC encoded nucleotides precluded the use of likelihood (ML) or classic distance based methods, such as neighbor–joining (NJ). However, a sequence–structure profile neighbor–joining (PNJ) tree was determined according to MÜLLER et al. (2004) using ProfDistS (FRIEDRICH et al. 2005; RAHMANN et al. 2006; WOLF et al. 2008), in combination with 12 letter encoded sequences. We replaced the set of taxa forming a known subclade by single supertaxa (in our case Sphaeropleales, including Sphaeropleaceae, Volvocales (with or without Treubarinia and *Hafniomonas*), *Golenkinia+Jenufa*, and the OCC outgroup), each of which we represented by a sequence profile (GRIBSKOV et al. 1987).

Table 2. Phylogenetic position of the *Golenkinia+Jenufa* clade. The table shows the sister taxon to the *Golenkinia+Jenufa* clade depending on different approaches used in phylogenetic reconstruction. *Golenkinia+Jenufa* sister either to Sphaeropleales+Volvocales (S+V), or to the Sphaeropleales, or to the Volvocales, depending on the used marker (18S/V4) and the selected tree reconstruction method (maximum parsimony or profile neighbor–joining). All trees were rooted at the OCC–clade or by midpoint rooting. In tree calculations, the Volvocales were represented with or without Treubarinia and *Hafniomonas*.

		Maximum Parsimony (MP)		Profile Neighbor–Joining (PNJ)	
		4–taxa tree (rooted)	3–taxa tree (midpoint)	4–taxa tree (rooted)	3–taxa tree (midpoint)
with Treubarinia/ <i>Hafniomonas</i>	18S	S+V*	Sphaeropleales	Volvocales	S+V*
	V4	S+V*	Sphaeropleales	Sphaeropleales	Sphaeropleales
without Treubarinia/ <i>Hafniomonas</i>	18S	S+V*	Sphaeropleales	Volvocales	S+V*
	V4	S+V*	Sphaeropleales	Sphaeropleales	S+V*

*Sphaeropleales+Volvocales (in other words, *Golenkinia+Jenufa* is attracted towards the outgroup, represented by the OCC–clade)

A sequence profile is a stochastic model of a sequence family. According to MÜLLER et al. (2004), “[...] it can also be pictured as a fuzzy or “smeared-out” sequence. Formally, a profile is also a sequence, but it is composed of probability distribution vectors instead of characters. Each position k specifies its own nucleotide distribution $\alpha_k = (\alpha_{k,A}, \alpha_{k,C}, \alpha_{k,G}, \alpha_{k,T})$. Nucleotide sequences are special profiles, where each α_k is given by one of the unit vectors $A = (1, 0, 0, 0)$, $C = (0, 1, 0, 0)$, and so on. Because we are interested in the “center of gravity” of the sequence family distribution, we simply take the position-specific, relative nucleotide frequencies over all sequence family members. This results in a robust estimate that is independent of estimated subclade topologies [...]”

All analyses were preceded by overall proof-of-concept trees (NJ and ML), reconstructed as described in BUCHHEIM et al. (2017), to demonstrate the alliance between *Golenkinia* and *Jenufa* alongside the monophyly of Volvocales (with or without Treubarinia) and Sphaeropleales, respectively (data not shown).

RESULTS

Reconstructing overall proof-of-concept trees (NJ and ML) as described in BUCHHEIM et al. (2017), but including *Jenufa*, yields low support (<50) for monophyletic Sphaeropleales, monophyletic Volvocales (including Treubarinia and *Hafniomonas*) and high support (>75) for the monophyletic *Golenkinia+Jenufa* assemblage (data not shown). The phylogenetic position of the *Golenkinia+Jenufa* clade remains completely unresolved. Whereas in BUCHHEIM et al. (2017) *Golenkinia* sisters to either Volvocales (NJ) or Sphaeropleales (ML), in this study (data not shown) *Golenkinia+Jenufa* are related to the Volvocales (NJ and ML), but with low support. However, if one superimposes the sequence-structure consensus for each of Sphaeropleales, Volvocales and *Golenkinia+Jenufa*, a different picture emerges. The *Golenkinia+Jenufa* sequence-structure consensus shows 330 matches to the Sphaeropleales (sequence-structure consensus, Fig. 1, Table 1, Data S2) but only 214 matches to the Volvocales (sequence-structure consensus, Fig. 1, Table 1). With 82 matches to the Sphaeropleales and only 30 matches to the Volvocales, the V4 region in structural domain II of the 18S rRNA gene provides particularly strong support for a sister group relationship between *Golenkinia+Jenufa* and the Sphaeropleales (Fig. 1, Table 1). Phylogenetic trees inferred from consensus data corroborate the latter alliance in particular for midpoint rooted three-taxon trees obtained by parsimony (MP) (Fig. 2a, Table 2). Profile-Neighbor-joining (PNJ) yields this midpoint rooted alliance only for the V4 region (Fig. 2a, Table 2). Rooted four-taxon trees (most with only low to moderate support) show either a sister group relation between *Golenkinia+Jenufa* and the Sphaeropleales (PNJ, when only the V4 region is used, Fig. 2b, Table 2), or a sister group relation between *Golenkinia+Jenufa* and the Volvocales (PNJ, when full length sequences are used, Fig. 2b, Table 2); or the *Golenkinia+Jenufa* clade

(probably because of long branch attraction, LBA) sisters to Sphaeropleales+Volvocales (in MP analysis, Fig. 2b, Table 2). Similar results were obtained for analyses exclusive of Treubarinia and *Hafniomonas* (Fig. 2b, Table 2). LBA must be taken into account, since the number of identical nucleotides and IUPAC encoded characters (influenced by the number of sequences in each clade) between *Golenkinia+Jenufa* and the OCC-clade (outgroup) is comparable to the number of identities between *Golenkinia+Jenufa* and Volvocales or Sphaeropleales (1454, 1200, and 1272 respectively).

DISCUSSION

On the one hand, this algal case study shows that 18S rDNA consensus data (IUPAC encoded sequences and consensus structures) and profile distances between groups of sequences can be used to reduce a phylogenetic problem to the essentials of a three-taxon analysis. On the other, rooting the three-taxon scenario for the clades in question remains problematic. The extremely long branches of *Golenkinia* and *Jenufa* are attracted to a chosen outgroup, reducing bootstrap support in any conceivable four-taxon tree obtained by parsimony or profile neighbor-joining analysis, both of which are known to be susceptible to LBA. Due to the ambiguity of IUPAC encoded characters, unfortunately, maximum likelihood methods cannot be used at present. Anyway, LBA is an indicator of noisy data, which in turn contradicts the data obtained for the V4 region in domain II of the 18S rDNA. A majority of the various assessments for the V4 region (258 nucleotides) support an alliance of the Sphaeropleales with a *Golenkinia+Jenufa* assemblage (Figs 1, 2). In this context, the V4 region could be treated as a synapomorphic character for Sphaeropleales and a *Golenkinia+Jenufa* assemblage (Fig. 1). In addition, analyses of the V4 region clearly corroborate recent studies that used chloroplast data to infer green algal phylogeny (cf. LEMIEUX et al. 2015; FANG et al. 2018; HE et al. 2018; FUČIKOVÁ et al. 2019). Although our analysis, like so many others, does not provide a definitive placement for the *Golenkinia+Jenufa* clade, it shows that 18S rDNA and chloroplast data are not necessarily in conflict. Using consensus data, the synapomorphic V4 region of the 18S rRNA gene as well as midpoint rooted 18S rDNA trees without an outgroup provide an interpretation of the 18S rDNA data that favors an alliance between the Sphaeropleales and the *Golenkinia+Jenufa* clade. In our study, the secondary structure helped us to locate a sequence region (V4) with a clear and unambiguous phylogenetic signal. Nevertheless, this study addresses a complex problem surrounding ancient divergences, long branches and incomplete taxon sampling; uncertainty is to be expected, in particular because of homoplasy in sequence data. Further studies could additionally focus on homoplasy in structural changes. Finally yet importantly, we hope that future studies will further strengthen

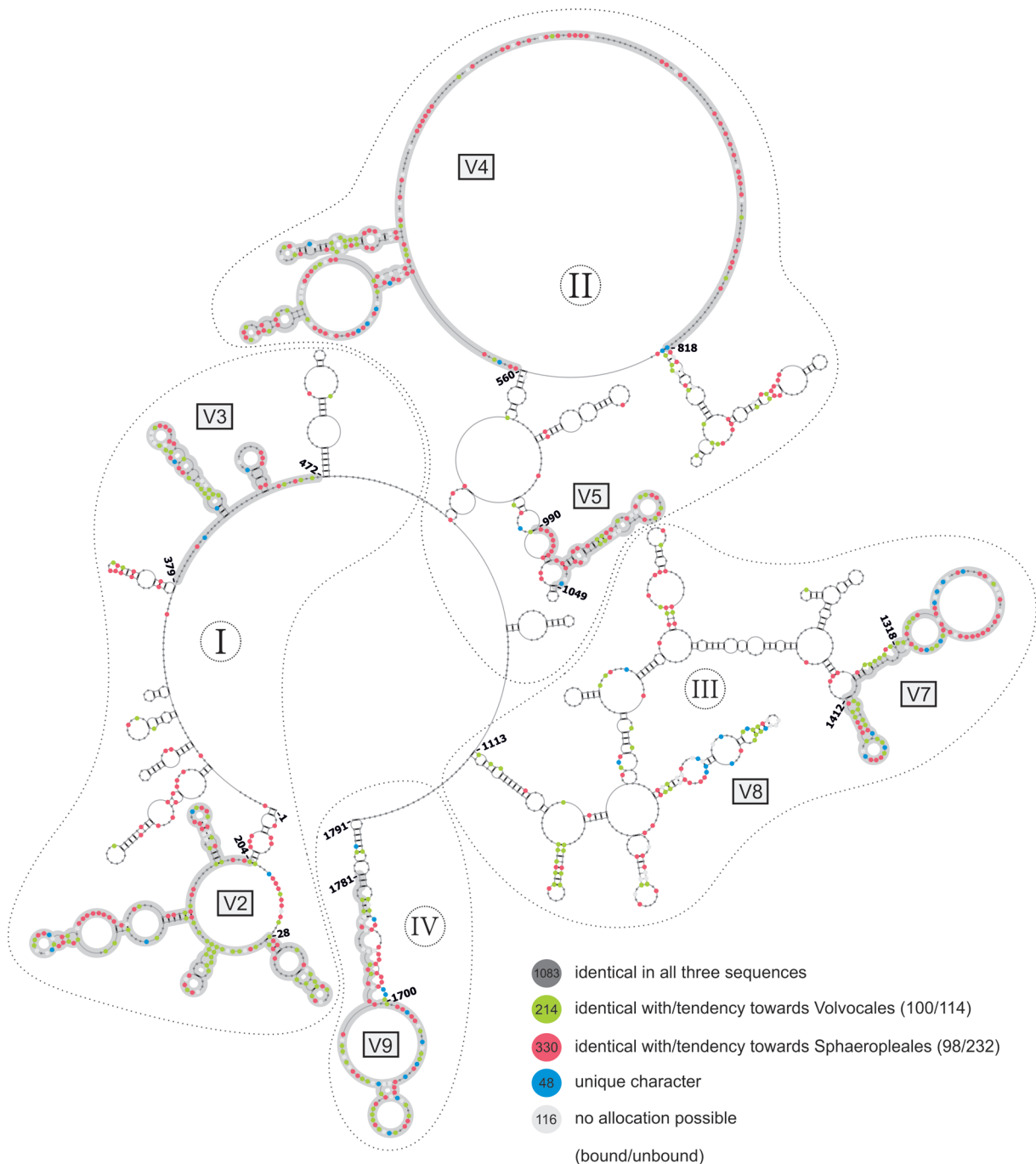


Fig. 1. 18S rDNA variability map. The figure was plotted using PseudoViewer (BYUN & HAN, 2006) and further edited in CoreIDRAW (www.corel.com). The figure shows the IUPAC encoded consensus sequence and the 100% consensus secondary structure from *Golenkinia+Jenufa* aligned against two further consensus sequences and consensus structures each from Sphaeropleales and from Volvocales. Structural domains are represented by Roman numerals I–IV according to STIEGLER et al. (1981). Hypervariable regions have a grey background and are numbered V2–V9 (according to K1 2012). V1 is not available (missing data). V6 is missing in eukaryotes. Phylogenetically informative nucleotides are highlighted and visualized in their structural context. Identical nucleotides in all three sequence–structure pairs are indicated by grey dots. Nucleotides identical with or with a tendency (according to the IUPAC code) towards Volvocales are shown in light green. Tendencies are explained in the supplementary material, Fig. S1. Nucleotides identical with or with a tendency (according to the IUPAC code) towards Sphaeropleales are shown in red. Unique characters for *Golenkinia+Jenufa* are highlighted in blue and nucleotides where no allocation was possible (according to the IUPAC code) are visualized in light grey. V4 within the structural domain II is typically used in biodiversity studies. Pseudoknots known in V4 could not be modeled by homology modeling.

our knowledge concerning other chlorophycean genera with uncertain phylogenetic affinities (e.g. *Microspora*, *Parallela*, *Dictyochloris*, or *Spermatozopsis*), as they have already been discussed by FUČIKOVÁ et al. (2019).

ACKNOWLEDGEMENTS

We thank Mark A. Buchheim (University of Tulsa, USA) for pre-review and critical reading of the manuscript. We further would like to thank Gerrit Wehner (University of Würzburg, Germany) for preliminary analyses inferring consensus structures.

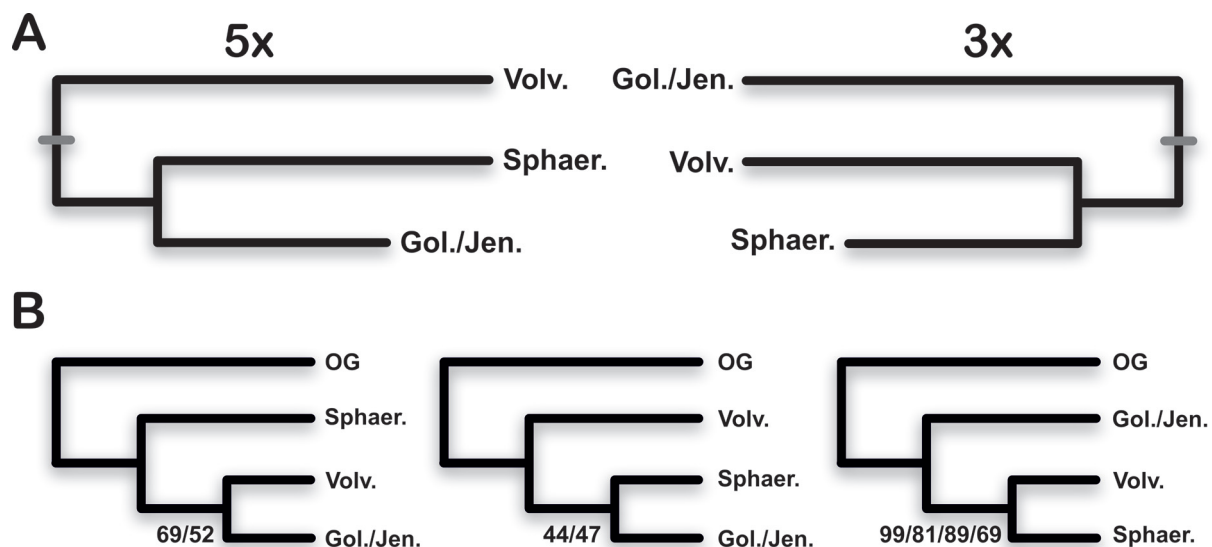


Fig. 2. Phylogenetic trees. (A) Midpoint rooted three-taxon trees: Five of the analyses (5X) favor Sphaeropleales and *Golenkinia+Jenufa* and only three (3X) favor Volvocales and Sphaeropleales (see Table 2). (B) Rooted four-taxon trees: (OG) outgroup from the OCC-clade. Bootstrap values: (69) PNJ, full set; (52) PNJ, without *Treubarinia/Hafniomonas*; (44) PNJ, full set (V4 region); (47) PNJ, without *Treubarinia/Hafniomonas* (V4 region); (99) Parsimony, full set; (81) Parsimony, full set (V4 region); (89) Parsimony, without *Treubarinia/Hafniomonas*; (69) Parsimony, without *Treubarinia/Hafniomonas* (V4 region).

REFERENCES

- BUCHHEIM, M. A.; MICHALOPULOS, E. A. & BUCHHEIM, J. A. (2001): Phylogeny of the Chlorophyceae with special reference to the Sphaeropleales: a study of 18S and 26S rDNA data. – *Journal of Phycology* 37: 819–835.
- BUCHHEIM, M. A.; MÜLLER, T. & WOLF, M. (2017): 18S rDNA sequence–structure phylogeny of the Chlorophyceae with special emphasis on the Sphaeropleales. – *Plant Gene* 10: 45–50.
- BUCHHEIM, M. A.; SUTHERLAND, D. M.; SCHLEICHER, T.; FÖRSTER, F. & WOLF, M. (2012): Phylogeny of Oedogoniales, Chaetophorales and Chaetopeltidales (Chlorophyceae): inferences from sequence–structure analysis of ITS2. – *Annals of Botany* 109: 109–116.
- BYUN, Y. & HAN, K. (2006): PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures. – *Nucleic Acids Research* 34: W416–W422.
- CANNONE, J. J.; SUBRAMANIAN, S.; SCHNARE, M. N.; COLLETT, J. R.; D’SOUZA, L. M.; DU, Y.; FENG, B.; LIN, N.; MADABUSI, L. V.; MÜLLER, 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. – *BMC Bioinformatics* 3: 2.
- FANG, L.; LELIAERT, F.; NOVIS, P. M.; ZHANG, Z.; ZHU, H.; LIU, G.; PENNY, D. & ZHONG, B. (2018): Improving phylogenetic inference of core Chlorophyta using chloroplast sequences with strong phylogenetic signals and heterogeneous models. – *Molecular Phylogenetics and Evolution* 127: 248–255.
- FELSENSTEIN, J. (1989): PHYLIP – phylogeny inference package (Version 3.2). – *Cladistics* 5: 164–165.
- FELSENSTEIN, J. (1985): Confidence limits on phylogenies: an approach using the bootstrap. – *Evolution* 39: 783–791.
- 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.
- FUČIKOVÁ, K.; LEWIS, P. O. & LEWIS, L. A. (2014): Putting incertae sedis taxa in their place: a proposal for ten new families and three new genera in Sphaeropleales (Chlorophyceae, Chlorophyta). – *Journal of Phycology* 50: 14–25.
- FUČIKOVÁ, K.; LEWIS, P. O. & LEWIS, L. A. (2016): Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution. – *Molecular Phylogenetics and Evolution* 98: 176–183.
- FUČIKOVÁ, K.; LEWIS, P. O.; NEUPANE, S.; KAROL, K. G. & LEWIS, L. A. (2019): Order, please! Uncertainty in the ordinal-level classification of Chlorophyceae. – *PeerJ* 7: e6899.
- HE, L.; WANG, Z.; LOU, S.; LIN, X. & HU, F. (2018): The complete chloroplast genome of the green alga *Hariotina reticulata* (Scenedesmeaceae, Sphaeropleales, Chlorophyta). – *Genes and Genomics* 40: 543–552.
- HEPPERLE, D. (2004): Align Ver.07/04©. Multisequence alignment–editor and preparation/manipulation of phylogenetic datasets. Win32–Version. – Distributed by the author via: <http://www.sequentix.de>.
- HODAČ, L. (2016): Green algae in soil: assessing their biodiversity and biogeography with molecular–phylogenetic methods based on cultures [PhD thesis]. – 185 pp., Georg August University of Göttingen.
- HODAČ, L.; HALLMANN, C.; ROSENKRANZ, H.; FASSHAUER, F. & FRIEDL, T. (2012): Molecular evidence for the wide distribution of two lineages of terrestrial green algae (Chlorophyta) over tropics to temperate zone. – *ISRN Ecology*: Article ID 795924.
- KELLER, A.; SCHLEICHER, T.; FÖRSTER, F.; RUDERISCH, B.; DANDEKAR, T.; MÜLLER, T.; WOLF, M. (2008): ITS2 data corroborate a monophyletic chlorophycean DO–group (Sphaeropleales). – *BMC Evolutionary Biology* 8: 218.

- KELLER, A.; FÖRSTER, F.; MÜLLER, T.; DANDEKAR, T.; SCHULTZ, J.; WOLF, M. (2010): Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. – *Biology Direct* 5: 4.
- KI, J. S. (2012): Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. – *Journal of Applied Phycology* 24: 1035–1043.
- LARKIN, M. A.; BLACKSHIELDS, G.; BROWN, N. P.; CHENNA, R.; MCGETTIGAN, P. A.; MCWILLIAM, H.; VALENTIN, F.; WALLACE, I. M.; WILM, A.; LOPEZ, R.; THOMPSON, J. D.; GIBSON, T. J. & HIGGINS, D. G. (2007): Clustal W and Clustal X version 2.0. – *Bioinformatics* 23: 2947–2948.
- LELIAERT, F.; SMITH, D. R.; MOREAU, H.; HERRON, M. D.; VERBRUGGEN, H.; DELWICHE, C. F. & DECLERCK, O. (2012): Phylogeny and molecular evolution of the green algae. – *Critical Reviews in Plant Sciences* 31: 1–46.
- LEMIEUX, C.; VINCENT, A. T.; LABARRE, A.; OTIS, C. & TURMEL, M. (2015): Chloroplast phylogenomic analysis of chlorophyte green algae identified a novel lineage sister to the Sphaeropleales (Chlorophyceae). – *BMC Evolutionary Biology* 15: 264.
- MARIN RODRIGUES, M. V.; MÜLLER, T.; BUCHHEIM, M. A.; CANBÄCK, B. & WOLF, M. (2017): THESES db: the algae 18S rDNA sequence–structure database for inferring phylogenies. – *Phycologia* 56: 186–192.
- MÜLLER, T.; RAHMANN, S.; DANDEKAR, T. & WOLF, M. (2004): Accurate and robust phylogeny estimation based on profile distances: a study of the Chlorophyceae (Chlorophyta). – *BMC Evolutionary Biology* 4: 20.
- NĚMCOVÁ, Y.; ELIÁŠ, M.; ŠKALOUD, P.; HODAČ, L. & NEUSTUPA, J. (2011): *Jenufa*, gen. nov.: a new genus of coccoid green algae (Chlorophyceae, incertae sedis) previously recorded by environmental sequencing. – *Journal of Phycology* 47: 928–938.
- PROCHÁZKOVÁ, K.; NĚMCOVÁ, Y. & NEUSTUPA, J. (2015): A new species *Jenufa aroterrestrica* (Chlorophyceae incertae sedis, Viridiplantae), described from Europe. – *Preslia* 87: 403–416.
- RAHMANN, S.; MÜLLER, T.; DANDEKAR, T. & WOLF, M. (2006): Efficient and robust analysis of large phylogenetic datasets. – In: HUI–HUANG, H. (ed.): *Advanced Data Mining Technologies in Bioinformatics*. – pp. 104–117, Idea Group Inc, USA.
- SEIBEL, P. N.; MÜLLER, T.; DANDEKAR, T.; SCHULTZ, J. & WOLF, M. (2006): 4SALE – a tool for synchronous RNA sequence and secondary structure alignment and editing. – *BMC Bioinformatics* 7: 498.
- SEIBEL, P. N.; MÜLLER, T.; DANDEKAR, T. & WOLF, M. (2008): Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. – *BMC Research Notes* 1: 91.
- SONG, H.; LI, S.; LIU, X.; WANG, Q.; ZHU, H.; LIU, G. & HU, Z. (2017): *Jenufa lobulosa* sp. nov. (Chlorophyceae, Chlorophyta), a new epilithic, terrestrial species described from China. – *Phycologia* 57: 52–60.
- STIEGLER, P.; CARBON, P.; ZUKER, M.; EBEL, J. P. & EHRESMANN, C. (1981): Structural organization of the 16S ribosomal RNA from *E. coli*. Topography and secondary structure. – *Nucleic Acids Research* 9: 2153–72.
- TIPPERY, N. P.; FUČIKOVÁ, K.; LEWIS, P. O. & LEWIS, L. A. (2012): Probing the monophyly of the Sphaeropleales (Chlorophyceae) using data from five genes. – *Journal of Phycology* 48: 1482–1493.
- TURMEL, M.; BROUARD, J. S.; GAGNON, C.; OTIS, C. & LEMIEUX, C. (2008): Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses. – *Journal of Phycology* 44: 739–750.
- WOLF, M.; ACHTZIGER, M.; SCHULTZ, J.; DANDEKAR, T. & MÜLLER, T. (2005): Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. – *RNA* 11: 1616–1623.
- WOLF, M.; BUCHHEIM, M.; HEGEWALD, E.; KRIENITZ, L. & HEPPERLE, D. (2002): Phylogenetic position of the Sphaeropleaceae. – *Plant Systematics & Evolution* 230: 161–171.
- WOLF, M. & BUCHHEIM, M. A. (2018): The internal transcribed spacer 2 of *Jenufa* (Chlorophyta, Chlorophyceae) is extraordinarily long: A hypothesis. – *Gene* 678: 100–104.
- WOLF, M.; HEGEWALD, E.; HEPPERLE, D. & KRIENITZ, L. (2003): Phylogenetic position of the Golenkiniaceae (Chlorophyta) as inferred from 18S rDNA sequence data. – *Biologia* 58: 433–436.
- WOLF, M.; KOETSCHAN, C. & MÜLLER, T. (2014): ITS2, 18S, 16S or any other RNA – simply aligning sequences and their individual secondary structures simultaneously by an automatic approach. – *Gene* 546: 145–149.
- WOLF, M.; RUDERISCH, B.; DANDEKAR, T.; SCHULTZ, J. & MÜLLER, T. (2008): ProfDistS: (profile–) distance based phylogeny on sequence – structure alignments. – *Bioinformatics* 24: 2401–2402.

Supplementary material

the following supplementary material is available for this article:

Table S1. Species list used in Buchheim et al. (2017) plus *Jenufa aroterrestrica*.

Fig. S1. (A) Consensus structure prediction. The figure shows how a 100% consensus structure, valid for the pairings, was transcribed from the global multiple sequence–structure alignment as obtained by 4SALE (SEIBEL et al. 2006, 2008). (B) 2D plot of individual sequence–structure information (left) vs. the final sequence–structure consensus (right). (C) Assignments (tendencies) of the IUPAC code in our three–taxa comparison (see Fig. 1).
Data S1. Sequence–structure alignment (69 taxa).
Data S2. Three–taxa alignment (consensus sequences and consensus structures).

This material is available as part of the online article (<http://fottea.czechphycology.cz/contents>)