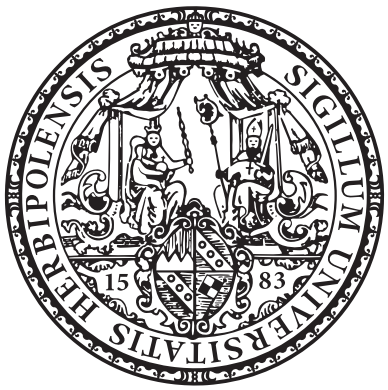
Secondary (and tertiary) structure of the ITS2 and its application for phylogenetic tree reconstructions and species identification



vorgelegt von
Dipl. Biol. Alexander Keller
Würzburg, 2010

Kumulative Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades (Dr. rer. nat.)
der Bayerischen Julius-Maximilians-Universität Würzburg

Einreichung: _____ in Würzburg

Mitglieder der Promotionskommission:

Vorsitzender: Prof. Thomas Dandekar
1. Gutachter: Prof. Thomas Dandekar
2. Gutachter: Prof. Ingolf Steffan-Dewenter

Promotionskolloquium: _____ in Würzburg

Aushändigung Doktorurkunde: _____ in Würzburg

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Preceding the actual thesis, I want to acknowledge the support of many persons that directly or indirectly enabled me to manage the projects, and of course lately the final writing. First to mention are my supervisors, Prof. Dr. Thomas Dandekar, Dr. Matthias Wolf, Prof. Dr. Jörg Schultz and Dr. Tobias Müller, who designed interesting and prosperous projects for my graduate studies and were always helpful, where needed. Further, I would like to thank all the other companions of our working group on the way to the PhD: these were Dr. Frank Förster, Tina Schleicher, Christian Koetschan, and the bachelor students Benjamin Merget and Gregor Schalk. The remaining members of the Department of Bioinformatics can be added to this list for all their fruitful discussions and useful ideas.

In addition to these persons of our department, I want to thank all my external collaborators, which are Prof. Dr. Mark Buchheim, Dr. Martin Wiemers, Dr. Eberhard Hegewald, Robert Junker, Dr. Nico Blüthgen, Christina Loewel, Prof. Dr. Curtis Daehler, Dr. Stefan Dötterl, Prof. Dr. Thomas Friedl, Prof. Dr. Roy Gross, Prof. Dr. Lothar Krienitz, Katrin Schlegelmilch, and Prof. Dr. Norbert Schütze, for all their efforts and their investments into our collaborative projects. Furthermore, I would like to especially thank Prof. Dr. Ulf Sorhannus as a collaborator and as a very obliging host (together with Prof. Dr. Marty Mitchell and Linda Kightlinger) during my visit at the Edinboro University of Pennsylvania.

Many thanks for agreeing to review this thesis to my universitary advisory committee Prof. Dr. Thomas Dandekar and Prof. Dr. Ingolf Steffan-Dewenter and as well the advisory committee of my graduate school BIGSS consisting of Prof. Dr. Thomas Dandekar, Prof. Dr. Jörg Schultz and Prof. Dr. Heinrich Sticht. For financial sponsoring and meeting the costs of my travels and research equipment, I gratefully acknowledge the Elite Network Bavaria graduate school BIGSS. Furthermore, as this thesis has been awarded with the Biocenter Science Award, I would like to thank the Biocenter Würzburg Commitee for their choice and I appreciate very much the interest in my work.

Especially, I am deeply grateful to my family including Katrin for supporting and encouraging me during the complete time of my studies; and even more for the reviving times away from these! They and several good friends had to cope with less visits by me and other cut-backs; I really appreciate your patience and hope to see you more in the time after thesis submission!

SUMMARY

Biodiversity may be investigated and explored by the means of genetic sequence information and molecular phylogenetics. Yet, with ribosomal genes, information for phylogenetic studies may not only be retained from the primary sequence, but also from the secondary structure. Software that is able to cope with two dimensional data and designed to answer taxonomic questions has been recently developed and published as a new scientific pipeline. This thesis is concerned with expanding this pipeline by a tool that facialiates the annotation of a ribosomal region, namely the ITS2. We were also able to show that this states a crucial step for secondary structure phylogenetics and for data allocation of the ITS2-database. This resulting freely available tool determines high quality annotations. In a further study, the complete phylogenetic pipeline has been evaluated on a theoretical basis in a comprehensive simulation study. We were able to show that both, the accuracy and the robustness of phylogenetic trees are largely improved by the approach.

The second major part of this thesis concentrates on case studies that applied this pipeline to resolve questions in taxonomy and ecology. We were able to determine several independent phylogenies within the green algae that further corroborate the idea that secondary structures improve the obtainable phylogenetic signal, but now from a biological perspective. This approach was applicable in studies on the species and genus level, but due to the conservation of the secondary structure also for investigations on the deeper level of taxonomy. An additional case study with blue butterflies indicates that this approach is not restricted to plants, but may also be used for metazoan phylogenies. The importance of high quality phylogenetic trees is indicated by two ecological studies that have been conducted. By integrating secondary structure phylogenetics, we were able to answer questions about the evolution of ant-plant interactions and of communities of bacteria residing on different plant tissues.

Finally, we speculate how phylogenetic methods with RNA may be further enhanced by integration of the third dimension. This has been a speculative idea that was supplemented with a small phylogenetic example, however it shows that the great potential of structural phylogenetics has not been fully exploited yet. Altogether, this thesis comprises aspects of several different biological disciplines, which are evolutionary biology and biodiversity research, community and invasion ecology as well as molecular and structural biology. Further, it is complemented by statistical approaches and development of informatical software. All these different research areas are combined by the means of bioinformatics as the central connective link into one comprehensive thesis.

ZUSAMMENFASSUNG

Biologische Diversität kann mit Hilfe molekularer Sequenzinformation und phylogenetischen Methoden erforscht und erfasst werden. Bei ribosomalen Genen kann man jedoch wertvolle Information nicht nur aus der Primärsequenz beziehen, sondern auch aus der Sekundärstruktur. In den letzen Jahren wurde Software entwickelt, die solche Daten für taxonomische Fragestellung verwerten kann. Diese Arbeit beschäftigt sich mit einer Erweiterung dieser Methodik durch eine Software-Anwendung, die die Annotation des ribosomalen Genes ITS2 deutlich vereinfacht. Mit dieser Studie konnten wir zeigen, dass dies einen entscheidenden Schritt der Sequenz-Struktur-Phylogenie und der Datenerfassung der ITS2-Datenbank darstellt. Die daraus resultierende und frei verfügbare Anwendung ermöglicht Annotationen von hoher Güte. In einer weiteren Studie wurde mittels Simulationen der gesamte Arbeitsfluß der Sequenz-Struktur Phylogenie auf theoretischer Ebene evaluiert. Dabei zeigte sich, dass sich sowohl die Genauigkeit, als auch die Robustheit von phylogenetischen Stammbäumen durch diesen Ansatz deutlich verbessern.

Der zweite große Teil der Arbeit befasst sich mit Fallbeispielen, in denen dieser Arbeitsfluß zur Aufklärung von taxomonischen and ökologischen Fragestellungen Anwendung fand. In diesem Rahmen konnten wir mehrere und voneinander unabhängige Phylogenien ermitteln, welche die theoretischen Ergebnisse einer Verbesserung phylogenetischer Bäume auch von biologischer Seite aus bekräftigen. Der Ansatz war anwendbar in sehr feinskaligen Studien auf Art bzw. Gattungsniveau, aber durch die starke Konservierung der Sekundärstruktur auch an sehr weit von einander entfernten taxonomischen Gruppen. Eine weitere Studie, die sich mit der Phylogenie von Bläulingen befasst, zeigt deutlich, dass dieser Ansatz nicht nur für Fragestellungen bei Pflanzen, sondern auch im Tierreich angewandt werden kann. Die Bedeutung von qualitativ hochwertigen Stammbäumen auch für andere Fachbereiche wird an zwei unserer ökologischen Studien deutlich: Mit Hinzunahme von Sekundärstruktur war es uns möglich Fragestellungen über die Evolution von Ameisen-Pflanzen Interaktionen sowie über ökologische Gemeinschaften von Bakterien auf verschiedenen Pflanzenteilen zu beantworten.

Zuletzt gehen wir spekulativ auf die Frage ein, wie Strukturphylogenie um die dritte Dimension erweitert werden kann. Dies bleibt zwar spekulativ und wurde nur um ein kleines Fallbeispiel ergänzt, jedoch zeigt sich deutlich, dass das Potential von Strukturphylogenie noch nicht erschöpft ist. Insgesamt befasst sich diese Arbeit mit Aspekten aus verschiedenen biologischen Disziplinen: Evolutionsbiologie und Biodiversitätsforschung, sowie Gemeinschafts- und Invasionsökologie, aber auch Molekular- und Strukturbiologie. Dies wurde ergänzt durch statistische Ansätze und Entwicklung von informatischer Software. Diese verschiedenen Forschungsrichtungen wurden mit Hilfe der Bioinformatik als zentrales Bindeglied vereint.

Part I.

General Introduction

Species Diversity and Evolutionary Biology

Biodiversity is a hot topic that regularly hits the headlines. This fascination about the manifoldness of living organisms that inhabit a given ecosystem and their interactions has already been established in the early human days as demonstrated by stone carvings and paintings throughout the world. This allure has been a constant companion throughout all human cultures and eras. Nowadays, in an industrialized and globally interconnected society, the United Nations declared this year (2010) to the "International Year of Biodiversity" to emphasize the importance of interest in biologcal diversity and to keep this fascination alive:

> » *It is a celebration of life on earth and of the value of biodiversity for our lives.* «
>
> — United Nations (2010, accessed 17th August 2010) —

The need of such a declaration is justified as human activities nowadays pose a severe thread to biodiversity with irreversible effects on essential living networks that provide vital services to all existing organisms including mankind. It is now of major importance to assess, explore and protect biological diversity.

Despite the need of profound knowledge, current estimates of global biological diversity are very rough ranging from 7 to 100 million living species on earth (Costello et al. 2010; Erwin 2002; European Distributed Institute of Taxonomy 2010; Global Taxonomy Initiative 2010; May 1988). Since Carolus Linnaeus in the 18th century, a species is one of the most basic units used in taxonomic classification of earth's organisms. In modern biology, the biological species concept defines a species as a set of actually or potentially interbreeding organisms, which is capable of producing fertile offspring (Mayr 1970; Poulton 1903). However, this definition bears several controversal aspects, so that a lot of different approaches have been suggested to describe species (e.g. Blaxter 2004; Cracraft 1989; Darwin 1859; Hennig 1966; Mahner and Bunge 1997; Ridley 1989; Simpson 1951; Templeton 1992; van Valen 1976). Thus, inevitably different concepts of the term "species" have been used in scientific history. Recent advantages in technology and biological knowledge have revealed that many so-called species are in fact complexes of taxa that can be most readily distinguished using e.g. genetic, behavioural or ecological characters (e.g. Blaxter 2004; Hebert et al. 2004; Jones and van Parijs 1993; Leaché and Fujita 2010; Rissler and Apodaca 2007).

The previous year, namely the "Darwin Year 2009", has been tributed to Charles Darwin for his achievements in the general understanding of biological diversity and evolution by the International Union of Biological Sciences. His argumentation in the ongoing debate on species concepts was that

> » *no one definition has satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species.* «
>
> — C. Darwin (1859, On the Origin of Species, p. 44) —

Further, he argues that a species is an arbitrary hypothetical construct given

> » *for the sake of convenience to a set of individuals closely resembling each other.* «
>
> — C. Darwin (1859, On the Origin of Species, p. 52) —

Thus, clusters of individuals with resembling characteristics are merged and hypothesized to form distinct groups of organisms that are in the best case reproductively isolated from others as fundamental taxonomic units for further discussions between naturalists. As a result, the term species is in the most cases used pragmatically by mixing several of the

various existing definitions that match best the particular characteristics of the concerned group of organisms.

Despite the various and precise theoretical definitions, such pragmatical attempts are in the most cases favourable in their practical application. Adaptive methods for identifying and distinguishing particular species are essential for measuring biodiversity and testing biological hypotheses. The traditional attempt to find resembling characteristics and to cluster organisms into species is to use morphological and in special cases behavioural as well as spatial and ecological factors. This is still the rule in nearly all groups of organisms. However, nowadays new approaches base on similarity of genetic information of individuals to define conspecific boundaries (Bickford et al. 2007; Casiraghi et al. 2010; Hebert et al. 2004). In the nearby future, such genetically defined species will likely outnumber the quantity of species defined by morphology.

Species diversity can be explored at a number of different levels and in principle may be quantified seperately at each. Beside that there is still no consensus in what a species really is, nowadays however it is commonly agreed by naturalists that current species represent a stage in the process of evolution, with their diversity the product of a long series of speciation and extinction events. Darwin mentions that

> *» characters which naturalists consider as showing true affinity between one or more species, are those which have been inherited from a common parent, and, in so far, all true classification being genealogical. «*
>
> — C. Darwin (1859, On the Origin of Species, p. 420) —

One of the methods to assess biodiversity is thus to reconstruct evolutionary relationships by the sequential clustering of organisms using the individuals' characteristics. All the mentioned types of characteristics have been used for this approach in scientific history, today however molecular data are predominantly preferred to reconstruct such phylogenies. Phylogenetic patterns have the potential to quantify and estimate biodiversity at the finest scale, that is, variation among species in features or attributes.

Molecular Phylogenetics

The awareness that every living organism bears genetic information in the form of deoxyribonucleic acid (DNA) has provided a fundamental cornerstone for diversity research with phylogenetics. Early studies were performed using gel electrophoresis of proteins to cluster organisms by size of these proteins (please see the review of Suárez-Díaz and Anaya-Muñoz 2008, and the references therein for further details on the early history of molecular phylogenies). However, the possibility to retain genetic information directly by sequencing DNA increased the effectiveness of phylogenetic studies. The DNA is composed of a four letter alphabet of the nucleotides Adenine (A), Cytosine (C), Guanine (G) and Thymine (T), which compose – in the correct sequential order – "blueprints" of complete organisms. This sequence of nucleotides is nowadays used for phylogenetic inferences (Felsenstein 2004).

DNA accumulates mutations over evolutionary time, so that closely related organisms are expected to show fewer changes in their nucleotide sequence than distantly related taxa, likewise to morphological features. The major advantage of genetic information compared to morphological data is that it inherits – in the most cases and with respect to the regarded genomic region – a multitude of such differences, which are easily obtainable by standard laboratory procedures. This also accounts for "difficult" taxa with small sizes and few morphological differences between species (Felsenstein 2004).

The general objective of molecular phylogenetics is to compare such nucleotide sequences obtained from the laboratory to make implications about the regarded biodiversity with

respect to its evolutionary history (Felsenstein 2004). To be able to receive significant clues from nucleotide sequences, these have to be made comparable prior to analyses. This step includes that it must be assured that the same fragment of the complete genomic DNA is regarded and that the nucleotides of these fragments are arranged according to their evolutionary history. This can be done either by pairwise or by multiple alignment procedures. Methods of determining differences between organisms range from simple counting of nucleotide changes (substitutions) to complex intrinsic or extrinsic mathematical models, which consider different substitution rates between nucleotides (Felsenstein 2004; Lanave et al. 1984; Rodriguez et al. 1990; Tavaré 1986). Gaps are inserted at locations of the sequences where no corresponding bases are found (Felsenstein 2004). They are considered to represent historic insertion and deletion events. Various techniques exist, with which phylogenetic trees can be reconstructed using this data and evolutionary models. These methods base on different mathematical concepts, the most important base on clustering (e.g. neighbor joining (NJ) (Saitou and Nei 1987)) of organisms, maximum parsimony (MP) (Camin and Sokal 1965), maximum likelihood (ML) (Felsenstein 1981) or bayesian analyses (Huelsenbeck and Ronquist 2001; Huelsenbeck et al. 2001).

In the best case the resulting phylogenetic tree represents the real evolutionary tree of life completely as a dichotom branching diagram. Outer nodes are the investigated organisms as operational taxonomic units (OTUs). Internal nodes of the diagram are expected to represent their common ancestors as hypothetical taxonomic units (HTUs). As measurements of confidence in the reconstruted tree, usually the accuracy and the robustness are considered. However, practically the accuracy is in the most cases undeterminable as the real tree of life is unknown. Thus, the robustness of each internal node, as a statistical measure of stability and determined e.g. by bootstrapping algorithms or as bayesian posterior probabilities (Alfaro et al. 2003; Hillis and Bull 1993), remains the only practical mean of confidence for phylogenetic trees. Different fragments of DNA have been used to infer phylogenies dependent on the level of relatedness between investigated organisms and their taxonomic affiliation. For example, mitochondrial genes have been historically prefered for studies in animals, whereas most regions of interests for plants and fungi as well as most single cell eukaryotic and obviously prokaryotic organisms originate from the nuclear genome.

Ribosomal Genes and the Internal Transcribed Spacer 2 for Phylogenetics

Ribonucleic acid (RNA), as a transcript of the DNA, bears the nucleotide Uracil (U) instead of its DNA-equivalent T. One of the many examples, where RNA occurs within organisms and has essential vital roles is the ribosomal RNA (rRNA). For phylogenetics, the nuclear rRNA cistron is generally considered to be an important region. Initial studies with this region stated milestones in acquiring knowledge about the fundamental tree of life (Woese and Fox 1977; Woese et al. 1990). The small subunit (SSU) and large subunit (LSU) of the ribosome (Fig. P.1.1) present highly conserved markers that can be used in phylogenetic reconstructions at a high taxonomic level (Hershkovitz and Lewis 1996). They are conserved as they make up the general backbone topology of the ribosomes that is supplemented with ribosomal proteins (Thomson and Tollervey 2005; Venema and Tollervey 2004). Small substitutional changes may lead to unfunctionality of the ribosomes, what results in cell or organismal death (Venema and Tollervey 1999). With this, a strong selection pressure lies upon these regions during the evolutionary process. Thus, even in far related organisms, the nucleotide sequences are comparable with few substitutions. However, it lacks the power to detect phylogenetic signal on a low level, as the number of substitions are not enough in closely related organisms.

In contrast, the internal transcribed spacer 2 (ITS2), which separates the nuclear riboso-

mal genes 5.8S and 28S (correspondingly 25S or 26S in e.g. several fungi), provides completely different associated characteristics for phylogenetic analyses (Alvarez and Wendel 2003; Baldwin et al. 1995; Coleman 2003; Cronn et al. 2002; Feliner and Rosselló 2007; Small et al. 1998). The actual nucleotide sequence is in contrast to the adjacent regions not, or to be more precise, not directly of importance for survival of an organism. It is neither coding for proteins nor is it present as an essential subunit of the mature ribsome (Venema and Tollervey 2004). Other functionalities of non-coding RNAs as e.g. the ability to act as microRNAs has not been detected for the spacer region. As it does not seem to be of importance to evolutionary maintain the nucleotide sequence, substitution rates are very high in comparison to the surrounding regions. Thus, from a phylogeneticists point of view, it is very useful for inferences of phylogenies at the species and genus level (Alvarez and Wendel 2003; Coleman 2000, 2003; Coleman and Vacquier 2002).

Secondary Structure Phylogenetics

The approach to combine two or more markers with different substitution rates is a common way to infer phylogenies that range from low- to high-level relationships between organisms (e.g. Dunn et al. 2008; Schoch et al. 2009). However, such attempts face different problems, which are not easily resolved (Huelsenbeck et al. 1996): erroneous results may be obtained in applied model-based inferences with sequences containing these markers concatenated together (e.g. conserved 5.8S and fast evolving ITS2). This is due to the fact that general substitution models obviously match only substitution rates of some of the used markers on the one hand or are very unprecise on the other hand.

In this context, another feature of the ITS2 renders possible an alternate way to infer phylogenies from a low to an high level of species relationships that does not face the mentioned problems in model based approaches. Even if the nucleotide sequence is not necessarily evolutionary maintained, some restrictions still exist for mutations during the evolution of the ITS2 and are necessary for organismal survival (Côté et al. 2002; Lafontaine and Tollervey 2001; Mitchell et al. 1996; Peculis and Greer 2002; Venema and Tollervey 2004). In general, RNA is capable to fold into a secondary structure. This as well applies to ribosomal genes and the interjacent spacers. Usually, such ribosomal secondary structures are evolutionary conserved as they represent fundamental organismal cell features. Althought its nucleotide sequence is not conserved, this is also true for the ITS2 (Côté et al. 2002; Liu and Schardl 1994; Mai and Coleman 1997; Schlötterer et al. 1994; Torres et al. 1990). Even if it is not present in the mature ribosome, its secondary structure folding is necessary during ribogenesis. This implicates that ITS2 secondary structure is well conserved across large parts of the tree of life (Coleman 2007; Joseph et al. 1999; Schultz et al. 2005).

In the last decades, several biological observations indicated that phylogenetic studies on the nucleotide level of ribosomal RNA may be supplemented with information of the secondary structure (Coleman 2003; Hillis and Dixon 1991; Mai and Coleman 1997; Schultz et al. 2006; Selig et al. 2008; Soltis et al. 1998; Wheeler and Honeycutt 1988; Wolf et al. 2005b). It has been speculated that this would result in enhanced reconstructions of evolutionary trees (Alvarez and Wendel 2003). This structural feature can thus be used in taxonomic studies as it may serve as a conserved margin for alignment, whereas there is enough substitutial information for phylogenetics due to sequences substitutions. It can be incorporated into substitution models for phylogenetic inferences (Gowri-Shankar and Rattray 2006; Havgaard et al. 2005; Jow et al. 2002; Schöniger and von Haeseler 1994; Seibel et al. 2006; Wolf et al. 2008). Thus, it provides both advantages in one genetic marker due to the ambivalency of its inherent evolutionary patterns.

Species Identification

Furthermore, other fortunate characteristics make the ITS2 an interesting marker for biodiversity research beside its usability in phylogenetic studies. The highly conserved flanking regions can be used as an anchor for universal primers what eases sequence amplification in the laboratory and keeps them consistent between studies (White et al. 1990). With that and the high amount of nucleotide substitutions even on a generic level, it meets the requirements to be used as a DNA barcoding marker, i.e. to distinguish species by a short fragment of the genomic sequence.

On the secondary structure level, its conservation leads to compensatory base changes (CBCs) and hemi compensatory base changes (hCBCs) (Dixon and Hillis 1993; Gutell et al. 1994). CBCs are mutations that occur in both nucleotides of a paired structural position while retaining the paired nucleotide bond. Such a motif may have two different evolutionary origins: usally, substitutions that are located within bound stem regions of the RNA affect the folding of the general structure. Some substitutions however have less impact: in RNA (in contrast to DNA) bonding between G and U is energetically favorable (additionally to the typical Watson-Crick Hydrogen (H)-bonds). Such changing events state hCBCs. Two such events may result in a complete CBC. The second possible method to evolutionary obtain a CBC is the substitution of a nucleotide that results in the loosing of the bond, which is compensated in a second step by mutation of the complementary base. However, this second scenario is less likely, as the structure will go through a period with a less stable or even false secondary structure. The interesting point for phylogenetic studies is, that it has been recently claimed that structural differences in the shape of CBCs in ITS2 are predictive of species limits. In this view, pairings of CBCs provide an indication for sexual incompatibility (Coleman 2003, 2009; Müller et al. 2007; Sorhannus et al. 2009), while their absence may indicate intercrossing ability (Coleman 2009). The latter is unsupported or rather supported with very low predictive power in the large scale analyses performed by Müller et al. (Müller et al. 2007)

Objectives of this Thesis

The idea that phylogenetic studies on the nucleotide level of ribosomal RNA may be supplemented with information of the secondary structure has been supported by several biological observations (Coleman 2003; Hillis and Dixon 1991; Mai and Coleman 1997; Soltis et al. 1998; Wheeler and Honeycutt 1988). Suggestions made in this context often promoted the internal transcribed spacers (ITS) region as a suitable marker for secondary structure phylogenetics (Alvarez and Wendel 2003; Coleman 2003). However, during that time, software lacked that was able to perform sequence-structure phylogenetics, whereas it was generally expected that this marker might be very useful for advanced approaches on a multidimensional level (Coleman 2003; Côté et al. 2002; Liu and Schardl 1994; Mai and Coleman 1997; Schlötterer et al. 1994; Torres et al. 1990).

So these improvements remained mostly theoretical and only few small manual examples were performed (e.g. Coleman 2003; Dixon and Hillis 1993). These biological observations and the novelty of the approach motivated Dr. J. Schultz, Dr. M. Wolf, Dr. T. Müller and Dr. T. Dandekar of the Department of Bioinformatics (University of Würzburg) to found the "ITS2-working-group" and to start the "ITS2-database and pipeline project" in 2005, which inherits, beside biological questions, the development of tools for phylogenetics with secondary structures and building up a database for structure deposition.

As a member of this group, it was one of my goals to address the issue of identification and delineation of the ITS2, what states a crucial task for secondary structure predictions

and a challenge due to the high substitution rates. The task was to compare traditional techniques and development of a novel method for annotation, plus the preparation as an online-tool freely available to web users (Publication P.1). This procedure was furthermore intended for automated ITS2-database data processing and the updating pipeline, to serve as a fundamental and crucial step during allocation of the complete underlying data of the database (Publication P.2).

With an established general phylogenetic pipeline and a published description of the general workflow (Schultz and Wolf 2009), it still remained unclear, how large the benefit for phylogenetic studies is by inclusion of secondary structure information, if at all. The next section of this thesis addresses the evaluation of secondary structure phylogenetics by simulation experiments (Publication P.3). It was of major concern, whether phylogenetic trees resolved with secondary structures are more accurate and/or more robust than those obtained by traditional techniques.

With the evaluated methods, we obtained a striking opportunity to resolve phylogenies for taxa where phylogenetic implications where hard to obtain with traditional ITS2 techniques, other markers or morphology. We thus performed two phylogenetic studies (Publications P.4 and P.5) and one large scale approach for phylogenetics and species identification (Publication P.6) with different groups of green algae. As an applied study within the animal kingdom, we reconstructed a phylogenetic tree for the blue butterfly subgenus *Agrodiaetus* and closly related species (Publication P.7).

The next sections of this thesis concentrate on ecological questions, for whose examination knowledge about species diversity and their evolutionary relationships have been of major importance and which was determined by the aforementioned methods. This method has not only been applied in this context with the eukaryotic ITS2 marker (Publication P.8), but furthermore transferred to phylogenetic studies in bacteria with the ribosomal 16S gene (Publication P.9).

Finally, we hypothesize how structural phylogenetics in the future may be expanded from the second to the third dimension (Publication P.10). This scenario is at the moment speculatory, however it shows the massive potential that is inherent in structural phylogenetics on the different levels.

In general, within this thesis I am adressing several different aspects of ITS2 secondary structure phylogenetics, which broadly range from the phylogenetic pipeline and their evaluation over applied phylogenetic studies and their usage for ecological questions to hypothetical future prospects of the methods.

Part II.

Materials and Methods

CHAPTER

ONE

MATERIALS

Hardware

The majority of analyses was performed with an Apple®MacBook®Pro running Mac OSX 10.5 Leopard®later Mac OSX 10.6 Snow Leopard®as operating system. The system was constituted with a 2.53 GHz Intel®Core 2 Duo computational processing unit (CPU) and 4 GB 1067 MHz double data rate (DDR) 3-synchronous dynamic random access memory (SDRAM). However some analyses (RNA2D3D (Martinez et al. 2008), RNAstructure (Mathews et al. 2004, until the release of a Mac version in 2009) required a SuSE™ 10.2 environment or Microsoft®Windows XP®and were performed on a computer using an Intel®Core 2 6600 CPU with 2.40 GHz and 2GB DDR 2-SDRAM.

Additionally, for the simulation study I used a high performance computing (HPC) cluster system. This cluster consists of 40 nodes using each two dual core Intel®5140 CPUs with 2.33 GHz. Eight to 16 GB of DDR 2-SDRAM were allocated for each of the nodes. Each node had a local 20 GB hard drive. Furthermore, 777 GB of network storage are used by the HPC.

Software

Most software tools used are licensed under general public license (GPL) or related open source distribution models. Major software packages used for all analyses and writing of this dissertation were Vienna Tools (Hofacker 2003), EMBOSS (Rice et al. 2000), HMMer (Eddy 1998), BLAST+ (Camacho et al. 2009), Phylip (Felsenstein 1989), R (R Development Core Team 2009), Perl (Perl Foundation 2010) and LaTeX (Latex Project Team 2010). Standalone tools were RNAstructure (Mathews et al. 2004), 4SALE (Seibel et al. 2006, 2008), CB-Analyzer (Wolf et al. 2005a), ProfDist (Friedrich et al. 2005; Wolf et al. 2008), SISSI (Gesell and von Haeseler 2006), ASSEMBLE (Jossinet and Westhof 2010), RNA2D3D (Martinez et al. 2008), Chimera (Pettersen et al. 2004), DIVA (Ronquist 1997), Figtree (Rambaut 2007), iTol (Letunic and Bork 2007), NJPlot (Gouy 1995), Pseudoviewer (Byun and Han 2009), Cytoscape (Shannon et al. 2003), Wordle (Feinberg 2009) and Fugu (Mortensen 2010). Commercial products used were Papers (Griekspoor 2010), Paup* (Swofford 2002), Adobe®Photoshop®and Illustrator®, CorelDRAW®, Microsoft®Office for Mac, Textmate (Macromates 2010). Software versions are stated within the corresponding sections as they have been regularly updated. Online tools frequently used were Leo (LEO GmbH 2010), PubMed (NCBI 2010), Google Scholar™ (Google Inc. 2010) and Wikipedia®(Wikimedia Foundation Inc. 2010).

DNA sequences and Databases

Raw data, as e.g. DNA sequences was either determined by collaboration partners in the lab or retrieved from public databases. In the former case, please see the corresponding manuscripts for laboratory conditions and primers. Publicly stored data was retrieved in the case of DNA / RNA sequences from GenBank (Benson et al. 1999), the ITS2-database (Koetschan et al. 2010), the European ribosomal RNA database (Wuyts et al. 2004) or Rfam (Griffiths-Jones et al. 2003). All ITS2 secondary structures were retained from the ITS2-database (Koetschan et al. 2010) or manually folded. Secondary structures of 18S data originated from the Strand-Database (Andronescu et al. 2008). Tertiary structure motifs were taken from the PDB (Henrick et al. 2008). Please refer to the corresponding publications regarding which data has been used from each of the databases.

CHAPTER
TWO

BIOINFORMATIC TOOLS

2.1. Annotation Tool

Types of HMMs

We estimated hidden Markov models (HMMs) with HMMer 2.3.2 (Eddy 1998) in order to define the borders of the ITS2. Separate HMMs were trained for animals, plants, and fungi. For taxon sampling, we downloaded all sequences from GenBank (Benson et al. 1999) for each of the three taxonomic groups matching a specific search pattern (e.g. animals: "Metazoa[ORGN] AND (ITS2 OR 'internal transcribed spacer 2')"). For unspecific HMMs usable for the vast majority of eukaryotes, we combined the taxon-specific alignments and estimated eukaryote HMMs for start and end of the ITS2.

Locations of the HMMs

We defined the boundaries of the ITS2 in accordance with the European ribosomal RNA database (Wuyts et al. 2004), Rfam (Griffiths-Jones et al. 2003) and the structural characteristics of the ribosomal cistron of *Apis* (Gillespie et al. 2006).

Procedure

Start and end HMMs were each comprised at 25 nucleotides preceding (3' end of 5.8S) and following (5' end of 28S) the ITS2, respectively. Of the retained sequences, 200 were chosen at random with at most one sequence per genus to avoid dominance of intensively studied genera. Taxa present twice in the dataset due to synonymous names in GenBank, as well as unidentified and undescribed species and sequences with less than 25 nucleotides of the ribosomal subunits, were manually removed. Finally, all sequences were manually aligned and cropped. All HMMs were calibrated with hmmcalibrate of the HMMer package (Eddy 1998). We embedded the HMMs with Perl (Perl Foundation 2010) into a web interface with a flexible graphical user interface.

CHAPTER

THREE

BIOINFORMATIC APPROACHES

3.1. HMM-Annotation

In all phylogenetic studies using the ITS2 and following the publication "5.8S/28S interaction and HMM based ITS2 annotation" (Keller et al. 2009a), all sequences were prior to secondary structure prediction annotated and delineated with a local perl version of the HMM-annotation tool accessible at the ITS2 database (Koetschan et al. 2010) and described in Section P.1 and Section 2.1.

3.2. Secondary Structure Prediction

Secondary structures were either directly folded with the help of RNAstructure (Mathews et al. 2004) or predicted via homology modeling (Wolf et al. 2005b). Independent of the method of acquisition they were displayed with Pseudoviewer 3 (Byun and Han 2009) and if necessary manually corrected for missing bonds in stem regions. This manual correction was replaced in the article "ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus Agrodiaetus (Lepidoptera: Lycaenidae: Polyommatus)" (Wiemers et al. 2009, Section P.7) by an automated perl script that applies as a standalone version the Nussinov algorithm as implemented for ITS2 homology modeling at the ITS2-database (Wolf et al. 2005b).

Direct folding

For direct folding, standard parameters for RNA folding in RNAstructure (Mathews et al. 2004) were used. Usually, the optimal structure was retained from the output, however in some cases where one of the suboptimal fitted ITS2 characteristics better, a suboptimal structure was chosen.

Homology modelling

Homology modelling was performed by using the custom modelling option as provided with the ITS2 Database (identity matrix and 50% threshold for the helix transfer) (Koetschan et al. 2010; Wolf et al. 2005b). A template was either obtained by direct folding or extracted from the ITS2 database. Templates for the individual studies are mentioned in the corresponding article sections.

3.3. Tertiary Structure Prediction

We applied two bioinformatics methods to determine the ITS2 (including 25 nucleotides of each 5.8S and 28S ribosomal RNA as a proximal stem) three-dimensional structure for the model organism *Chlamydomonas rheinhardtii*. With RNA2D3D (Martinez et al. 2008) furthermore two closely related organisms were used for investigation (*C. debaryana* and *Gonium pectorale*).

ASSEMBLE

The first tool was ASSEMBLE (Jossinet and Westhof 2010) as part of the S2S platform (Jossinet and Westhof 2005). Tertiary structure models are generated by splitting paired and unpaired regions in separate building blocks. Helical properties are calculated so that stem regions result in a double helix, whereas bulges and loops result in single stranded helical regions. Information from the PDB database can be applied to selections so that the topologies are adapted according to structural motives (Henrick et al. 2008). During or after such processing, the building blocks may be stacked to a single three-dimensional model of the complete molecule. Furthermore, the software allows alignment and homology modeling of homologous molecules.

RNA2D3D

As a second tool, we used RNA2D3D (Martinez et al. 2008), which is a more automated attempt for three-dimensional model prediction of a complete molecule. Unpaired regions are simple estimations of a planar topology and thus no further manipulation is necessary to receive a continuous structure. However, further manipulations are possible if the knowledge is present for the molecule of interest (Martinez et al. 2008). In a comparison with laboratory-verified structures, it is described within this publication that models are good initial estimations.

CHAPTER

FOUR

PHYLOGENETIC PROCEDURES

4.1. Alignments

Sequences and sequence-structure-pairs were in all studies automatically and synchronously aligned with 4SALE 1.5 (Seibel et al. 2006, 2008) as the standard software for alignments. 4SALE translates sequence-structure tuple information prior to alignment into pseudo-proteins. Pseudo-proteins were coded such that each of the four nucleotides may be present in three different states: unpaired, opening base-pair and closing base-pair. Thus, an ITS2 specific 12 x 12 scoring matrix was used for calculation of the alignments (Seibel et al. 2006, 2008). Sequence-structure alignments calculated for this dissertation are available at the ITS2 database supplements page (Koetschan et al. 2010).

4.2. Substitution model selection

Sequence only analyses

For the complete alignment we tested for appropriate models of nucleotide substitution using the Akaike information criterion (AIC) as implemented in Modeltest (Posada and Crandall 1998). Resulting models were general time reversible models, which were used for PAUP* (Swofford 2002) ML (Felsenstein 1981), MP (Camin and Sokal 1965) and NJ analyses (Saitou and Nei 1987). MrBayes (Huelsenbeck et al. 2001) and RAxML (Stamatakis et al. 2008) do not require estimated substitution rates, since they estimate these during the tree reconstruction procedure.

Sequence-Structure analyses

For reconstructions that integrate secondary structures we used a general time reversible (GTR) model working on a 12 letter alphabet. It inherits the four nucleotides in three structural states (unpaired, paired left, paired right), equivalent to the 12 letter alphabet used in 4SALE (Seibel et al. 2006, 2008). This GTR model using ML distances is included within the ProfDistS (Wolf et al. 2008) distribution.

4.3. Tree reconstructions

Maximum likelihood

ML (Felsenstein 1981) analysis were performed with a heuristic search (ten random taxon addition replicates) and nearest neighbour interchange (NNI). ML was usually performed within PAUP*. However, in the publication "ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales)" (Keller et al. 2008c) we additionally used RAxML (Stamatakis et al. 2008) at the CIPRES portal (Cyberinfrastructure for phylogenetic research 2010) to achieve 1.000 bootstraps with a substitution model estimated by RAxML.

Maximum Parsimony

MP (Camin and Sokal 1965) was accomplished with gaps treated as missing data and all characters coded as "unordered" and equally weighted. MP was as well performed within PAUP*.

Baysesian analyses

Furthermore, with MrBayes (Huelsenbeck and Ronquist 2001; Huelsenbeck et al. 2001) a Bayesian analysis was carried out for tree reconstruction using a GTR substitution model with substitution rates estimated by MrBayes (nst = 6).

Neighbor Joining

We clustered taxonomic units with NJ in PAUP*. Furthermore, using ProfDist (Friedrich et al. 2005), profile neighbor joining (PNJ) trees (Müller et al. 2007) were calculated according to the BioNJ algorithm (Gascuel 1997). Analysis with PNJ were also performed with predefined profiles. For profile definitions in each of the studies please refer to the corresponding sections. Secondary structure trees were exclusively reconstructed by PNJ with ProfDistS (Wolf et al. 2008).

Bootstrapping and Consensus

Usually 1.000 bootstrap pseudoreplicates (Felsenstein 1985) were generated in all analyses unless stated otherwise. One hundred bootstrap replicates were generated for the ML analyses in PAUP*. Consensus trees were build according to the extended majority rule implemented in the Phylip (Felsenstein 1989) package.

4.4. CBC analyses

We utilized CBCanalyzer 1.1 (Müller et al. 2007; Wolf et al. 2005a) to detect CBCs and hCBCs between sequence-structure pairs. Corresponding CBC distances were ported with this software as well into the Newick tree format (Felsenstein et al. 1986) and displayed with tree displaying software.

4.5. Tree viewers

Several different tree viewers were used for the individual studies. These were iTol (Letunic and Bork 2007), FigTree (Rambaut 2007) and NJPlot (Gouy 1995). All trees were after dis-

play exported in portable document format (PDF) or as scalable vector graphics (SVG) und refined with Adobe Illustrator ®or Corel Draw ®.

CHAPTER
FIVE

SIMULATIONS

5.1. Simulations

Simulations of ITS2 sequences were performed with SISSI v0.98 (Gesell and von Haeseler 2006). Secondary structures were included in the simulation process of coevolution by application of two separate GTR models (unpaired regions: Q_{seq} Tab. 5.1; stem regions: Q_{struc} Tab. 5.2). Simulations were started given an ancestral sequence, a reference tree and a certain number of taxa so that 300 different evolutionary scenarios were examined (Tab. 5.3).

Table 5.1.: *ITS2 specific nucleotide relative rate matrix Q_{seq}. These correspond to the rates that are used in ProfDist (Friedrich et al. 2005) for sequence only data.*

	A	C	G	U
A	0.000	0.945	2.297	1.117
C	0.945	0.000	1.040	2.973
G	2.297	1.040	0.000	1.000
U	1.117	2.973	1.000	0.000

Table 5.2.: *ITS2 specific dinucleotide relative rate matrix Q_{struct}. These rates base on the same alignments as the rate estimation for the 16 x 16 GTR in ProfDist (Wolf et al. 2008).*

	AA	AC	AG	AU	CA	CC	CG	CU	GA	GC	GG	GU	UA	UC	UG	UU
AA	0.000	0.000	0.000	0.000	0.039	0.000	0.000	0.000	0.522	0.000	0.000	0.000	1.056	0.000	0.000	0.000
AC	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	1.919	0.000	0.000	0.000	0.338	0.000	0.000
AG	0.000	0.000	0.000	0.000	0.000	0.000	1.023	0.000	0.000	0.000	0.327	0.000	0.000	0.000	1.839	0.000
AU	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.934	0.000	0.000	0.000	0.090	0.000	0.000	0.000	1.049
CA	0.039	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	1.082	0.000	0.000	0.000
CC	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.611	0.000	0.000	0.000	0.248	0.000	0.000
CG	0.000	0.000	1.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.042	0.000	0.000	0.000	0.155	0.000
CU	0.000	0.000	0.000	0.934	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.934	0.000	0.000	0.000	0.111
GA	0.522	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.061	0.000	0.000	0.000
GC	0.000	1.919	0.000	0.000	0.000	1.611	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.910	0.000	0.000
GG	0.000	0.000	0.327	0.000	0.000	0.000	1.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.735	0.000
GU	0.000	0.000	0.000	0.090	0.000	0.000	0.000	0.934	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
UA	1.056	0.000	0.000	0.000	1.082	0.000	0.000	0.000	1.061	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UC	0.000	0.338	0.000	0.000	0.000	0.248	0.000	0.000	0.000	1.910	0.000	0.000	0.000	0.000	0.000	0.000
UG	0.000	0.000	1.839	0.000	0.000	0.000	0.155	0.000	0.000	0.000	1.735	0.000	0.000	0.000	0.000	0.000
UU	0.000	0.000	0.000	1.049	0.000	0.000	0.000	0.111	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000

Table 5.3.: *Factors varied for evolutionary scenarios. Each of the 150 possible combinations between the factors (F1-F3) represented one simulated scenario. GI = GenBank (Benson et al. 1999) Identifier.*

F1 (Ancestor	F1 Clade	GI)	F2 Number of Taxa	F3 Branch Lengths
Achlya	Water molds	3941302	10	0.025
Arabidopsis	Plants	1245677	14	0.050
Gigaspora	Fungi	3493494	18	0.100
Gonium	Green Algae	3192577		0.150
Haliotis	Animals	15810877		0.200
				0.250
				0.300
				0.350
				0.400
				0.450

5.2. Datasets

Sequence data-set

For each scenario, the order of the 2,000 simulated sequence-sets retained from SISSI was shuffled. The first 1,000 were chosen and used as a sequence data-set.

Sequence-structure data-set

For each of the sequence-sets used in the sequence data-set, we determined the individual secondary structure of each sequence by homology modeling with at least 75% helix transfer (Wolf et al. 2005b). The ancestral sequence was used as a template. Thus, for the sequence-structure data-set we combined sequences with their respective secondary structures according to the 4SALE methodology (Seibel et al. 2006, 2008). Note, this approach using individual secondary structures is in contrast to alignments only guided by a consensus structure.

Doubled nucleotide data-set

The remaining 1,000 simulated sequence-sets were used to exemplify effects on phylogenetic analyses of a hypothetical ITS2 gene size duplication. Each sequence of these sets was concatenated with a corresponding sequence of the sequence data-set (same taxon in the simulation trees). Thus we received a data-set of doubled nucleotide content that includes as well 1,000 sequence-sets.

5.3. Robustness and Accuracy

For all sequence sets, PNJs trees were calculated and bootstrapped with 100 pseudo-replicates to retain information about the stability of the resulting tree. Bootstrap support values of all tree branches obtained from the 1,000 sequence-sets of a certain scenario were extracted

and pooled. Furthermore, the resulting trees were compared to the respective reference tree. In this regard, two tree distance quantification methods were applied, Robinson-Foulds distances using the Phylip Package v3.68 (Felsenstein 1989) and Quartet distances using Qdist v1.0.6 (Mailund and Pedersen 2004). Results of all sequence-sets were combined for a given scenario to receive the distributions of bootstrap values, Quartet distances and Robinson-Foulds distances, respectively.

The result of each 14-taxa-scenario was plotted as a boxplot with notches using R v2.9.0 (R Development Core Team 2009). An interpolating spline curve with three degrees of freedom was added. For the remaining scenarios (10 and 18 taxa) only spline curves were added for the sake of clarity.

Part III.

Results

CHAPTER
SIX

CONTRIBUTIONS TO THE METHODOLOGICAL PIPELINE

P.1. 5.8S-28S rRNA interaction and HMM-based ITS2 annotation

Authors: A. Keller*, T. Schleicher, J. Schultz, T. Müller, T. Dandekar, M. Wolf

published 2009

Journal: *Gene*
Volume: 430, Pages: 50–57

*First author

Author's Contributions:
I, J. Schultz and M. Wolf designed and coordinated the study. I and T. Schleicher constructed the alignments. I prepared and calculated the HMMs with support by T. Müller. I developed the webinterface with contributions by J. Schultz. I drafted the manuscript. All authors contributed to the final manuscript and approved it.

Gene 430 (2009) 50–57

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

5.8S-28S rRNA interaction and HMM-based ITS2 annotation

Alexander Keller, Tina Schleicher, Jörg Schultz, Tobias Müller, Thomas Dandekar, Matthias Wolf *

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

ARTICLE INFO

Article history:
Received 10 July 2008
Received in revised form 30 September 2008
Accepted 16 October 2008
Available online 30 October 2008

Received by L. Marino-Ramirez

Keywords:
Hidden Markov models
Internal transcribed spacer 2
Non-coding RNA
Phylogenetics
Ribosomal RNA
Secondary structure

ABSTRACT

The internal transcribed spacer 2 (ITS2) of the nuclear ribosomal repeat unit is one of the most commonly applied phylogenetic markers. It is a fast evolving locus, which makes it appropriate for studies at low taxonomic levels, whereas its secondary structure is well conserved, and tree reconstructions are possible at higher taxonomic levels. However, annotation of start and end positions of the ITS2 differs markedly between studies. This is a severe shortcoming, as prediction of a correct secondary structure by standard *ab initio* folding programs requires accurate identification of the marker in question. Furthermore, the correct structure is essential for multiple sequence alignments based on individual structural features. The present study describes a new tool for the delimitation and identification of the ITS2. It is based on hidden Markov models (HMMs) and verifies annotations by comparison to a conserved structural motif in the 5.8S/28S rRNA regions. Our method was able to identify and delimit the ITS2 in more than 30 000 entries lacking start and end annotations in GenBank. Furthermore, 45 000 ITS2 sequences with a questionable annotation were re-annotated. Approximately 30 000 entries from the ITS2-DB, that uses a homology-based method for structure prediction, were re-annotated. We show that the method is able to correctly annotate an ITS2 as small as 58 nt from *Giardia lamblia* and an ITS2 as large as 1160 nt from humans. Thus, our method should be a valuable guide during the first and crucial step in any ITS2-based phylogenetic analysis: the delineation of the correct sequence. Sequences can be submitted to the following website for HMM-based ITS2 delineation: http://its2.bioapps.biozentrum.uni-wuerzburg.de.

1. Introduction

Since Woese and Fox (1977), the nuclear rRNA cistron is an important region for phylogenetic studies. The small subunit (SSU) and large subunit (LSU) of the ribosome (Fig. 1) present highly conserved markers that can be used in phylogenetic reconstructions at a high taxonomic level (Hershkovitz and Lewis, 1996). In contrast, the fast evolving adjacent spacers have larger variations in their sequences and are thus more widely used for inferences of phylogenies at the species and genus level (Coleman, 2000, 2003; Coleman and Vacquier, 2002; Álvarez and Wendel, 2003; Müller et al., 2007). Application of substitution models in model-based inference to sequences containing these markers concatenated together (e.g. conserved 5.8S and fast evolving ITS2) may lead to erroneous results as the levels of substitutions differ significantly between the markers (Huelsenbeck et al., 1996).

However, one of the spacers, the internal transcribed spacer 2 (ITS2), provides both advantages in one genetic marker. It is increasingly applied to approach not only low-level phylogenetic analyses but also inferences at higher taxonomic levels due to the conservation of the secondary structure across large parts of the tree of life (Coleman, 2003, 2007; Schultz et al., 2005; Wolf et al., 2005; Schultz et al., 2006; Selig et al., 2008). In the field of phylogenetic analyses, methods that make use of secondary structures have been shown to yield more robust alignments and trees than methods that do not include structural information (Biffin et al., 2007; Keller et al., 2008). However, to maximally benefit from the information residing in structural features, it is imperative that the marker in question is correctly identified and delimited. In our experience, an offset of even a few nucleotides may result in inconsistent structures from *ab initio* predictions.

The ITS2 has rapidly gained importance in the biosciences. This is exemplified by the observation that the annual number of PubMed publications with ITS2 in the title has increased from 26 to 155 per year between 1998 and 2008. Furthermore, the ITS2 has even been proposed for use in species barcoding and array technologies (Cangelosi et al., 1997; Ben-David et al., 2007; Landis and Gargas, 2007; Park et al., 2007; Engelmann et al., in press). It is thus essential that delimitation of ITS2 is consistent throughout the bioscience community so that direct comparisons of the resulting sequences and secondary structures can be made. Identification and delimitation of the ITS2 can be difficult and time-consuming, however, owing primarily to its high variability in length and lack of sequence conservation at the nucleotide level. It is preferable to delineate the

Abbreviations: bp, basepairs; CBC, compensatory base change; C2-site, cleavage site 2; HMM, Hidden Markov Model; ITS2, Internal Transcribed Spacer 2; LSU, large subunit; nt, nucleotides; rRNA, ribosomal RNA; SSU, small subunit.

* Corresponding author. Tel.: +49 931 888 4562; fax: +49 931 888 4552.
E-mail address: Matthias.Wolf@biozentrum.uni-wuerzburg.de (M. Wolf).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.10.012

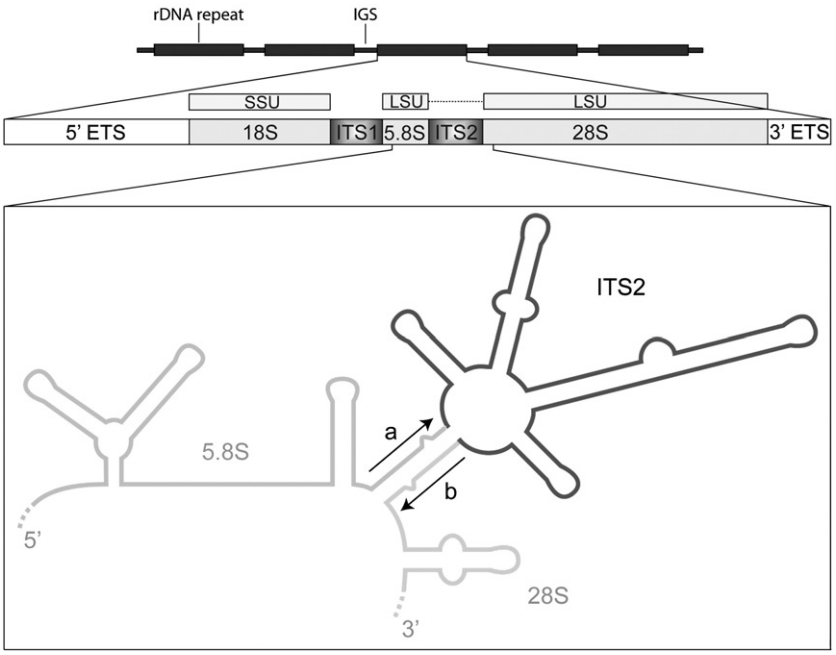


Fig. 1. Location and structural characterization of the ITS2 within the ribosomal repeat (after Coleman, 2003). The proximal stem of the ITS2 is imperfect with one free nucleotide on each of the complementary strands. The regions specified by (a) and (b) are used for HMM modeling (start and end, respectively). The IGS (intergenic spacer) and ETS (external transcribed spacer) are not further treated in this study.

ITS2 by examining the 3′ and 5′ termini of the ribosomal 5.8S and 28S rRNA, respectively, which has been performed only in a few studies. In this paper we present a method based on Hidden Markov Models (HMMs) to delimit ITS2 sequences and verify their annotations, that is related to the procedure used by Nilsson et al. (2008). Furthermore, we evaluate its performance against relevant entries in the international sequence databases.

2. Materials and methods

Since the ITS2 evolves rapidly, the process of identification and delimitation of its boundaries is a complicated task, particularly when there are no highly similar and correctly annotated reference sequences present in the public databases. The high rate of sequence evolution of the ITS2 also means that a eukaryote wide, sequence-based identification of the ITS2 itself is not possible with ordinary methods of sequence pattern recognition. Yet the sequences preceding and following the ITS2 (5.8S and 28S of the LSU) are well conserved. For these we estimated HMMs with HMMer 2.3.2 (Eddy, 1998) in order to define the borders of the ITS2. Start and end HMMs were each comprised at 25 nucleotides preceding (3′ end of 5.8S) and following (5′ end of 28S) the ITS2, respectively. This length is a compromise between precision and usability: 25 nucleotides seem to be long enough to detect only the desired sequences using an *E*-value for detection of significant hits below 0.001, and at the same time short enough to match the DNA product resulting from amplification with most of the commonly used ITS region primers (e.g. White et al., 2001).

Separate HMMs were trained for animals, plants, and fungi. For taxon sampling, we downloaded all sequences from GenBank (Benson et al., 2008) for each of the three taxonomic groups matching a specific search pattern (e.g. animals: ‘ Metazoa[ORGN] AND (ITS2 OR “internal transcribed spacer 2”)’). Of the retained sequences, 200 were chosen at random with at most one sequence per genus to avoid dominance of intensively studied genera. Taxa present twice in the dataset due to synonymous names in GenBank, as well as unidentified and undescribed species and sequences with less than 25 nucleotides of the ribosomal subunits, were manually removed. The resulting sample sizes (start/end) used for HMM modeling are for plants 199/141, animals 195/151, and fungi 177/126. Finally, all sequences were manually aligned and cropped. For more unspecific HMMs usable for the vast majority of eukaryotes, we combined the three taxon-specific alignments and estimated eukaryote HMMs for the start and end of the ITS2. We defined the boundaries of the ITS2 in accordance with the European ribosomal RNA database (Wuyts et al., 2004), Rfam (Griffiths-Jones et al., 2003) and Gillespie et al. (2006). All HMMs were calibrated with *hmmcalibrate* of the HMMer package. We embedded the HMMs with Perl into a web interface with a flexible graphical user interface.

To estimate the performance of the HMMs (*E*-value < 0.001) we randomly chose 100 sequences from GenBank (supplement available at the ITS2-DB) and calculated relative frequencies (P) of positive and negative results for each of the following cases: (1) HMMs were capable to delimit both ends, (2) only the 5.8S start, but not the 28S end, was found, (3) only the 28S end was detectable and (4) no border has been retrieved. To evaluate the results of (1) we manually checked the hybridization of the proximal stem to estimate P (no hybridization | found) and P (hybridization | found). For (2), (3) and (4) we manually examined GenBank entries for ITS2 limits and determined P (present | not found), P (improved search | not found) and P (not present | not found). With an improved search several manipulations were allowed that are not used and recommended for automated annotation and must be manually applied and verified: (a) usage of taxon specific HMMs, (b) ITS2 sequences below 150 nucleotides, (c) reverse complementary and (d) acceptance of *E*-values higher than 0.001. All annotations resulting from this search were manually verified.

Further, we used the unspecific eukaryote HMMs to evaluate all ITS2 sequences matching the search pattern 'ITS2 OR "internal transcribed spacer 2"' in GenBank. In addition, we compared all sequences in the ITS2-DB with the same HMMs (Wolf et al., 2005; Schultz et al., 2005, 2006; Selig et al., 2008). We re-annotated their position and length where necessary. Databases were accessed at the 4th of June 2008.

Finally, we examined various popular sequences and secondary structures of Chlorophyta to compare annotations differing between

GenBank, the ITS2-DB and the HMMs. All *ab initio* secondary structure predictions in this manuscript were performed with RNAstructure 4.6 (Mathews et al. 2004).

3. Results and discussion

3.1. The 5.8S-28S ribosomal RNA interaction

The conservation of the secondary structure of ITS2 sequences is explained by the crucial role of ITS2 during ribogenesis, although ITS2 is subsequently spliced away and thus absent in mature ribosomes (van Nues et al., 1995; Venema and Tollervey, 1995; Mitchell et al., 1996). Several studies pointed out that conserved structural motifs of the ITS2 are necessary for various aspects of ribosome processing, such as the U/C-U pyrimidine-pyrimidine mismatch (Coleman, 2003, 2007; Schultz et al., 2005), the general topology (Joseph et al., 1999; Schultz et al., 2005; Wolf et al., 2005), and the conserved C2-site (Côté et al., 2002; Thomson and Tollervey, 2005). In this study we focus on the essential regions preceding and following the ITS2 (i.e., the 5.8S and 28S, respectively) as they can be used to identify the correct position of the ITS2 sequence (Peculis and Greer, 1998; Côté and Peculis, 2001). The proximal part and the distal part hybridize during ribogenesis into an approximately 15bp imperfect helix and thereby isolate the ITS2 with its typical four-fingered hand structure (Fig. 1).

The hybridized 5.8S and 28S rRNA parts have a free nucleotide on each side with approximately six base pairs in between. When regarding three-dimensional properties (helical turns), the free nucleotides are on the same face of the helix. In exceptional cases the free nucleotide of the 28S part may vary to two free nucleotides mismatching one free nucleotide of the 5.8S rRNA. The structural pattern of this proximal stem is necessary for successful detection of the processing machinery (Venema and Tollervey, 1995, 1999; Côté and Peculis, 2001) and has been proposed for detection of pseudogenes (Harpke and Peterson, 2007, 2008). As the proximal stem is of major importance to the ribosomal machinery and thus well conserved, the structure of this part of the RNA provides maximum certainty for the verification of ITS2 annotations and is a feature easy to spot with common folding algorithms. We recommend that phylogenetic studies using ITS2 sequences should not miss this step of data verification in their preliminary analyses.

3.2. HMM-based annotation of the internal transcribed spacer 2

Apart from the use of the proximal stem for the purpose of verifying the delimitations of the ITS2, it may also contribute to the identification of the spacer region with HMMs. Sequence motifs and degrees of conservation of the sequences used in our HMM-based modeling are displayed in Fig. 2. The proximal stem is remarkably conserved within the three eukaryote kingdoms plants, fungi, and animals. The figure also shows the extensive presence of compensatory base changes (CBCs) and hemi-CBCs (A-U to G-U and vice-versa) within the proximal stem as proofs of secondary structure (Gutell et al., 1994).

Table 1
Number of ITS2 annotations obtained from three different sources (GenBank, ITS2-DB, and HMMs)

	Start	End	Complete
Sequences present in GenBank — HMM hits in GenBank	162 703	81 178	193 708[a] 75 441
Annotations in GenBank — HMM overlap with GenBank-annotation	109 239	48 391	120 179[b] 45 068
Annotations in the ITS2-DB — HMM overlap with ITS-DB-annotation	76 658	34 170	86 084[c] 34 031

[a] matching the search pattern 'ITS2 OR "internal transcribed spacer 2"'.
[b] with either start, end or both annotations.
[c] with start and end annotations.

3.3. Comparison of the HMM-based annotation

To evaluate the quality of current ITS2 annotations, we analysed in a first step all ITS2 sequences present in GenBank matching the search pattern 'ITS2 OR "internal transcribed spacer 2"' with the unspecific eukaryote HMMs. Of the 193 708 sequences only 62% included annotations (Table 1). In total, our HMMs located 162 703 starts of the ITS2 region, 81 178 ends, and 75 441 complete ITS2 sequences with an *E*-value below 0.001. Annotations differed with medians +4 nt (96 416 sequences, more nucleotides with HMM-based annotation)/−14 nt (8 671 sequences, fewer nucleotides with HMM-based annotation) and +2 nt (35 111 sequences)/−30 nt (11 488 sequences) for starts and ends, respectively (Table 2 and Fig. 3). Several unpaired nucleotides (usually 1–10nt) that border the first and last helix of the ITS2 help to preserve the secondary structure. About 80% (starts) and 67% (ends) were less divergent than 5 nucleotides, which could be deemed acceptable in that predictions based on such start and end positions result in similar secondary structures. With HMMer we were able to annotate about 30 000 previously un-annotated GenBank sequences from the ITS region.

In a second step, we compared our predictions with the ITS2-DB (Schultz et al., 2006; Selig et al., 2008) which predicts ITS2 anno-

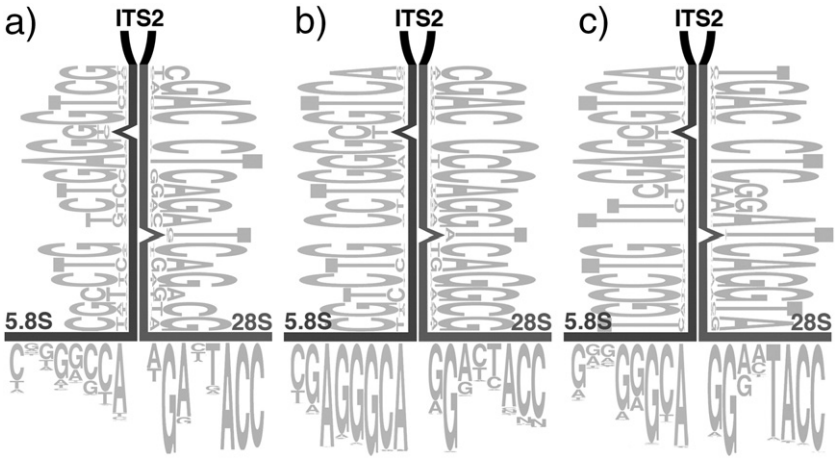


Fig. 2. RNA logos of HMM alignments for three taxonomic groups: (a) animals, (b) plants, and (c) fungi. The proximal stem of the regions adjacent to the ITS2 is well conserved in its sequence and secondary structure. The high saturation of compensatory base changes (CBCs) and hemi-CBCs prove the common secondary structure (Gutell et al., 1994). Created with the help of WebLogo 2.8.2 (Crooks et al., 2004).

Table 2
Medians of positive differences (more nucleotides with HMM-based annotation), identical annotations, and negative differences (fewer nucleotides with HMM-based annotation) with their respective sample sizes in parenthesis resulting from comparisons of HMM-based annotations with GenBank and the ITS2-DB

	HMM start	HMM end
ITS2-DB	+3 (67682)	+3 (21991)
	0 (1841)	0 (817)
	−5 (7135)	−19 (11362)
GenBank	+4 (96416)	+2 (35111)
	0 (4152)	0 (1792)
	−14 (8671)	−30 (11488)

tations by homology modeling (Wolf et al., 2005). Of the 86084 ITS2 sequences with structural information, HMMer was able to annotate 76658 starts and 34170 ends with an *E*-value below 0.001 (Table 1). Out of these, 33031 sequences were annotated with both start and end. The relatively small amount of hits in the 28S region is best explained by the large amount of sequences including fewer than the 25 nucleotides of this part of the subunit necessary for detection. Comparing the results from the homology modeling and the HMM, 67682 annotations differed with a median of +3 nt positively and 7135 sequences negatively with −5 nt as median of deviation for the start (Table 2 and Fig. 3). The corresponding values for the end part were +3 nt (21991 sequences) and −19 nt (11362 sequences). Approximately 74% and 52% of the annotations differed by less than ±5 nucleotides for start and end, respectively. The strong differences in the end annotations may be the result of the homology modeling including more nucleotides to enforce the folding of a short fourth stem. In support of this notion, several reliable structures have been published which lack a fourth helix (Coleman, 2007). In these cases, enforced fourth helices result in erroneous delimitations.

3.4. Examples of deviating annotations

To illustrate the differences between (i) the ITS2 homology modeling, (ii) the HMM-based annotations and (iii) the manually annotated sequences on GenBank we provide examples of green algae (Chlorophyta) in Fig. 4: (a) Since the ITS2 of *Trebouxia glomerata* (GI:187469844) has manually been well annotated, *ab initio* predictions of the secondary structure are possible. Homologous structures are present in the ITS2-DB and thus homology modeling is possible as well. The annotations are in accordance with the HMMs. (b) During the annotation of *Enteromorpha flexuosa* (GI:5834548) a presumable typo caused the annotated sequence to start 100 nucleotides before the true start and increased the sequence length from 193 to 293 nucleotides. This results in structures not predictable with *ab initio* folding algorithms. Yet the correct delimitation is detectable by both homology modeling and the HMMs. Removing nucleotides successively, we found that in this example, even ten nucleotides surplus resulted in a non-homologous structure. It switched back to the correct secondary structure with only five nucleotides more than annotated by HMMs and homology modeling.

The sequence of (c) *Desmodesmus* spec. (GI:169798019) has not been annotated at all and correct prediction by *ab initio* software is not possible. Homology modeling, too, is intractable since the secondary structure differs from the usual four-fingered hand by branching of the first helix (van Hannen et al, 2002; Hegewald and Wolf, 2003; Keller et al. 2008). After correction of the position by the HMMs, correct secondary structure prediction is possible with *ab initio* folding software. This is one example of many instances where annotations are improved by the HMMs (Table 2). (d) An advantage of homology modeling in contrast to the HMMs is that sequences missing the 25 nucleotides of each part of the LSU (5.8S or 28S rRNA) necessary for detection by HMMs or even nucleotides of the ITS2 are nevertheless detectable. They are declared at the ITS2-DB as partial structures (e.g. *Ulva linza*, GI: 157889127).

3.5. Annotation capabilities

The results of the performance tests yielded the following results (Table 3): In case, start and ends were detected by HMMs, all instances resulted in a correct hybridization of the proximal stem (*P* (no hybridization | found)=0.00). Where either the start or the end was detected, no sequence with both tails present remained undetected by an improved search (*P* (present | not found)=0.00). For the last case (both ends not annotated) only in one instance the HMMs failed to detect the boundaries (*P* (present | not found)=0.01). This sequence was from an euglenoid organism, for which no HMMs have been trained. The overall performance is improvable by available user specified modifications (improved search), which is not recommended for an automated and unattended annotation.

In *Giardia lamblia*, a species assumed to resemble the first eukaryotes, the HMMs identified a short ITS2 sequence of 58 nucleotides also recognized by Edlind et al. (1990). This is a major deviation from typical ITS2 sequences averaging at approximately 210 nucleotides. However, the structural motif of the proximal stem is present together with a distinctive lack of one of the unpaired nucleotides (Côté and Peculis, 2001). The 58 ITS2 nucleotides are capable of folding into a stem, which may be a simple elongation of the LSU stem with its structural motif. This could represent an ancestral ITS2 in early eukaryotes, and one that serves to illustrate the transition from the fused prokaryote 5.8S/23S to the eukaryote 5.8S-ITS2-28S region (Lafontaine and Tollervey, 2001). By contrast, the human ITS2 comprises 1160 nucleotides and is an extremely long sequence and complicated secondary structure (Gonzalez et al. 1990). Yet it, too, was

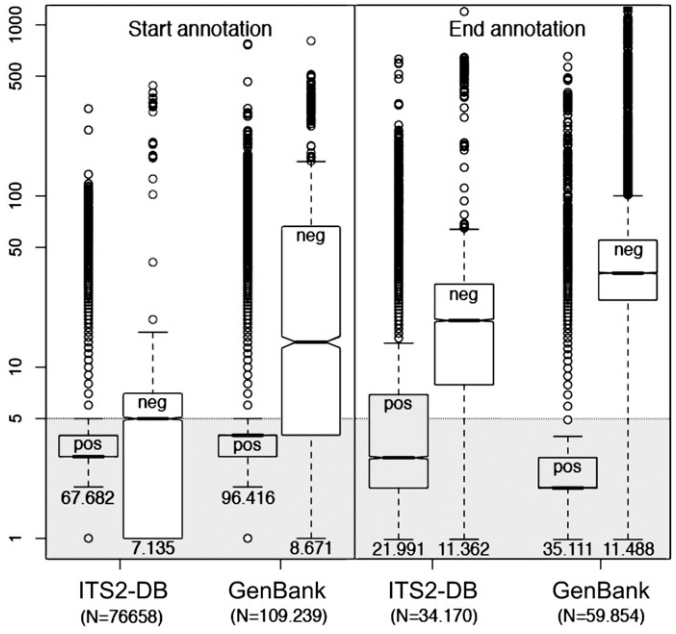


Fig. 3. Distribution of positive and negative deviations of HMM-based annotations from GenBank or the ITS2-DB. Left: start annotation and right: end annotation. Plus-minus five nucleotides are in most cases acceptable as such a difference is small enough not to interfere with the structure prediction procedure. Positive deviations indicate that the use of the HMMs results in a larger ITS2 as compared with GenBank or the ITS2-DB, whereas negative values indicate the opposite. Numbers in parenthesis represent the sample size as the number of sequences with annotations by both HMMs and the respective database. Numbers below the boxplots are the total number of species with positive or negative deviation from the HMMs, respectively.

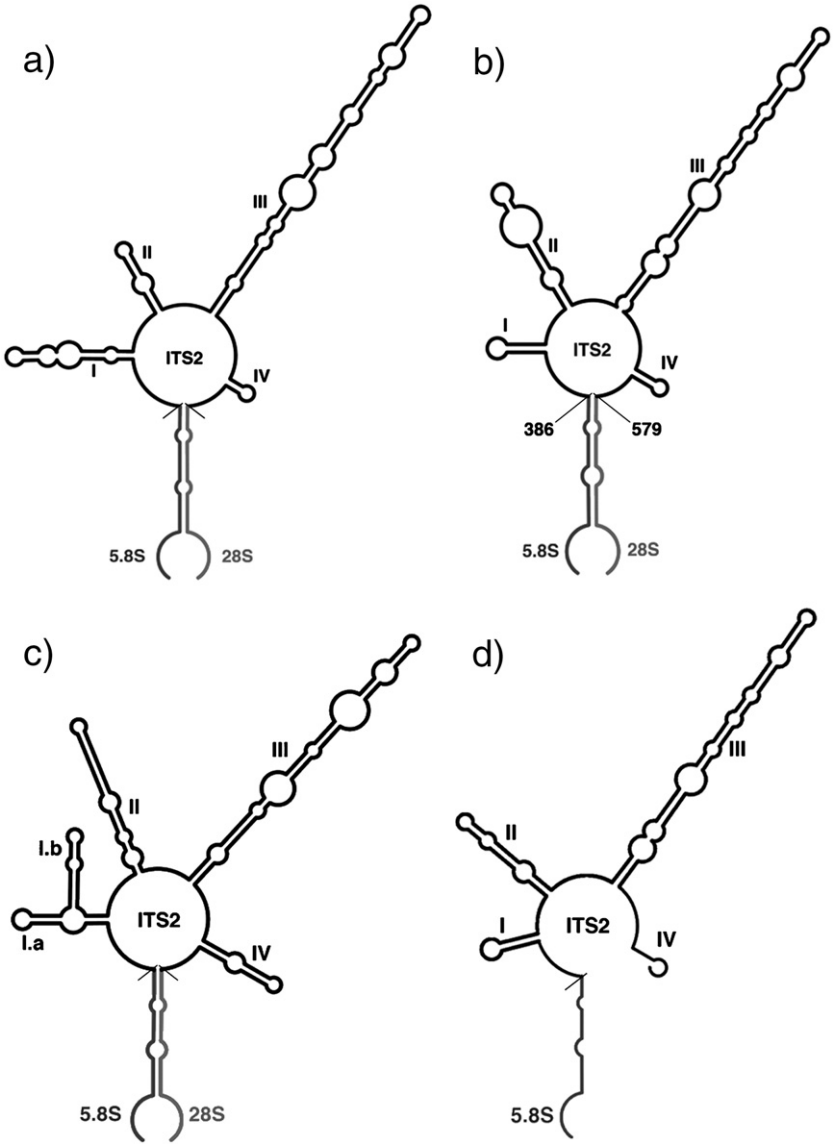


Fig. 4. Comparison of methods for ITS2 annotation: (a) *Ab initio* prediction of the sequence (*Trebouxia glomerata*, GI:187469844) was possible as the entry has been manually annotated with the correct positions. Homology modeling, too, was feasible, since homologous structures are available in the ITS2-DB. The result of the HMM-based annotation was almost identical to the other two methods and refined the annotations only by few nucleotides. (b) Misannotations lead to unpredictable structures by *ab initio* predictors in *Enteromorpha exuosa* (GI:5834548). Homology modeling was able to annotate and fold the sequence correctly. After HMM-based re-annotation, *ab initio* methods were able to fold the sequence into the typical four-fingered hand structure. (c) ITS2-wise misannotated sequences as here of *Desmodesmus* spec (GI:169798019) deviating from the usual four-fingered model were foldable with neither homology modeling nor *ab initio* prediction software. A correction of the sequence by HMMs leads to successful performance by the *ab initio* software. (d) Results for partial sequences can at present only be obtained through homology modeling (*Ulva linza*, GI: 157889127).

annotatable with the web utility and resulted in the common structural motive. As the HMMs of the proximate stem were capable of identifying and verifying both extremes (*Giardia lamblia* and human ITS2), we assume them to be widely applicable throughout the Eukaryota for ITS2 identification and verification.

Table 3
Estimation of conditional probabilities P (event | condition) to estimate the amount of erroneous annotations or undetected borders of 5.8S and 28S limits of the ITS2 by HMMs

	5.8S and 28S	5.8S only	28S only	Neither
P (no hybridization | found)	0.00	–	–	–
P (hybridization | found)	1.00	–	–	–
P (present | not found)	–	0.00	0.00	0.01
P (improved search | not found)	–	0.14	0.72	0.34
P (not present | not found)	–	0.86	0.28	0.65

3.6. A web utility for HMM-based annotation

To provide access to the HMM-based ITS2 identification, we created a web interface for ITS2 delimitation accessible at the ITS2-DB (http://its2.bioapps.biozentrum.uni-wuerzburg.de). The site integrates an HMMer search with five available taxon-specific HMMs: eukaryotes, plants, animals, flies, and fungi. The eukaryote HMM is a combined set of plants, animals, and fungi. We trained a special model for dipterans, because they differ markedly in their architecture of their ribosomal repeat. This taxon is of special scientific interest, since it contains a large amount of disease vectors regularly investigated for their phylogenetic relationships with the ITS2 marker. This is demonstrated by the fact that approximately a fifth of all metazoan ITS2 sequences on GenBank come from this taxonomic group.

At our web service, all alignments used for the HMM modeling are displayed and are downloadable in the FASTA-format (Pearson and Lipman, 1988). For any ITS region sequence pasted into the annotation

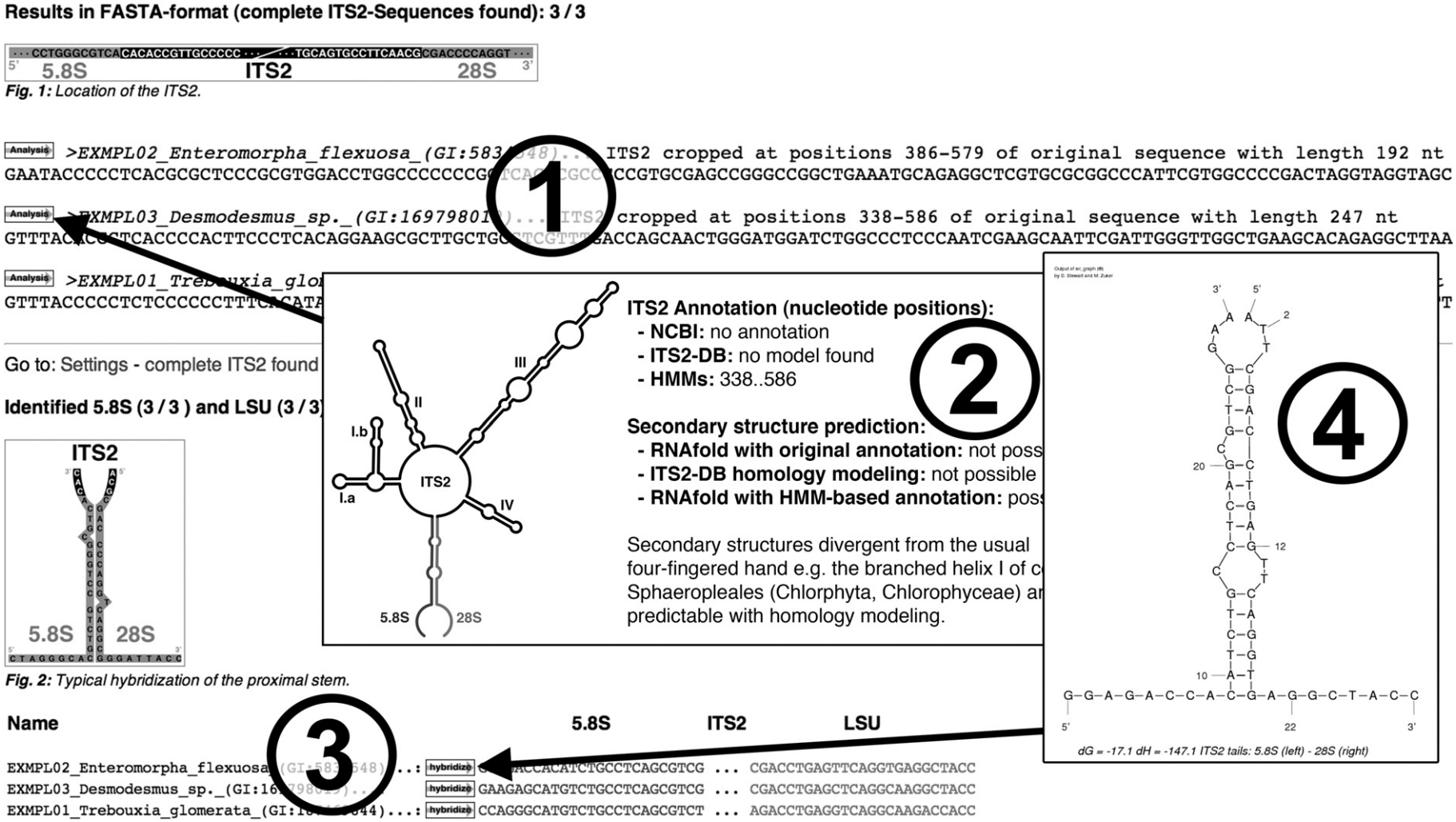


Fig. 5. Screenshot of the results returned by the web utility for HMM-based annotation. Section 1 contains the cropped ITS2 sequences and additional information about the position within the original sequence and the length. In case of the plant examples, the "Analyse" button reveals tooltip information about the annotations of GenBank and the ITS2-DB as well as those of the HMMs as shown in section 2. The preceding and following 25 nucleotides of the ITS2 are displayed in section 3. With a correct annotation, these regions hybridize to form the proximal stem. This can be evaluated by pressing the "Hybridize" button, which will display a tooltip of the stem folding as shown in section 4.

interface of the site, the cropped ITS2 and the bordering regions of the 5.8S rRNA and the 28S rRNA are specified (Fig. 5). For the latter two, a visual representation of the hybridization of the proximal stem is instantly displayed and allows direct verification of the correct annotation. We included an online tutorial by providing examples for plants, metazoans, and fungi. In case of the plant examples, a button will reveal tooltip information about the annotations in GenBank, the ITS2-DB, and by the HMMs.

4. Conclusions

A multitude of ITS2 sequences is available in current nucleotide databases. Yet many of these sequences are not annotated at all or have inconsistent or otherwise compromised annotations. The unreliability of public DNA sequences is another compounding factor (e.g. Koonin et al., 1996; Kyrpides and Ouzounis, 1999; Nilsson et al., 2006; Lin et al., 2008). Sequences lacking annotation, as well as sequences with incorrect annotation, may easily be re-annotated with our web utility proposed in this study.

By contrast, the ITS2-DB uses homology modeling of secondary structure and thus provides a functional criterion for sequence identification. Its reliability is high for structures for which homologous structures are known and integrated into the database. A major advantage of the ITS2–DB is that homologous secondary structures are automatically predicted and are directly usable for phylogenetic studies (Seibel et al., 2006, 2008; Wolf et al., 2008). A problem inherent in the database is that structures that differ from known structures may be difficult to predict. For example several nucleotides belonging to the 28S rRNA are sometimes included to enforce a fourth helix for species lacking it. By verifying the proximal stem with an HMM such mistakes are easily detected and may be manually corrected. Furthermore, branching of helices or additional helices may restrain homology modeling from predicting the correct secondary structure.

We conclude that the application of HMMs for the region of the 5.8S-28S rRNA interaction is of major importance in identification and verification of ITS2 sequences. Since the number of publications with ITS2 structures deviating from the usual four-fingered hand increases (Coleman, 2007), we suggest that such structures are annotated and verified by the integration of the hybridized proximal stem with the web utility for HMM identification to prove their reliability.

Acknowledgments

Funds for the study and the authors were provided by the German federal state Bavaria and the Deutsche Forschungsgemeinschaft (DFG) grant (Mu-2831/1-1). AK additionally appreciates the financial support of the BIGSS graduate school. Further we appreciate the comments of Henrik Nilsson and two anonymous reviewers, which improved the manuscript. We further acknowledge Abram Calderon (University of California, San Francisco) for a language check.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2008.10.012.

References

Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29, 417–434.
Ben-David, T., Melamed, S., Gerson, U., Morin, S., 2007. ITS2 sequences as barcodes for identifying and analyzing spider mites (Acari: Tetranychidae). Exp. Appl. Acarol. 41, 169–181.
Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2008. Genbank. Nucl. Acids Res. 36, D25–D30.
Biffin, E., Harrington, M., Crisp, M., Craven, L., Gadek, P., 2007. Structural partitioning, paired-sites models and evolution of the ITS transcript in *Syzygium* and Myrtaceae. Mol. Phylogenet. Evol. 43, 124–139.
Cangelosi, G.A., Hamlin, A.M., Marin 3rd, R., Scholin, C.A., 1997. Detection of stable pre-rRNA in toxigenic *Pseudo-nitzschia* species. Appl. Environ. Microbiol. 63, 4859–4865.
Coleman, A.W., 2000. The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. Protist 151, 1–9.
Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. TIG 19, 370–375.
Coleman, A.W., 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucl. Acids Res. 35, 3322–3329.
Coleman, A.W., Vacquier, V., 2002. Exploring the phylogenetic utility of ITS sequences for animals: a test case for Abalone (*Haliotis*). J. Mol. Evol. 54, 246–257.
Côté, C., Peculis, B., 2001. Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. Nucl. Acids Res. 29, 2106–2116.
Côté, C., Greer, C., Peculis, B., 2002. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. RNA 8, 786–797.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. Weblogo: A sequence logo generator. Genome Res. 14, 1188–1190.
Eddy, S., 1998. Profile hidden Markov models. Bioinformatics 14, 755–763.
Edlind, T.D., Sharetzky, C., Cha, M.E., 1990. Ribosomal RNA of the primitive eukaryote *Giardia lamblia*: large subunit domain I and potential processing signals. Gene 96, 289–293.
Engelmann, J.C., et al., in press. Modeling cross-hybridization on phylogenetic DNA microarrays increases the detection power of closely related species. Mol. Ecol. Res.
Gillespie, J.J., Johnston, J.S., Cannone, J.J., Gutell, R.R., 2006. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements. Insect Mol. Biol. 15, 657–686.
Gonzalez, I.L., Chambers, C., Gorski, J.L., Stambolian, D., Schmickel, R.D., Sylvester, J.E., 1990. Sequence and structure correlation of human ribosomal transcribed spacers, J. Mol. Biol. 212, 27–35.
Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy, S.R., 2003. Rfam: an RNA family database. Nucl. Acids Res. 31, 439–441.
Gutell, R.R., Larsen, N., Woese, C.R., 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. 58, 10–26.
Harpke, D., Peterson, A., 2007. 5.8S motifs for the identification of pseudogenic ITS regions. Botany 86, 300–305.
Harpke, D., Peterson, A., 2008. Extensive 5.8S nrDNA polymorphism in *Mammillaria* (Cactaceae) with special reference to the identification of pseudogenic internal transcribed spacer regions. J. Plant. Res. 121, 261–270.
Hegewald, E., Wolf, M., 2003. Phylogenetic relationships of *Scenedesmus* and *Acutodesmus* (Chlorophyta, Chlorophyceae) as inferred from 18S rDNA and ITS-2 sequence comparisons. Plant Syst. Evol. 241, 185–191.
Hershkovitz, M.A., Lewis, L.A., 1996. Deep-level diagnostic value of the rDNA-ITS region. Mol. Biol. Evol. 13, 1276–1295.
Huelsenbeck, J.P., Bull, J.J., Cunningham, C.W., 1996. Combining data in phylogenetic analysis. Trends Ecol. Evolut. 11, 152–158.
Joseph, N., Krauskopf, E., Vera, M., Michot, B., 1999. Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. Nucl. Acids Res. 27, 4533–4540.
Keller, A., et al., 2008. ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales). BMC Evol. Biol. 8, 218.
Koonin, E.V., Mushegian, A., Bork, P., 1996. Non-orthologous gene replacement. TIG 12, 334–336.
Kyrpides, N.C., Ouzounis, C.A., 1999. Whole-genome sequence annotation: 'going wrong with confidence'. Mol. Microbiol. 32, 886–887.
Lafontaine, D.L.J., Tollervey, D., 2001. The function and synthesis of ribosomes. Nat. Rev. Mol. Cell. Biol. 2, 514–520.
Landis, F.C., Gargas, A., 2007. Using ITS2 secondary structure to create species-specific oligonucleotide probes for fungi. Mycologia 99, 681–692.
Lin, Y.-H., Chang, B.C.H., Chiang, P.-W., Tang, S.-L., 2008. Questionable 16S ribosomal RNA gene annotations are frequent in completed microbial genomes. Gene 416, 44–47.
Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., Turner, D.H., 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. U. S. A. 101, 7287–7292.
Mitchell, P., Petfalski, E., Tollervey, D., 1996. The 3′ end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. Genes Dev. 10, 502–513.
Müller, T., Philippi, N., Dandekar, T., Schultz, J., Wolf, M., 2007. Distinguishing species. RNA 13, 1469–1472.
Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.-H., Kõljalg, U., 2006. Taxonomic reliability of DNA sequences in public sequence databases: A fungal perspective. PLoS ONE 1, e59.
Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., Larsson, K.-H., 2008. Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. Evol. Bioinform. 4, 193–201.
Park, M.-H., Sim, C.-J., Baek, J., Min, G.-S., 2007. Identification of genes suitable for DNA barcoding of morphologically indistinguishable Korean Halichondriidae sponges. Mol. Cells 23, 220–227.
Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A. 85, 2444–2448.
Peculis, B., Greer, C., 1998. The structure of the ITS2-proximal stem is required for pre-rRNA processing in yeast RNA. RNA 4, 1610–1622.
Schultz, J., Maisel, S., Gerlach, D., Müller, T., Wolf, M., 2005. A common core of secondary

structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11, 361–364.
Schultz, J., Müller, T., Achtziger, M., Seibel, P., Dandekar, T., Wolf, M., 2006. The internal transcribed spacer 2 database-a web server for (not only) low level phylogenetic analyses. Nucl. Acids Res. 34, 704–707.
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.
Selig, C., Wolf, M., Müller, T., Dandekar, T., Schultz, J., 2008. The ITS2 Database II: homology modelling RNA structure for molecular systematics. Nucl. Acids Res. 36, D377–D380.
Thomson, E., Tollervey, D., 2005. Nop53p is required for late 60S ribosome subunit maturation and nuclear export in yeast. RNA 11, 1215–1224.
van Hannen, E.J., Fink, P., Lürling, M., 2002. A revised secondary structure model for the internal transcribed spacer 2 of the green algae *Scenedesmus* and *Desmodesmus* and its implication for the phylogeny of these algae. Eur. J. Phycol. 37, 203–208.
van Nues, R.W., et al., 1995. Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J. Mol. Biol. 250, 24–36.
Venema, J., Tollervey, D., 1995. Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. Yeast 11, 1629–1650.
Venema, J., Tollervey, D., 1999. Ribosome synthesis in *Saccharomyces cerevisae*. Annu. Rev. Genet. 33, 261–311.
White, T.J., Bruns, T.D., Lee, S.B., Taylor, J.W., 2001. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: a Guide to Methods and Applications. Academic Press, London.
Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U. S. A. 74, 5088–5090.
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., Ruderisch, B., Dandekar, T., Schultz, J., Müller, T., 2008. ProfDistS: (Profile-) distance based phylogeny on sequence — structure alignments. Bioinformatics 24, 2401–2402.
Wuyts, J., Perriere, G., van de Peer, Y., 2004. The European ribosomal RNA database. Nucl. Acids Res. 32, D101–D103.

P.2. The ITS2 Dababase III-sequences and structures for phylogeny

Authors: C. Koetschan, F. Förster, A. Keller*, T. Schleicher, B. Ruderisch, R. Schwarz, T. Müller, M. Wolf and J. Schultz

published 2010

Journal: *Nucleic Acids Research*
Volume: 38, Pages: D275–D279

*Co-author

Author's Contributions:
F. Förster, J. Schultz and C. Koetschan did a complete redesign of the database model. F. Förster and C. Koetschan also redesigned the generation and update pipeline including programming and testing. I contributed by adding the annotation procedure to the web-interface and calculating the necessary hidden Markov Models. F. Förster estimated new scoring matrices and gap costs for different alignment methods for ITS2 sequences, sequence-structure pairs together with T. Müller. All authors approved the final version.

Published online 17 November 2009 *Nucleic Acids Research, 2010, Vol. 38, Database issue* ***D275–D279***
doi:10.1093/nar/gkp966

The ITS2 Database III—sequences and structures for phylogeny

Christian Koetschan, Frank Förster, Alexander Keller, Tina Schleicher, Benjamin Ruderisch, Roland Schwarz, Tobias Müller*, Matthias Wolf* and Jörg Schultz*

Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland 97074 Wuerzburg, Germany

Received September 15, 2009; Revised October 12, 2009; Accepted October 13, 2009

ABSTRACT

The internal transcribed spacer 2 (ITS2) is a widely used phylogenetic marker. In the past, it has mainly been used for species level classifications. Nowadays, a wider applicability becomes apparent. Here, the conserved structure of the RNA molecule plays a vital role. We have developed the ITS2 Database (http://its2.bioapps.biozentrum.uni-wuerzburg.de) which holds information about sequence, structure and taxonomic classification of all ITS2 in GenBank. In the new version, we use Hidden Markov models (HMMs) for the identification and delineation of the ITS2 resulting in a major redesign of the annotation pipeline. This allowed the identification of more than 160 000 correct full length and more than 50 000 partial structures. In the web interface, these can now be searched with a modified BLAST considering both sequence and structure, enabling rapid taxon sampling. Novel sequences can be annotated using the HMM based approach and modelled according to multiple template structures. Sequences can be searched for known and newly identified motifs. Together, the database and the web server build an exhaustive resource for ITS2 based phylogenetic analyses.

INTRODUCTION

The internal transcribed spacer 2 (ITS2) of the nuclear rDNA cistron is a widely used phylogenetic marker. In its early years it was specifically used for low-level phylogenetic analyses, i.e. of species within the same genus. At that time, only nucleotide information of the fast evolving sequence was used. With analyses of the two-dimensional structure of the molecule it became evident that the structure is highly conserved throughout the eukaryotes (1–3). The combination of a fast evolving sequence with a slow evolving structure within one molecule suggested its capability for higher level classifications (4). In the last years, the ITS2 has been revealed to be more than just an excellent phylogenetic marker. Its applications include usage as a marker for species identification in environmental samples (phylochips) (5,6), as a target molecule for barcoding (7,8) and for distinguishing species (9). In many of these cases, the structure plays a fundamental role.

Even though sequence databases typically include a large quantity of ITS2 sequences, no coherent information source existed so far including both sequence and structure information, with ITS2 specific annotations. As a consequence of this lack, every scientist had to predict the structure of each molecule in his/her dataset more or less manually. Even worse, in the majority of phylogenetic procedures as e.g. alignment or tree calculation the structure could not be used at all as the corresponding software was not capable of integrating the structure information. In order to tackle these problems and to be better able to exploit the power of this intriguing molecule, we have developed the ITS2 Database. Its goal is to provide a valid structure for every ITS2 sequence within GenBank and thereby to become an exhaustive data source for sequence/structure based phylogenetic analyses, as well as offering tools capable of exploiting the information surplus obtained by these secondary structures. In this article, we describe additions to the ITS2 Database in terms of (i) new developments in automated structure

*To whom correspondence should be addressed. Tel: +49 931 318 4553; Fax: +49 931 318 4552; Email: joerg.schultz@biozentrum.uni-wuerzburg.de
Correspondence may also be addressed to Tobias Müller, Matthias Wolf. Tel: +49 931 318 4562; Fax: +49 931 318 4552;
Email: matthias.wolf@biozentrum.uni-wuerzburg.de
Present address:
Roland Schwarz, Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

prediction, (ii) new features for the access to the data via the Web interface and (iii) new tools for the analysis of ITS2 sequences.

DATA GENERATION

In the previous version of the database, we used a BLAST (10) based approach for the detection of ITS2 in un-annotated GenBank (11) sequences. We were able to predict the structure of more than 35 000 ITS2 where the start- and end-positions were either lacking or misidentified. As BLAST per se is a local alignment tool (hence the name) and the sequence length is very variable throughout the eukaryotes, heuristics had to be implemented to identify the start and end points of the ITS2. To improve this approach, we have recently developed a Hidden Markov Model (HMM) based method for the correct delineation of the ITS2 (12). Start and end position are inferred from the surrounding 5.8S and 28S regions, that are highly conserved. This method initiated a complete re-design of data generation for the ITS2 Database (Figure 1). In the initial step, we searched through the complete nucleotide database (nt) of GenBank for potential ITS2 sequences using hmmsearch (13). Simultaneously, all annotated ITS2 were extracted from GenBank. In cases where both methods were informative about the position of the ITS2, the HMM based information superseded that from GenBank. This led to 196 697 sequences with positional information of the ITS2 (Database accessed at the 22 June 2009). In the second step, all retained sequences were folded using UNAfold (14). Typical ITS2 features were shown by 63 645 structures, namely the conserved core of four helices with the third as the longest. This was a substantial increase compared to the previous approach where only GenBank annotations were taken into account. This indicated the necessity of a correct delineation for the folding step. In the next step, these structures served as templates in the homology modelling process. In contrast to the previous approach, we iterated the homology modelling process until no further new correct structures were identified. This resulted in an additional 99 010 predicted full-length structures, further underlining the presence of a conserved structural core of the ITS2 throughout all eukaryotes. Remaining sequences which could either not be homology modelled or where start and end position could not be predicted run through a final step resulting in partial structures. A BLAST search against all identified sequence structure pairs was performed. All significant hits (E-value $< 10^{-10}$) were extended in both directions by five bases. Finally, we applied a less strict homology modelling which required at least two concatenated helices with a transfer larger than 75% each. This resulted in more than 50 000 partial structures. Using the modified pipeline, which would run in a single core 1221 days, we now provide structural information for over 210 000 ITS2, doubling the number of the previous version. As a detailed taxonomic breakdown (Table 1) the best coverage is found in fungi and plants with 80 and 93%, respectively. Only for ~25% of the metazoan ITS2 sequences, a structure could be predicted. This could indicate a deviation from the 'common core'. It could also be caused by problems of UNAfold to identify the correct fold, leading to a paucity of templates for homology modelling. Additionally, the ITS2 Database now contains a record for each GenBank entry which was identified either via textual annotation or our HMM based annotation tool, rendering it as an exhaustive resource for ITS2 sequences and structures.

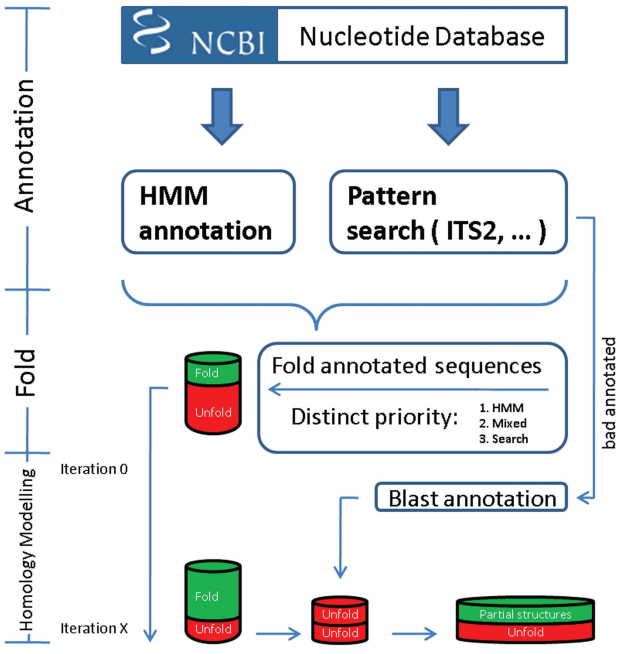


Figure 1. Flow chart of the new pipeline for the ITS2 annotation.

WEB INTERFACE

Search tab

In addition to a search for sequences and structures with GenBank identifiers or species information, we now also provide a BLAST based search. However, standard BLAST procedures are frequently not able to identify distantly related ITS2 sequences because of their high sequence divergence. To overcome this hindrance, we have implemented a sequence and structure based BLAST search that includes information about the highly conserved structure for the homology search. The sequence-structure BLAST uses an ITS2 specific 12×12 scoring matrix representing each nucleotide/structure combination as tuple. This matrix is also used in 4SALE (15) and, as corresponding rate matrix, in ProfDistS (16) for automatic sequence-structure alignment and phylogenetic reconstruction, respectively. Thus, species sampling that starts with any sequence of interest and covers broad taxonomic ranges has become as simple as a BLAST search.

Annotate tab

The web interface does not only present access to the information stored in the database. Further, it provides

Table 1. Taxonomic breakdown of predicted ITS2 structures

	Structure		Partials		All	
	Count	Percentage	Count	Percentage	Count	Percentage
Alveolata	1750	34.67	947	18.76	5048	53.43
Amoebozoa	19	13.01	9	6.16	146	19.18
Apusozoa	0	0.00	0	0.00	35	0.00
Choanoflagellida	0	0.00	0	0.00	1	0.00
Cryptophyta	25	38.46	17	26.15	65	64.62
Environmental samples	26	28.26	7	7.61	92	35.87
Euglenozoa	3	0.62	191	39.71	481	40.33
Fornicata	0	0.00	0	0.00	3	0.00
Fungi	79 251	59.14	28 124	20.99	134 005	80.13
Fungi/Metazoa incertae sedis	2	2.86	0	0.00	70	2.86
Haptophyceae	6	19.35	3	9.68	31	29.03
Heterolobosea	1	0.59	1	0.59	170	1.18
Metazoa	4754	20.14	1357	5.75	23 603	25.89
Nucleariidae	0	0.00	0	0.00	2	0.00
Parabasalidea	1	0.51	0	0.00	197	0.51
Rhizaria	12	2.66	2	0.44	451	3.10
Rhodophyta	27	3.52	28	3.65	768	7.16
Stramenopiles	4441	52.01	2537	29.71	8539	81.72
Viridiplantae	72 322	72.95	20 488	20.67	99 141	93.61
Sum	162 640	59.61	53 711	19.69	272 848	79.29

tools for researchers to process newly determined sequences and to integrate them with already published ones. As shown in the data generation pipeline, correct delineation of the ITS2 sequence can be crucial for structure prediction. We therefore have implemented a web-based interface for the HMM based annotation. It integrates five taxon-specific HMMs for searches and several individually selectable parameters, as e.g. cut-off *E*-value or size limitation. As a result, delimited ITS2 sequences are shown as well as the predicted hybrid of 5.8S and 28S rRNA as a confirmation of the HMM annotation's accuracy (12).

Model tab

After annotation of newly retained ITS2 sequences and selection of a taxon sampling from the ITS2 Database, secondary structures may be determined by two means: First, prediction may be accomplished by homology modelling with the complete set of sequences and structures of the database serving as templates (Predict tab). A second approach is to identify the best template structure within the taxon sampling and use it for homology modelling of the remainders (Model tab). To date, one had to manually run through all possible templates and select the one which resulted in the highest helix transfer percentages. To avoid this tedious and somewhat arbitrary procedure, we now provide the possibility to use multiple sequence-structure pairs to model multiple target sequences. The database will calculate all against all structures and select the template which resulted in the homology prediction with highest percentages of helix transfers for all target sequences.

Similarly, suboptimal structures of a sequence as e.g. retained from minimum free energy folding software, may be given as template input for a set of sequences. As a result, the database will model the structure for all requested sequences with the best fitting suboptimal secondary structure. This is needed, as sometimes the energetically best structure is not the biologically correct one. As the complete homology modelling approach is independent of the ITS2, it may be used to predict the secondary structure of any RNA given a homologous molecule with a known structure.

Motif tab

In addition to the overall structure, conserved motifs like an UGGU sequence preceding the apex of the third helix and a pyrimidine–pyrimidine mismatch in the second helix have been described for the ITS2 (2). In the aforementioned study, identification of these motifs was based on a small dataset and performed mainly by manual inspection. With the availability of the large set of ITS2 sequences in our database, we searched in an automatic way (17) for highly conserved motifs in the ITS2. From our pool of homology models, we randomly extracted a set of unique species. Analysing separately fungal and plant alignments, known and novel motifs were identified. Although the UGGU motif 5′ side to the apex of helix III differs in its composition for fungi, it is located in a corresponding position. For both kingdoms, the U–U mismatch is surrounded by two motifs: one to the left of helix II and one to the right between helix II and III with additional AAA (Figure 2). Having transformed these sequence motifs into HMMs, we now provide identification of these motifs in sequences of interest (Motif tab).

The ITS2 of *Dahlia brevis* as an example

As an example to illustrate the information that can be extracted from the database and the Web interface we analysed the ITS2 of *D. brevis* (18). Looking up the entry for the GenBank identifier 31281745 in the ITS2

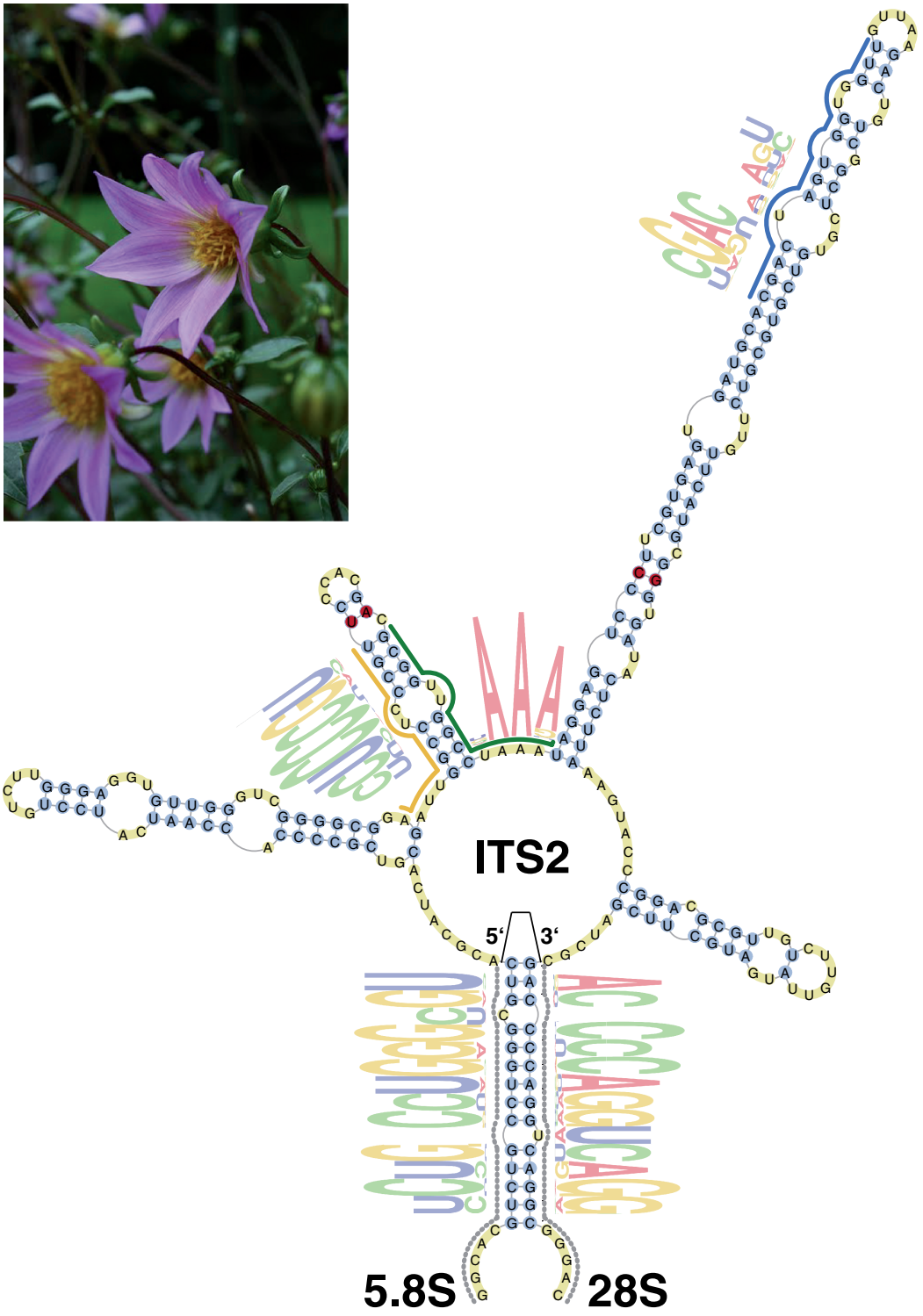


Figure 2. General ITS2 topology and visualization of plant HMM motifs for the secondary structure of *D. brevis* (gi: 31281745). Annotation from HMMs of 5.8S and 28S are displayed as dotted lines tracing the outline of their position, whereas the ITS2 motif HMMs are represented by coloured lines. In parts of these motifs, nucleotide frequencies are presented (21,22). Nucleotides are coloured yellow in unpaired regions, whereas paired nucleotides are blue. CBCs between secondary structures of *D. brevis* and *D. scapigeroides* (gi: 31281755) are shown in red.

Database revealed a stereotypical ITS2 structure (Figure 2). It adopts the common four helix structure with the third as the longest. Additionally, all sequence motifs characteristic for plants are present. In a comparison with another species, here *D. scapigeroides* (gi: 31281755), two Compensatory Base Changes (CBCs) could readily be identified. Indeed, two sequences belong with a probability of 93% to two different species, if at least one CBC is present (9). It should be mentioned, that the CBC criterion works only in one direction. The presence of more than one CBCs indicates with high probability two different species, if there is no CBC, there still could be two species. As *D. brevis* follows all the stereotypes of an ITS2 as the best scoring sequence resulting from all motif searches, it was selected as the 'May 2009' ITS2 in the newly added rubric 'ITS2 of the Month'.

CONCLUSIONS

With the new pipeline for structure prediction, the ITS2 Database now provides information about the structure of more than 210 000 ITS2 molecules, nearly 80% of all ITS2 sequences in GenBank, covering all major taxonomic units. Having the structure available is only the first step for a successful phylogenetic analysis. It would be a pity to use the structure only for the manual refinement of an alignment and neglect it in all other steps. We thus have developed additional stand-alone programs for the entire procedure, which includes automatic alignment calculation [4SALE (15)] as well as tree reconstruction [ProfDistS (16)] considering both, sequences AND secondary structures (these programs have to be downloaded separately). Together, they are seamlessly integrated into a pipeline from sequence through structure and finally to the phylogenetic tree (19). Finally, species boundaries in the dataset can be estimated using the CBCanalyzer [(20), meanwhile also implemented in 4SALE].

The application of secondary structures for the reconstruction of phylogenies improves not only the stability of resulting trees, but more importantly increases the accuracy of phylogenetic estimations (muanuscript under preparation). Thus, it would be desirable to include structural information not only for the ITS2, but also for other frequently used phylogenetic RNA markers.

ACKNOWLEDGEMENTS

We would like to thank Art Vogel (Hortus botanicus Leiden) for permission to use his beautiful photography of *Dahlia brevis*.

FUNDING

Deutsche Forschungsgemeinschaft (DFG) (grant Mu-2831/1-1 to C.K. and T.S.); BIGSS graduate school (to A.K.); Bundesministerium für Bildung und Forschung (BMBF) (grant FUNCRYPTA to F.F.).

Conflict of interest statement. None declared.

REFERENCES

1. Coleman,A.W. (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res.*, **35**, 3322–3329.
2. Mai,J.C. and Coleman,A.W. (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J. Mol. Evol.*, **44**, 258–271.
3. Schultz,J., Maisel,S., Gerlach,D., Muller,T. and Wolf,M. (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA*, **11**, 361–364.
4. Coleman,A.W. (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet*, **19**, 370–375.
5. Engelmann,J., Rahmann,S., Wolf,M., Schultz,J., Fritzilas,E., Kneitz,S., Dandekar,T. and Muller,T. (2008) Modeling cross-hybridization on phylogenetic rDNA microarrays increases the detection power of closely related species. *Mol. Ecol. Res.*, **9**, 83–93.
6. Landis,F.C. and Gargas,A. (2007) Using ITS2 secondary structure to create species-specific oligonucleotide probes for fungi. *Mycologia*, **99**, 681–692.
7. Moniz,M.B. and Kaczmarska,I. (2009) Barcoding of diatoms: nuclear encoded ITS revisited. *Protist*, Epub ahead of print.
8. Ben-David,T., Melamed,S., Gerson,U. and Morin,S. (2007) ITS2 sequences as barcodes for identifying and analyzing spider mites (Acari: Tetranychidae). *Exp. Appl. Acarol.*, **41**, 169–181.
9. Muller,T., Philippi,N., Dandekar,T., Schultz,J. and Wolf,M. (2007) Distinguishing species. *RNA*, **13**, 1469–1472.
10. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
11. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2008) GenBank. *Nucleic Acids Res.*, **36**, D25–D30.
12. Keller,A., Schleicher,T., Schultz,J., Muller,T., Dandekar,T. and Wolf,M. (2009) 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. *Gene*, **430**, 50–57.
13. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
14. Markham,N.R. and Zuker,M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.
15. Seibel,P.N., Muller,T., Dandekar,T. and Wolf,M. (2008) Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res. Notes*, **1**, 91.
16. Wolf,M., Ruderisch,B., Dandekar,T., Schultz,J. and Muller,T. (2008) ProfDistS: (profile-) distance based phylogeny on sequence—structure alignments. *Bioinformatics*, **24**, 2401–2402.
17. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
18. Saar,D.E., Polans,N.O. and Sorensen,P.D. (2003) A phylogenetic analysis of the genus Dahlia (Asteraceae) based on internal and external transcribed spacer regions of nuclear ribosomal DNA. *Syst. Bot.*, **28**, 627–639.
19. Schultz,J. and Wolf,M. (2009) ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics. *Mol. Phylogenet. Evol.*, **52**, 520–523.
20. Wolf,M., Friedrich,J., Dandekar,T. and Muller,T. (2005) CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. *In Silico Biol.*, **5**, 291–294.
21. Byun,Y. and Han,K. (2009) PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics*, **25**, 1435–1437.
22. Gorodkin,J., Heyer,L.J., Brunak,S. and Stormo,G.D. (1997) Displaying the information contents of structural RNA alignments: the structure logos. *Comput. Appl. Biosci.*, **13**, 583–586.

CHAPTER
SEVEN

EVALUATION OF SECONDARY STRUCTURE PHYLOGENETICS

P.3. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees

Authors: A. Keller*, F. Förster*, T. Müller, T. Dandekar, J. Schultz, M. Wolf

published 2010

Journal: *Biology Direct*
Volume: 4, Pages: 5.1–5.12

*Equally contributing first author

Author's Contributions:
Together with M. Wolf, J. Schultz and T. Dandekar, I designed the study. Furthermore, I performed together with F. Förster the simulation studies and statistical analyses. I, F. Förster and M. Wolf drafted the manuscript. F. Förster and T. Müller estimated the substitution models for simulations and reconstructions. All authors contributed to the writing of the manuscript and approved the final version.

Keller et al. Biology Direct 2010, 5:4
http://www.biology-direct.com/content/5/1/4

BIOLOGY DIRECT

RESEARCH Open Access

Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees

Alexander Keller†, Frank Förster†, Tobias Müller, Thomas Dandekar, Jörg Schultz*, Matthias Wolf*

Abstract

Background: In several studies, secondary structures of ribosomal genes have been used to improve the quality of phylogenetic reconstructions. An extensive evaluation of the benefits of secondary structure, however, is lacking.

Results: This is the first study to counter this deficiency. We inspected the accuracy and robustness of phylogenetics with individual secondary structures by simulation experiments for artificial tree topologies with up to 18 taxa and for divergency levels in the range of typical phylogenetic studies. We chose the internal transcribed spacer 2 of the ribosomal cistron as an exemplary marker region. Simulation integrated the coevolution process of sequences with secondary structures. Additionally, the phylogenetic power of marker size duplication was investigated and compared with sequence and sequence-structure reconstruction methods. The results clearly show that accuracy and robustness of Neighbor Joining trees are largely improved by structural information in contrast to sequence only data, whereas a doubled marker size only accounts for robustness.

Conclusions: Individual secondary structures of ribosomal RNA sequences provide a valuable gain of information content that is useful for phylogenetics. Thus, the usage of ITS2 sequence together with secondary structure for taxonomic inferences is recommended. Other reconstruction methods as maximum likelihood, bayesian inference or maximum parsimony may equally profit from secondary structure inclusion.

Reviewers: This article was reviewed by Shamil Sunyaev, Andrea Tanzer (nominated by Frank Eisenhaber) and Eugene V. Koonin.

Open peer review: Reviewed by Shamil Sunyaev, Andrea Tanzer (nominated by Frank Eisenhaber) and Eugene V. Koonin. For the full reviews, please go to the Reviewers' comments section.

Background

In the last decades, traditional morphological systematics has been augmented by novel molecular phylogenetics. One advantage of molecular data is the increased amount of parsimonious informative characters retained from genes that are usable for the inference of evolutionary relationships. This transition from few morphological features to abundant nucleotide or amino acid information has been a breakthrough for investigations of species relationships [1].

However, genetic data often inherits ambiguous information about phylogenetic relationships. Especially for very closely or distantly related taxa, certain parts of data sets may contradict each other or carry insufficient information. Phylogeneticists counter such problems e.g. by increase of the marker's size by inclusion of more nucleotides, thus increasing the amount of available data [2]. Moreover, different markers are combined, so that for example nuclear or mitochondrial genes are concatenated to increase the power of phylogenetic inferences [3,4]. These methods however face new problems. Increase of the number of nucleotides does not necessarily improve the accuracy of a tree reconstruction. Stochastically, only the robustness of the results is increased, if the complete elongated sequence evolved

* Correspondence: Joerg.Schultz@biozentrum.uni-wuerzburg.de; matthias.wolf@biozentrum.uni-wuerzburg.de
† Contributed equally
Department of Bioinformatics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

under the same evolutionary constraints [5]. The second method, marker concatenation, combines genes that result from different evolutionary processes and thus indeed include different evolutionary signals that may improve accuracy. However, they need to be investigated with marker-specific phylogenetic procedures as e.g. varying substitution models [6-8].

In this study we evaluate an alternative method applicable to ribosomal RNA (rRNA) genes that increases information content without addition of nucleotides. As non-coding RNA fragments of the genome, the rRNA gene is generally capable of folding into a secondary structure. In most cases, these structures are necessary for cell function and are thus evolutionarily conserved. Accordingly, structural information may be treated as a conserved marker. Secondary structures of ribosomal RNA therefore offer an additional source of information for tree reconstruction. In particular this is a major advantage in cases where secondary structures are very conserved, yet mutations of nucleotides occur frequently. This applies to the internal transcribed spacer 2 (ITS2) of the eukaryote ribosomal cistron [9,10]. Its secondary structure is evolutionarily maintained as it is of importance in ribogenesis. By contrast, the evolutionary rate of its sequence is relatively high and it is not present in the mature ribosome.

ITS2 sequences have been commonly used to infer phylogenies. Moreover, several studies already included secondary structures in their analyses either by morphometrical matrices or by sequence-structure alignments [11-16]. All these studies agree that the resulting reconstructions are improved by the secondary structures. However, no study has investigated and evaluated this benefit in detail. Evaluations of phylogenetic procedures are typically performed by two different means: the most commonly applied confidence measure in phylogenetics is non-parametric bootstrapping. Bootstrap support values are a measure of robustness of the tree and allow identification of trees or parts of trees that are not unambiguously supported by the data [17,18]. The second point of interest is accuracy measured by the distance between the real and the reconstructed tree. As the 'real' biological tree of life is not available, a switch to sequence simulations along 'real' artificial trees is necessary [19]. In this study we (1) simulate ITS2 sequences along evolutionary trees and (2) compare the results of tree reconstructions by sequence only data and combined sequence-structure data. Additionally, (3) the benefit of structural data is compared with that of sequence elongation. Furthermore, (4) a small biological example of plant phylogeny is presented in which reconstructions that either base on sequence-only or sequence-structure data are compared.

Results

The overall calculation time took 80,000 processor hours on our 40 nodes network cluster. Each node comprised four Xeon 2.33 GHz cores. In total 448 GB RAM were used by the cluster.

The shapes of bootstrap, Quartet distance and Robinson-Foulds distance distributions were similar for equidistant and variable distance trees. However, the branches of the trees for each underlying data set (sequence, sequence-structure and doubled sequence) received higher bootstrap support values and fewer false splits with constant branch lengths compared to variable distances, though differences were minimal (Figs. 1, 2, 3 and 4). Only Quartet distances are shown, since they are congruent with the results of the Robinson-Foulds distance (Additional file 1). Additionally, we included a relative per-branch representation of accuracy divided by the number of internal nodes in the Additional file 1. Bootstrap values and tree distances obtained by differing ancestor sequences were similar in their distributions and thus combined for each scenario during the analysis process. Naturally, with increasing branch lengths, all three investigated data sets (sequences, doubled sequences and sequence-structure) became less accurate and robust, i.e. Quartet distances increased and bootstrap support of nodes decreased. This effect was also observable with an increasing number of external nodes.

Differences between the three methods also increased with evolutionary distance and number of taxa. Thus, the three methods (especially sequence-structure and doubled sequence) yielded almost similar results with low divergence (e.g. branch length 0.05) and few taxa (e.g. 10 taxa), whereas the results were different with branch lengths above 0.25 and at least 14 taxa.

For the lowest branch length we simulated, i.e. 0.025, in comparison to medium divergences a decreased accuracy and bootstrap support was observable with all three methods. This is explainable by too few base changes as providing information for phylogenetic tree reconstruction.

Sequence data performed best in reconstruction of trees (as the maximum and minimum of the spline-curves for bootstraps and tree distances, respectively) at a divergence level between 0.05 and 0.1. Sequence-structure shifted the optimal performance to higher divergences. This effect was also observable for doubled sequence, however it was not as prominent as for sequence-structure.

In general, the robustness of recalculated trees was highest for doubled sequence information contents. However, inclusion of secondary structures largely increased the bootstrap support values of nodes in contrast to normal sequence data. There is thus a

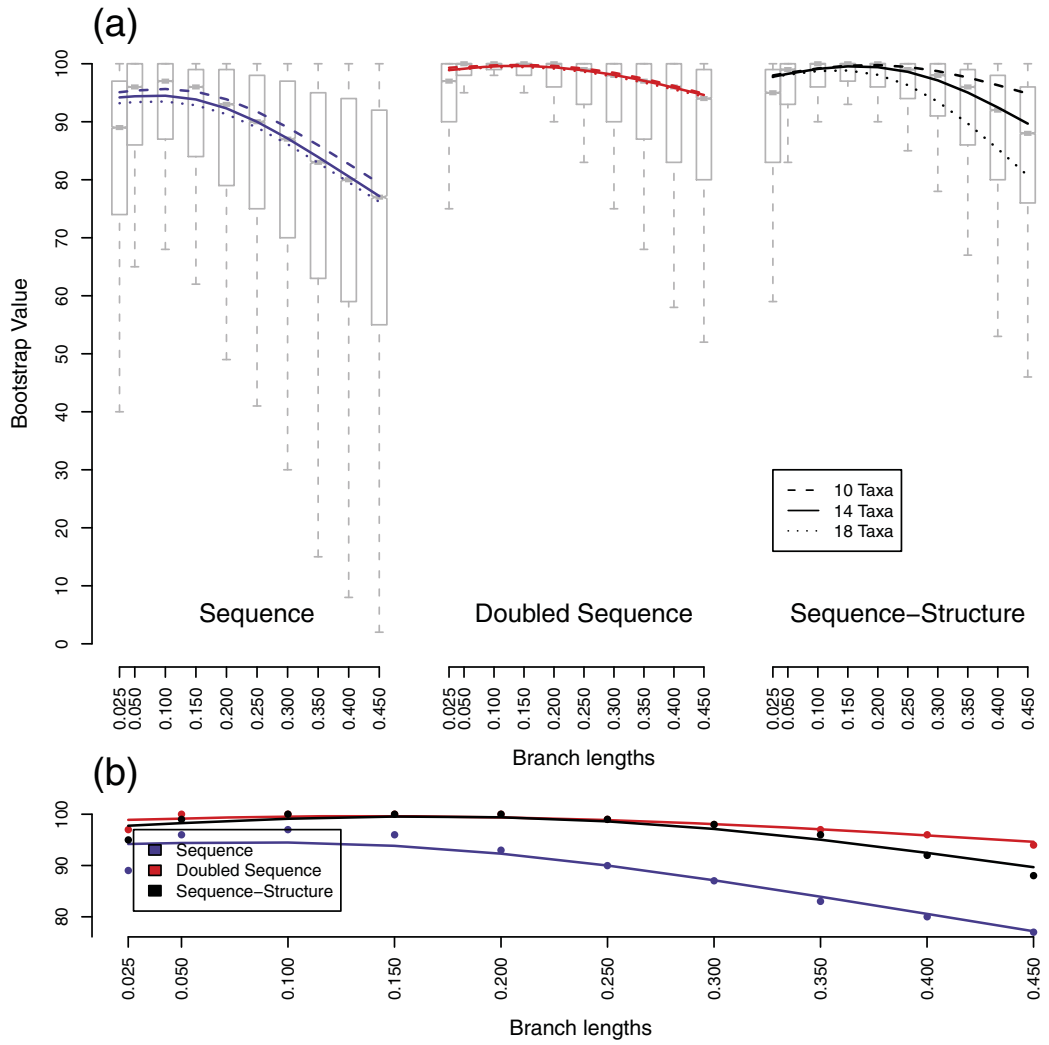


Figure 1 Bootstrap support values for equidistant trees. All five ancestral sequences were combined for a given scenario. (a) Boxplot and solid splines are for 14 taxa scenarios of the three methods. Dashed lines and dotted lines are splines of ten and 18 taxa, respectively. (b) Direct comparison of the 14 taxa splines and medians of all three methods. Sample sizes are 7,000, 11,000 and 15,000 for each of the ten, 14 and 18 taxa scenarios, respectively. Splines show a decrease of robustness with increased number of taxa used and increased branch lengths. Secondary structure and doubled sequences show an improvement in robustness in contrast to normal sequence information.

robustness benefit to using secondary structure that is not directly comparable to benefits achieved by marker elongation.

Additionally, the accuracy of the trees benefitted from secondary structures: the number of false splits was significantly reduced compared to sequence as well as doubled sequence data. Thus sequences-structures yielded the most accurate results in our comparisons.

The results of trees reconstructed with sequence data and sequence-structure data for the plant example were very different. Sequence only information resulted in a correct topology reconstruction of genera (Fig. 5). However, the family of the Malvaceae could not be resolved. This supports the notion that the optimum divergence level of ITS2 sequences is at the species/genus level (see as well Additional file 2). By contrast, all genera and families could be resolved with secondary structures. This results in a flawless tree topology and highlights the improved accuracy. Furthermore, the robustness of the tree has been enhanced and the optimal divergence level has been widened.

Discussion

Number of Taxa and Divergence

Based on the simulations, we draw several conclusions regarding phylogenetic tree reconstructions with and without secondary structures. First of all, the robustness of a tree and its accuracy were significantly negatively

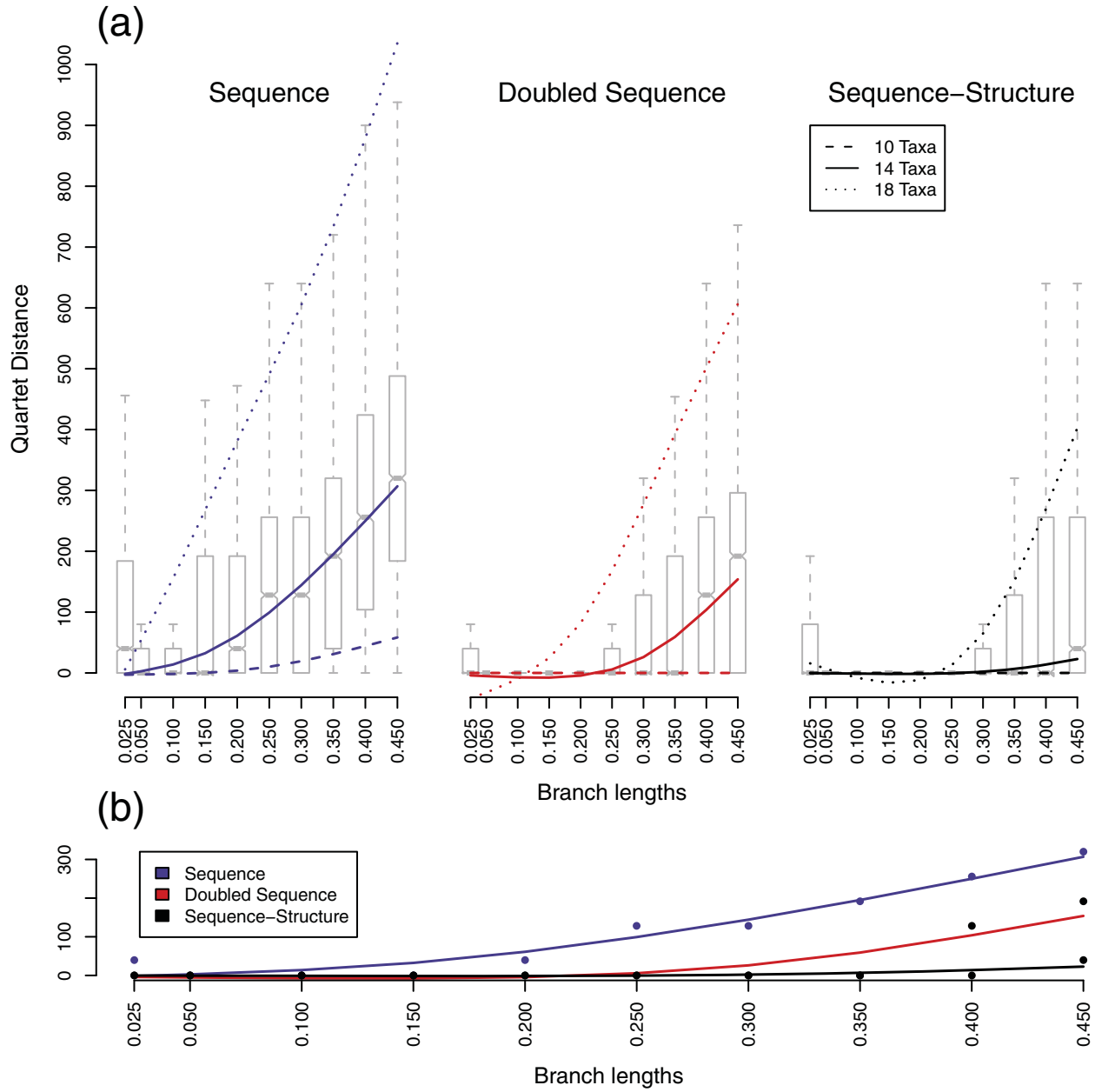


Figure 2 Quartet distances values for equidistant trees. All five ancestral sequences were combined for a given scenario. (a) Boxplot and solid splines are for 14 taxa scenarios of the three methods. Dashed lines and dotted lines are splines of ten and 18 taxa, respectively. (b) Direct comparison of the 14 taxa splines and medians of all three methods. The samples size of each scenario is 1,000. The accuracy of tree topologies decreases with more taxa and greater evolutionary distances between sequences. Trees calculated with secondary structures or doubled sequences show greater accuracy than those determined with normal sequences.

correlated with number of taxa. This is the case even for normalized per-branch accuracy data (Additional file 1). Graybeal [20] argues that an increased taxon sampling enhances accuracy of a resolved tree in the 'Felsenstein zone'. We argue that such an enhancement is the case for special occurrences of long branch attraction, but not, according to our study, for general tree topologies. This is in accordance with Bremer et al. [2] as well as Rokas and Carroll [21], who also notice a slight decrease in accuracy with increased taxon sampling.

Secondly, according to Yang [22], a gene has an optimum level of sequence divergence for phylogenetic studies. The upper limits are reached when the observed difference is saturated, whereas the lower boundary is

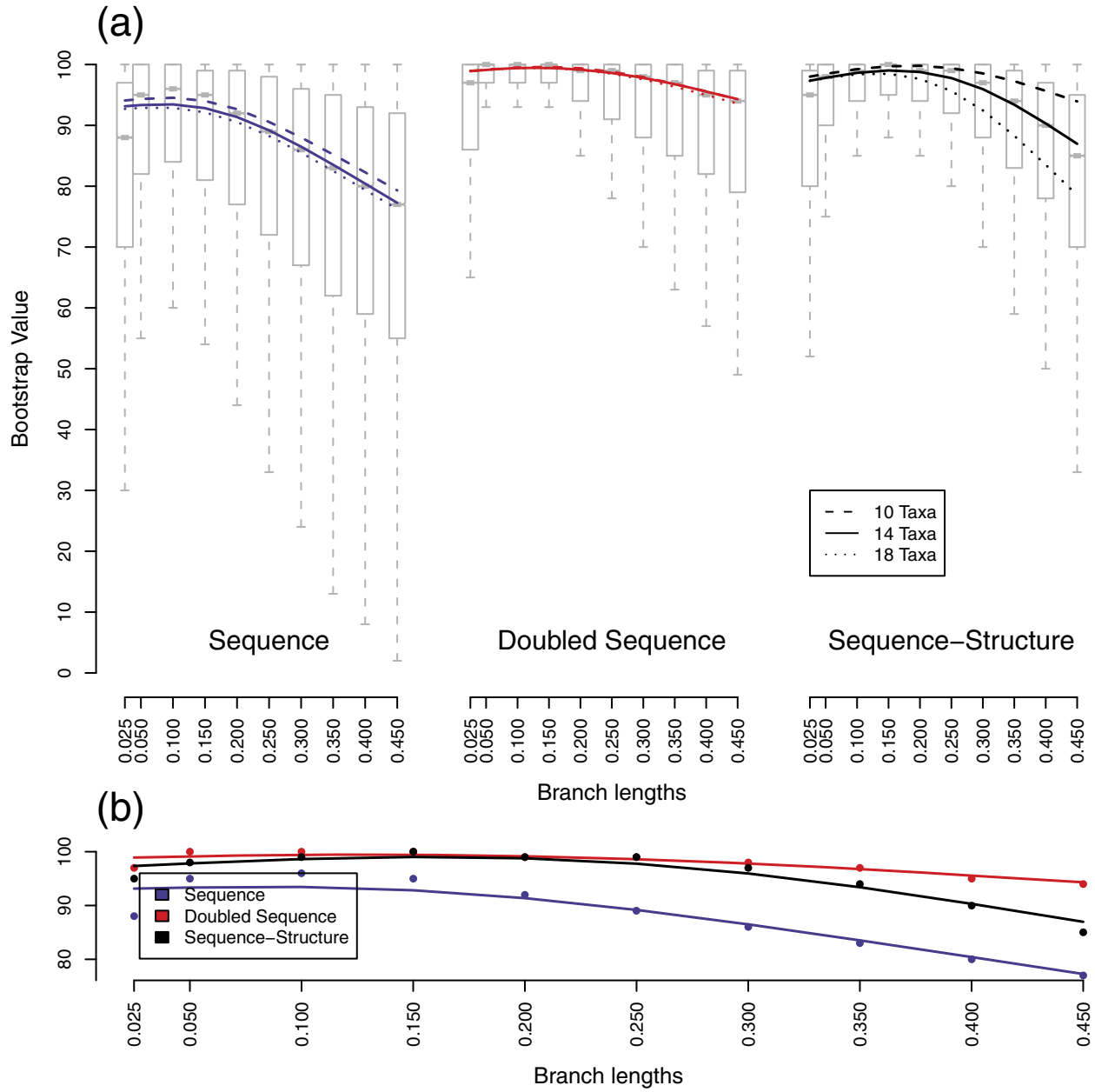


Figure 3 Bootstrap support values for trees with variable branch lengths. Subfigures are explained in Figure 1. Sample sizes are 7,000, 11,000 and 15,000 for each of the ten, 14 and 18 taxa scenarios, respectively.

lack of information content caused by too few substitutions. We observed a similar pattern so that we are able to estimate the divergence level of best performance for ITS2 sequences with and without secondary structures. However, these differ for sequence data and sequence-structure data in two ways: inclusion of secondary structures shifted the best performance to a higher level of divergence. Thus, organisms that are more distantly related can be included in phylogenies. Furthermore, the range of optimal performance is wider for sequence-structure data. A shift to more distantly related sequences does not necessarily mean that relationships of closely related taxa are not any more resolvable. In a review Coleman [9] also identified this potential of ITS2 secondary structures by discussing several case studies. The small biological example of the Malvales and Sapindales in this study supports this notion. Our study mainly covers artificial data: a large scale comparison with biological data regarding the extension of the performance span is still desirable.

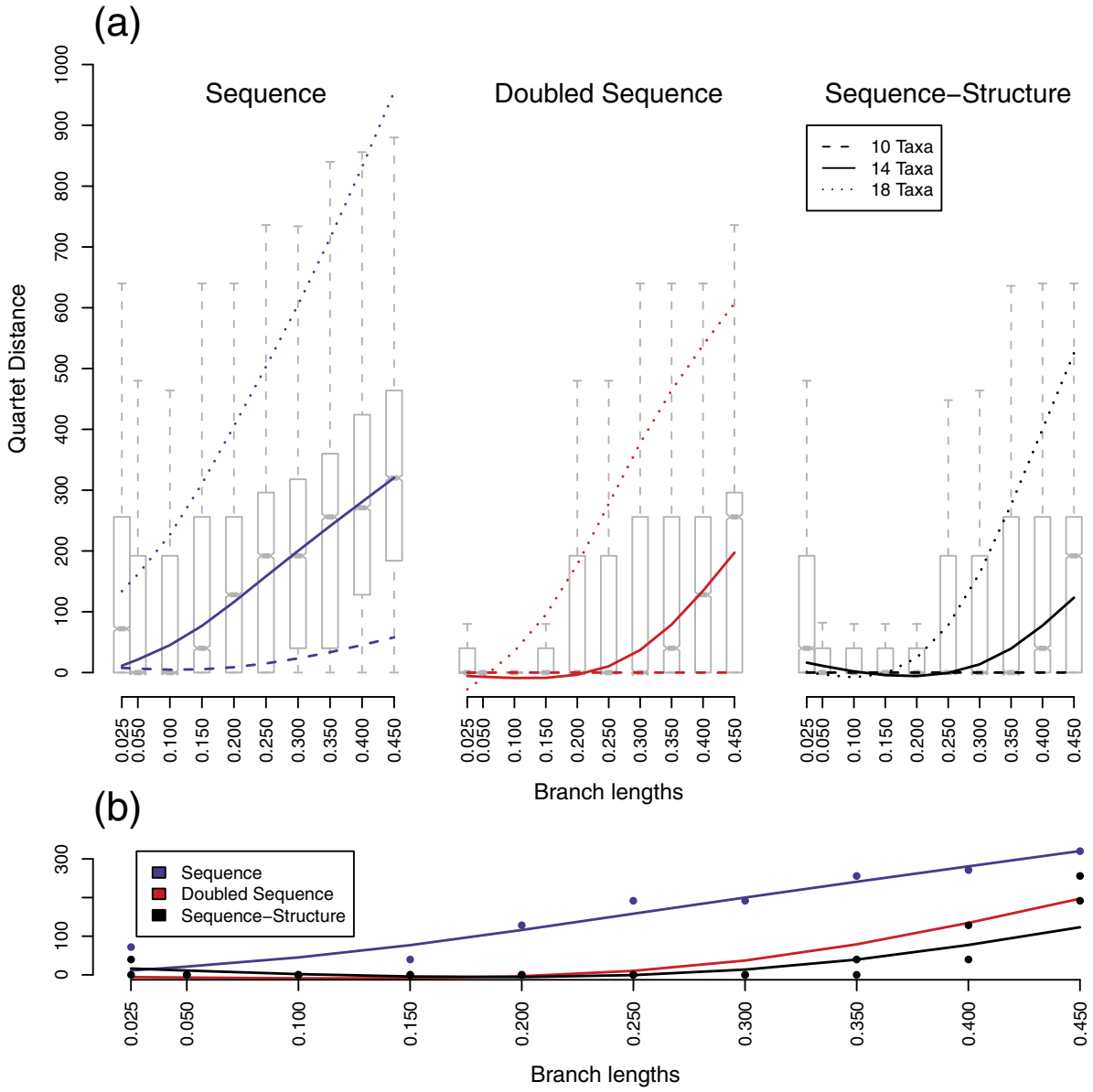


Figure 4 Quartet distances values for trees with variable branch lengths. Subfigures are explained in Figure 2. The samples size of each scenario is 1,000.

Robustness and Accuracy

A substantial benefit to tree robustness was observable when including secondary structure information. Trees reconstructed with secondary structures are generally better bootstrap-supported by the data than those resulting from sequence only data [18]. This is caused by a gain of information content due to increased number of states possible for each nucleotide (unpaired, paired). This information is extractable with a suitable combined score matrix as implemented in 4SALE [23] or similar by site partitioning as in PHASE [24].

The major benefit we identified for phylogenetics is the improvement of accuracy. Sequences-structures performed far better than sequences alone in matching the 'real' tree, especially for high divergences. The resulting immense profit for phylogeneticists is obvious. It is the most crucial property of a phylogenetic tree to be as accurate as possible.

Secondary structure vs. Marker elongation

Both, inclusion of secondary structures and increase of the number of nucleotides improved the reconstructed phylogenetic trees. However, inclusion of secondary

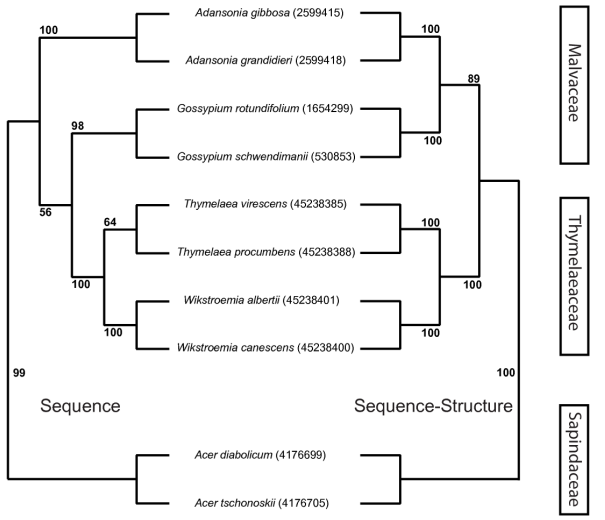


Figure 5 Tree topology of the plants example. Left side: topology and bootstrap values of sequence only data. Right side: corresponding tree with inclusion of secondary structure. Families of the species are given at the right end. GenBank identifiers are in parenthesis after the species names.

structure in the reconstruction process is not equivalent to marker elongation. The major effect of more nucleotides is to increase the bootstrap support values. This has already been demonstrated by other authors [2,5]. With a theoretical increase of marker's length to infinitely large, corresponding bootstraps within a tree will stochastically be maximized as they exactly represent the data. In contrast, the benefit of secondary structures is predominantly the improvement of a tree's accuracy. Thus, additional sequence elongation and secondary structures represent different types of information increase. As the secondary structure analysis already covers the whole marker region of the ITS2 sequence, sequence elongation is not possible for real biological data.

The results retained in this study for the ITS2 region may be transfered to other ribosomal genes. However, the combination of a conserved secondary structure with a variable sequence seems to be of major benefit in phylogenetic studies. Other ribosomal markers, as e.g. 5.8S or 28S rRNA genes may profit less from addition of secondary structures than the ITS2, as the markers themselves are relatively conserved.

Conclusions

Secondary structures of ribosomal RNA provide a valuable gain of information content that is useful for phylogenetics. Both, the robustness and accuracy of tree reconstructions are improved. Furthermore, this enlarges the optimal range of divergence levels for taxonomic inferences with ITS2 sequences. Thus, the usage of ITS2 sequence together with secondary structure for taxonomic inferences is recommended [25]. This pipeline is theoretically as well applicable to other reconstruction methods as maximum likelihood, bayesian inference or maximum parsimony. They may equally profit from secondary structure inclusion.

Methods

Simulation of ITS2 Sequences

Simulations of ITS2 sequences were performed with SISSI v0.98 [26]. Secondary structures were included in the simulation process of coevolution by application of two separate substitution models (Fig. 6, Additional file 3: Tab. 1 and Tab. 2): firstly we used a nucleotide 4 × 4 GTR substitution model Q_{seq} for the evolution of unpaired nucleotides and secondly a dinucleotide 16 × 16 GTR substitution model Q_{struct} for substitution of bases that form stem regions [11,27]. Q_{seq} and Q_{struct} were both estimated by a manually verified alignment based on 500 individual ITS2 sequences and structures with a variant of the method described by Müller and Vingron [28]. For lack of information about insertion and deletion events in the ITS2 region, such were not included into the simulations.

Simulations were started given (a) an ancestral sequence and (b) a reference tree that contained (c) specific branch lengths and (d) a certain number of taxa. In total, we used 10 different branch lengths, 5 ancestral sequences and 6 different trees (3 topologies for equal and variable branch length) resulting in 300 different combinatory conditions as evolutionary scenarios. (a) Ancestral sequences and structures were taken from the ITS2 database after HMM annotation [29-31]. They represented a cross section of the Eukaryota i.e. *Arabidopsis* (Plants) [GenBank:1245677], *Babesia* (Alveolata) [GenBank:119709754], *Gigaspora* (Fungi) [GenBank:3493494], *Gonium* (Green Algae) [GenBank:3192577] and *Haliotis* (Animals) [GenBank:15810877]. (b) The complete procedure was accomplished for two trees that shared a similar topology (Fig. 7). Tree shapes were chosen to resemble trees of a previously published simulation study [32]. The first was a tree that included constant branch lengths, whereas the second tree alternately varied +/- 50% of a given branch length. (c) The used branch lengths were 0.025, 0.05, 0.01, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4 and 0.45. For comparison, pairwise distances of a typical phylogenetic study with ITS2 sequences have been added as Additional file 2. (d) Reference trees were calculated for 10, 14 and 18 taxa. The ancestral sequence served as an origin of the simulated sequences, but was not included in the reconstruction process and resulting tree.

Each simulated sequence set contained sequences according to the number of taxa. Sequence sets were

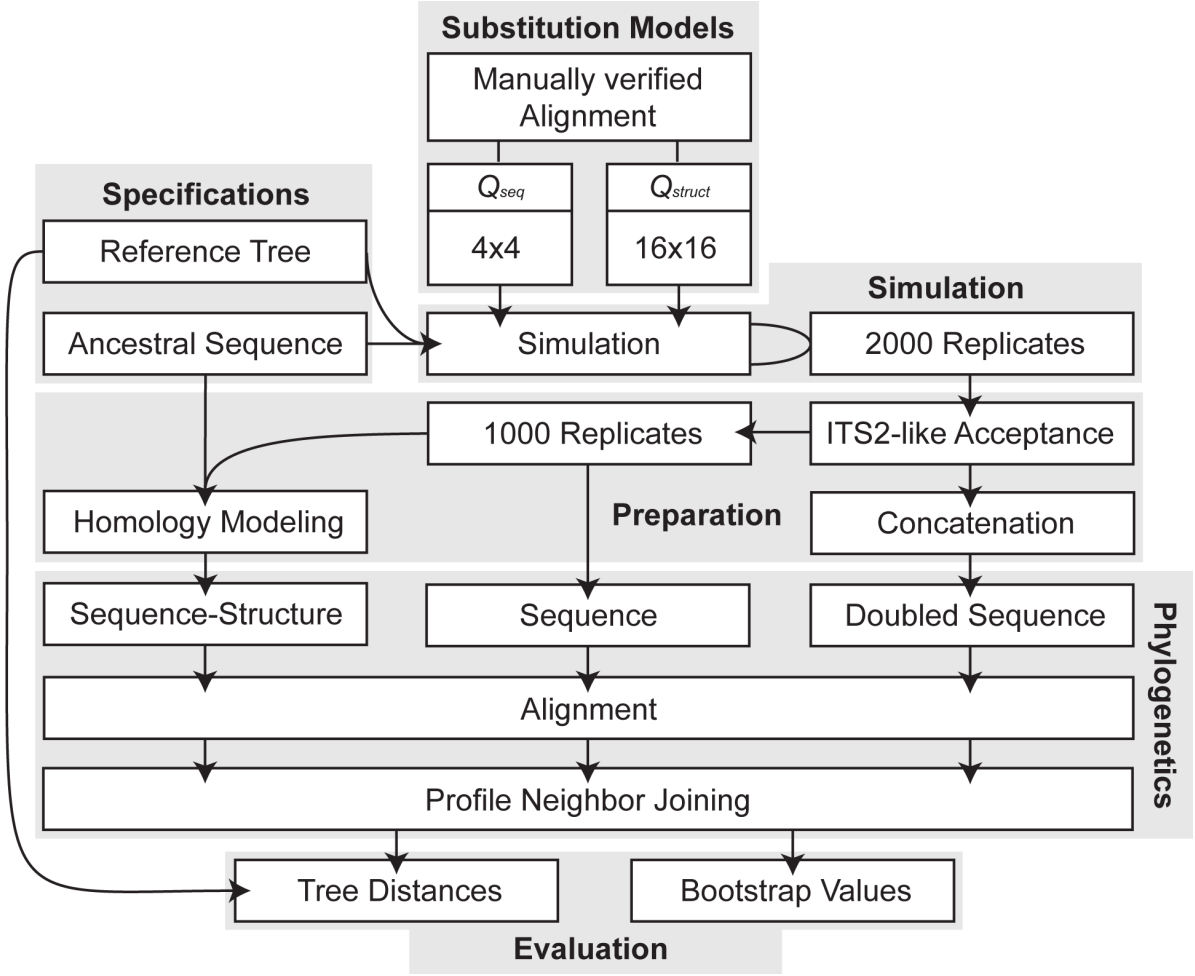


Figure 6 Flowchart of simulation and phylogenetic reconstruction process. Simulation of 2000 replicates of sequence sets was performed along the reference trees with an ancestral ITS2 sequence. Out of these, 1000 sequence-structure, sequence and concatenated sets were generated. Multiple sequence alignments were created for each of these sets and evolutionary distances were estimated with Profile Neighbor Joining. Resulting trees were afterwards compared with the reference trees and regarding their bootstrap support values. Q_{seq} and Q_{struct} are the substitution models for unpaired and paired regions, respectively.

accepted as composed of ITS2-like sequences if the structure of each sequence had been determinable by homology modeling with a threshold of 75% helix transfer [33]. For homology modeling, the ancestral sequence served as a template. Thus, each structure had four helices with the third helix as the longest. This acceptance scheme has been introduced for two reasons: the data is very similar to biological samples [10] and the structure prediction method is equal to that used at the ITS2 database [30] as well as phylogenetic reconstructions [25]. In total, 2,000 valid sequence sets were obtained for each scenario, what corresponds to 600,000 sequence sets summarized over all scenarios.

The complete sequence set is downloadable at the Supplements section of the ITS2 Database http://its2.bioapps.biozentrum.uni-wuerzburg.de/.

Sequences and Structures of the Data Sets

Sequence data set: for each scenario, the order of the 2,000 simulated sequence sets retained from SISSI was shuffled. The first 1,000 were chosen and used as a sequence data set.

Sequence-structure data set: for each of the sequence sets used in the sequence data set, we determined the individual secondary structure of each sequence by homology modeling with at least 75% helix transfer [33]. The ancestral sequence was used as a template. Thus, for the sequence-structure data set we combined sequences with their respective secondary structures according to Seibel et al. [23]. Note, this approach using individual secondary structures is in contrast to alignments only guided by a consensus structure. Doubled nucleotide data set: The remaining 1,000 simulated

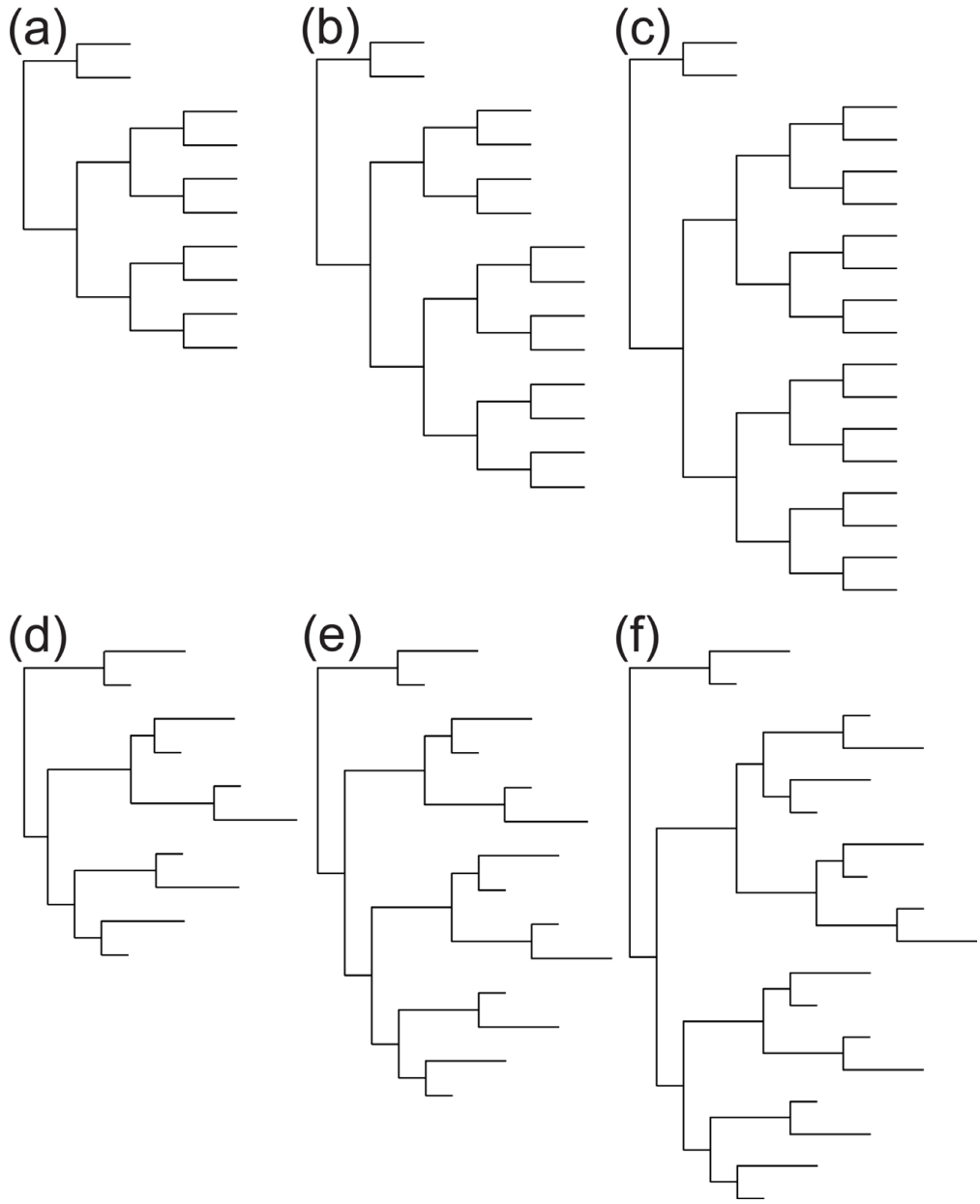


Figure 7 Reference tree topologies used for simulation process. Trees (a), (b) and (c) were trees with equidistance of branches. Trees (d), (e) and (f) were the corresponding variable trees with varying branch lengths. Trees (a) and (d) include ten taxa, (b) and (e) 14 taxa and (c) and (f) 18 taxa.

sequence sets were used to exemplify effects on phylogenetic analyses of a hypothetical ITS2 gene size duplication. Each sequence of these sets was concatenated with a corresponding sequence of the sequence data set (same taxon in the simulation trees). Thus we received a data set of doubled nucleotide content that includes as well 1,000 sequence sets.

Reconstruction of Simulated Phylogenetic Trees

For each simulated sequence set, ClustalW v2.0.10 [34] was used for calculation of multiple sequence alignments. In the cases of sequences and doubled sequences we used an ITS2 specific 4 × 4 scoring matrix [29,30]. For secondary structures, we translated sequence-structure information prior to alignment into pseudoproteins as described for 4SALE v1.5 [23,35]. Pseudoproteins were coded such that each of the four nucleotides may be present in three different states: unpaired, opening base-pair and closing base-pair. Thus, an ITS2 specific 12 × 12 scoring matrix was used for calculation of the alignment [23].

Reconstruction of phylogenetic trees for all trees has been performed with Profile Neighbor Joining (PNJ) of a console version of ProfDistS 0.9.8 [36,37]. With this we estimated improvements due to secondary structures, but keep the method of reconstruction constant. We decided in favor of PNJ and against other methods like maximum likelihood, Bayesian inference and parsimony for several reasons: the distance matrices are independent of insertion and deletion events, the algorithm is very fast and a pipeline for reconstructions with PNJ using secondary structures is already published [25]. However beneficial effects may be transferable to these methods. Profile building was allowed with default settings. General time reversible models (GTRs) were applied with the corresponding 4 × 4 and 12 × 12 substitution matrices for sequences and sequences-structures, respectively.

Robustness and Accuracy

Profile Neighbor Joining trees were bootstrapped with 100 pseudo-replicates to retain information about the stability of the resulting tree. Bootstrap support values of all tree branches obtained from the 1,000 sequence sets of a certain scenario were extracted and pooled. Furthermore, the resulting trees were compared to the respective reference tree. In this regard, two tree distance quantification methods were applied, Robinson-Foulds distances using the Phylip Package v3.68 [38] and Quartet distances using Qdist v1.0.6 [39]. Results of all sequence sets were combined for a given scenario to receive the distributions of bootstrap values, Quartet distances and Robinson-Foulds distances, respectively. The result of each 14-taxa-scenario was plotted as a boxplot with notches using R v2.9.0 [40]. An interpolating spline curve was added. For the remaining scenarios (10 and 18 taxa) only spline curves were added for the sake of clarity.

Short biological case study

Here we provide a short example of ITS2 secondary structure phylogeny, applied to biological data: we sampled sequences of three plant families using the ITS2-database browse feature (database accessed: June 2009): Thymelaeaceae (Malvales), Malvalceae (Malvales) and Sapindaceae (Sapindales). For each family we chose two sequences of the first two appearing genera. Tree reconstruction followed the methods described by Schultz and Wolf [25] and is equivalent to the reconstruction procedure used for the simulated sequence sets. Furthermore, the same procedure was applied without secondary structure information for comparison.

Reviewers' comments

Reviewer's report 1

Shamil Sunyaev, Division of Genetics, Dept. of Medicine, Brigham & Women's Hospital and Harvard Medical School

This manuscript demonstrates the utility of taking into account secondary structure in the phylogenetic analysis. Using comprehensive simulations and a real dataset of ITS2 sequences the authors demonstrated that for higher sequence divergence trees constructed with the help of secondary structure information improve accuracy and robustness. Another interesting result is that addition of taxa may reduce accuracy of tree reconstruction at least in terms of quartet distance between reconstructed and true trees.

Author's response

Thanks a lot for this positive report!

Reviewer's report 2

Andrea Tanzer, Institute for Theoretical Chemistry, University of Vienna (nominated by Frank Eisenhaber, Bioinformatics Institute (BII) Agency for Science, Technology and Research, Singapore)

General comments:

The manuscript "Ribosomal Secondary Structures improve Accuracy and Robustness in Reconstruction of Phylogenetic Trees" compares different methods to improve the quality of phylogenetic analysis. RNA secondary structure information has been included in a variety of previous phylogenetic analysis, but this is the first study exploring the effect on the resulting trees in detail.

The authors use internal transcribed spacer 2 of ribosomal RNAs, a well established set of markers, to simulate a broad spectrum of 300 different scenarios. In addition, they compare their results from the simulations to a set of biological examples from selected plant species.

Overall, the manuscript is carefully written and the authors chose analysis and method appropriately. The simulated sequence set could be used for future studies.

Minor comments:

*) The title might be a little bit miss-leading since 'Ribosomal Secondary Structures' do not improve the 'Accuracy and Robustness in Reconstruction of Phylogenetic Trees' in general and the method should be applicable to other RNA markers. Therefore, I suggest something like "Including Secondary Structures improve Accuracy and Robustness in Reconstruction of Phylogenetic Trees".

*) The setup for the simulations is quite complex. It might help the reader if you add a table or figure to the supplemental material that summarizes the individual conditions for each data set produced.

Alternatively, you could just add to the text that you use 10 different branch length, 5 ancestral sequences and 6 different trees (3 topologies for equal and variable branch length) resulting in 300 different conditions. If I understand this correctly, then you retrieved for each of these 300 conditions 2,000 sequence sets (a total of

600,000 sets), where each set contains 10, 14 and 18 taxa, resp., depending on the tree topology used. These numbers should be mentioned in the text.

*) The set of simulated sequences should be accessible, such that it can be downloaded and used by the community for further studies. Maybe put a link on the website of the ITS2 database.

*) Predicting secondary structures of single sequences occasionally results in (mfe) structures of unexpected shapes. One way to get around this problem is the calculation of consensus structures of a set of related sequences. The resulting consensus structures can then be used for contraint folding of those sequences that could not be folded correctly in the first place. Furthermore, the sequences might fold into a number of equally good structures, but folding programs present only the first result (under default settings). The 'true' structure could as well be among the best folds, but not necessarily the optimal one (suboptimal folding). After all, folding algorithms only make the most plausible predictions. In this study, prediction of RNA secondary structures includes homology modelling. It is of question weather this is the most efficient method. However, since the structures deposited at the ITS2 database were created that way, it seems legitimate to apply it here a well.

Author's response

Thank you for carefully reading the manuscript. We addressed the minor comments regarding text changes and included the necessary information within the text. The set of simulated sequences is now downloadable at the Supplement section of the ITS2 Database http://its2.bioapps.biozentrum.uni-wuerzburg.de/. We totally agree that there are other possibly more efficient methods concerning structure prediction. However, as already stated by Dr. Tanzer 'structures deposited at the ITS2 database were created that way [homology modelling], it seems legitimate to apply it here as well'. The big advantage of the ITS2 is, that the core folding pattern is already known. Therefore, we have an external criterium to check for the correctness of the predicted structures.

Reviewer's report 3

Eugene V. Koonin, National Center for Biotechnology Information, NIH, Bethesda

This is a useful method evaluation work that shows quite convincingly the inclusion of RNA secondary structure information into phylogenetic analysis improves the accuracy of neighbor-joining trees. My only regrets are about a certain lack of generality. It would be helpful to see a similar demonstration for for at least two different kinds of nucleic acid sequences not only ITS2. Also, at the end of the Conclusion section, the authors suggest that secondary structure could help also with other phylogenetic approaches (ML etc). Showing this explicitly would be helpful, especially, given that NJ is hardly the method of choice in today's phylogenetics.

Author's response

Thank you for your encouraging report. For ITS2 the core structure is well known and there are about 200,000 individual secondary structures available. However, it is absolutely right that it would be helpful to perform an analysis also on other types of phylogenetic RNA markers. Unfortunately, today there is no comparable amount of data available concerning secondary structures of other RNAs. Similarily, there are no programs to run an analysis on other methods such as parsimony, maximum likelihood and/or bayesian methods simultanously considering sequence and secondary structure information.

Additional file 1: Normalized Quartet distance and Robinson-Foulds plots. Similar to Figures 2 and 4, but showing per-branch Quartet distances as a normalized standard i.e. divided by number of splits. Robinson-Foulds Distances are given in absolute and normalized versions.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1745-6150-5-4-S1.PDF]

Additional file 2: Empirical pairwise distances. Pairwise distances of an ITS2 case study that integrates secondary structure.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1745-6150-5-4-S2.PDF]

Additional file 3: Substitution matrices. Nucleotide 4 × 4 GTR substitution model Q_{seq} for the evolution of unpaired nucleotides and a dinucleotide 16 × 16 GTR substitution model Q_{struct}.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1745-6150-5-4-S3.PDF]

Acknowledgements

The assistance of Richard Copley (Oxford, United Kingdom) in language correction is greatly appreciated. Financial support for this study was provided by the Deutsche Forschungsgemeinschaft (DFG) grant (Mu-2831/1-1). AK was supported by the BIGSS graduate school of the land Bavaria. FF was supported by the Bundesministerium für Bildung und Forschung (BMBF) grant FUNCRYPTA.

Authors' contributions

AK, JS, MW and TD designed the study. FF and AK performed the simulation experiments and analyses. FF and TM estimated the substitution models used for simulations and reconstructions. AK, FF and MW drafted the manuscript. All authors contributed to writing the paper, read the final manuscript and approved it.

Competing interests

The authors declare that they have no competing interests.

Received: 21 December 2009
Accepted: 15 January 2010 Published: 15 January 2010

References

1. Woese C, Kandler O, Wheelis M: **Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.** *Proc Natl Acad Sci USA* 1990, **87(12)**:4576-4579.

2. Bremer B, Jansen R, Oxelman B, Backlund M, Lantz H, Kim KJ: **More characters or more taxa for a robust phylogeny-case study from the Coffee family (Rubiaceae).** *Syst Biol* 1999, **48(3)**:413-435.
3. van Oppen M, McDonald B, Willis B, Miller D: **The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: reticulation, incomplete lineage sorting, or morphological convergence?.** *Mol Biol Evol* 2001, **18(7)**:1315-1329.
4. Slowinski J, Lawson R: **Snake phylogeny: evidence from nuclear and mitochondrial genes.** *Mol Phylogenet Evol* 2002, **24(2)**:194-202.
5. Erixon P, Svennblad B, Britton T, Oxelman B: **Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics.** *Syst Biol* 2003, **52(5)**:665-73.
6. Whelan S, Liò P, Goldman N: **Molecular phylogenetics: state-of-the-art methods for looking into the past.** *Trends Genet* 2001, **17(5)**:262-72.
7. Posada D, Crandall KA: **The effect of recombination on the accuracy of phylogeny estimation.** *J Mol Evol* 2002, **54(3)**:396-402.
8. Egger B, Koblmüller S, Sturmbauer C, Sefc K: **Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*.** *BMC Evol Biol* 2007, **7**:137.
9. Coleman AW: **ITS2 is a double-edged tool for eukaryote evolutionary comparisons.** *TIG* 2003, **19(7)**:370-375.
10. Coleman AW: **Pan-eukaryote ITS2 homologies revealed by RNA secondary structure.** *Nucleic Acids Res* 2007, **35(10)**:3322-3329.
11. Schöniger M, von Haeseler A: **A stochastic model for the evolution of autocorrelated DNA sequences.** *Mol Phylogenet Evol* 1994, **3(3)**:240-7.
12. Tillier ERM, Collins RA: **High apparent rate of simultaneous compensatory base-pair substitutions in ribosomal RNA.** *Genetics* 1998, **148(4)**:1993-2002.
13. Young I, Coleman AW: **The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example.** *Mol Phylogenet Evol* 2004, **30**:236-242.
14. Biffin E, Harrington M, Crisp M, Craven L, Gadek P: **Structural partitioning, paired-sites models and evolution of the ITS transcript in *Syzygium* and Myrtaceae.** *Mol Phylogenet Evol* 2007, **43**:124-139.
15. Grajales A, Aguilar C, Sanchez J: **Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals.** *BMC Evol Biol* 2007, **7**:90.
16. Keller A, Schleicher T, Förster F, Ruderisch B, Dandekar T, Müller T, Wolf M: **ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales).** *BMC Evol Biol* 2008, **8**:218.
17. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39(4)**:1993-2002.
18. Hillis D, Bull J: **An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis.** *Syst Biol* 1993, **42(2)**:182-192.
19. Hillis DM, Huelsenbeck JP, Cunningham CW: **Application and accuracy of molecular phylogenies.** *Science* 1994, **264(5159)**:671-7.
20. Graybeal A: **Is it better to add taxa or characters to a difficult phylogenetic problem?.** *Syst Biol* 1998, **47**:9-17.
21. Rokas A, Carroll SB: **More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy.** *Mol Biol Evol* 2005, **22(5)**:1337-44.
22. Yang Z: **On the best evolutionary rate for phylogenetic analysis.** *Syst Biol* 1998, **47**:125-33.
23. Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M: **4SALE - a tool for synchronous RNA sequence and secondary structure alignment and editing.** *BMC Bioinformatics* 2006, **7**:498.
24. Jow H, Hudelot C, Rattray M, Higgs P: **Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution.** *Mol Biol Evol* 2002, **19(9)**:1591-1601.
25. Schultz J, Wolf M: **ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics.** *Mol Phylogenet Evol* 2009, **52**:520-523.
26. Gesell T, von Haeseler A: **In silico sequence evolution with site-specific interactions along phylogenetic trees.** *Bioinformatics* 2006, **22(6)**:716-722.
27. Meyer S, von Haeseler A: **Identifying site-specific substitution rates.** *Mol Biol Evol* 2003, **20(2)**:182-189.
28. Müller T, Vingron M: **Modeling amino acid replacement.** *J Comput Biol* 2000, **37(6)**:761-776.
29. Schultz J, Müller T, Achtziger M, Seibel PN, Dandekar T, Wolf M: **The internal transcribed spacer 2 database-a web server for (not only) low level phylogenetic analyses.** *Nucleic Acids Res* 2006, **34(Supp 2)**:W704-707.
30. Selig C, Wolf M, Muller T, Dandekar T, Schultz J: **The ITS2 Database II: homology modelling RNA structure for molecular systematics.** *Nucleic Acids Res* 2008, , **36 Database:** D377-80.
31. Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M: **5.8S-28S rRNA interaction and HMM-based ITS2 annotation.** *Gene* 2009, **430(1-2)**:50-7.
32. Alfaro ME, Zoller S, Lutzoni F: **Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence.** *Mol Biol Evol* 2003, **20(2)**:255-266.
33. Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T: **Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures.** *RNA* 2005, **11(11)**:1616-1623.
34. Thompson J, Higgins D, Gibson T: **ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**:4673-4680.
35. Seibel PN, Müller T, Dandekar T, Wolf M: **Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE.** *BMC Res Notes* 2008, **1**:91.
36. Friedrich J, Dandekar T, Wolf M, Müller T: **ProfDist: a tool for the construction of large phylogenetic trees based on profile distances.** *Bioinformatics* 2005, **21(9)**:2108-2109.
37. Wolf M, Ruderisch B, Dandekar T, Schultz J, Müller T: **ProfDistS: (profile-) distance based phylogeny on sequence - structure alignments.** *Bioinformatics* 2008, **24**:2401-2402.
38. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166.
39. Mailund T, Pedersen CNS: **QDist-quartet distance between evolutionary trees.** *Bioinformatics* 2004, **20(10)**:1636-7.
40. R Development Core Team: *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria 2009http://www.R-project.org.

doi:10.1186/1745-6150-5-4
Cite this article as: Keller *et al.*: Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biology Direct* 2010 **5**:4.

CHAPTER
EIGHT

PHYLOGENETIC CASE STUDIES

P.4. ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales)

Authors: A. Keller*, T. Schleicher*, F. Förster, B. Ruderisch, T. Dandekar, T. Müller, M.Wolf

published 2008

Journal: *BMC Evolutionary Biology*
Volume: 8, Pages: 218.1–218.12

*Equally contributing first author

Author's Contributions:
M. Wolf designed the study. F Förster determined the new sequences in the laboratory. B. Ruderisch implemented the strPNJ within ProfDist. I and T. Schleicher performed sequence analyses, structure prediction and phylogenetic analyses. T. Müller developed the ITS2 sequence-structure substitution model and the ITS2 sequence-structure scoring matrix. T. Schleicher, M. Wolf and I drafted the manuscript. All authors contributed to writing the paper, read the final manuscript and approved it.

BMC Evolutionary Biology

BioMed Central

Research article

Open Access

ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales)

Alexander Keller†, Tina Schleicher†, Frank Förster, Benjamin Ruderisch, Thomas Dandekar, Tobias Müller and Matthias Wolf*

Address: Department of Bioinformatics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

Email: Alexander Keller - a.keller@biozentrum.uni-wuerzburg.de; Tina Schleicher - tina.schleicher@biozentrum.uni-wuerzburg.de; Frank Förster - frank.foerster@biozentrum.uni-wuerzburg.de; Benjamin Ruderisch - b.ruderisch@gmx.de; Thomas Dandekar - dandekar@biozentrum.uni-wuerzburg.de; Tobias Müller - tobias.mueller@biozentrum.uni-wuerzburg.de; Matthias Wolf* - matthias.wolf@biozentrum.uni-wuerzburg.de

* Corresponding author †Equal contributors

Published: 25 July 2008

BMC Evolutionary Biology 2008, **8**:218 doi:10.1186/1471-2148-8-218

Received: 12 March 2008
Accepted: 25 July 2008

This article is available from: http://www.biomedcentral.com/1471-2148/8/218

Abstract

Background: Within Chlorophyceae the ITS2 secondary structure shows an unbranched helix I, except for the '*Hydrodictyon*' and the '*Scenedesmus*' clade having a ramified first helix. The latter two are classified within the Sphaeropleales, characterised by directly opposed basal bodies in their flagellar apparatuses (DO-group). Previous studies could not resolve the taxonomic position of the '*Sphaeroplea*' clade within the Chlorophyceae without ambiguity and two pivotal questions remain open: (1) Is the DO-group monophyletic and (2) is a branched helix I an apomorphic feature of the DO-group? In the present study we analysed the secondary structure of three newly obtained ITS2 sequences classified within the '*Sphaeroplea*' clade and resolved sphaeroplealean relationships by applying different phylogenetic approaches based on a combined sequence-structure alignment.

Results: The newly obtained ITS2 sequences of *Ankyra judayi*, *Atractomorpha porcata* and *Sphaeroplea annulina* of the '*Sphaeroplea*' clade do not show any branching in the secondary structure of their helix I. All applied phylogenetic methods highly support the '*Sphaeroplea*' clade as a sister group to the 'core Sphaeropleales'. Thus, the DO-group is monophyletic. Furthermore, based on characteristics in the sequence-structure alignment one is able to distinguish distinct lineages within the green algae.

Conclusion: In green algae, a branched helix I in the secondary structure of the ITS2 evolves past the '*Sphaeroplea*' clade. A branched helix I is an apomorph characteristic within the monophyletic DO-group. Our results corroborate the fundamental relevance of including the secondary structure in sequence analysis and phylogenetics.

Background

Taxonomists face inconsistent or even contradictory clues when they examine the affiliation of organisms to higher taxonomic groupings. Several characters may yield alternative hypotheses explaining their evolutionary background. This also applies to the taxonomic position of the Sphaeropleaceae [1-23]. Different authors affiliate the green algal family by morphological characters to either ulvophytes or chlorophytes, until amendatory Deason et al. [10] suggested that the Neochloridaceae, the Hydrodic-

tyaceae and the Sphaeropleaceae should be grouped as Sphaeropleales within the chlorophytes, since all of them have motile biflagellate zoospores with a direct-opposite (DO) confirmation of basal bodies.

Subsequently, other taxonomic lineages (the '*Ankistrodesmus*' clade, the '*Bracteacoccus*' clade, the '*Pseudomuriella*' clade, *Pseudoschroederia*, the '*Scenedesmus*' clade, *Schroederia* and the '*Zofingiensis*' clade) were added to this biflagellate DO group, because they show molecular affiliation to either Neochloridaceae or Hydrodictyaceae [24].

Although nowadays most authors agree that the DO group is monophyletic, until now no study pinpointed the taxonomic linkage of the name-giving '*Sphaeroplea*' clade to the remaining 'core Sphaeropleales' persuasively with genetic evidence [6,23], i.e. the sister clade remains unclear [15,24]. Likewise, with respect to morphology, studies of 18S and 26S rRNA gene sequences neither resolve the basal branching patterns within the Chlorophyceae with high statistical power nor corroborate a monophyletic biflagellate DO group without ambiguity [6,23].

Müller et al. [25] obtained moderate statistical support for the close relationship of the '*Sphaeroplea*' clade and the 'core Sphaeropleales' with profile distances of 18S and 26S rDNA. In this study we followed and expanded their methodology with a very different phylogenetic marker. The internal transcribed spacer 2 (ITS2), the region of ribosomal RNA between the 5.8S rRNA gene and the large subunit (26S rDNA) has proven to be an appropriate marker for the study of small scale phylogenies of close relatives [26-29]. The sequence is in contrast to the bordering regions of ribosomal subunits evolutionary not conserved, thus genetic differentiation is detectable even in closely related groups of organisms. By contrast, the secondary structure seems to be well conserved and thus provides clues for higher taxonomic studies [27,30-33]. Secondary structure information is furthermore especially interesting within the Chlorophyceae, because van Hannen et al. [34] described an uncommon branching of ITS2 helix 1 within the genera *Desmodesmus*, *Hydrodictyon* [35] and *Scenedesmus*. It is not known when this feature evolved and whether it is, as we expect, an apomorphic feature for the DO-group. It is obvious that phylogenetic statements should be improvable by inclusion of structural information in common sequence analysis. For example, Grajales et al. [36] calculated morphometric matrices from ITS2 secondary structures for phylogenetic analyses, but treated information of sequence and structure as different markers. Here we combine sequence with structural information in just one analysis. Aside from the biological problem, we address the pivotal question of a methodological pipeline for sequence-structure phylogenetics using rDNA data.

Methods

DNA extraction, amplification and sequencing

Extraction of genomic DNA from cultured cells of *Ankyra judayi*, *Atractomorpha porcata* and *Sphaeroplea annulina* was done using Dynabeads® (DNA DIRECT Universal, Dynal Biotech, Oslo, Norway) according to the manufacturer's protocol. PCR reactions were performed in a 50 μl reaction volume containing 25 μl FastStart PCR Master (Roche Applied Science), 5 μl gDNA and 300 nM of the primers ITS3 (5'-GCA TCG ATG AAG AAC GCA GC-3') and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') designed by White et al. [37].

Cycling conditions for amplification consisted of 94°C for 10 min, 30 cycles of 94°C for 30 s, 50°C for 30 s and 72°C for 45 s, followed by a final extension step of 10 min at 72°C. PCR products were analysed by 3% agarose gel electrophoresis and ethidium bromide staining.

PCR probes where purified with the PCR Purificaton Kit (Qiagen) and where quantified by spectrometry. Each sequencing probe was prepared in an 8 μl volume containing 20 ng DNA and 1.25 μM Primer. Sequencing was carried out using an annealing temperature of 50°C with the sequencer Applied Biosystems QST 3130 Genetic Analyzer by the Institute of Hygiene and Microbiology (Würzburg, Germany).

ITS2 secondary structure prediction

ITS2 secondary structures of the three newly obtained sequences were folded with the help of RNAstructure [38] and afterwards manually corrected. All available 788 chlorophycean ITS2 sequences were obtained from the NCBI nucleotide database. The ITS2 secondary structure of *Atractomorpha porcata* was used as template for homology modelling. Homology modelling was performed by using the custom modelling option as provided with the ITS2-Database [30-33] (identity matrix and 50% threshold for the helix transfer). Forty-nine species representing the chlorophycean diversity were retained and used as comparative taxa in inferring phylogenies (Table 1). For this taxon sampling, accurate secondary structures of sequences were now folded by RNAstructure and additionally corrected using Pseudoviewer 3 [39]. We standardized start and end of all helices according to the optimal folding of the newly obtained sequences.

Alignment and phylogenetic analyses

Using 4SALE [40,41] with its ITS2 specific scoring matrix, we automatically aligned sequences and structures simultaneously. Sequence-structure alignment is available at the ITS2 database supplements page. For the complete

Table 1: Chlorophyte species used for this investigation.

Clade	Species	Strain	GenBank
'*Sphaeroplea* '	*Ankyra judayi* (G.M. Smith) Fott 1957	SAG 17.84	EU352800
	Atractomorpha porcata Hoffman 1984 strain	SAG 71.90	EU352803
	Sphaeroplea annulina (Roth) C. Agardh 1824	SAG 377.1a	EU352801
	Sphaeroplea annulina (Roth) C. Agardh 1824	SAG 377.1e	EU352802
'*Dunaliella*'	*Haematococcus droebakensis* Wollenweber 1908	-	U66981
	Dunaliella parva Lerche 1937	-	DQ116746
	Dunaliella salina (Dunal) Teodoresco 1905	CCAP 19/18	EF473746
'*Hydrodictyon* '	*Hydrodictyon africanum* Yamanouchi 1913	UTEX 782	AY779861
	Hydrodictyon patenaeforme Pocock	CCAP 236/3	AY577736
	Hydrodictyon reticulatum (Linnaeus) B. de St.-Vincent 1824	CBS	AY779862
	Pediastrum braunii Wartmann 1862	SAG 43.85	AY577756
	Pediastrum duplex Meyen 1829	UTEX 1364	AY779868
	Pseudopediastrum boryanum (Raciborski) Sulek 1969	UTEX 470	AY779866
	Sorastrum spinulosum Nägeli 1849	UTEX 2452	AY779872
	Stauridium tetras (Ehrenberg) Ralfs 1844	EL 0207 CT	AY577762
'*Oedogonium*'	*Bulbochaete hiloensis* (Nordstedt) Tiffany 1937	-	AY962677
	Oedogonium cardiacum (Hassall) Wittrock 1870	-	AY962675
	Oedogonium nodulosum Wittrock 1872	-	DQ078301
	Oedogonium oblongum Wittrock 1872	-	AY962681
	Oedogonium undulatum (Brébisson) A. Braun 1854	-	DQ178025
'*Reinhardtii*'	*Chlamydomonas incerta* Pascher 1927	SAG 81.72	AJ749625
	Chlamydomonas komma Skuja 1934	-	U66951
	Chlamydomonas petasus Ettl	SAG 11.45	AJ749615
	Chlamydomonas reinhardtii Dangeard 1888	CC-620	AJ749638
	Chlamydomonas typica Deason & Bold 1960	SAG 61.72	AJ749622
	Eudorina elegans Ehrenberg 1831	ASW 107	AF486524
	Eudorina unicocca G.M. Smith 1930	UTEX 1215	AF486525
	Gonium octonarium Pocock 1955	Tex	AF054424
	Gonium pectorale O.F. Müller 1773	Chile K	AF054440
	Gonium quadratum E. G. Pringsheim ex H. Nozaki	Cal 3-3	AF182430
	Pandorina morum (O.F. Müller) Bory de Saint-Vincent 1824	Chile	AF376737
	Volvox dissipatrix (Shaw) Printz	-	U67020
	Volvox rousseletii G.S.West	-	U67025
	Volvulina steinii Playfair 1915	-	U67034
	Yamagishiella unicocca (Rayburn & Starr) Nozaki 1992	ASW 05129	AF098181
'*Scenedesmus*'	*Desmodesmus abundans* (Kirchner) Hegewald 2000	UTEX 1358	AJ400494
	Desmodesmus bicellularis (Chodat) An, Friedl & Heg. 1999	CCAP 276/14	AJ400498
	Desmodesmus communis (Hegewald) Hegewald 2000	UTEX 76	AM410660
	Desmodesmus elegans (Hortobágyi) Heg. & Van. 2007	Heg 1976–28	AM228908
	Desmodesmus opoliensis (P.G. Richter) Hegewald 2000	EH 10	AM410655
	Desmodesmus pleiomorphus (Hindák) Hegewald 2000	UTEX 1591	AM410659
	Desmodesmus quadricauda (Turpin) Hegewald	-	AJ400495
	Scenedesmus acuminatus (Lagerheim) Chodat 1902	UTEX 415	AJ249511
	Scenedesmus acutiformis (B. Schröder) F. Hindák 1990	SAG 276.12	AJ237953
	Scenedesmus basiliensis Chodat 1926	UTEX 79	AJ400489
	Scenedesmus dimorphus (Turpin) Kützing 1833	UTEX 417	AJ400488
	Scenedesmus longus Meyen 1829 ex Ralfs	NIOO-MV5	AJ400506
	Scenedesmus obliquus (Turpin) Kützing 1833	Tow 9/21P-1W	DQ417568
	Scenedesmus pectinatus Meyen 1828	An 111a	AJ237954
	Scenedesmus platydiscus (G.M. Smith) Chodat 1926	UTEX 2457	AJ400491
	Scenedesmus raciborskii Woloszynska 1914	An 1996–5	AJ237952
	Scenedesmus regularis Svirenko	Heg 1998–2	AY170857
	Scenedesmus wisconsinensis (G.M. Smith) Chodat 1996	An 41	AJ237950

Listed is the current clade classification of the species [69,70,24] and the GenBank accession numbers of the analyzed sequences. The four newly obtained sequences are of the '*Sphaeroplea*' clade.

alignment we tested for appropriate models of nucleotide substitution using the Akaike Information Criterion (AIC) as implemented in Modeltest [42]. The following PAUP-block was used for all maximum likelihood based phylogenetic analyses with PAUP* [43]: Lset Base = (0.2299 0.2415 0.2152) Nst = 6 Rmat = (1.4547 3.9906 2.0143 0.1995 3.9906) Rates = gamma Shape = 1.1102 Pinvar = 0.0931;. A maximum likelihood (ML) analysis was performed with a heuristic search (ten random taxon addition replicates) and nearest neighbour interchange (NNI) [44].

Maximum parsimony (MP) [45] was accomplished with gaps treated as missing data and all characters coded as "unordered" and equally weighted. Additionally, we clustered taxonomic units with neighbour-joining (NJ) [46] using maximum likelihood distances. Furthermore, with MrBayes [47] a Bayesian analysis (B) was carried out for tree reconstruction using a general time reversible substitution model (GTR) [48-50] with substitution rates estimated by MrBayes (nst = 6). Moreover, using ProfDist, a profile neighbour-joining (PNJ) tree [51,25] was calculated using the ITS2 specific substitution model available from the ITS2 Database. PNJ was also performed with predefined profiles (prePNJ) of all the clades given in Table 1.

For clade '*Scenedesmus*' two profiles were used for groups 'true *Scenedesmus*' (*Scenedesmus* except *S. longus*) and '*Desmodesmus*' (*Desmodesmus* and *S. longus*). We performed a sequence-structure profile neighbour-joining (strPNJ) analysis with a developmental beta version of ProfDist (available upon request). The tree reconstructing algorithm works on a 12 letter alphabet comprised of the 4 nucleotides in three structural states (unpaired, paired left, paired right). Based on a suitable substitution model [40], evolutionary distances between sequence structure pairs have been estimated by maximum likelihood. All other applied analyses were computed only on the sequence part of the sequence-structure alignment. For MP, NJ, PNJ, prePNJ and strPNJ analyses 1.000 bootstrap pseudoreplicates [52] were generated. One hundred bootstrap replicates were generated for the ML analysis. Additionally we used RAxML at the CIPRES portal to achieve 1.000 bootstraps with a substitution model estimated by RAxML [53]. All methods were additionally applied to a 50% structural consensus alignment cropped with 4SALE (data not shown). The individual steps of the analysis are displayed in a flow chart (Fig. 1).

Results

New ITS2 sequences

GenBank accession numbers for newly obtained nucleotide sequences are given in Table 1 (entries 1–4). The two ITS2 sequences of *Sphaeroplea annulina* (Roth, Agardh) strain SAG 377-1a and strain SAG 377-1e were identical and thus only the first one was used for further analysis. According to folding with RNAstructure, ITS2 secondary structures of the three newly obtained sequences did not exhibit any branching in their helix I (Fig. 2) as it is described for the 'core Sphaeropleales', i.e. helix I was more similar to those of the CW-group and the '*Oedogonium*' clade. Helix I of *Sphaeroplea* annulina was explicitly longer (9 nucleotides) than those of the other newly obtained algae. Due to this insertion, for *Sphaeroplea*, a branching pattern was enforceable, but would have lower energy efficiency. However, the additional nucleotides are not homologous to the insertion capable of making an additional stem (Y-structure) found in the '*Scenedesmus*' and the '*Hydrodictyon*' clade (approximately 25 bases).

ITS2 sequence and secondary structure information

ITS2 sequence lengths of all studied species ran from 202 to 262 nucleotides (nt), 235 nt on average. The GC contents of ITS2 sequences ranged from 36.84% to 59.92%, with a mean value of 52.42%. The number of base pairs (bp) varied between 64 and 89 bp and averaged 77 bp. The cropped alignment (50% structural consensus) showed that 23% of the nucleotides had at least a 50% consistency in their pairings. Compensatory base changes (CBCs) as well as hemi-CBCs (all against all) range from 0 to 16 with a mean of 6.6 CBCs (Fig. 2). Sequence pairs lacking CBCs were exclusively found within the same major clade.

Characteristics in a conserved part of alignment

In agreement with Coleman [28], the 5' side part near the tip of helix III was highly conserved including the UGGU motif [54,55,30], likewise the UGGGU motif in case of Chlorophyceae. We selected a part of the alignment at this position with adjacent columns (Fig. 2) to verify the suggested conservation. Having a closer look at this part of helix III, in our case, it showed typical sequence and structural characteristics for distinct groups. Studied species of the '*Oedogonium*' clade possess at position 3 in the selected part of the alignment an adenine and in addition at positions 3–5 paired bases. In contrast, the CW-group solely possessed three consecutively paired bases in this block, but not the adenine. A typical pattern for clades of the DO-group was a twofold motif of 3 bases: uracile, adenine and guanine at positions 7–9, which is repeated at positions 11–13. This could be a duplication, which results in a modified secondary structure. In addition, the '*core Sphaeropleales*' ('*Hydrodictyon*' clade and '*Scenedesmus*' clade) showed an adenine base change at position 6, compared to all other clades.

Phylogenetic tree information

The PAUP* calculation applying maximum Parsimony included a total of 479 characters, whereas 181 characters were constant, 214 variable characters were parsimony-

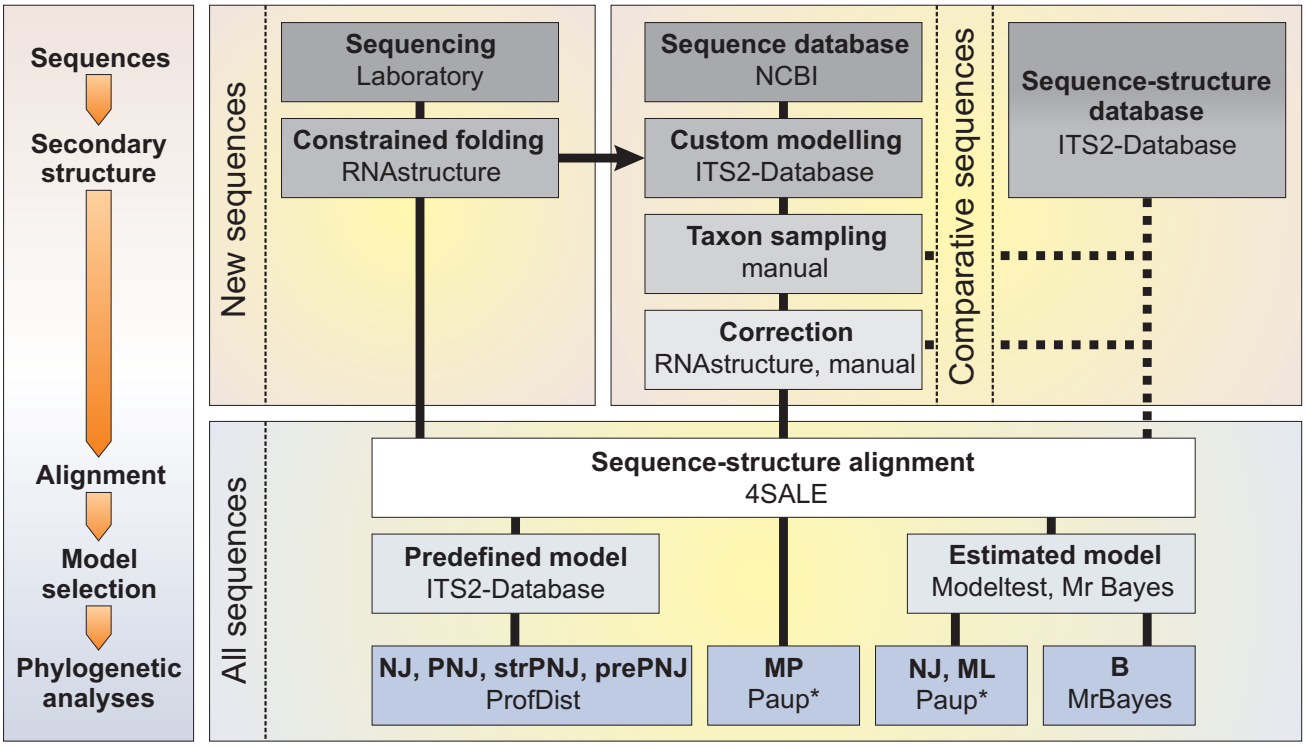


Figure 1

Flowchart of the methods applied in this study. Sequences were obtained from the laboratory and from NCBI and afterwards folded with RNAstructure [38] or custom modelling of the ITS2 Database [30-33]. An alternative way may pose to directly access sequences and structures deposed at the ITS2 Database. The sequence-structure alignment was derived by 4SALE [40]. Afterwards several phylogenetic approaches were used to calculate trees: NJ = neighbour-joining, PNJ = profile neighbour-joining, strPNJ = sequence-structure neighbour-joining, prePNJ = predefined profiles profile neighbour-joining, MP = maximum Parsimony, ML = maximum likelihood and B = Bayesian analysis.

informative compared to 84 parsimony-uninformative ones.

The resulting trees (Fig. 3 and 4, Table 2) of all performed analyses (NJ [PAUP* and ProfDist], PNJ, prePNJ, strPNJ, ML [PAUP* and RAxML], MP, B) yielded six major clades: the '*Dunaliella*', the '*Hydrodictyon*', the '*Oedogonium*', the '*Reinhardtii*', the '*Scenedesmus*', and the '*Sphaeroplea*' clade. All of them were separated and – except for the '*Scenedesmus*' clade – highly supported by bootstrap values of 83–100%, respectively by Bayesian posterior probabilities of 0.86–1.0.

The 'Hydrodictyon' clade, the 'Scenedesmus' clade and the 'Sphaeroplea' clade form one cluster that was strongly supported by high bootstrap values of 67–96% (node "g"). The three clades composed the DO-group. The opposite cluster included the 'Dunaliella' and the 'Reinhardtii' clade, forming the CW-group. The 'Oedogonium' clade was chosen as the outgroup [56]. Both clusters (CW-group and 'Oedogonium' clade) were strongly supported by bootstrap values of 84–100% (nodes "i" and "h").

Except for the Bayesian analysis (least support for node "c"), all applied methods yielded node "e" as the weakest point within the basal (labelled) branches (Table 2), which presents the relationship between the '*Hydrodictyon*' and the '*Scenedesmus*' clade on the one hand and the '*Dunaliella*', the '*Oedogonium*', the '*Reinhardtii*' and the '*Sphaeroplea*' clade on the other hand. The phylogenetic tree resulting from neighbour-joining analysis by PAUP* (Fig. 3) did not support node "e" at all, but strongly supported the remaining labelled branches. The maximum likelihood analysis by PAUP* (Fig. 4) did not encourage node "e" either. Both maximum likelihood methods did not even support nodes "a" ('true *Scenedesmus*' compared to remaining clades) and "c" ('*Scenedesmus*' opposite to remaining clades). All other basal branches were supported by this method.

Varying neighbour-joining analyses by ProfDist (NJ, PNJ, prePNJ, strPNJ) supported all basal branches – except for the weakest node "e" (average support) – with very high bootstrap support values of 84–100%. The maximum Parsimony method gave average support (63 and 62%) for

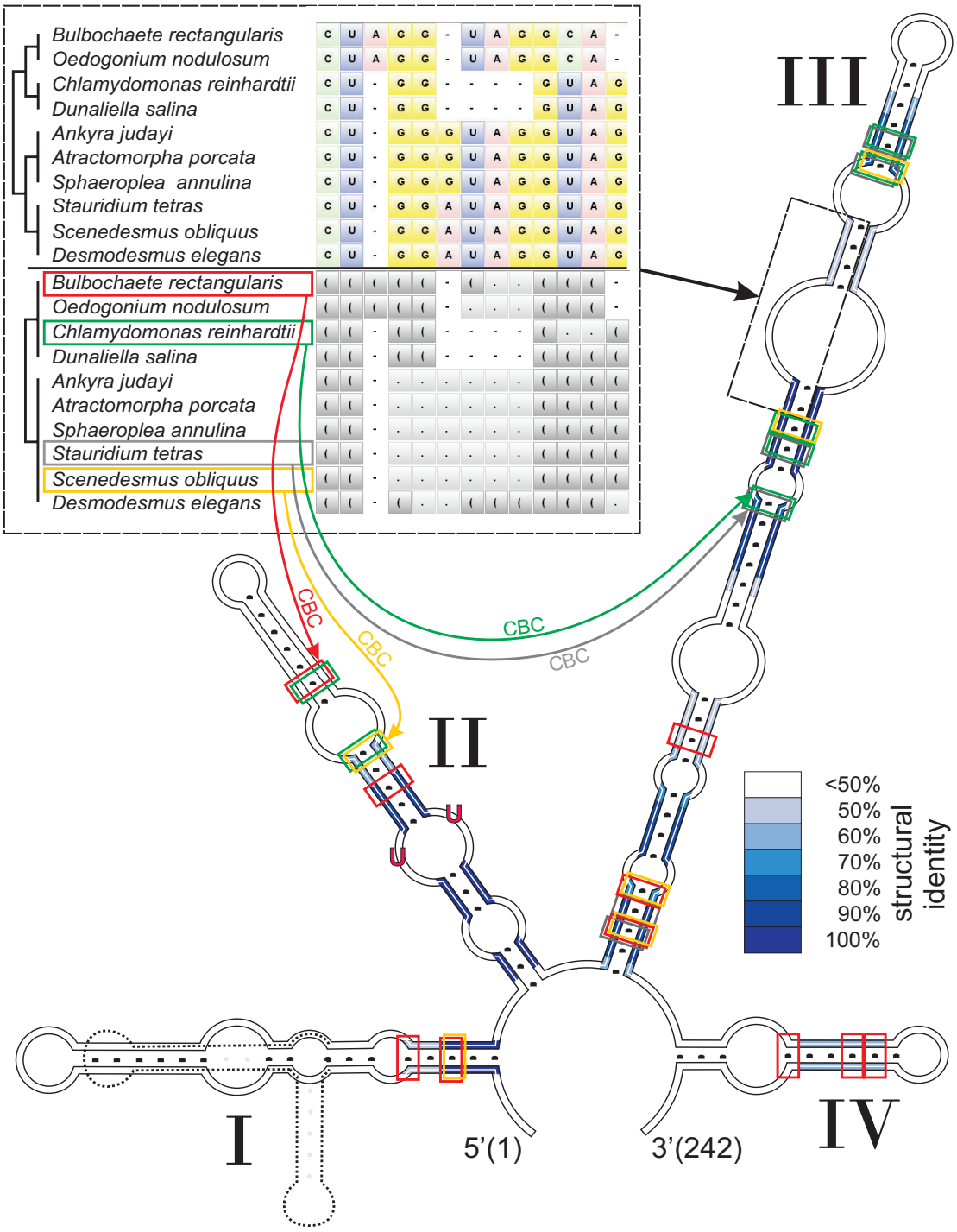


Figure 2
ITS2 structure of *Sphaeroplea annulina*, degrees of conservation and structure alignment. The structure of the internal transcribed spacer 2 of *Sphaeroplea annulina* shows the common four helices. Helix I is unbranched. Helix I of *Scenedesmus obliquus* with its branch is underlain in grey. The degree of conservation over the whole alignment is indicated in blue with different degrees of colour saturation. The structural consensus function of 4SALE [40] returns nucleotides on given percentages. In the upper left corner is the sequence-structure alignment of the conserved distal part of helix III showing a differentiation of the major clades with sequence and/or structure.

Table 2: Bootstrap support values for basal branches of all methods applied.

Software		**ProfDist**				**PAUP***			**MrBayes**	**RAxML**
Model		ITS2				Modeltest		-	Estimated	
Analysis		NJ	PNJ	prePNJ	strPNJ	NJ	ML	MP	B	ML
Nodes	a	99	95	100[1]	100	91	-	82	*0.86*	-
	b	96	96	100[1]	96	99	93	86	*1.00*	98
	c	88	88	95	88	90	-	63	*0.72*	-
	d	100	99	100[1]	100	100	92	100	*1.00*	96
	e	62	55	53	60	-	-	62	*0.97*	64
	f	100	100	100[1]	100	100	99	100	*1.00*	100
	g	**87**	**91**	**88**	**96**	**86**	**67**	**80**	***0.98***	**93**
	h	99	99	100[1]	99	100	100	100	*1.00*	100
	i	90	90	92	84	93	88	85	*0.99*	89
	j	97	98	100[1]	98	93	91	91	*0.99*	98
	k	97	96	100[1]	95	96	88	83	*1.00*	99
Figure		3							4	

The table supplements Fig. 3 and Fig. 4. Node "g" supports a monophyletic DO group and is printed in bold letters. Software used: ProfDist and PAUP*. Models of substitution: ITS2 = GTR with ITS2 substitution matrix, Modeltest: TVM+I+G with estimated parameters. Phylogenetic analysis: NJ = neighbour-joining, PNJ = profile neighbour-joining, prePNJ = profile neighbour-joining with predefined profiles, strPNJ = sequence-structure profile neighbour-joining, ML = maximum likelihood, B = Bayesian analysis (posterior probabilities), MP = maximum Parsimony. [1]Predefined profiles for profile neighbour-joining.

node "c" and "e" and high bootstrap values (80–100%) for the remaining basal clades. The Bayesian analysis offered posterior probabilities of 0.72 for node "c" and 0.86–1.0 for the remaining basal nodes. For further sister group relations see Fig. 3 and 4.

In comparison, the topology of the phylogenetic tree based on the 50% cropped alignment did not change, but the bootstrap support values were lower in all cases (data not shown).

Discussion

The internal transcribed spacer 2 (ITS2) is required in ribosome biogenesis [57-59] and its gradual removal from mature rRNA is driven by its specific secondary structure [60,59].

Using three newly obtained ITS2 sequences from *Ankyra judayi*, *Atractomorpha porcata* and *Sphaeroplea annulina* (Sphaeropleaceae) in this study we aimed to pursue two consecutive questions concerning the phylogenetic relationships within Chlorophyceae. (1) What is the phylogenetic position of the newly sequenced algae relative to the 'core Sphaeropleales' and could the biflagellate DO-group be regarded as monophyletic? (2) How does the secondary structure of the new ITS2 sequences look like and is an autapomorphic feature of the secondary structure associated with the monophyletic DO-group?

Considering the question (1) Buchheim et al. [6] and Wolf et al. [23] approached the problem with 18S + 26S rDNA and 18S rDNA data, but the relationship between the 'core Sphaeropleales' and the Sphaeropleaceae remained unclear. However, in their studies, *Ankyra*, *Atractomorpha* and *Sphaeroplea* clustered in a monophyletic clade named Sphaeropleaceae. We confirm this '*Sphaeroplea*' clade with all three genera being strongly separated from other clades. As a result of a Bayesian analysis on a combined 18S and 26S rDNA dataset Shoup and Lewis [61] also found the Sphaeropleaceae as the most basal clade within the Sphaeropleales, but again the analysis lacked a strong backing. Beside these difficulties the 'core Sphaeropleales' were already shown to be monophyletic with high certainty [6,25,62,61,23].

The DO-group (Sphaeropleales including the '*Sphaeroplea*' clade) as emended by Deason et al. [10], for which the directly opposed basal body orientation and basal body connection features are verified [63-65], is now strongly supported by molecular phylogenetic analyses. There was already evidence of an extended DO-group [6,66,67], however, for some groups ultrastructural results are still lacking, and even though the collective basal body orientation and connection imply a monophyletic DO-group, until now no molecular phylogenetic analysis could show this with solid support [6,62,24,23]. We demonstrate for the first time with robust support values for the equivocal nodes that the 'core Sphaeropleales', the '*Sphaeroplea*' clade, and the Sphaeropleales are monophyletic.

Regarding question (2), for all structures of the '*Hydrodictyon*' and the '*Scenedesmus*' clade, helix I shows the typical

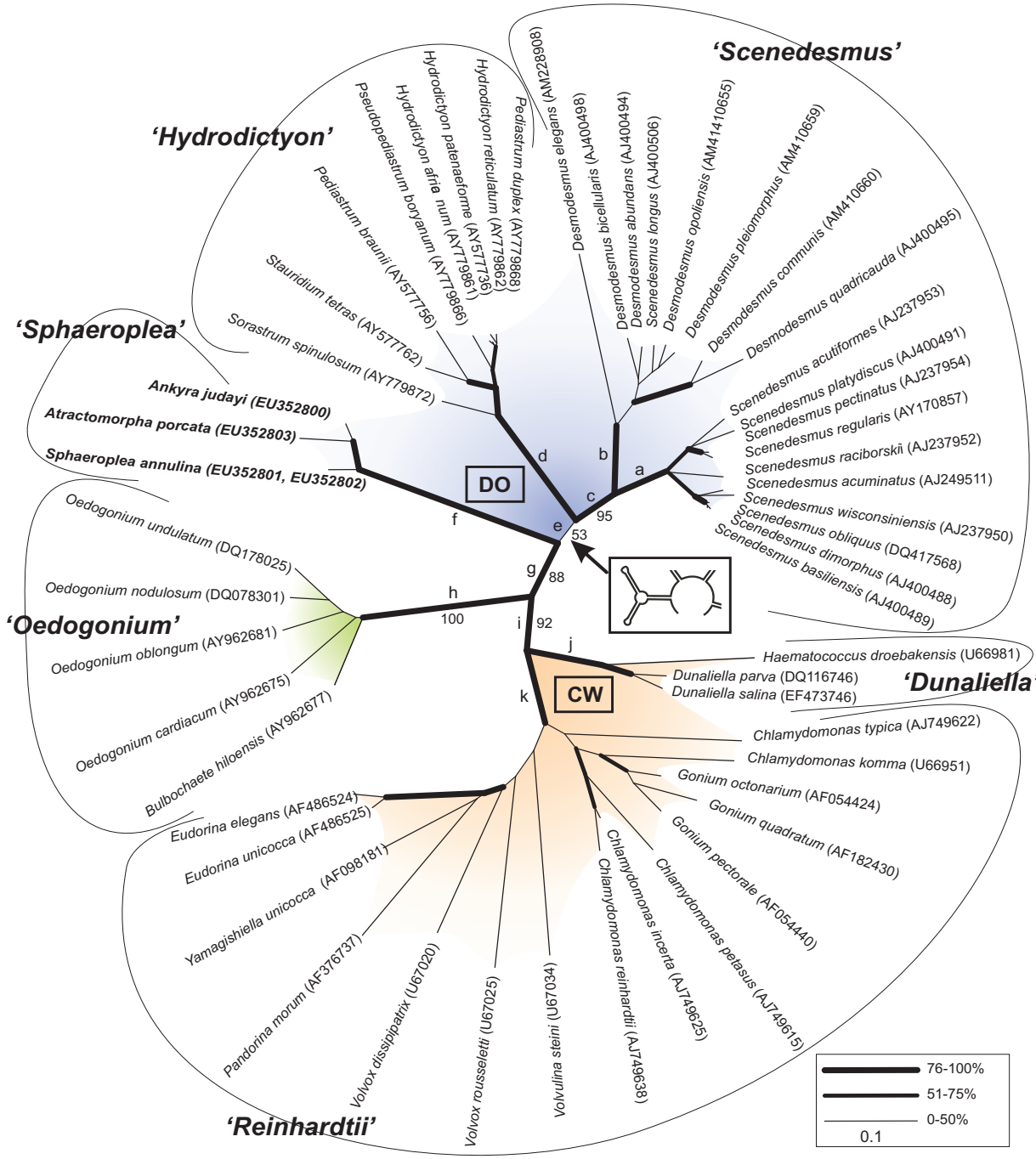


Figure 3

Neighbour-joining phylogeny of the Chlorophyceae based on comparison of ITS2 rRNA sequences and structures. The tree is unrooted, but the '*Oedogonium*' clade is most likely appropriate as outgroup [56]. Sequences of the '*Sphaeroplea*' clade were sequenced for this study and shown in bold letters. The phylogenetic tree is calculated by neighbour-joining with PAUP* [46,43] for an alignment with 52 taxa and 479 characters. The substitution model was set to TVM+I+G with parameters estimated by Modeltest [42]. Bootstrap values of basal branches are given for profile neighbour-joining with predefined profiles (ProfDist with ITS2 substitution model) [51,31]. Branch thickness is dependant of Bootstrap values calculated with four distance methods: neighbour-joining (PAUP*), neighbour-joining, complete profile neighbour-joining and sequence-structure profile neighbour-joining (all three ProfDist with ITS2 substitution model).

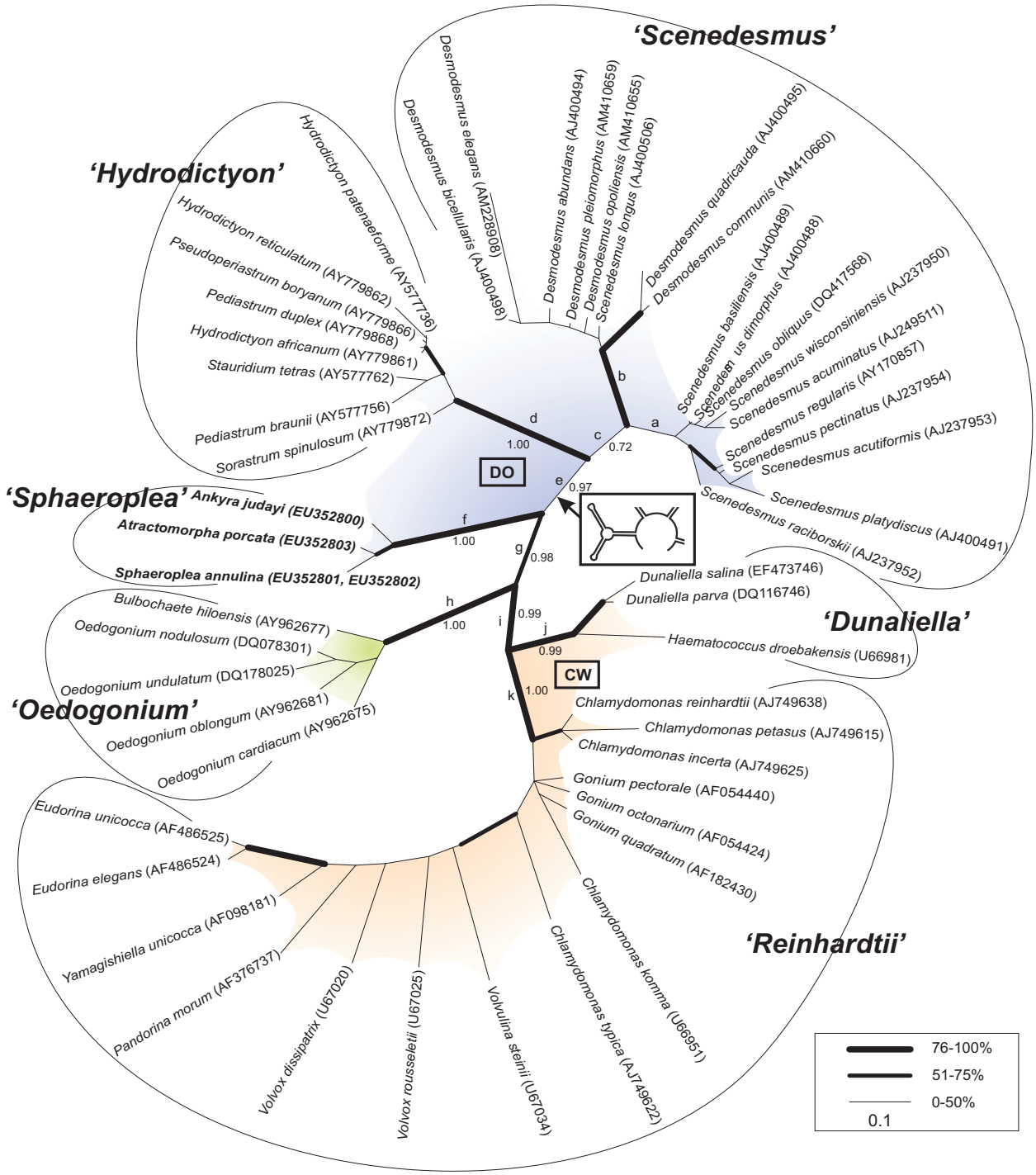


Figure 4

Phylogeny of chlorophyte ITS2 sequences and structures based on distances of a Bayesian analysis. The alignment contained 52 taxa and 479 characters. The suggested outgroup is the '*Oedogonium*' clade [56]. Sequenced species are shown in bold ('*Sphaeroplea*' clade). Substitution models and tree distances were calculated with MrBayes [47]. Posterior probabilities are shown for basal branches. Branch thickness is dependant of Bootstrap values calculated with maximum likelihood (PAUP* with TVM+I+G, RAxML) [42,53,43] and maximum Parsimony (PAUP*) (see legend). Resulting parameter of performing MP are L = 1231, CI = 0.4427, HI = 0.5573, RI = 0.7264, RC = 0.3216.

branching (Y-structure). Initially, An et al. [68] proposed a secondary structure model with an unbranched helix I for ITS2 sequences of '*Scenedesmus*' clade members. Thereafter, van Hannen et al. [34] updated the model by folding the nucleotide sequences based upon minimum free energy and found a branched helix I as the most energetically stable option. The branching is result of an insertion of approximately 25 nucleotides capable of folding as an individual stem within the 5' end of the first helix. However, ITS2 sequence and secondary structure information of further '*core Sphaeropleales*' members, e.g. the '*Ankistrodesmus*' clade and the '*Bracteacoccus*' clade, lacks hitherto. In contrast, the Y-structure is absent within the '*Sphaeroplea*' clade and any other investigated group so far. Thus this feature is – contrary to our expectation – not an autapomorphic character for the biflagellate DO-group as a whole but for the 'core Sphaeropleales'.

Regarding future work, the resolution among the main clades of Chlorophyceae was statistically poorly supported in previous studies [68,15,6,23]. Pröschold and Leliaert [24] reviewed the systematics of green algae by applying a polyphasic approach, but did not yield a clear resolution regarding a sister taxon to the Sphaeropleales. Since they are not yet available, ITS2 sequences of chaetopeltidalean and chaetophoralean taxa could not be included in the present study and therefore the phylogenetic relationships between the main Chlorophyceae clades remain open. We recommend involving sequence and secondary structure information of chaetopeltidalean and chaetophoralean ITS2 sequences in future studies to find out if the monophyletic biflagellate DO-group could be further extended to a general monophyletic DO-group containing quadri- and biflagellate taxa. A genome-wide approach indicates that Sphaeropleales and Chlamydomonadales are sister taxa, however only a few organisms are included in this study [56]. An additional uprising question is when the Y has evolved within the 'core Sphaeropleales'. This could be resolved by inclusion of other members (e.g. *Bracteacoccus*) in further studies.

The two major reasons contributing to the robust results presented here are the change of the phylogenetic marker and the inclusion of secondary structure information. In contrast to previous phylogenetic work concerning Chlorophyceae, this study is based on the ITS2, which offers a resolution power for relationships from the level of subspecies up to the order level, because of their variable sequence but conserved secondary structure [26,30-33]. Hitherto commonly used markers in contrast are a lot more restricted. Using 4SALE [40] with implemented structure consideration, we could achieve for the first time a global simultaneously generated sequence-structure alignment (c.f. Fig. 1) yielding specific sequence and structural features distinguishing different algae lineages (c.f. Fig. 2).

Conclusion

In summary, the powerful combination of the ITS2 rRNA gene marker plus a multiple global alignment based synchronously on sequence and secondary structure yielded high bootstrap support values for almost all nodes of the computed phylogenetic trees. Thus, the relationship of Sphaeropleaceae is here resolved, being a part of the Sphaeropleales representing the monophyletic biflagellate DO-group. Furthermore, we could elucidate a branched helix I of ITS2 as an autapomorphic feature within the DO-group. This feature could be found only in the '*Hydrodictyon*' and the '*Scenedesmus*' clade. Our results corroborate the presented methodological pipeline, the fundamental relevance of secondary structure consideration, as well as the elevated power and suitability of ITS2 in phylogenetics. For a methodological improvement it is suitable to ameliorate the alignment algorithm in further considering horizontal dependencies of paired nucleotides, and moreover in future ITS2 studies it is suggested to include sequence and secondary structure information of hitherto not regarded taxa to resolve the chlorophycean phylogeny.

Authors' contributions

MW designed the study. FF determined the new sequences in the laboratory. BR implemented the strPNJ within ProfDist. TS and AK performed sequence analyses, structure prediction and phylogenetic analyses. TM developed the ITS2 sequence-structure substitution model and the ITS2 sequence-structure scoring matrix. TS, AK and MW drafted the manuscript. All authors contributed to writing the paper, read the final manuscript and approved it.

Acknowledgements

Financial support for AK and TS was provided by the Deutsche Forschungsgemeinschaft (DFG) grant (Mu-2831/1-1). AK was additionally supported by BIGSS (Elite graduate school). FF was supported by the Bundesministerium für Bildung und Forschung (BMBF) grant FUNCRYPTA. The newly obtained sequences originate from SAG cultures (Göttingen, Germany). We thank the Institute of Hygiene and Microbiology (Würzburg, Germany) for sequencing. We thank T. Ulmar Grafe (University of Brunei Darussalam) for proof-reading English.

References

1. Agardh CA: *Systema algarum XXV* Lund, Sweden: Soc Physiogr; 1824.
2. Bold HC, Wynne MJ: *Introduction to the algae: structure and reproduction* Englewood Cliffs, NJ: Prentice Hall; 1985.
3. Booton GC, Floyd GL, Fuerst PA: **Polyphyly of tetrasporalean green algae inferred from nuclear small-subunit ribosomal DNA.** *J Phycol* 1998, **34:**306-311.
4. Booton GC, Floyd GL, Fuerst PA: **Origins and affinities of the filamentous green algal orders Chaetophorales and Oedogoniales based on 18S rRNA gene sequences.** *J Phycol* 1998, **34:**312-318.
5. Buchheim MA, Hoffman LR: **Ultrastructure of male gametes of *Sphaeroplea robusta* (Chlorophyceae).** *J Phycol* 1986, **22:**176-185.

6. Buchheim MA, Michalopulos EA, Buchheim JA: **Phylogeny of the Chlorophyceae with special reference to the Sphaeropleales: a study of 18S and 26S rDNA data.** *J Phycol* 2001, **37:**819-835.
7. Cáceres EJ, Robinson DG: **Ultrastructural studies on *Sphaeroplea annulina* (Chlorophyceae). Vegetative structure and mitosis.** *J Phycol* 1980, **16:**313-320.
8. Cáceres EJ, Robinson DG: **Ultrastructural studies on *Sphaeroplea annulina* (Chlorophyceae). II. Spermatogenesis and male gamete structure.** *J Phycol* 1981, **17:**173-180.
9. Chapman R, Buchheim MA, Delwiche CF, Friedl T, Huss VA, Karol KG, Lewis LA, Manhart J, McCourt RM, Olsen JL, Waters DA: **Molecular systematics of the green algae.** In *Molecular systematics of plants II: DNA sequencing* Edited by: Soltis DE, Soltis PS, Doyle J. Boston, MA: Kluwer Acad. Publ; 1998:508-540.
10. Deason TR, Silva P, Watanabe S, Floyd GL: **Taxonomic status of the species of the green algal genus *Neochloris* .** *Plant Sys Evol* 1991, **177:**213-219.
11. Engler A, Prantl K: *Die natürlichen Pflanzenfamilien I* Leipzig, Germany: Verlag von Wilhelm Engelmann; 1897.
12. Hoffman LR: ***Atractomorpha echinata* gen. et ap.-nov., a new anisogamous member of the Sphaeropleaceae (Chlorophyceae).** *J Phycol* 1983, **19:**76-86.
13. Hoffman LR: ***Atractomorpha porcata* ap. nov., a new member of the Sphaeropleaceae (Chlorophyceae) from California.** *J Phycol* 1984, **20:**225-236.
14. Kützing FT: *Species Algarum* Leipzig, Germany: Brockhaus Verlag; 1849.
15. Lewis LA, McCourt RM: **Green algae and the origin of land plants.** *Am J Bot* 2004, **91:**1535-1556.
16. Mattox KR, Stewart KD: **Classification of the green algae: a concept based on comparative cytology.** In *Systematics of the green algae* Edited by: Irvine DEG, John DM. London, UK: Academic Press; 1984:29-42.
17. Pascher A: **Systematische Übersicht über die mit Flagellaten in Zusammenhang stehenden Algenreihen und Versuch einer Einreihung dieser Algenstämme in die Stämme des Pflanzenreiches.** *Beihefte Bot Centralbl* 1931, **48:**317-332.
18. Pascher A: **Über geißlbewegliche Eier, mehrköpfige Schwärmer und vollständigen Schwärmverlust bei *Sphaeroplea*.** *Beihefte Bot Centralbl* 1939, **59:**188-213.
19. Rieth A: **Über die vegetative Vermehrung bei *Sphaeroplea wilmani* Fritsch et Rich.** *Flora* 1952, **139:**28-38.
20. Rieth A: **Zur Kenntnis der Gattung *Sphaeroplea, Sphaeroplea caubrica* Fritsch.** *Flora* 1953, **140:**130-139.
21. Stewart KD, Mattox KR: **Comparative cytology, evolution and classification of the green algae with some consideration of other organisms with chlorophylls a and b.** *Botanical Rev* 1975, **41:**104-135.
22. West LW, Fritsch FE: *A treatise on the british freshwater algae, Reviewed edition* Cambridge, UK: Cambridge University Press; 1927.
23. Wolf M, Buchheim MA, Hegewald E, Krienitz L, Hepperle D: **Phylogenetic position of the Sphaeropleaceae (Chlorophyta).** *Plant Sys Evol* 2002, **230:**161-171.
24. Pröschold T, Leliaert F: **Systematics of the green algae: conflict of classic and modern approaches.** In *Unravelling the algae: the past, present, and future of algal systematics* Edited by: Brodie J, Lewis J. London, UK: CRC Press; 2007:123-135.
25. Müller T, Rahmann S, Dandekar T, Wolf M: **Accurate and robust phylogeny estimation based on profile distances: a study of the Chlorophyceae (Chlorophyta).** *BMC Evol Biol* 2004, **4:**20.
26. Coleman AW: **ITS2 is a double-edged tool for eukaryote evolutionary comparisons.** *TIG* 2003, **19:**370-375.
27. Coleman AW, Vacquier VD: **Exploring the phylogenetic utility of ITS sequences for animals: a test case for Abalone (*Haliotis*).** *J Mol Evol* 2002, **54:**246-257.
28. Coleman AW: **Pan-eukaryote ITS2 homologies revealed by RNA secondary structure.** *Nucl Acids Res* 2007, **35:**3322-3329.
29. Young I, Coleman AW: **The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example.** *Mol Phylogenet Evol* 2004, **30:**236-242.
30. Schultz J, Maisel S, Gerlach D, Müller T, Wolf M: **A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota.** *RNA* 2005, **11:**361-364.
31. Schultz J, Müller T, Achtziger M, Seibel P, Dandekar T, Wolf M: **The internal transcribed spacer 2 database – a web server for (not only) low level phylogenetic analyses.** *Nucl Acids Res* 2006, **34(Suppl 2):**W704-707.
32. Selig C, Wolf M, Müller T, Dandekar T, Schultz J: **The ITS2 Database II: homology modelling RNA structure for molecular systematics.** *Nucl Acids Res* 2008, **36:**D377-D380.
33. Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T: **Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures.** *RNA* 2005, **11:**1616-1623.
34. van Hannen EJ, Fink P, Lürling M: **A revised secondary structure model for the internal transcribed spacer 2 of the green algae *Scenedesmus* and *Desmodesmus* and its implication for the phylogeny of these algae.** *Eur J Phycol* 2002, **37:**203-208.
35. Buchheim MA, Buchheim JA, Carlson T, Braband A, Hepperle D, Krienitz L, Wolf M, Hegewald E: **Phylogeny of the Hydrodictyaceae (Chlorophyceae): inferences from rDNA data.** *J Phycol* 2005, **41:**1039-1054.
36. Grajales A, Aguilar C, Sanchez J: **Phylogenetic reconstruction using secondary structures of internal transcribed spacer 2 (ITS2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals.** *BMC Evol Biol* 2007, **7:**90.
37. White TJ, Bruns T, Lee S, Taylor J: **Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics.** In *PCR protocols: a guide to methods and applications* Edited by: Innis MA, Gelfand DH, Sninsky JJ, White TJ. San Diego, CA: Academic Press; 1990:315-322.
38. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH: **Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.** *Proc Natl Acad Sci USA* 2004, **101:**7287-7292.
39. Byun Y, Han K: **PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures.** *Nucl Acids Res* 2006, **34:**416-422.
40. Seibel P, Müller T, Dandekar T, Schultz J, Wolf M: **4 SALE – A tool for synchronous RNA sequence and secondary structure alignment and editing.** *BMC Bioinformatics* 2006, **7:**498.
41. Thompson JD, Higgins DG, Gibson TJ: **ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22:**4673-4680.
42. Posada D, Crandall KA: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14(9):**817-818.
43. Swofford DL: *PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4.0b10* Sunderland, MA: Sinauer Associates; 2002.
44. Felsenstein J: **Evolutionary trees from DNA sequences: a maximum likelihood approach.** *J Mol Evol* 1981, **17:**368-376.
45. Camin JH, Sokal RR: **A method for deducing branching sequences in phylogeny.** *Evolution* 1965:311-326.
46. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4(4):**406-425.
47. Huelsenbeck JP, Ronquist F: **MrBayes: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17:**754-755.
48. Lanave C, Preparata G, Sacone C, Serio G: **A new method for calculating evolutionary substitution rates.** *J Mol Evol* 1984, **20:**86-93.
49. Rodriguez F, Oliver JL, Marin A, Medina J: **The general stochastic model of nucleotide substitution.** *J Theor Biol* 1990, **142:**485-501.
50. Tavaré S: **Some probabilistic and statistical problems in the analysis of DNA sequences.** In *Some mathematical questions in biology: DNA sequence analysis* Edited by: Lipman D, Miura RM. Providence, RI: American Mathematical Society; 1986:57-86.
51. Friedrich J, Dandekar T, Wolf M, Müller T: **ProfDist: a tool for the construction of large phylogenetic trees based on profile distances.** *Bioinformatics* 2005, **21(9):**2108-2109.
52. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985:783-791.
53. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAxML web-servers.** *Systematic Biology* 2008 in press.
54. Liu J, Schardl CL: **A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes.** *Plant Molecular Biology* 1994, **26:**775-778.
55. Mai JC, Coleman AW: **The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants.** *J Mol Evol* 1997, **44:**258-271.

56. Turmel M, Brouard JS, Gagnon C, Otis C, Lemieux C: **Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses.** *J Phycol* 2008, **44:**739-756.
57. Côté CA, Peculis BA: **Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast.** *Nucleic Acids Research* 2001, **29:**2106-2116.
58. Peculis BA, Greer CL: **The structure of the ITS2-proximal stem is required for pre-rRNA processing in yeast RNA.** *RNA* 1998, **4:**1610-1622.
59. Nues RW van, Rientjes JM, Morré SA, Mollee E, Planta RJ, Venema J, Raué Hendrik A: **Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA.** *J Mol Biol* 1995, **250:**24-36.
60. Sande CA van der, Kwa MR, van Nues RW, van Heerikhuizen H, Raué Hendrik A, Planta RJ: **Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA.** *J Mol Biol* 1992, **223:**899-910.
61. Shoup S, Lewis LA: **Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta).** *J Phycol* 2003, **39:**789-796.
62. Hegewald E, Hepperle D, Wolf M, Krienitz L: **Phylogenetic placement of *Chlorotetraedron incus*, *C. polymorphum* and *Polyedriopsis spinulosa* (Neochloridaceae, Chlorophyta).** *Phycologia* 2001, **40:**399-402.
63. Melkonian M: **Structural and evolutionary aspects of the flagellar apparatus in green algae and land plants.** *Taxon* 1982, **31:**255-265.
64. Melkonian M, Surek B: **Phylogeny of the Chlorophyta: congruence between ultrastructural and molecular evidence.** *Bull Soc Zool Fr* 1995, **120:**191-208.
65. Wilcox LW, Floyd GL: **Ultrastructure of the gamete of *Pediastrum duplex* (Chlorophyceae).** *J Phycol* 1988, **24:**140-146.
66. Lewis LA: **Diversity and phylogenetic placement of *Bracteacoccus tereg* (Chlorophyceae, Chlorophyta) based on 18S ribosomal RNA gene sequence data.** *J Phycol* 1997, **33:**279.
67. Krienitz L, Hegewald E, Hepperle D, Wolf M: **The systematics of coccoid green algae: 18S rRNA gene sequence data versus morphology.** *Biologia* 2003, **58:**437-446.
68. An SS, Friedl T, Hegewald E: **Phylogenetic relationships of *Scenedesmus* and *Scenedesmus*-like coccoid green algae as inferred from ITS-2 rDNA sequence comparisons.** *Plant biol* 1999, **1:**418-428.
69. **AlgaeBase** [http://www.algaebase.org]
70. 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:**265-300.

P.5. ITS2 sequence-structure phylogeny in the Scenedesmaceae with special reference to *Coelastrum* (Chlorophyta, Chlorophyceae), including the new genera *Comasiella* and *Pectinodesmus*

Authors: E. Hegewald, M. Wolf, A. Keller*, T. Friedl, L. Krienitz

published 2010

Journal: *Phycologia*
Volume: 49, Pages: 325–335

*Co-author

Author's Contributions:
The study was designed and lead by E. Hegewald, M. Wolf and L. Krienitz. Laboratory work (sequencing procedure and electron microscopy) was performed by E. Hegewald and T. Friedl. Phylogenetic analayses including secondary structure predictions were done by me and M. Wolf. E. Hegewald and L. Krienitz drafted the manuscript with contributions of me and M. Wolf. All authors contributed to the final version of the manuscript and approved it.

Phycologia (2010) Volume 49 (4), 325–335 Published 2 July 2010

ITS2 sequence-structure phylogeny in the Scenedesmaceae with special reference to *Coelastrum* (Chlorophyta, Chlorophyceae), including the new genera *Comasiella* and *Pectinodesmus*

EBERHARD HEGEWALD[1]*, MATTHIAS WOLF[2], ALEXANDER KELLER[2], THOMAS FRIEDL[3] AND LOTHAR KRIENITZ[4]

[1]*Institute of Chemistry and Dynamics of the Geosphere 3, Research Centre Jülich, D-52425 Jülich, Germany*
[2]*Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany*
[3]*Experimental Phycology and Culture Collection of Algae (SAG), Department of the Albrecht-von-Haller-Institute for Plant Sciences at the University of Göttingen, Untere Karspüle 2, D-37073 Göttingen, Germany*
[4]*Leibniz-Institute of Freshwater Ecology & Inland Fisheries, Alte Fischerhütte 2, D-16775 Stechlin, Germany*

HEGEWALD E., WOLF M., KELLER A., FRIEDL T. AND KRIENITZ L. 2010. ITS2 sequence-structure phylogeny in the Scenedesmaceae with special reference on *Coelastrum* (Chlorophyta, Chlorophyceae). *Phycologia* 49: 325–335. DOI: 10.2216/09-61.1

Sequences and secondary structures of the nuclear-encoded internal transcribed spacer 2 (ITS2) ribosomal RNA of nine *Coelastrum* taxa, *Asterarcys quadricellulare* (Coelastraceae), *Westella botryoides* (hitherto Oocystaceae) and *Dimorphococcus lunatus* (Scenedesmaceae) were determined and compared with existing GenBank entries of scenedesmacean taxa (*Desmodesmus*, *Enallax*, *Neodesmus*, *Scenedesmus*). Phylogenetic analyses showed that the studied *Coelastrum* taxa belong to several different lineages within the Scenedesmaceae: five *Coelastrum* taxa (*Coelastrum microporum*, *Coelastrum astroideum*, *C. astroideum* var. *rugosum* = *Coelastrum rugosum*, *Coelastrum pseudomicroporum* and *Coelastrum sphaericum* incl. *Coelastrum proboscideum*) form monophyletic clades, whereas two strains labeled *Coelastrum morum* belong to different genera. The African strain of *C. morum* clusters with *Coelastrum cambricum*. The Finnish strain labeled *C. morum* clusters with *Asterarcys*, *Dimorphococcus* and *Hariotina*. According to its morphology this strain belongs to *Coelastrella*, related to *Coelastrella saiponensis*. *Westella botryoides* belongs to a separate clade within the Scenedesmaceae. *Coelastrum reticulatum* is positioned in the clade with *Asterarcys*, *Dimorphococcus* and *Coelastrella*; hence its separation in a separate genus, as originally described (*Hariotina*), is justified. In general, the phylogenetic analysis of ITS2 data shows that the Coelastraceae are included in the monophyletic Scenedesmaceae, and thus the splitting into two families is not justified, but they belong to the monophyletic subfamily Coelastroidea. The genera *Comasiella* and *Pectinodesmus* are newly erected, and several new combinations are proposed.

KEY WORDS: ITS2, new combinations, phylogeny, secondary structure, *Acutodesmus*, Coelastraceae, *Coelastrella*, *Coelastrum*, *Comasiella* gen. nov., *Hariotina*, *Pectinodesmus* gen. nov., Scenedesmaceae, *Scenedesmus*, *Westella*

INTRODUCTION

The family Scenedesmaceae (Oltmanns 1904), which belongs to the class Chlorophyceae, was described for flat or curved coenobia of different cell shape (ovate to spindle shaped) and later (see: Komárek & Fott 1983) expanded to genera with three-dimensional coenobia or syncoenobia (*Makinoella* Okada, now Oocystaceae, *Tetrallantos* Teiling and *Dimorphococcus* A. Braun). Komárek & Fott (1983) included in the family 28 genera, but recently the genera *Crucigeniella* Lemmermann (Krienitz *et al.* 2004), *Dicloster* Jao, Wei & Hu (Hegewald & Hanagata 2000), *Didymocystis* Korshikov (Hegewald & Deason 1989), *Didymogenes* Schmidle (Krienitz *et al.* 2004), *Makinoella* (Hepperle *et al.* 2000), and *Tetrachlorella* Korshikov were transferred to the class Trebouxiophyceae. Furthermore, a new genus was added: *Pseudodidymocystis* Hegewald & Deason (Hegewald & Deason 1989). The genera *Desmodesmus* S.S. An, E. Hegewald & Friedl (An *et al.* 1999) and *Acutodesmus* Tsarenko (Tsarenko & Petlevanny 2001) were split from the genus *Scenedesmus* Meyen. The genus *Tetradesmus* G.M. Smith was included in *Acutodesmus* by Tsarenko & Petlevanny (2001). However, according to molecular data *Acutodesmus*, as originally described, is polyphyletic (Hegewald & Wolf 2003).

The Coelastraceae (Wille 1909) were erected because of the three-dimensionally arranged, more or less spherical coenobia. Smith (1920) added a genus with star-like arranged spindle-like cells: *Actinastrum* Lagerheim. According to Komárek & Fott (1983), the family comprises five genera that are morphologically very different: *Actinastrum* (six species), *Asterarcys* Comas (one species), *Coelastropsis* Fott & Kalina (one species), *Coelastrum* Nägeli (depending on the author: 18 to 40 species), *Ducellieria* Teiling (two species) and *Soropediastrum* Wille (two species). The genus *Ducellieria* was also classified within the Xanthophyceae (Teiling 1957; Ettl 1978; Couté 1984), but recently it was shown to belong to the Oomycetes (Kusel-Fetzmann & Nouak 1981, Hesse *et al.* 1989). The genus *Actinastrum* was transferred to the Trebouxiophyceae on the basis of molecular evidence (Wolf *et al.* 2002). Phylogenetic analyses revealed that two strains originally described as members of *Scotiella* Fritsch but later transferred to *Coelastropsis* (Puncochárová & Kalina 1981) or *Scenedesmus* (Hanagata 1998) clustered within the Scenedesmaceae (Hegewald & Hanagata 2000).

* Corresponding author (e.hegewald@googlemail.com).

The morphological diversity of the coelastracean genera raises the question of whether these algae are a natural assemblage or an artificial group. The question of the systematic status even applies to single species of the type genus, *Coelastrum*, and has already been debated (summarized in Chan 1973). In *Coelastrum* there is some degree of polymorphism (Rayss 1915; Fenwick 1962, 1968; Fenwick *et al.* 1966) that prevents clear taxonomic conclusions solely on the basis of morphological features.

A close relationship of the Coelastraceae and the Scenedesmaceae was suggested by several authors and the family Coelastraceae was treated e.g. as subfamily Coelastreae within the Scenedesmaceae by Smith (1920) or as subfamily Coelastroideae by Printz (1927). Johnson *et al.* (2007) also discussed the phylogenetic position of *Coelastrum* and *Scenedesmus*.

In this study we were primarily interested in the phylogeny of the Coelastraceae and its relationship with the Scenedesmaceae, as inferred from analysis of the nuclear encoded internal transcribed spacer of the ribosomal RNA (ITS2 rRNA). Further genera of Scenedesmaceae were included as well as some newly isolated *Coelastrum* strains, e.g. strains with a rugose cell wall surface. The taxonomic status of these rugose *Coelastrum* strains is not definitively clarified, as they are commonly classified as varieties of *Coelastrum* species [*Coelastrum astroideum* var. *rugosum* (Rich) Sodomková or *Coelastrum sphaericum* var. *rugulosum* (Thomasson) Sodomková]. So far, only a few ITS2 sequences of *Coelastrum* strains have been published in GenBank (Benson *et al.* 2008). The *Coelastrum* strains in the strain collections are often wrongly identified; for example, the strain SAG 217-1c labelled *Coelastrum microporum* Nägeli was identified as *Coenochloris polycocca* (Korshikov) Korshikov by Wolf *et al.* (2003).

In addition, we were interested in the phylogenetic positions of other Scenedesmaceae: *Dimorphococcus lunatus* A. Braun and *Westella botryoides* (W. West) De Wildeman. These are aberrant scenedesmacean species characterized by the formation of three-dimensional coenobia or syncoenobia. We were also interested in the phylogenetic position of the genus *Asterarcys*, which was placed in the Coelastraceae; however, when revised by Hegewald & Schmidt (1992), the scenedesmacean genus *Suxenella* Shrivastava & Nizamuddin was put in synonymy with *Asterarcys*. In the past *Asterarcys* has also been considered a subgeneric taxon of the genera *Coelastrum*, *Crucigenia* Morren and *Tetrastrum* Chodat (Hegewald & Schmidt 1992); its phylogenetic position is therefore in need of clarification. Finally the phylogenetic positions of additional scenedesmacean taxa are discussed.

Although the 18S DNA shows in the Scenedesmaceae only few base-pair differences, the ITS2 with its common core of secondary structure in the Eukaryota (Schultz *et al.* 2005) has proved to be a more helpful tool for discrimination at the species level (e.g. An *et al.* 1999; van Hannen *et al.* 2002; Coleman 2003, 2009; Hegewald *et al.* 2005; Jeon & Hegewald 2006; Schultz *et al.* 2006; Vanormelingen *et al.* 2007; Schultz & Wolf 2009). Here we demonstrate its usefulness at the genus level also.

MATERIAL AND METHODS

Taxon sampling, DNA extraction, polymerase chain reaction (PCR) and sequencing

Strains newly sequenced were obtained from the SAG Culture Collection of Algae (Göttingen, Germany). Two strains were also obtained from the algal collection of the Leibniz-Institute of Freshwater Ecology & Inland Fisheries (Stechlin, Germany), one strain from the collection Tsarenko (algal strain collection, Kiev, Ukraine) and one strain from the Culture Collection of Algae at the University of Texas at Austin (UTEX, Austin, TX) (Table 1). For additional taxon sampling we downloaded all sequences matching the search pattern 'Scenedesmaceae and (ITS2 OR ''internal transcribed spacer 2'')' from GenBank (Benson *et al.* 2008). Additionally, two outgroup sequences of Hydrodictyaceae were chosen (*Hydrodictyon reticulatum* (L.) Lagerh. AY577747 and *Pediastrum duplex* Meyen AY577757). All retained ITS2 sequences were delimited and cropped with the hidden Markov model (HMM)-based annotation tool present at the ITS2 database (Keller *et al.* 2009; E-value < 0.001, Viridiplantae HMMs). DNA extraction, PCR and sequencing were performed as previously described in Hegewald & Wolf (2003).

Alignment and phylogenetic analyses

The phylogenetic analyses followed the procedure outlined in Schultz & Wolf (2009). Sequences and secondary structures were automatically aligned with 4SALE 1.5 using an ITS2-specific scoring matrix for sequences and structures (Seibel *et al.* 2006, 2008). To determine evolutionary distances between organisms simultaneously on sequences and secondary structures we used profile neighbor joining (PNJ) as implemented in ProfDistS 0.98 (Friedrich *et al.* 2005; Wolf *et al.* 2008). For this, we applied an ITS2-specific general time-reversible substitution model (Seibel *et al.* 2006). The resulting tree was displayed with iTol 1.3.1 (Letunic & Bork 2007) and further processed with CorelDRAW X3 (Corel Corporation, Ottawa, Canada). The alignment was further investigated with substitution rate calibration (SRC) to estimate the impact of long-branch attraction (data not shown, Van de Peer & De Wachter 1994). Additionally, we performed a maximum likelihood and a Bayesian analysis for which structural information was omitted. To keep these analyses time efficient, we included only a subset of sequences that represented all major clades. These analyses resulted in similar but less robust phylogenies than the tree recalculated by PNJ with secondary structures (supplementary data); therefore, they are not presented in the results.

Secondary structure prediction

The secondary structure of the ITS2 of *Scenedesmus obtusus* was predicted with RNAstructure 4.6 (Mathews *et al.* 2004) and exported to Vienna format with CBCanalyzer 1.0.3 (Wolf *et al.* 2005b). Structures of the remaining sequences were predicted by homology modeling in the ITS2 database (Wolf *et al.* 2005a; Schultz *et al.* 2006; Selig *et al.* 2008) with

Table 1. Strains sequenced in the study and one additional strain for which the sequence was obtained from GenBank.

Original name	Recent strain number	Original strain number	Original site of collection	GenBank accession number for ITS2	New name
Asterarcys cubensis Comas	SAG2195	Comas 1977/75	Cuba, Escaleras de Jaruco, basin	GQ375088	*Asterarcys quadricellulare* (Behre) E. Hegewald & A. Schmidt
Coelastrum astroideum De Notaris	SAG65.81	Hegewald 1973-233	Peru, Laguna Pacucha	GQ375089	
Coelastrum astroideum De Notaris	Krienitz 2005-45	Krienitz 2005-45	Tunisia, Jerba, oxidation pond	GQ375090	*Coelastrum microporum* Nägeli
Coelastrum astroideum var. *rugosum* (Rich) Sodomková	Tsarenko 1995-61	Tsarenko 1995-61	Germany, Lake Tollense	GQ375092	*Coelastrum rugosum* Rich
Coelastrum astroideum var. *rugosum* (Rich) Sodomková	UTEX 2442	Hegewald 1971-138	Hungary, fish pond at Babat	GQ375093	*Coelastrum rugosum* Rich
Coelastrum cambricum Archer	SAG7.81, UTEX 2446	Hegewald 1973-202	Peru, Iquitos, fish pond	GQ375106	
Coelastrum microporum Nägeli	SAG2292	Krienitz 1988/11	Germany, river Elbe near Aken	GQ375095	
Coelastrum sp.	Tow6/3P-9W	Tow6/3P-9W	USA, Minnesota, Itasca State Park	DQ417575	*Coelastrum microporum* Nägeli
Coelastrum morum W. West & G.S. West	SAG217-5	Droop (1950)	Finland, Brennskar	GQ375096	*Coelastrella* sp.
Coelastrum morum W. West & G.S. West	SAG2078	Hegewald 1999-5	Namibia, Windhoek, pond	GQ375097	
Coelastrum proboscideum (var. *dilatatum* Vischer)	SAG217-2	Vischer 13 (1924)	Switzerland, Neudorf	GQ375098	*Coelastrum sphaericum* Nägeli
Coelastrum proboscideum (var. *gracile* Vischer)	SAG217-3	Vischer 15 (1924)	Switzerland, Neudorf	GQ375099	*Coelastrum sphaericum* Nägeli
Coelastrum pseudomicroporum Korshikov	SAG33.88	Vodenicarov 710	Bulgaria, Lake Srebarna	GQ375100	
Coelastrum reticulatum (Dangeard) Senn	SAG8.81	Hegewald 1977-101	Germany, Aschau, pond	GQ375101	*Hariotina reticulata* Dangeard
Coelastrum sphaericum Nägeli	SAG32.81	Hegewald 1974-71	Hungary, Budapest aquarium	GQ375102	
Dimorphococcus lunatus Braun	SAG224-1	Bourrelly 90 (1945)	Unknown, acid water	GQ375103	
Westella botryoides (W. West) De Wildeman	SAG2094	Hegewald 2002-6	Germany, Nürnberg, channel	GQ375104	

the aforementioned structure as a template and at least 75% helix transfer (identity matrix). Sequences for which no correct annotation or secondary structure could be obtained were omitted.

Scanning electron microscopy (SEM)

For SEM, the samples were fixed with formaldehyde or glutaraldehyde and dehydrated in 20, 40, 60, 80 and 100% acetone, critical-point dried, sputtered with gold and studied under a Zeiss Gemini 1550VP electron microscope (Figs 2–7). For Fig. 2 a cryostage at −120°C was used.

RESULTS

The phylogenetic analyses recovered *Coelastrum* as paraphyletic (Fig. 1). *Coelastrum* was nested within the family Scenedesmaceae, in a clade in which strains of *Hariotina*, *Asterarcys*, *Coelastrella* Chodat and *Dimorphococcus* also occurred. The results of the SRC corroborated these results and rejected an erroneous effect caused by long-branch attraction of rapidly evolving taxa (e.g. *Desmodesmus*). Although the positions of many clades in the tree were not resolved because of the low support of the most internal branches, most clades in the terminal parts of the tree were well supported in their monophyly. A subclade containing the type species of *Coelastrum* (*C. sphaericum* Nägeli) and two strains of *Coelastrum proboscideum* Bohlin was well supported, as well as a subclade containing *C. microporum* and *C. astroideum* De-Notaris, a subclade with strains of *Coelastrum morum* W. & G.S. West and *Coelastrum cambricum* Archer, and a subclade with two strains of *Coelastrum rugosum* Rich, while the position of *Coelastrum pseudomicroporum* Korshikov was not resolved. The strains of *C. rugosum* (Figs 2–7) did not fall into a well-supported monophyletic group with *C. astroideum*; hence, *C. rugosum* should not be considered a variety of *C. astroideum*, as proposed by Sodomková (1972). The secondary structures of these two taxa differ from each other by four compensatory base changes (CBCs), two of which are located in the conserved base part of helix 1, one in the second helix and one in the fourth helix. The first and the latter are in common with *C. sphaericum*. The characteristic rugose wall was not always visible in light microscopy (Fig. 8), but it was well visible under the SEM, although

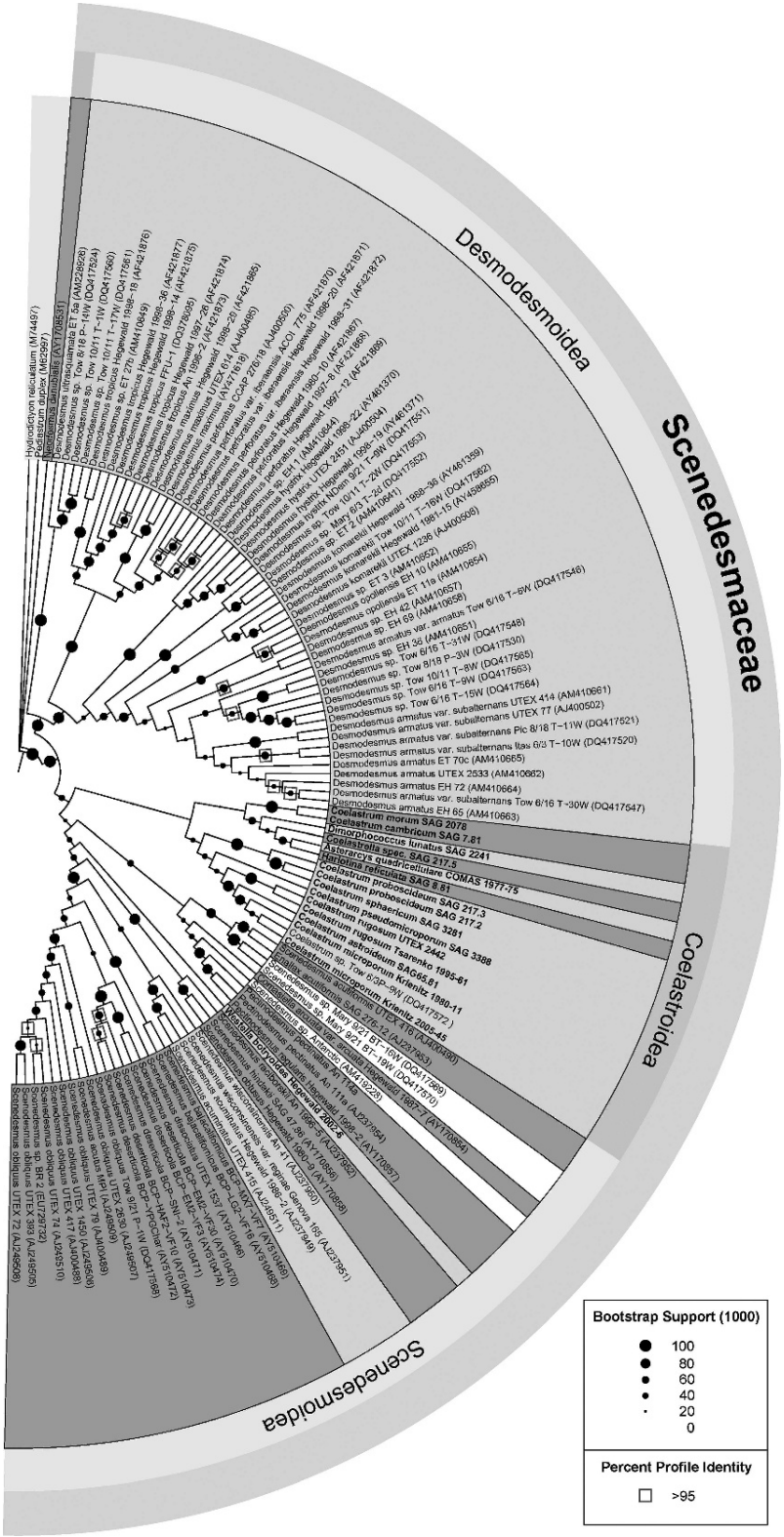
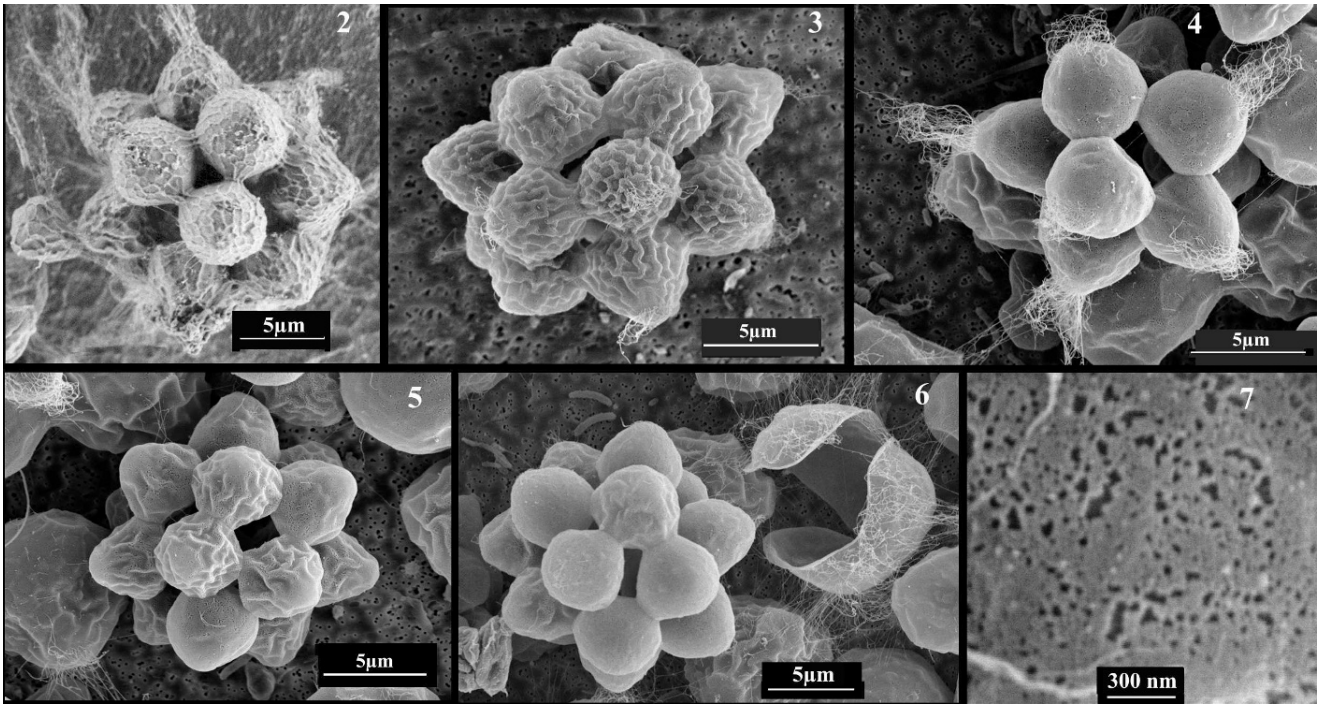


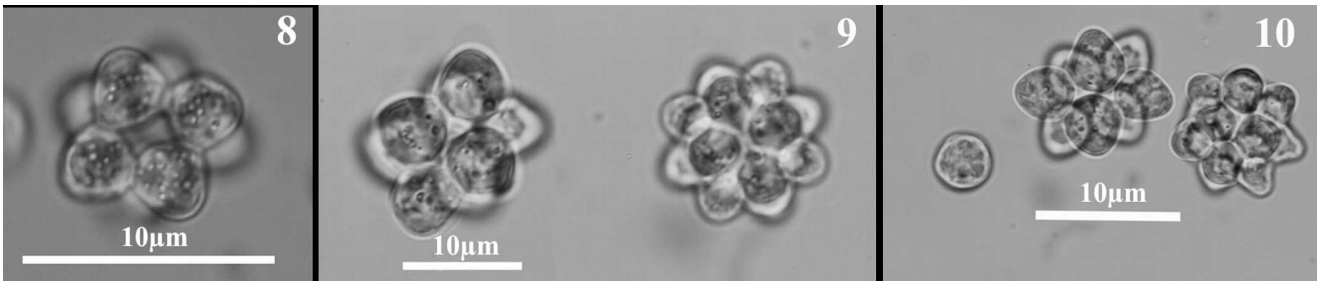
Fig. 1. ITS2-based profile-neighbor-joining tree including all major genera of Scenedesmaceae and two outgroup taxa (*Hydrodictyon reticulatum* and *Pediastrum duplex*). Dot thickness at branches represents bootstrap support values (1000 replicates, see explanation in figure). Sequences produced in this study are in bold. Neighbor-joining profiles created for sequences and structure with more than 95% identity are boxed. The names of the strains of UTEX are corrected according to Hegewald (1989).



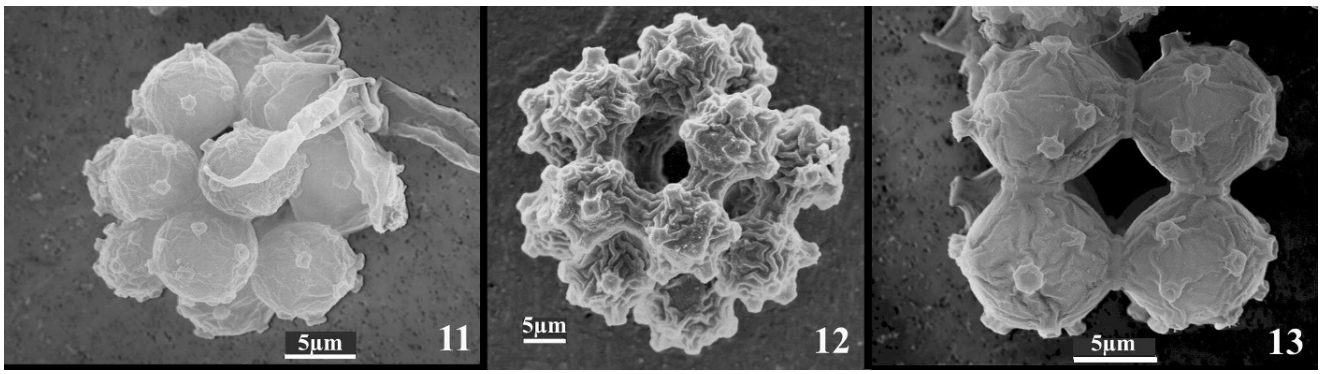
Figs 2–7. SEM images of *Coelastrum rugosum* strain UTEX 2442 (stub UTEX 2442); Figs 3–7 illustrate coenobia after critical-point drying.
Fig. 2. Sixteen-celled coenobium after deep-freezing treatment with typical rugose cell wall. Some cells with bristle excretion.
Fig. 3. Coenobium with rugose cell wall.
Fig. 4. Coenobium with nearly smooth cell walls and bristle excretion.
Fig. 5. Coenobium with cell walls variably corrugated.
Fig. 6. Mother cell wall with daughter coenobium just released.
Fig. 7. Detail of cell wall at higher magnification.

there were also coenobia that were lacking the cell wall ridges (Figs 4–7). To prove that the folds were not artifacts of the preparation for the SEM, we verified their presence using the deep freezing method (Fig. 2). It was difficult to observe morphological differences between *C. astroideum*, *C. pseudomicroporum* and the studied strain of *C. cambricum* (Figs 8–10).

Coelastrum morum strain SAG 2078 (Figs 11–13) and *C. cambricum* formed a highly supported monophyletic group. This group was included in a clade with several other genera placed hitherto in the Coelastraceae and also with the morphologically different *Dimorphococcus* (which was hitherto placed in the subfamily Scenedesmoideae), but with very low statistical support. Morphologically the taxa *C. morum* and *C. morum* f. *capensis* Fritsch seem not to be closely related but, as visible in the strain SAG 2078, both types of coenobia occur in the same strain (Figs 11, 12); this suggests that the f. *capensis* is not distinct from *C. morum* f. *morum*. The second strain labeled *C. morum* (SAG 217-5) was located in the sister cluster of *C. morum/cambricum*. In the culture collection SAG this strain is labeled *C. morum* ("*morus*"), in the strain collection CCAP (Culture Collection of Algae and Protozoa, SAMS Research Services Ltd, Dunstaffnage Marine Laboratory, OBAN, Argyll PA37 1QA, United Kingdom) and CCALA (Culture Collection of Autotrophic Organisms, Centre of Phycology, Tíeboá, Czech Republic) as *Coelastropsis costata* (Korshikov) Fott & Kalina. Light microscopical observations (Fig. 14, see



Figs 8–10. *Coelastrum* species observed in light microscopy.
Fig. 8. *Coelastrum rugosum* strain UTEX 2442.
Fig. 9. *Coelastrum microporum* strain Krienitz 1980-11.
Fig. 10. *Coelastrum cambricum* strain SAG 7.81.



Figs 11–13. SEM images of *Coelastrum morum* strain SAG 2078 (stub SAG 2078).
Fig. 11. Older 16-celled coenobium with cell connections not visible.
Fig. 12. Younger 16-celled coenobium with cell connections well visible.
Fig. 13. Eight-celled coenobium.

also http://www.butbn.cas.cz/ccala/col_images/310.jpg) and especially electron microscopical studies (Figs 15–17) show ridges on cell walls indicating that this strain belongs to the genus *Coelastrella*. The closest relative, the taxon *Coelastrella saiponensis* Hanagata, shows morphological similarities but differs by six nucleotides in the 18S rDNA sequence (Hegewald, unpublished observations).

Coelastrum reticulatum (Dangeard) Senn is found in the same cluster that contains the morphologically very different *Coelastrum morum/cambricum*, *Asterarcys*, *Dimorphococcus* and *Coelastrella*. Fig. 1 also supports the placement of this species in a separate genus. The other genera of that clade are totally different in cell shape and arrangement. Finally, we can show that the former family Coelastraceae has its phylogenetic level as a subfamily in Scenedesmaceae. The subfamily Scenedesmoideae includes the genera *Comasiella* nov. gen., *Enallax* Pascher, *Scenedesmus*, *Pectinodesmus* nov. gen. and *Westella* De-Wildeman, but their support at the base of the branching is weak.

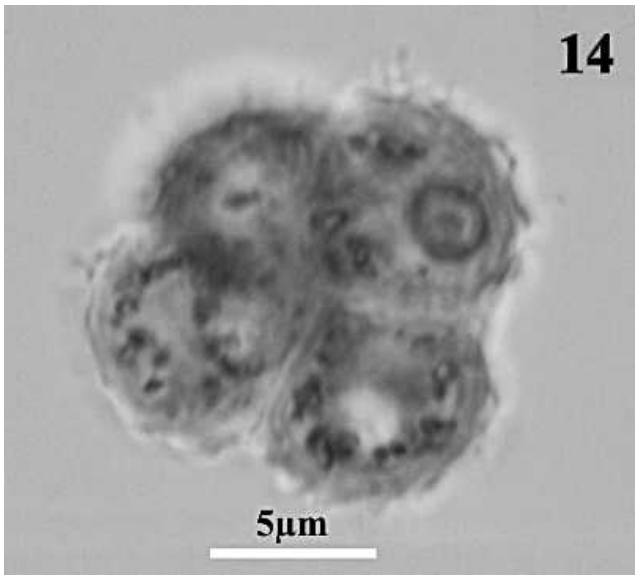


Fig. 14. *Coelastrella* sp. strain SAG 217-5, as observed in light microscopy. Note ridges appearing as protuberances.

The subfamily Desmodesmoideae includes the species-rich genus *Desmodesmus* and the genus *Neodesmus* Hindák and *Pseudodidymocystis*.

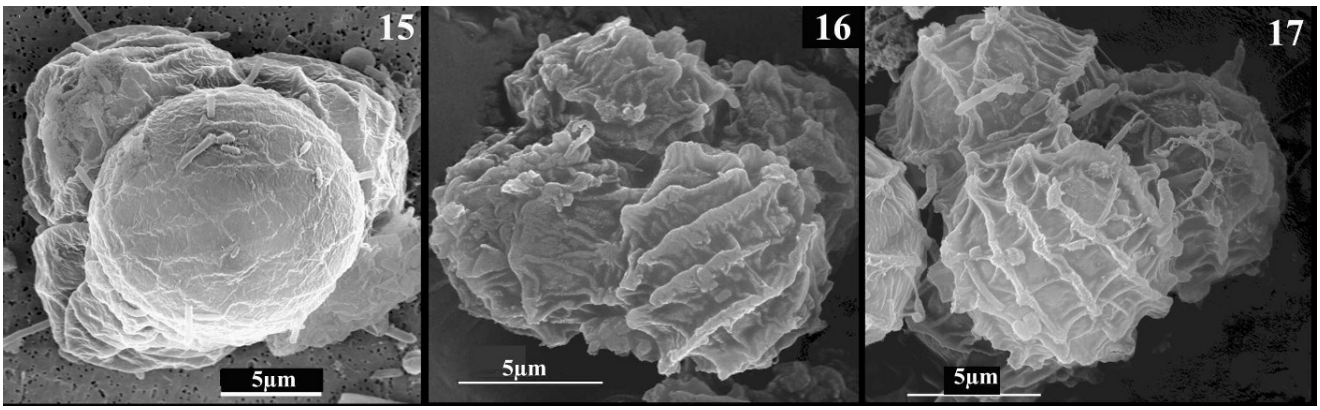
Westella was not clearly assigned to any clade. However, it was placed between *Pectinodesmus* and *Scenedesmus*, suggesting that it probably belongs to the Scenedesmaceae. Morphologically it is quite different from both genera because of its globular cells, embedded in mucilage (Fig. 18). Within *Scenedesmus*, *Scenedesmus rotundus* Lewis & Flechtner has also globular cells but it is not embedded in mucilage and does not form that type of coenobium.

DISCUSSION

The family Coelastraceae was erected by Wille (1909) because of the spherical coenobia, which are strikingly different from the flat coenobia of the Scenedesmaceae.

The idea that *Coelastrum* and its relatives could represent a subfamily in the family Scenedesmaceae was already expressed by Smith (1920) and Printz (1927). Our ITS2 study support this view – *Coelastrum* is grouped together with *Hariotina*, *Asterarcys*, *Coelastrella* and *Dimorphococcus* in a separate cluster that could be regarded as a subfamily.

So far the delineation of taxa within the genus *Coelastrum* has been exclusively on the basis of morphological criteria, which exhibit a broad range of variability. Here we present the first results supporting the morphological evidence by molecular evidence on the basis of CBCs within the secondary structure of ITS2. CBCs can be used to distinguish species (Müller *et al.* 2007). For example, it was shown that strains of *C. rugosum* do not form a monophyletic group with *C. astroideum* and differ by four CBCs. The first and the last are common to *C. sphaericum*. We suggest treating the taxon *C. rugosum* as a species and not as a variety, as chosen by Sodomková (1972). The cell wall structures of *C. sphaericum* var. *rugosum* were illustrated by Tell & Couté (1979); because no cultured strains of this alga are available, its phylogenetic position cannot be presently clarified. The subclade with the strains of the type species *C. sphaericum* includes two strains



Figs 15–17. SEM images of *Coelastrella* sp. strain SAG 217-5 (stub SAG 217-5), showing variability in the habit of ridges. Fig. 15 shows the most common design.

labeled *C. proboscideum*, but this taxon was already treated as a synonym of *C. sphaericum* by Hajdu et al. (1976).

The taxonomic relationships of *Coelastrum morum*

This taxon was described in the 19th century, but its descriptions and illustrations were puzzling. The first illustration of *C. morum* in our sense is given by Reinsch (1877) and labeled '*C. verrucosus* Reinsch', a taxon that this author described 2 years earlier in 1875. This taxon was treated e.g. by Komárek & Fott (1983) as a new combination of *Sphaerastrum verrucosum* Reinsch (Reinsch 1875), but in the glossary they labeled it '*Coelastrum* (*Sphaerastrum*) *verrucosum*', whereas Comas (1989) and John *et al.* (2002) interpreted it as a newly described species.

The nomenclatural and taxonomical circumscription of Reinsch's species (*S. verrucosus*, *C. verrucosus*, and *Colastrum scabrum* Reinsch) and of the genus name *Sphaerastrum* Reinsch, which has already been used by Greeff (1873) for a Heliozoan, is very complicated. We discard all these names for our strain and use the name *C. morum* of West & West (1896). *Coelastrum morum* was treated by Comas (1989) as a variety of *C. verrucosum*. *Coelastrum morum* was illustrated as a spherical colony of globular cells with protuberances regularly distributed all over the surface, which is in good agreement with our Fig. 11. The connecting strands were not visible, but it can be suspected that two to three of these protuberances connected the cells (e.g. Sodomková 1972). Illustrations similar to our Figs 12 and 13 of the strain SAG 2078 are given by Fritsch (1918) as *C. morum* f. *capensis*. The same taxon is excellently illustrated in Rino (1972). Both illustrations differ from *C. morum*, as described by West & West (1896), for the cell shape and the cell-connecting strands, but coenobia with this cell type fall within the range of variability of the species (Figs 11–13). *Coelastrum morum* seems to have subtropical distribution (e.g. Fritsch 1918; West & West 1896; Rino 1972; Komárek & Fott 1983; Comas 1989) and the strain sequenced in this study was isolated from Namibia (Table 1).

We suspect that the *Coelastrum* taxa with few and loosely arranged cells such as *C. morum* var. *acutiverrocosum* Bourr. & Manguin and the recently described *Coelastrum pascheri* Lukavský (Lukavský 2006), and possibly also *C. verrucosum*, are *Coelastrella* taxa.

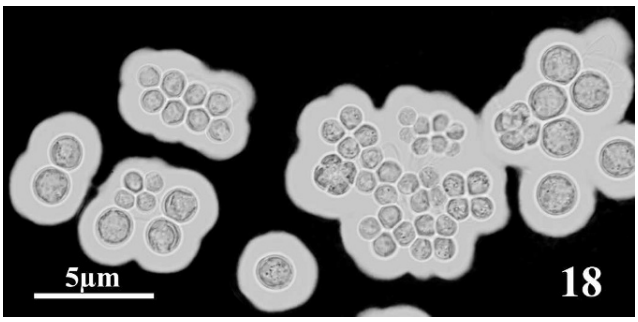


Fig. 18. *Westella botryoides* strain SAG 2094 under the light microscope, mainly four-celled coenobia, some stages of division, also single cells, two- and eight-celled coenobia. Negative staining with India ink, hence mucilage visible.

The *Hariotina reticulata* relationships

Coelastrum reticulatum was originally included in a separate genus, *Hariotina* P.A. Dangeard (Dangeard 1889) and subsequently transferred to *Coelastrum* by Senn (1899). Since then this species has been treated as a member of this genus. Recently Hegewald et al. (2002) recommended treating it as a member of a separate genus and for this reintroduced the name *Hariotina*. Similar observations were reported by Krienitz et al. (2003). The genus is characterized by one to three thin elongated connection strands on the top of the cells and the coenobium is embedded in mucilage, whereas *Coelastrum* has its short wide connection strands at the base of the cells and has no mucilage envelope.

The genera of the subfamily Desmodesmoideae and Scenedesmoideae

The monographic treatment of the Scenedesmaceae of Komárek & Fott (1983) resulted in 28 genera. Some genera were transferred to other families or even classes on the

basis of DNA sequence data (e.g. *Makinoella* Okada and *Tetrachlorella* to the Oocystaceae, Trebouxiophyceae; Hepperle *et al.* 2000) or on electron microscopical evidence (e.g. *Didymocystis*; Hegewald & Deason 1989). The genera *Crucigeniella* (Krienitz et al. 2003), *Dicloster* (An *et al.* 1999; Hegewald & Hanagata 2000; Proeschold *et al.* in press) and *Didymogenes* (Schnepf & Hegewald 1993, Proeschold *et al.* in press) were also transferred to the Trebouxiophyceae. The genera *Pseudotetradesmus* Hirose & Akiyama, *Raysiella* Edelstein & Prescott and *Schroederiella* Wołłosz. were merged into *Scenedesmus* by Hegewald (1989). The scenedesmacean genus *Suxenella* Srivastava & Nizamuddin was placed in synonymy with the coelastracean genus *Asterarcys* (Hegewald & Schmidt 1992), but this genus groups with the Scenedesmaceae.

The new genus *Pseudodidymocystis* was added to the Scenedesmaceae by Hegewald & Deason (1989). The genus *Coelastrella* was shown to belong to the Scenedesmaceae and not to the Chlorellaceae (Hanagata 1998, Hegewald & Hanagata 2000). Hanagata (1998) also transferred two species of *Scotiellopsis* Vinatzer to *Scenedesmus*. Although a variety of the type species was studied and not the type species itself, we suspect that the whole genus *Scotiellopsis* belongs to the Scenedesmaceae. The genus *Scotiellopsis* was erected for aggregate-building species and *Coelastrella* for unicellular species. We do not consider this as a valid character for separation at the genus level (as it is also not the case in *Scenedesmus* or *Desmodesmus*) and therefore include *Coelastrella* in *Scotiellopsis*. *Desmodesmus* (An *et al.* 1999) and *Acutodesmus* (Tsarenko & Petlevanny 2001) were split from *Scenedesmus*, whereas *Tetradesmus* was included in *Acutodesmus* (Tsarenko & Petlevanny 2001). Krienitz *et al.* (2003) realized that *Pectodictyon pyramidale* Akiyama & Hirose groups together with *H. reticulata*, and therefore it is placed not in the Radiococcaceae but in the Scenedesmaceae subfamily Coelastroideae. Because *P. pyramidale* is not the type of *Pectodictyon* Taft, this genus should be provisionally retained in the Radiococcaceae. A morphological similarity of *P. pyramidale* with *Coelastrum proboscideum* was mentioned by Bourrelly (1966). But, on the basis of 18S rRNA sequence data, *P. pyramidale* is more closely related to *Hariotina* and *C. morum*. *Pectodictyon pyramidale* is similar in morphology to *Hariotina*, especially because of the apical cell connections, as well visible on the illustrations in the Protist Information Server (1995–2009) and in Yamagishi & Akiyama (1994). A morphological similarity with *C. cambricum* is represented by the central protuberance on each cell, as illustrated by Yamagishi & Akiyama (1994).

From the former Coelastraceae, the genera *Coelastropsis* was already transferred to the Scenedesmaceae (Hanagata 1998; Hegewald & Hanagata 2000) and the genus *Actinastrum* was transferred to the Trebouxiophyceae (Wolf *et al.* 2002).

It was recommended by Hegewald & Wolf (2003) that the genus *Scenedesmus* should be revised and split into several genera. The phylogenetic results of this study, illustrated in Fig. 1, support this viewpoint. The clades obtained in our analysis show the erection of *Comasiella* as a separate genus from *Scenedesmus* and a further splitting of the new genus *Pectinodesmus* from the subgenus *Acutodesmus* as appropriate taxonomic decisions. Hegewald & Hanagata (2000) hesitated to treat the subgenera as genera and our present results, which are based on a larger set of strains, support their decision of not splitting the genus *Acutodesmus*. The subgenus *Scenedesmus* is a monophyletic taxon, which was also separated from the subgenus *Acutodesmus* in the phylogenetic trees in Lewis & Flechtner (2004) and by the G/C composition also in Paschma & Hegewald (1986) and Hegewald (1997).

According to our phylogenetic tree (Fig. 1) the subfamily Coelastroideae includes the genera *Coelastrum*, *Coelastrella*, *Hariotina*, *Asterarcys* and *Dimorphococcus*.

The subfamily Desmodesmoideae (Hegewald & Hanagata 2000, 2002) includes the large genus *Desmodesmus*, as also the small genera *Neodesmus* and *Pseudodidymocystis*.

In conclusion, the family Scenedesmaceae includes now 29 genera: *Acutodesmus*, *Asterarcys*, *Coelastrella*, *Coelastropsis*, *Coelastrum*, *Comasiella*, *Coronastrum* Thompson, *Crucigenia*, *Danubia* Hindák, *Desmodesmus*, *Dimorphococcus*, *Enallax*, *Gilbertsmithia* Iyengar, *Hariotina*, *Komarekia* Fott, *Lauterborniella* Schmidle (doubtful genus), *Neodesmus*, *Pectinodesmus*, *Pseudodidymocystis*, *Pseudotetrastrum* Hindák, *Scenedesmus*, *Schmidleia* Wolosz., *Tetrallantos*, *Tetranephris* Leite et Bicudo, *Tetrastrum*, *Westella*, *Westellopsis* Jao, *Willea* Schmidle (probably also a member of Trebouxiophyceae because of similarity with *Crucigeniella*) and the species *Pectodictyon pyramidale*. The coelastraceaen genus *Soropediastrum* Wille is excluded because it is doubtful according to Komárek & Fott (1983).

CONCLUSIONS

New taxa and new combinations

Comasiella E. Hegewald, M. Wolf, Al. Keller, Friedl & Krienitz *gen. nov.*

Cellulae oblongae et curvatae, cum obtusis polis, in coenobiis 4–8(–16) cellularum. Coenobia cum tegumento mucoso. Distinctae de aliis Scenedesmaceis sunt per differentias in ITS2 consensibus sequentiarum.

Cells elongate and curved with obtuse cell poles in 4–8 (–16)-celled coenobia. The coenobia are surrounded by mucilage. Distinguished from other Scenedesmaceae due to differences in ITS2 consensus sequences.

ETYMOLOGY: Named to honour the phycologist A.G. Comas from Cuba, who studied intensively the Scenedesmaceae.

TYPE SPECIES: *Comasiella arcuata* (Lemmermann) E. Hegewald, M. Wolf, Al. Keller, Friedl & Krienitz

Comasiella arcuata (Lemmermann) E. Hegewald, M. Wolf, Al. Keller, Friedl & Krienitz *comb. nov.*

BASIONYM: *Scenedesmus bijugatus* var. *arcuatus* Lemmermann, 1898, *Bot. Centralbl.* 76: 150.

According to the tree in Fig. 2 of Hegewald & Wolf (2003) a second taxon belongs to that genus:

Comasiella arcuata var. *platydisca* (G.M. Smith) E. Hegewald & M. Wolf *comb. nov*

BASIONYM:*Scenedesmus arcuatus* var. *platydiscus* G.M.Smith, 1916, Trans. Wisc. Acad. Sci. Arts & Lett. 18: 451.

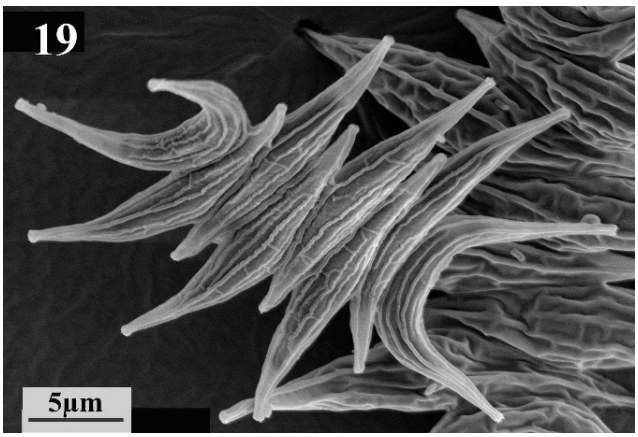


Fig. 19. *Pectinodesmus pectinatus* strain Hegewald 2001-2 (CCAP 276/68) under the SEM (stub Heg 01-2).

Pectinodesmus E. Hegewald, M. Wolf, Al. Keller, Friedl & Krienitz *gen. nov.*

Cellulae fusiformes in planis coenobiis 4 vel 8; cellulatis, vel inter eas torquatae (90°). Cellulae in linearibus ordinibus vel alternae. Magnitudo cellularum > 10 µm. In SEM observatione cellulae cristatae videntur (Fig. 19). Tegumentum mucosum non habent. Distinctae de aliis Scenedesmaceis sunt per differentiis in ITS2 consensibus sequentiarum.

Cells spindle-like in four- or eight-celled flat coenobia or cells twisted up to 90° to each other, cells linearly or alternatingly arranged. Cell sizes > 10 µm. Under the SEM cells with longitudinally arranged ridges (Fig. 19). No mucilage envelope. Distinguished from other Scenedesmaceae because of differences in ITS2 consensus sequences.

TYPE SPECIES: *Pectinodesmus pectinatus* (Meyen) E. Hegewald, M. Wolf, Al. Keller, Friedl & Krienitz

Pectinodesmus pectinatus (Meyen) E. Hegewald, M. Wolf, Al. Keller, Friedl & Krienitz *comb. nov.*

BASIONYM: *Scenedesmus pectinatus* Meyen, 1829 Verh. K. Leopold.-Carol. Akad. Naturf. 14: 775.

Pectinodesmus regularis (Svir.) E. Hegewald , M. Wolf, Al. Keller, Friedl & Krienitz *comb. nov.*

BASIONYM: *Scenedesmus regularis* Svirenko, 1924, Russk. Arkh. Protistol. 3: 178.

ACKNOWLEDGEMENTS

We thank L. Wei for the alignment of strain Krienitz 2005-45, D. Hepperle for the alignments of six strains and T. Holtmann for translating the diagnosis into Latin language. We thank W. Schroeder (Figs 2–7, 11–13) and H.-P. Bochem for SEM photos (Figs 15–17) and the reviewers and the associate editor for valuable comments and language corrections. L. Krienitz and M. Wolf were supported by grants of the Deutsche Forschungsgemeinschaft (DFG). A. Keller acknowledges the support of the BIGSS graduate school. This work was supported by a grant extended by the German Ministry of Education and Research (BMBF) to TF, contract 01LC0026, project BIOLOG 'AlgaTerra AT4'.

REFERENCES

AN S.S., FRIEDL T. & HEGEWALD E. 1999. Phylogenetic relationships of *Scenedesmus* and *Scenedesmus*-like coccoid green algae as inferred from ITS-2 rDNA sequence comparisons. *Plant Biology* 1: 418–428.

BENSON D.A., KARSCH-MIZRACHI I., LIPMAN D.J., OSTELL J. & WHEELER D.L. 2008. *Genbank*. *Nucleic Acids Research* 36: D25–D30.

BOURRELLY P. 1966. *Les Algues dXeau douce. I. Les algues vertes*. Bourbée & Cie, Paris. 511 pp.

CHAN K. 1973. A review on the genus *Coelastrum* (Chlorophyceae). *Journal of the Chinese University of Hong Kong* 1: 271–281.

COLEMAN A.W. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* 19: 370–375.

COLEMAN A.W. 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* 50: 197–203.

COMAS A. 1989. Taxonomische Übersicht der zönobialen Chlorokokkalalgen von Kuba. II. Fam. Coelastraceae. *Archiv für Hydrobiologie/Suppl.* 82*, Algological Studies* 56: 347–364.

COMAS A. 1996. Las Chlorococcales dulciacuicolas de Cuba. *Bibliotheca Phycologica* 99: 1–192.

COUTÉ A. 1984. PremiPres observations au M.E.T. et au M.E.B. sur la cytologie de *Ducellieria chodatii* (Ducell.) Teiling (Xanthophyceae, Mischococcales, Chlorobotrydaceae). *Nova Hedwigia* 39: 651–662.

DANGEARD P.A. 1889. Mémoire sur les algues. *Le Botaniste* 1: 127–174.

ETTL H. 1978. *Xanthophyceae 1. Teil*. In: *Süßwasserflora von Mitteleuropa,* vol. 3 (Ed. by H. Ettl, J. Gerloff & H. Heynig), G. Fischer Verlag, Stuttgart, New York. 530 pp.

FENWICK M.G. 1962. Some interesting algae from Lake Huron. *Transactions of the American Microscopic Society* 81: 72–76.

FENWICK M.G. 1968. Review of the status of some green algae in the genus *Coelastrum*. *Michigan Botanist* 7: 129–131.

FENWICK M.G., HANSEN J.O. & LYNCH D.L. 1966. Polymorphic forms of *Coelastrum proboscideum* Bohn. *Transactions of the American Microscopic Society* 85: 579–581.

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.

FRITSCH F.E. 1918. A first report on the fresh-water algae mostly from the Cape Peninsula in the herbarium of the South African Museum. *Annales of the South African Museum* 9: 483–611.

GREEFF R. 1873. Über Radiolarien und radiolarienartige Rhizopoden des süssen Wassers. Sitzungsberichte der Gesellschaft zur Beförderung der gesammten Naturwissenschaften zu Marburg. 1873, No. 5: 47–64.

GUNNISON D. & ALEXANDER M. 1975. Basis for the resistance of several algae to microbial decomposition. *Applied Microbiology* 29: 729–738.

HAJDU L., HEGEWALD E. & CRONBERG G. 1976. Beiträge zur Taxonomie der Gattung *Coelastrum*/Chlorophyta, Chlorococcales. *Annales Historico Naturales Musei Nationalis Hungarici* 68: 31–37.

HANAGATA N. 1998. Phylogeny of the subfamily Scotiellocystoideae (Chlorophyceae, Chlorophyta) and related taxa inferred from 18S ribosomal RNA gene sequence data. *Journal of Phycology* 34: 1049–1054.

HANAGATA N. 2001. New species of *Coelastrella* and *Scenedesmus* (Chorophyceae, Chlorophyta). *The Journal of Japanese Botany* 76: 129–136.

HEGEWALD E. 1989. The *Scenedesmus* strains of the Culture Collection of the University of Texas at Austin, Texas (UTEX). *Arch. Hydrobiologie*/Suppl. 82*, Algological Studies* 55: 153–189.

HEGEWALD E. 1997. Taxonomy and phylogeny of *Scenedesmus*. *Algae (Korean Journal of Phycology)* 12: 235–246.

HEGEWALD E. & DEASON T. 1989. *Pseudodidymocystis*, a new genus of Scenedesmaceae (Chlorophyceae). *Archiv für Hydrobiologie*/Suppl. 82*, Algological Studies* 55: 119–127.

HEGEWALD E. & HANAGATA N. 2000. Phylogenetic studies on Scenedesmaceae (Chlorophyta). *Algological Studies* 100: 29–49.

HEGEWALD E. & SCHMIDT A. 1992. *Asterarcys* Comas, eine weit verbreitete tropische Grünalgengattung. *Archiv für Hydrobiologie*/Suppl. 94*, Algological Studies* 66: 25–30.

HEGEWALD E. & WOLF M. 2003. Phylogenetic relationships of *Scenedesmus* and *Acutodesmus* (Chlorophyta, Chlorophyceae) as inferred from 18S rDNA and ITS-2 sequence comparisons. *Plant Systematics and Evolution* 241: 185–191.

HEGEWALD E., COESEL P.F.M. & HEGEWALD P. 2002. A phytoplankton collection from Bali, with the description of a new *Desmodesmus* species (Chlorophyta, Scenedesmaceae). *Algological Studies* 105: 51–78.

HEGEWALD E., SCHMIDT A., BRABAND A. & TSARENKO P. 2005. Revision of the *Desmodesmus* (Sphaeropleales, Scenedesmaceae) species with lateral spines. 2. The multi-spined to spineless taxa. *Archiv für Hydrobiologie*/Suppl. 157*, Algological Studies* 116: 1–38.

HEPPERLE D., HEGEWALD E. & KRIENITZ L. 2000. Phylogenetic position of the Oocystaceae (Chlorophyta). *Journal of Phycology* 36: 590–595.

HESSE M., KUSEL-FETZMANN E. & CARNIEL K. 1989. Life cycle and ultrastructure of *Ducellieria chodati* (Oomycetes). *Plant Systematics and Evolution* 165: 1–15.

JEON S.L. & HEGEWALD E. 2006. A revision of the species *Desmodesmus perforatus* and *D. tropicus* (Chlorophyta, Chlorococcales, Scenedesmaceae). *Phycologia* 45: 567–584.

JOHN D.M., WHITTON B.A. & BROOK A.J. 2002. *The freshwater algal flora of the British Isles*. Cambridge University Press, New York. 714 pp.

JOHNSON J.L., FAWLEY M.W. & FAWLEY K.P. 2007. The diversity of *Scenedesmus* and *Desmodesmus* (Chlorophyceae) in Itasca State Park, Minnesota, USA. *Phycologia* 46: 214–229.

KELLER A., SCHLEICHER T., SCHULTZ J., MÜLLER T., DANDEKAR T. & WOLF M. 2009. 5.8S–28S rRNA interaction and HMM-based ITS2 annotation. *Gene* 430: 50–57.

KOMÁREK J. & FOTT B. 1983. Chlorophyceae (Grünalgen), Ordnung Chlorococcales. In: *Die Binnengewässer. Das Phytoplankton des Süßwassers, 7. Teil, 1. Hälfte* (Ed. by G. Huber-Pestalozzi), vol. XVI. Schweizerbart, Stuttgart. 1044 pp.

KRIENITZ L., HEGEWALD E., HEPPERLE D. & WOLF M. 2003. The systematics of coccoid green algae: 18S rRNA gene sequence data versus morphology. *Biologia, Bratislava* 58: 437–446.

KRIENITZ L., HEGEWALD E., HEPPERLE D., HUSS V.A.R., ROHR T. & WOLF M. 2004. Phylogenetic relationship of *Chlorella* and *Parachlorella* gen. nov. (Chlorophyta, Trebouxiophyceae). *Phycologia* 43: 529–542.

KUSEL-FETZMANN E. & NOUAK H. 1981. *Ducellieria chodati* – Alge oder Pilz? *Plant Systematics and Evolution* 138: 199–207.

LEMMERMANN E. 1898. Beiträge zur Kenntnis der Planktonalgen. *Botanischen Centralblatt* 76: 150–156.

LETUNIC I. & BORK P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23: 127–128.

LEWIS L.A. & FLECHTNER V.R. 2004. Cryptic species of *Scenedesmus* (Chlorophyta) from desert soil communities of Western North America. *Journal of Phycology* 40: 1127–1137.

LUKAVSKÝ J. 2006. *Coelastrum pascheri* sp. n., a new green algae from lakes of the Bohemian Forest. *Biologia, Bratislava* 61, Suppl. 20: 485–490.

MATHEWS D.H., DISNEY M.D., CHILDS J.L., SCHROEDER S.J., ZUKER M. & TURNER D.H. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America* 101: 7287–7292.

MEYEN F.J.F. 1829. Beobachtungen über einige niedere Algenformen. *Verhandlungen der Kaiserlichen Leopoldinisch-Carololinischen Akademie der Naturforscher* 14: 769–778.

MÜLLER T., PHILIPPI N., DANDEKAR T., SCHULTZ J. & WOLF M. 2007. Distinguishing species. *RNA* 13: 1469–1472.

OLTMANNS F. 1904. *Morphologie und Biologie der Algen,* vol. 1: Spezieller Teil. G. Fischer, Jena. 733 pp.

PASCHMA R. & HEGEWALD E. 1986. DNA base composition within the genus *Scenedesmus* (Chlorophyta). *Plant Systematics and Evolution* 153: 171–180.

PRINTZ H. 1927. Chlorophyceae (nebst Conjugatae, Heterocontae und Charophyta). In: *Die Natürlichen Pflanzenfamilien nebst ihren Gattungen und wichtiger Arten insbesondere der Nutzpflanzen,* 3. Band (Ed. by A. Engler & K. Prantl), Wilhelm Engelmann, Leipzig. 463 pp.

PROESCHOLD T., BOCK C., LUO W. & KRIENITZ L. 2010. Polyphyletic origin of bristle formation in Chlorellaceae: *Micractinium*, *Didymogenes* and *Hegewaldia* gen. nov. (Trebouxiophyceae, Chlorophyta). *Phycological Research* 58: 1–8.

PROTIST INFORMATION SERVER, http://protist.i.hosei.ac.jp/pdb/images/Chlorophyta/Pectodictyon/sp_1a.jpg

PUNCOCHÁROVÁ M. & KALINA T. 1981. Taxonomy of the genus *Scotiellopsis* Vinatzer (Chlorococcales, Chlorophyta). *Archiv für Hydrobiologie*/Suppl. 60, 2*, Algological Studies* 27: 119–147.

RAYSS T. 1915. Le *Coelastrum proboscideum* Bohl. Étude de planctologie expérimentale suivide d'une révision des *Coelastrum* de la Suisse. Université Genève, Institute de Botanique 9: 1–65.

REINSCH P. 1875. Contributiones ad Algologiam et Fungologiam. Lipsiae 1875. 103 pp. + 131 plates pp.

REINSCH P.F. 1877. On freshwater algae from the Cape of Good Hope. *Journal of the Linnean Society London, Botany* 16: 232–248.

REYMOND O. 1975. La paroi cellulaire de *Coelastrum* (*Chlorophycées*). *Archiv für Microbiologie* 102: 95–101.

RINO J.A. 1972. Contribuiç o para o conhecimento das algas de água doce de Moçambique – III. *Revista de Sciencis Biologicas* 5, ser. A: 121–264 + pl. I–XXXII.

SCHNEPF E. & HEGEWALD E. 1993. *Didymogenes palatina* Schmidle and *Didymogenes anomala* (G. M. Smith) Hind. (Chlorococcales): taxonomy, ultrastructure, autosporogenesis and autospore wall assembly. *Archiv für Protistenkunde* 143: 41–53.

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.

SCHULTZ J., MAISEL S., GERLACH D., MÜLLER T. & WOLF M. 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11: 361–364.

SCHULTZ J., MÜLLER T., ACHTZIGER M., SEIBEL P., DANDEKAR T. & WOLF M. 2006. The internal transcribed spacer 2 database – a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Research* 34: 704–707.

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.

SELIG C., WOLF M., MÜLLER T., DANDEKAR T. & SCHULTZ J. 2008. The ITS2 Database II: homology modelling RNA structure for molecular systematics. *Nucleic Acids Research* 36: D377–D380.

SENN G. 1899. Über einige coloniebildende einzellige Algen. *Botanische Zeitung* 57: 39–105.

SMITH G.M. 1916. A monograph of the algal genus *Scenedesmus* based upon based pure culture studies. *Transactions of the Wisconsin Academy of Sciences, Arts & Letters* 18: 422–528, pl. 25–33.

SMITH G.M. 1920. Phytoplankton of the inland lakes of Wisconsin. *Wisconsin Geological and Natural History Survey, Bulletin* 57: 1–234.

SODOMKOVÁ M. 1972. Taxonomische Übersicht der Gattung *Coelastrum* Nägeli. *Acta Universitatis Carolinae-Biologica* 1970: 481–512.

SVIRENKO D.O. 1924. Al'gologiceskie nabljudenija. *Russkij Arkhiv Protistologii* 3: 175–182.
TEILING 1957. Some little known Swedish phytoplankters. *Svensk Botaniska Tidskrift* 51: 207–222.
TELL G. & COUTE A. 1979. Ultrastructure de la paroi cellulaire de *Coelastrum sphaericum* var. *rugulosum* (Thom.) Sodomkova en microscopie èlectronique a balayage. *Rev. Algol., N.S.* 14: 163–168.
TSARENKO P.M. & PETLEVANNY O.A. 2001. Addition to the diversity of algae of Ukraine. *Algologia,* Supplement, 1–130.
VAN DE PEER Y. & DE WACHTER Y. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Applications in the Biosciences* 10: 569–570.
VAN HANNEN E.J., FINK P. & LÜRLING M. 2002. A revised secondary structure model for the internal transcribed spacer 2 of the green algae *Scenedesmus* and *Desmodesmus* and its implication for the phylogeny of these algae. *European Journal of Phycology* 37: 203–208.
VANORMELINGEN P., HEGEWALD E., BRABAND A., KITSCHKE M., FRIEDL T., SABBE K. & VYVERMAN W. 2007. The systematics of a small spineless *Desmodesmus* taxon, *D. costato-granulatus* (Sphaeropleales, Chlorophyceae), based on ITS2rDNA sequence analyses and cell wall morphology. *Journal of Phycology* 43: 378–396.
WEST W. & WEST G.S. 1896. Algae from Central Africa. *Journal of Botany (London)* 35: 377–384.
WILLE N. 1909. Conjugatae und Chlorophyceae. In: *Die natürlichen Pflanzenfamilien* (Ed. by A. Engler & K. Prantl), Nachträge zum 1. Teil. Engelmann, Leipzig. 136 pp.
WOLF M., KRIENITZ L. & HEPPERLE D. 2002. Phylogenetic position of *Actinastrum hantzschii* Lagerheim 1882 (Chlorophyta, Trebouxiophyceae). *Algological Studies* 104: 59–67.
WOLF M., HEPPERLE D. & KRIENITZ L. 2003. On the phylogeny of *Planktospheria* and *Schizochlamydella* (Radiococcaceae, Chlorophyta). *Biologia, Bratislava* 58: 759–765.
WOLF M., ACHTZIGER M., SCHULTZ J., DANDEKAR T. & MÜLLER T. 2005a. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA* 11: 1616–1623.
WOLF M., FRIEDRICH J., DANDEKAR T. & MÜLLER T. 2005b. CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. *Silico Biology* 5: 291–294.
WOLF M., RUDERISCH B., DANDEKAR T., SCHULTZ J. & MÜLLER T. 2008. ProfDistS: (Profile-) Distance based phylogeny on sequence–structure alignments. *Bioinformatics* 24: 2401–2402.
YAMAGISHI T. & AKIYAMA M. (Eds). 1994. *Photomicrographs of the Freshwater Algae* 13. Uchida Rokakuho, 100 pp.

Received 15 July 2009; accepted 2 December 2009
Associate editor: Fabio Rindi

P.6. Internal transcribed spacer 2 (nu ITS2 rRNA) sequence-structure phylogenetics: Towards an automated reconstruction of the green algal tree of life

Authors: M.A. Buchheim, A. Keller*, C. Koetschan, F. Förster, B. Merget, M. Wolf

submitted 2010

*Co-author

Author's Contributions:
M. Wolf and M.A. Buchheim designed and coordinated the study. C. Koetschan calculated the alignments. I, F. Förster and B. Merget performed the phylogenetic analyses. B. Merget developed the Cartoon2Profile algorithm. M.A. Buchheim drafted the manuscript with contributions of me and M. Wolf.

Internal transcribed spacer 2 (nu ITS2 rRNA) sequence-structure phylogenetics: Towards an automated reconstruction of the green algal tree of life[1]

Mark A. Buchheim [2]

Alexander Keller [3]

Christian Koetschan [3]

Frank Förster [3]

Benjamin Merget [3]

Matthias Wolf [3, 4]

[1] Manuscript received _____; revision accepted _____.

[2] Department of Biological Science and the Mervin Bovaird Institute for Molecular Biology and Biotechnology, The University of Tulsa, 800 South Tucker Drive, Tulsa, OK 74104, USA

[3] Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

[4]Corresponding author

ABSTRACT

Sequence data from the nuclear-encoded, internal transcribed spacer 2 (ITS2) obtained from chlorophycean green algae were used to demonstrate the feasibility of an automated analytical approach to DNA barcoding and phylogenetics. Sequences and secondary structures from 591 algae classified as Chlorophyceae, obtained from the ITS2 Database, were synchronously aligned using an ITS2 sequence-structure-specific scoring matrix. Phylogenetic relationships, based on sequences and their secondary structure, were reconstructed by Profile Neighbor-Joining (PNJ), through the use of an ITS2 sequence-structure-specific, General Time Reversible (GTR) substitution model. Bootstrap support for manually pre-defined profiles was estimated based on 100 pseudo-replicates. Despite the fact that the ITS2 region is a relatively short gene fragment (128-483 bases across the Chlorophyta) and is generally characterized as exhibiting high rates of substitution that limit its' utility for broad phylogenetic analysis, results from our analyses of the ITS2 data are not only robust, but remarkably congruent with results from analyses of 18S rRNA, 26S rRNA, *rbc*L and *atp*B data. Given the successful application of a barcoding approach using the ITS2 data for the Chlorophyceae, the exercise was extended to include data from the green algal classes Ulvophyceae (938 sequences) and Trebouxiophyceae (741 sequences). These results offer "proof-of-concept" for the use of ITS2 in barcoding the Viridiplantae and confirm previous assessments which indicated that the ITS2 has the potential to serve as a powerful tool for assessing taxonomic diversity on a grand (i.e., eukaryotic) scale.

INTRODUCTION

Researchers for a host of organisms have turned to DNA barcoding as a powerful, new tool in the study of diversity. Although the literature is replete with cautionary statements regarding DNA barcoding (DeSalle et al. 2005; Ebach and Holdrege 2005; Smith 2005; Wheeler 2005; Will et al. 2005; Holdrege and Ebach 2006), a large number of studies have suggested that the benefits of barcoding either outweigh the problems or that most problems can be addressed (Blaxter 2004; Blaxter et al. 2005; Hebert and Gregory 2005; Savolainen et al. 2005; Chase and Fay 2009a; Engelmann et al. 2009; Hollingsworth et al. 2009a; Jakupciak and Colwell 2009; Seberg and Petersen 2009; Wolf and Schultz 2009).

Much of our own research interests have focused less on the issue of species delimitation but rather more on the phylogenetics of chlorophytan green algae (Buchheim and Chapman 1991, 1992; Buchheim et al. 1996; Buchheim et al. 1997a, 1997b; Buchheim et al. 2001; Buchheim et al. 2002; Wolf et al. 2002b; Wolf et al. 2002a; Hegewald and Wolf 2003; Krienitz et al. 2003; Wolf et al. 2003b; Wolf et al. 2003a; Krienitz et al. 2004; Buchheim et al. 2005; Buchheim et al. 2010). Nonetheless, our own work (Hegewald and Hanagata 2000; Buchheim et al. 2005; Müller et al. 2007; Buchheim et al. 2010; Hegewald et al. 2010) and the work of many others (Bakker et al. 1995; Pillmann et al. 1997; Coat et al. 1998; An et al. 1999; Fabry et al. 1999; Coleman 2001; Lewis and Flechtner 2004) have revealed the utility of the nu ITS2 rRNA (ITS2) gene in studies of closely related green algae. It has become abundantly clear that much of the data gathered in our purely phylogenetics efforts have tremendous potential for validating an approach to DNA barcoding for the Chlorophyta.

Barcoding efforts within the Viridiplantae (green plants) have, as one might expect, largely focused on vascular plants, in general, and flowering plants, in particular

(Chase et al. 2005; Kress et al. 2005; Newmaster et al. 2006; Chase et al. 2007; Kress and Erickson 2007; Fazekas et al. 2008; Ledford 2008; Hollingsworth et al. 2009a; Hollingsworth et al. 2009b; Seberg and Petersen 2009). Genomic targets for potential land plant barcodes have included chloroplast (*rbc*L, *atp*B, *mat*K, *psb*A, *rpo*C1, *rpo*B, *ndh*J, *acc*D), mitochondrial (COX [CO]1) and nuclear genes (various single copy genes, ITS1, ITS2, 5.8S) (Chen et al. 2010). Chen et al. (2010) concluded that many of these potential markers are inappropriate for barcoding due to low variability (e.g., *rpo*B, *ndh*J, *acc*D, *atp*B, COX1, 5.8S rRNA) or suffer from difficulties in amplification (e.g., ITS1 rRNA and nuclear, single copy genes). The chloroplast encoded *mat*K gene (with *rbc*L) has been formally selected as a DNA barcoding candidate for the land plants (CBOL Plant Working Group 2009). However, the absence of *mat*K from all green algae except the charophytes (Lemieux et al. 2000; Turmel et al. 2002; Sanders et al. 2003) renders moot, the question of its utility for the Chlorophyta.

It remains possible that one or more of the problematic genomic targets noted above could be useful for studies of chlorophytan barcoding. However, at present, only the 5.8S rRNA and ITS1 rRNA genes have been studied in more than fifty chlorophytan taxa (3025 GenBank citations). Moreover, if the goal is to identify and test a universal (at least for the Viridiplantae) barcoding candidate, it is important to target only those candidates that will be of use for the land plants. Of those potentially suitable genomic targets that remain, only the cp *rbc*L (2477 current GenBank citations) and nu ITS2 rRNA (3418 current GenBank citations) genes have been routinely targeted for assessing chlorophytan diversity. Investigations of the *rbc*L gene from Chlorophyta have failed to identify a set of universal primers that successfully yield amplicons for all Chlorophyta (Nozaki et al. 1995; Nozaki et al. 1999; Nozaki et al. 2000; Nozaki 2001; Buchheim et al. 2010). Moreover, attempts to obtain *rbc*L data from Cladophoralean green algae (Ulvophyceae) have largely been unsuccessful (only 3 GenBank citations as

of 10/10/2010). Because of the extreme heterogeneity in *rbc*L across the green algae, the *rbc*L is, effectively, a non-universal gene. In contrast, the nu ITS2 gene from virtually all Viridiplantae can be amplified with a single set of universal primers (White et al. 1999). Some have even suggested that the nu ITS2 rRNA may be useful for comparisons within much of the domain Eukarya (Hershkovitz and Lewis 1996; Mai and Coleman 1997; Coleman 2003; Schultz et al. 2005; Coleman 2007). On the basis of the efficiency of amplification, the nu ITS2 rRNA gene is preferable to the cp *rbc*L. In addition, as a nuclear gene, the nu ITS2 rRNA gene is likely to have broader taxonomic applicability (i.e., beyond Viridiplantae) should it be deemed a good DNA barcode.

Many of the limitations first associated with the nu ITS2 rRNA (e.g., too much variation, too few nucleotide sites) have been overcome by secondary structure analysis which has systematically identified regions of variability as well as areas of substantial conservation (Coleman 2003; Schultz et al. 2005; Wolf et al. 2005; Schultz et al. 2006; Coleman 2007; Schultz and Wolf 2009). Furthermore, a simulation study recently confirmed the benefit of a sequence-structure approach (Keller et al. 2010). Analyses of the simulated data resulted in the most robust trees, as assessed by the bootstrap, when secondary structure data were included (Keller et al. 2010). Moreover, the addition of sequence-structure permits the comparison of a much broader phylogenetic spectrum (Keller et al. 2010). Much of the progress in establishing a nu ITS2 rRNA tool for diversity assessment, including its potential use in DNA barcoding, has been accomplished as a consequence of new bioinformatics applications, concepts and resources (Müller et al. 2004; Friedrich et al. 2005; Schultz et al. 2005; Wolf et al. 2005; Rahmann 2006; Schultz et al. 2006; Seibel et al. 2006; Müller et al. 2007; Wolf et al. 2008; Koetschan et al. 2010). In particular, the ITS2 Database III has substantially advanced the effectiveness of phylogenetic analyses using ITS2 data. At present, the ITS2 Database III, mined from the NCBI database, comprises over 250,000 structures

(both partial and complete) that covers the range of eukaryotic diversity (Koetschan et al. 2010). One of the innovations that is coupled with the database is the use of Hidden Markov Models to more fully automate the annotation pipeline (Koetschan et al. 2010). The final stage of the pipeline involves homology-modeling that provides the user with a sequence-structure assessment that is the product of a phylogenetically broad, comparative approach (Koetschan et al. 2010). Given the bioinformatics support coupled with the relative ease of obtaining comparable data, the nu ITS2 rRNA appears to be a superior candidate for use as a DNA barcode for the Chlorophyta.

One goal of this study is to evaluate the use of an automated workflow that includes those analyses suggested by Schultz and Wolf (2009) and that can be accomplished within a reasonable time frame on an ordinary desktop computer. The need for automated procedures without further manual corrections in phylogenetics and species delineation is obvious, as the number of available sequences on public databases grows daily.

The ultimate goal of this investigation is, however, a demonstration of the utility (i.e., proof-of-concept) of the nu ITS2 rRNA as a DNA barcode for the Chlorophyta as tested against phylogenetic assessments based on other markers. The green algal class, Chlorophyceae, in particular, has been the target of a substantial number of phylogenetic investigations in which the nu ITS2 rRNA gene was included as a genomic target (Coleman and Mai 1997; Angeler et al. 1998; An et al. 1999; Angeler et al. 1999; Coleman 1999; Fabry et al. 1999; Schagerl et al. 1999; Hegewald and Hanagata 2000; van Hannen et al. 2000; Cifuentes et al. 2001; Coleman 2001; Gonzàlez et al. 2001; van Hannen et al. 2002; Hegewald and Wolf 2003; Wolf et al. 2003a; Lewis and Flechtner 2004; Pocock et al. 2004; Buchheim et al. 2005; Fawley et al. 2005; McManus and Lewis 2005; Keller et al. 2008; Yamada et al. 2008; Coleman 2009;

Buchheim et al. 2010). These chlorophycean investigations, which represent only a portion of the total body of work in which the nu ITS2 rRNA gene has been used to study chlorophytan diversity (>80 published manuscripts), clearly show the utility of this marker in addressing species level questions. Our challenge is to determine if the use of automated analytical methods with both primary and secondary structural analysis yield robust trees that are largely congruent with other sets of data (e.g., 18S rRNA, 26S rRNA, *rbc*L, *atp*B).

With our early results confirming the utility of ITS2 in a DNA barcoding investigation of the Chlorophyceae, we extended the test to include the whole of the phylum, Chlorophyta. Our test of this approach clearly indicates that the nu ITS2 rRNA data possess considerable power to reconstruct reasonably robust hypotheses that are congruent with past work that employed markers that have been deemed "more conservative" than the nu ITS2 rRNA gene. Our results indicate that the ITS2 gene has the potential to serve as a powerful tool for phylogenetics and DNA barcoding in an extraordinarily broad taxonomic context that may eventually encompass virtually the entirety of the domain Eukarya.

MATERIALS AND METHODS

All phylogenetic analyses followed the procedure outlined in Schultz and Wolf (2009). Data were obtained (2009/09/30) from the ITS2 Database (Schultz et al. 2006; Selig et al. 2008; Koetschan et al. 2010). A global, multiple sequence-structure alignment of all available (591) chlorophycean ITS2 sequences with available secondary structures was generated in 4SALE v1.5 (Seibel et al. 2006; Seibel et al. 2008). Sequences and secondary structures were synchronously aligned, making use of an ITS2 sequence-structure specific scoring matrix (Seibel et al. 2006; Seibel et al.

168 2008). Accordingly, alignments were calculated for the Ulvophyceae (938 sequences)
169 and Trebouxiophyceae (741 sequences). Further, a global Chorophyta tree was
170 calculated that includes all the sequences described above for the individual class-
171 specific trees. For each of the alignments, a set of all *Micromonas* (Prasinophyceae)
172 sequences available in the ITS2 database was used as the outgroup. Based on primary
173 and secondary structure information, phylogenetic relationships were reconstructed by
174 Profile Neighbor-Joining (PNJ) (Müller et al. 2004), through the use of an ITS2
175 sequence-structure-specific, General Time Reversible (GTR) substitution model, in
176 ProfDistS v0.9.8 (Friedrich et al. 2005; Rahmann 2006; Wolf et al. 2008). In addition to
177 the usual Windows/Mac/Linux GUIs, all of the methods described above may be used
178 from a UNIX command line shell and thus be incorporated in any type of automated
179 scripts. The complete procedure of data acquisition, alignment calculation and tree
180 reconstruction took less than one hour of computational time for the three class-specific
181 trees and 3.5 h for the complete Chlorophyta tree on a conventional 2.0 GHz single core
182 computer.

183 In a second manual step we obtained bootstrap support values (Felsenstein,
184 1985) for the major taxonomic clades within the trees. For this step, manual profiles
185 were set in ProfDistS with the Cartoon2Profile tool (http://profdist.bioapps.biozentrum.
186 uni-wuerzburg.de/cgi-bin/index.php?section=cart2prof), after rooting and visualizing
187 the distance trees with FigTree v1.2.3 (Rambaut 2009). Cartoon2Profile is a Perl script
188 that converts cartoons as set in FigTree into a ProfDistS compatible profile file.
189 Cartoon2Profile has been explicitly developed for this study, but may be used for any
190 investigation that uses FigTree and ProfDistS. Calculation of bootstrap values with
191 these profiles required less than 10 minutes of computational time using a desktop
192 computer. We visualized a concatenated topology of the three class-specific trees in a
193 hyperbolic tree based on the HyperGeny tree browser

(http://bioinformatics.psb.ugent.be/hypergeny). The hyperbolic tree is publicly available as a supplement to this study at the ITS2-Database Supplements Page and at http://hypertree.bioapps.biozentrum.uni-wuerzburg.de.

RESULTS

The aligned nu ITS2 rRNA data for the class Chlorophyceae yielded a tree (Fig. 1) that resolved data representing the orders Oedogoniales (*Oedogonium*, *Bulbochaete* and *Oedocladium*), Sphaeropleales (*Desmodesmus*, *Scenedesmus*, *Atractomorpha* and *Sphaeroplea*), and Chlamydomonadales/Volvocales (*Chlamydomonas* [three non-monophyletic clades], *Yamagishiella*, *Pandorina*, *Eudorina*, *Astrephomene*, *Gonium*, *Phacotus* and *Dunaliella*). Two distinct chlamydomonad alliances were resolved (with only weak bootstrap support) by the ITS2 data (Fig. 1). The Sphaeropleales were resolved as monophyletic with high bootstrap support (94%). Furthermore, distinct lineages corresponding to putative chlorophycean species are preserved by the analytical protocol utilized in this experiment (Fig. 1).

Given the success of the experiment with data from the Chlorophyceae, the test was extended to include a comprehensive sampling of nu ITS2 rRNA sequence data from the green algal classes, Trebouxiophyceae (741 sequences) and Ulvophyceae (938 sequences). These data were analyzed under the same analytical conditions as the Chlorophyceae, including the use of prasinophycean data as the outgroup. The PNJ analysis resolved three principal clades of trebouxiophycean taxa (Fig. 2) that correspond to two sets of microthamnialean taxa (the *Trebouxia* alliance [Microthamniales I] and the *Asterochloris* alliance [Microthamniales II) and the Chlorellales which includes *Chlorella*, *Parachlorella*, *Coccomyxa*, *Micractinium* and *Didymogenes*. Bootstrap values for these three clades are 99%, 94% and 96%,

219 respectively. Results of a third PNJ analysis (Fig. 3) revealed high bootstrap support for
220 a Bryopsidales clade (92% bootstrap support; *Halimeda* and *Caulerpa* alliances). A
221 *Urospora*/*Acrosiphonia* clade was resolved with 79% bootstrap support. Neither of the
222 two ulvalean alliances (Ulvales I: *Bolbocoelon, Blidingia, Monostroma, Umbraulva*
223 and one group of *Ulva* taxa; Ulvales II: a second group of *Ulva* taxa) were robustly
224 resolved. However, the Ulvales II clade formed a sister group with the
225 *Urospora*/*Acrosiphonia* alliance with 70% bootstrap support. As with the
226 chlorophycean data (Fig. 1), the trebouxiophycean (Fig. 2) and ulvophycean (Fig. 3)
227 data revealed numerous distinct branches that correspond to putative species.

228 A composite, phylum-level analysis of ITS2 data (Fig. 4) derived from each of
229 the class-level analyses reveals the same major clades for each class of green algae.
230 However, the branching order of some of these clades differs between class-level and
231 phylum-level analyses. The class level analyses, by default, present each class as
232 monophyletic (Figs. 1-3). In contrast, the phylum level analysis challenges, albeit
233 weakly, the monophyly of each of the classes (Fig. 4). For the Chlorophyceae, the
234 Oedogoniales are allied with Ulvales I and Chlorellales III (*Coccomyxa*), a subset of the
235 Sphaeropleales (Sphaeropleales II [Sphaeropleaceae]) are allied with Chlorellales I
236 (*Chlorella*, *Parachlorella*, *Micractinium*, *Didymogenes*, *Diacanthos*, *Closteriopsis*,
237 *Actinastrum*, *Dictyosphaerium*, *Auxenochlorella*, *Lobosphaeropsis*), II
238 (*Pseudochlorella*, *Koliella*), and Microthamniales II (Fig. 4), and Sphaeropleales I
239 (*Desmodesmus* and *Scenedesmus*) is sister to Ulvales I. The Chlamydomonadales are
240 resolved as a monophyletic sister group to the latter alliance (Fig. 4). The
241 Trebouxiophyceae form four distinct, non-monophyletic clades comprising the
242 Microthamniales I, Microthamniales II, Chlorellales III, and Microthamniales II +
243 Chlorellales I + Chlorellales II (Fig. 4). The Ulvophyceae also form four, non-

monophyletic clades comprising the Bryopsidales II (*Caulerpa*), Ulvales + *Urospora*/*Acrosiphonia*, Bryopsidales I (*Halimeda*), and Ulvales I (Fig. 4).

DISCUSSION

The independent analyses for each chlorophytan class generally recover phylogenetic signal that is consistent with studies of 18S rRNA (Buchheim et al. 1990; Buchheim and Chapman 1991, 1992; Buchheim et al. 1996; Friedl 1996; Nakayama et al. 1996a; Nakayama et al. 1996b; Buchheim et al. 1997a, 1997b; Hepperle et al. 2000; Buchheim et al. 2001; Hepperle et al. 2001; Krienitz et al. 2001; Pröschold et al. 2001; Buchheim et al. 2002; Wolf et al. 2002b; Wolf et al. 2002a; Hegewald and Wolf 2003; Krienitz et al. 2003; Wolf et al. 2003b; Wolf et al. 2003a; Krienitz et al. 2004; Lewis and Flechtner 2004; Buchheim et al. 2005; Mei et al. 2007; Nakada and Nozaki 2007; Nakada et al. 2008a; Nakada et al. 2008b; Sluiman et al. 2008; Friedl et al. 2009; Buchheim et al. 2010), 26S rRNA (Buchheim et al. 2001; Buchheim et al. 2002; Leliaert et al. 2003; Buchheim et al. 2005; Mei et al. 2007), *rbc*L (Daugbjerg et al. 1994, 1995; Nozaki et al. 1995; Nozaki et al. 1997b; Nozaki et al. 1997a; Nozaki et al. 1998; Nozaki et al. 1999; Nozaki et al. 2000; Nozaki 2001, 2003; Nozaki et al. 2003; Zechman 2003; Nakazawa et al. 2004; Nozaki et al. 2006; Loughnane et al. 2008; Buchheim et al. 2010) and *atp*B (Nozaki et al. 2000; Nozaki 2001; Nozaki et al. 2003; Buchheim et al. 2010).

Topological differences do exist between results with ITS2 data and other data sets. For example, analyses of the ITS2 data for the Chlorophyceae place the Chlamydomonadales as a basal, paraphyletic assemblage in the class (Fig. 1), whereas, both 18S and 26S rRNA data place the Oedogoniales, Chaetophorales and/or Chaetopeltidales as basal members of the class (Buchheim et al. 2001; Buchheim et al.

2002). However, these differences can be attributed to (1) weak support in one or both sets of data, (2) substantial differences in taxon sampling (e.g., no ITS2 data for Chaetopeltidales or Chaetophorales are available), (3) substantial differences in outgroup rooting, or (4) some combination of these influences. In addition to differences between phylogenetic results from ITS2 and other data sets, differences between results from class-level and phylum-level analyses of ITS2 data were also observed. For example, the class level analysis challenges the monophyly of Chlamydomonadales (Fig. 1), but the phylum level analysis (Fig. 4) resolves the order as monophyletic. Again, these differences are not robust and, thus, can be attributed to weak support, taxon sampling error or both.

These results represent further evidence that the ITS2 data can be aligned for a taxonomically broad set of organisms and that the alignment yields corroborated alliances of chlorophytan taxa. Most importantly, our results confirm that the analytic procedure does not lead to a loss of signal for the resolution of discrete, species level branches. The behavior of the ITS2 in conjunction with the automated, secondary-structure-based alignment compels us to conclude that the ITS2 data offer the best choice for DNA barcoding for the Chlorophyta.

The remarkable results for the ITS2 gene from chlorophytan taxa raise the question: can these data and approaches to DNA barcoding be applied to other organisms? Given that ITS2 data already exist for so many disparate groups of organisms, there is little doubt that this protocol could be easily extended to other members of the domain Eukarya. Recent work, which validates the use of ITS2 in barcoding embryophyte plants and animals, strongly supports this assertion (Yao et al. 2010). As with most tools, there will be situations that may negate the utility of the ITS2 as a DNA barcode. For example, some parasitic taxa have been identified as

possessing substantially shortened ITS2 genes (Edlind et al. 1990). The ability of the analytical method to recover data from shortened sequences has yet to be tested in a broad taxonomic context.

One of the more problematic issues for the use of ITS2 as a DNA barcode is that of heterogeneity. As part of the rDNA array, multiple, homogeneous copies of the ITS2 are presumed to exist within all eukaryotic organisms (ironically, making it an excellent barcode candidate due to greater ease of amplification). An assumption of homogeneity, as a consequence of concerted evolution (Zimmer et al. 1980; Arnheim 1983), may be unrealistic for a number of organisms (Harpke and Peterson 2006), including at least some chlorophytes (Pillmann et al. 1997; Famà et al. 2000). Since heterogeneity of the rDNA array is an issue for the use of ITS2 in an ordinary phylogenetic analysis, the problem is not merely a product of its use in DNA barcoding. Consequently, the same measures for identifying heterogeneity (cloning, mixing of multiple PCR reactions, see also below) can be applied for use in DNA barcoding. Nonetheless, addressing the problem of heterogeneity in the ITS2 clearly burdens the approach with additional time and expense. However, it is our contention that this extra burden is overshadowed by the significant savings in time and effort through the use of the automated analytical pipeline. No other DNA barcoding candidate is similarly equipped for analytical high-throughput. Furthermore, no other potential barcode exhibits the same level of universality (i.e., in primers for PCR) than the ITS2. Thus, the ITS2 meets criterion one of the recommendations for a standard plant barcode (CBOL Plant Working Group 2009). Furthermore, our current assessment of primary and secondary sequence structure among an exhaustive survey of chlorophytan diversity indicates that ITS2 also meets Criteria Two (bi-directional sequencing with few or no ambiguities) and Three (enables the most species to be distinguished) of the CBOL recommendations (CBOL Plant Working Group 2009).

Despite some notable exceptions (Wolf and Schultz 2009; Chen et al. 2010; Gile et al. 2010; Yao et al. 2010), the ITS2 gene has largely been shunned by those investigators that are designing or promoting DNA barcodes for the land plants (Chase et al. 2003; CBOL Plant Working Group 2009; Chase and Fay 2009a, 2009b). Concern about the confounding impact of pseudogenes and the potential presence of intraspecific or intra-individual variation (due to differing rates of homogenization of the rDNA tandem array or due to introgression) were cited as reasons for relegating ITS to, at best, a supporting role in DNA barcoding for the land plants (CBOL Plant Working Group 2009; Chase and Fay 2009a, 2009b). The confounding influence of pseudogenes (from the aberrant secondary structures produced by ITS pseudogenes that have accumulated a substantive number of indels as a consequence of the loss of function of the ITS gene) can be minimized or eliminated by the use of DMSO during the PCR (Chase et al. 2003). In addition, testing for the presence of conserved 5.8S rRNA motifs may be a relatively easy (i.e., amplifying the spacer region to include the 5.8S rRNA adds very little time and investment to an investigation of the ITS2) means of recognizing spacer pseudogenes (Harpke and Peterson 2008). At present, there have been no reports of ITS pseudogenes in the Chlorophyta, but this is likely to change as more chlorophytan taxa are scrutinized.

As was noted above, the issue of heterogeneity within a species or within an individual has the potential to be more problematic than the confounding issue of ITS pseudogenes. Regardless of the source, ITS heterogeneity has been deemed a liability for its use as a DNA barcode for the land plants (Chase and Fay 2009a, 2009b). However, life history differences between most Chlorophyta and the embryophytes may account, at least in part, for the antipathy towards the ITS2. Specifically, many Chlorophyta exhibit zygotic meiosis and, thus, are vegetatively haploid. All embryophytes exhibit sporic meiosis and, thus, are vegetatively diploid. Therefore, the

346 ITS2 in many Chlorophyta behaves more like organellar genes that exhibit uniparental
347 inheritance. Angiosperms will have two copies from each parent, thus doubling the
348 opportunities for introducing heterogeneity. Introgression, which may play a role in the
349 evolutionary history of a significant number of angiosperm taxa, is often cited as the
350 culprit in producing multiple ITS alleles which, in turn, would likely confound a
351 phylogenetic analysis (Chase et al. 2003; Chase and Fay 2009b). Except for some
352 marine macrophytes that may exhibit sporic meiosis (Kapraun 1993, 1994; Kapraun and
353 Buratti 1998; Durand et al. 2002), there seems to be little evidence of introgression
354 (Verbruggen et al. 2005) that could produce ITS2 heterogeneity in the Chlorophyta.
355 Moreover, the positive results from one of the most recent and extensive investigations
356 of ITS2 as a DNA barcode for plants (Yao et al. 2010) suggest that the concerns
357 regarding ITS2 may be overstated.

358 Lastly, we address the issue of pragmatism. As we stated in the Introduction,
359 virtually all of the other candidate genomic targets for DNA barcoding in the
360 Chlorophyta exhibit one or more serious deficiencies. The *rbc*L gene may be able to
361 play a role in DNA barcoding, but a lack of universal primers coupled with numerous
362 difficult or intractable chlorophytan groups negates the use of *rbc*L for the near term.
363 At present, the ITS2 gene is the only viable candidate for immediate use in DNA
364 barcoding for the Chlorophyta. Despite objections to the use of ITS2 for land plants,
365 our tests of the ITS2 data demonstrate that this marker resolves major green algal
366 lineages (some with high bootstrap support). Most importantly, our results dramatically
367 illustrate that ITS2 data from unknown chlorophytan organisms can be plugged into a
368 high resolution tool for taxonomic assessment. If, as we have asserted, the ITS2 gene
369 can serve as a powerful DNA barcode, then this approach has the potential to address
370 some of the most intractable problems in microbial ecology and diversity including

371 analyses of community structure, the paradox of plankton, issues of dispersal and the
372 nature or existence of biogeographical patterns among algal microbes.

373

374 **FIGURE LEGENDS**

375 Fig. 1. PNJ tree (with bootstrap values from 100 replicates) for sequence-structure data
376 from the nu ITS2 rRNA gene for a comprehensive sampling of the class Chlorophyceae.
377 Major taxonomic groups are labelled and highlighted using differential color coding. A
378 high resolution version of this tree is available as a supplemental file.

379 Fig. 2. PNJ tree (with bootstrap values from 100 replicates) for sequence-structure data
380 from the nu ITS2 rRNA gene for a comprehensive sampling of the class
381 Trebouxiophyceae. Major taxonomic groups are labelled and highlighted using
382 differential color coding. A high resolution version of this tree is available as a
383 supplemental file.

384 Fig. 3. PNJ tree (with bootstrap values from 100 replicates) for sequence-structure data
385 from the nu ITS2 rRNA gene for a comprehensive sampling of the class Ulvophyceae.
386 Major taxonomic groups are labelled and highlighted using differential color coding. A
387 high resolution version of this tree is available as a supplemental file.

388 Fig. 4. PNJ tree for sequence-structure data from the nu ITS2 rRNA gene for a
389 comprehensive sampling of the phylum Chlorophyta. Major taxonomic groups are
390 labelled and highlighted using differential color coding.

391

392

393

394

LITERATURE CITED

An SS, Friedl T, Hegewald E (1999) Phylogenetic relationships of *Scenedesmus* and *Scenedesmus*-like coccoid green algae as inferred from ITS-2 rDNA sequence comparisons. Plant Biology 1(4): 418-428.

Angeler DG, Schagerl M, Coleman AW (1998) Does the infraspecific disjunct occurrence of the carotenoid loroxanthin reflect phylogenetical relationships within syngens of *Pandorina morum* (Volvocales, Chlorophyta)? Biologia 53(4): 567-575.

Angeler DG, Schagerl M, Coleman AW (1999) Phylogenetic relationships among isolates of *Eudorina* species (Volvocales, Chlorophyta) inferred from molecular and biochemical data. Journal of Phycology 35(4): 815-823.

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn M, editors. Evolution of Genes and Proteins. Sunderland, MA: Sinauer Associates. pp. 38-61.

Bakker FT, Olsen JL, Stam WT (1995) Evolution of nuclear rDNA ITS sequences in the Cladophora albida/sericea clade (Chlorophyta). Journal of Molecular Evolution 40(6): 640-651.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C et al. (2005) Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society of London B Biological Sciences 360(1462): 1935-1943.

Blaxter ML (2004) The promise of a DNA taxonomy. Philosophical Transactions of the Royal Society of London B Biological Sciences 359(1444): 669-679.

Buchheim MA, Chapman RL (1991) Phylogeny of the colonial green flagellates: A study of 18S and 26S ribosomal RNA sequence data. BioSystems 25(1-2): 85-100.

421 Buchheim MA, Chapman RL (1992) Phylogeny of *Carteria* (Chlorophyceae) inferred
422 from molecular and organismal data. Journal of Phycology 28(3): 362-374.
423 Buchheim MA, Buchheim JA, Chapman RL (1997a) Phylogeny of the VLE-14
424 *Chlamydomonas* (Chlorophyceae) group. A study of 18S rRNA gene sequences.
425 Journal of Phycology 33(6): 1024-1030.
426 Buchheim MA, Buchheim JA, Chapman RL (1997b) Phylogeny of *Chloromonas*
427 (Chlorophyceae): A study of 18S ribosomal RNA gene sequences. Journal of
428 Phycology 33(2): 286-293.
429 Buchheim MA, Michalopulos EA, Buchheim JA (2001) Phylogeny of the
430 Chlorophyceae with special reference to the Sphaeropleales: A study of 18S and
431 26S rDNA data. Journal of Phycology 37(5): 819-835.
432 Buchheim MA, Turmel M, Zimmer EA, Chapman RL (1990) Phylogeny of
433 *Chlamydomonas* (Chlorophyta) based on cladistic analysis of nuclear 18S
434 ribosomal RNA sequence data. Journal of Phycology 26(4): 689-699.
435 Buchheim MA, Buchheim JA, Carlson T, Kugrens P (2002) Phylogeny of
436 *Lobocharacium* (Chlorophyceae) and allies: A study of 18S and 26S rDNA data.
437 Journal of Phycology 38(2): 376-383.
438 Buchheim MA, Kirkwood A, Buchheim JA, Verghese B, Henley WJ (2010)
439 Hypersaline soil supports a diverse community of *Dunaliella* (Chlorophyceae).
440 Journal of Phycology 46: 1038-1047.
441 Buchheim MA, Lemieux C, Otis C, Gutell RR, Chapman RL et al. (1996) Phylogeny of
442 the Chlamydomonadales (Chlorophyceae): A comparison of ribosomal RNA
443 gene sequences from the nucleus and the chloroplast. Molecular Phylogenetics
444 and Evolution 5(2): 391-402.

445 Buchheim MA, Buchheim JA, Carlson T, Braband A, Hepperle D et al. (2005)
446 Phylogeny of the Hydrodictyaceae (Chlorophyceae): Inferences from rDNA
447 data. Journal of Phycology 41(5): 1039-1054.
448 CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the
449 National Academy of Sciences of the United States of America 106: 12794-
450 12797.
451 Chase MW, Fay MF (2009a) Barcoding of plants and fungi. Science 325(5941): 682-
452 683.
453 Chase MW, Fay MF (2009b) Response to J. Schultz and M. Wolf's E-Letter. Science
454 Online.
455 Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP et al. (2005) Land
456 plants and DNA barcodes: short-term and long-term goals. Philosophical
457 Transactions of the Royal Society of London B Biological Sciences 360(1462):
458 1889-1895.
459 Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y et al. (2003) Molecular
460 systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae).
461 Annals of Botany (London) 92: 107-127.
462 Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinan S et al. (2007) A
463 proposal for a standardised protocol to barcode all land plants. Taxon 56(2):
464 295-299.
465 Chen S, Yao H, Han J, Liu C, Song J et al. (2010) Validation of the ITS2 Region as a
466 Novel DNA Barcode for Identifying Medicinal Plant Species. PLoS ONE 5(1):
467 e8613.
468 Cifuentes AS, Gonzàlez MA, Inostroza I, Aguilera A (2001) Reappraisal of
469 physiological attributes of nine strains of *Dunaliella* (Chlorophyceae): Growth

and pigment content across a salinity gradient. Journal of Phycology 37(2): 334-344.

Coat G, Dion P, Noailles MC, De Reviers B, Fontaine JM et al. (1998) *Ulva armoricana* (Ulvales, Chlorophyta) from the coasts of Brittany (France): II. Nuclear rDNA ITS sequence analysis. European Journal of Phycology 33(1): 81-86.

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(24): 13892-13897.

Coleman AW (2001) Biogeography and speciation in the *Pandorina*/*Volvulina* (Chlorophyta) superclade. Journal of Phycology 37(5): 836-851.

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends in Genetics 19(7): 370-375.

Coleman AW (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. Nucleic Acids Research 35(10): 3322-3329.

Coleman AW (2009) Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. Molecular Phylogenetics and Evolution 50(1): 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(2): 168-177.

Daugbjerg N, Moestrup O, Arctander P (1994) Phylogeny of the genus *Pyramimonas* (Prasinophyceae, Chlorophyta) inferred from the *rbc*L gene. Journal of Phycology 30(6): 991-999.

494 Daugbjerg N, Moestrup O, Arctander P (1995) Phylogeny of genera of Prasinophyceae
495 and Pedinophyceae (Chlorophyta) deduced from molecular analysis of the *rbc*L
496 gene. Phycological Research 43(4): 203-213.
497 DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species
498 delimitation and DNA barcoding. Philosophical Transactions of the Royal
499 Society of London B Biological Sciences 360(1462): 1905-1916.
500 Durand C, Manuel M, Boudouresque CF, Meinesz A, Verlaque M et al. (2002)
501 Molecular data suggest a hybrid origin for the invasive *Caulerpa recemosa*
502 (Caulerpales, Chlorophyta) in the Mediterranean Sea. Journal of Evolutionary
503 Biology 15: 122-133.
504 Ebach MC, Holdrege C (2005) More Taxonomy, Not DNA Barcoding. BioScience
505 55(10): 823-824.
506 Edlind TD, Sharetsky C, Cha ME (1990) Ribosomal RNA of the primitive eukaryote
507 *Giardia lamblia*: large subunit domain I and potential processing signals. Gene
508 96: 289-293.
509 Engelmann JC, Rahmann S, Wolf M, Schultz J, Fritzilas E et al. (2009) Modeling cross-
510 hybridization on phylogenetic DNA microarrays increases the detection power
511 of closely related species. Molecular Ecology Resources 9: 83-93.
512 Fabry S, Koehler A, Coleman AW (1999) Intraspecies analysis: Comparison of ITS
513 sequence data and gene intron sequence data with breeding data for a worldwide
514 collection of Gonium pectorale. Journal of Molecular Evolution 48(1): 94-101.
515 Famà P, Olsen JL, Stam WT, Procaccini G (2000) High levels of intra- and inter-
516 individual polymorphism in the rDNA ITS1 of *Caulerpa recemosa*
517 (Chlorophyta). European Journal of Phycology 35: 349-356.

518 Fawley MW, Fawley KP, Owen HA (2005) Diversity and ecology of small coccoid
519 green algae from Lake Itasca, Minnesota, USA, including Meyerella
520 planktonica, gen. et sp nov. Phycologia 44(1): 35-48.
521 Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG et al. (2008)
522 Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant
523 Species Equally Well. PLoS ONE 3(7): e2802.
524 Friedl T (1996) Evolution of the polyphyletic genus *Pleurastrum* (Chlorophyta):
525 Inferences from nuclear-encoded ribosomal DNA sequences and motile cell
526 ultrastructure. Phycologia 35(5): 456-469.
527 Friedl T, Proeschold T, Lewis LA, Letsch MR (2009) Symbiosis in green algae: Origin
528 and diversity of a successful life style. Phycologia 48(4, Suppl. S): 31-32.
529 Friedrich J, Dandekar T, Wolf M, Mueller T (2005) ProfDist: a tool for the construction
530 of large phylogenetic trees based on profile distances. Bioinformatics 21(9):
531 2108-2109.
532 Gile GH, Stern RF, James ER, Keeling PJ (2010) DNA Barcoding of the
533 Chlorarachniophytes using nucleomorph ITS sequences. Journal of Phycology
534 46: 743-750.
535 Gonzàlez MA, Coleman AW, Gomez PI, Montoya R (2001) Phylogenetic relationship
536 among various strains of *Dunaliella* (Chlorophyceae) based on nuclear ITS
537 rDNA sequences. Journal of Phycology 37(4): 604-611.
538 Harpke D, Peterson A (2006) Non-concerted ITS evolution in *Mammillaria*
539 (Cactaceae). Molecular Phylogenetics and Evolution 41: 579-593.
540 Harpke D, Peterson A (2008) 5.8S motifs for the identification of pseudogenic ITS
541 regions. Botany 86: 300-305.
542 Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy.
543 Systematic Biology 54(5): 852-859.

544 Hegewald E, Hanagata N (2000) Phylogenetic studies on Scenedesmaceae
545 (Chlorophyta). Archiv für Hydrobiologie Supplement 136: 29-49.
546 Hegewald E, Wolf M (2003) Phylogenetic relationships of *Scenedesmus* and
547 *Acutodesmus* (Chlorophyta, Chlorophyceae) as inferred from 18S rDNA and
548 ITS-2 sequence comparisons. Plant Systematics and Evolution 241(3-4): 185-
549 191.
550 Hegewald E, Wolf M, Keller A, Friedl T, Krienitz L (2010) ITS2 sequence-structure
551 phylogeny in the Scenedesmaceae with special reference to *Coelastrum*
552 (Chlorophyta, Chlorophyceae), including the new genera *Comasiella* and
553 *Pectinodesmus*. Phycologia 49(4): 325-335.
554 Hepperle D, Hegewald E, Krienitz L (2000) Phylogenetic position of the Oocystaceae
555 (Chlorophyta). Journal of Phycology 36(3): 590-595.
556 Hepperle D, Krienitz L, Hollibaugh T (2001) Molecular phylogeny of picoplanktonic
557 chlorophytes and the discovery of a new evolutionary lineage. Phycologia 40(4
558 Supplement): 36.
559 Hershkovitz MA, Lewis LA (1996) Deep-level diagnostic value of the rDNA-ITS
560 region. Molecular Biology and Evolution 13(9): 1276-1295.
561 Holdrege C, Ebach MC (2006) Response from Holdrege and Ebach: What about Taxa?
562 Bioscience 56(2): 93.
563 Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT et al. (2009a)
564 Selecting barcoding loci for plants: evaluation of seven candidate loci with
565 species-level sampling in three divergent groups of land plants. Molecular
566 Ecology Resources 9(2): 439-457.
567 Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S et al.
568 (2009b) A DNA barcode for land plants. Proceedings of the National Academy
569 of Sciences of the United States of America 106(31): 12794-12797.

570 Jakupciak JP, Colwell RR (2009) Biological agent detection technologies. Molecular
571 Ecology Resources 9(Suppl. 1): 51-57.
572 Kapraun DF (1993) Karyology of marine green algae. Phycologia 32: 1-21.
573 Kapraun DF (1994) Cytophotometric estimation of nuclear DNA contents in thirteen
574 species of the Caulerpales (Chlorophyta). Cryptogamic Botany 4: 410-418.
575 Kapraun DF, Buratti JR (1998) Evolution of genome size in the Dasycladales
576 (Chlorophyta) as determined by DAPI cytophotometry. Phycologia 37: 176-183.
577 Keller A, Förster F, Müller T, Dandekar T, Schultz J et al. (2010) Including RNA
578 secondary structures improves accuracy and robustness in reconstruction of
579 phylogenetic trees. Biology Direct 5: 4.
580 Keller A, Schleicher T, Förster F, Ruderisch B, Dandekar T et al. (2008) ITS2 data
581 corroborate a monophyletic chlorophycean DO-group (Sphaeropleales). BMC
582 Evolutionary Biology 8: 218.
583 Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B et al. (2010) The ITS2
584 Database III - sequences and structures for phylogeny. Nucleic Acids Research
585 38: D275-279.
586 Kress WJ, Erickson DL (2007) A Two-Locus Global DNA Barcode for Land Plants:
587 The Coding *rbc*L Gene Complements the Non-Coding *trn*H-*psb*A Spacer
588 Region. PLoS One 2(6): Article No.: e508.
589 Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA
590 barcodes to identify flowering plants. Proceedings of the National Academy of
591 Sciences of the United States of America 102: 8369-8374.
592 Krienitz L, Ustinova I, Friedl T, Huss VAR (2001) Traditional generic concepts versus
593 18S rRNA gene phylogeny in the green algal family Selenastraceae
594 (Chlorophyceae, Chlorophyta). Journal of Phycology 37(5): 852-865.

595 Krienitz L, Hegewald E, Hepperle D, Wolf M (2003) The systematics of coccoid green
596 algae: 18S rRNA gene sequence data versus morphology. Biologia 58(4): 437-
597 446.
598 Krienitz L, Hegewald EH, Hepperle D, Huss VAR, Rohr T et al. (2004) Phylogenetic
599 relationship of *Chlorella* and *Parachlorella gen. nov*. (Chlorophyta,
600 Trebouxiophyceae). Phycologia 43(5): 529-542.
601 Ledford H (2008) Botanical identities: DNA barcoding for plants comes a step closer.
602 Nature 451: 616.
603 Leliaert F, Rousseau F, de Reviers B, Coppejans E (2003) Phylogeny of the
604 Cladophorophyceae (Chlorophyta) inferred from partial LSU rRNA gene
605 sequences: Is the recognition of a separate order Siphonocladales justified?
606 European Journal of Phycology 38(3): 233-246.
607 Lemieux C, Otis C, Turmel M (2000) Ancestral chloroplast genome in *Mesostigma*
608 *viride* reveals an early branch of green plant evolution. Nature 403: 649-652.
609 Lewis LA, Flechtner VR (2004) Cryptic species of *Scenedesmus* (Chlorophyta) from
610 desert soil communities of Western North America. Journal of Phycology 40(6):
611 1127-1137.
612 Loughnane CJ, McIvor LM, Rindi F, Stengel DB, Guiry MD (2008) Morphology, *rbc*L
613 phylogeny and distribution of distromatic *Ulva* (Ulvophyceae, Chlorophyta) in
614 Ireland and southern Britain. Phycologia 47(4): 416-429.
615 Mai JC, Coleman AW (1997) The internal transcribed spacer 2 exhibits a common
616 secondary structure in green algae and flowering plants. Journal of Molecular
617 Evolution 44(3): 258-271.
618 McManus HA, Lewis LA (2005) Molecular phylogenetics, morphological variation and
619 colony-form evolution in the family Hydrodictyaceae (Sphaeropleales,
620 Chlorophyta). Phycologia 44(6): 582-595.

621 Mei H, Liu G-X, Hu Z-Y (2007) Phylogenetic studies of Oedogoniales (Chlorophyceae,
622 Chlorophyta) based on 28S rDNA sequences. Acta Hydrobiologica Sinica 31(4):
623 492-498.
624 Müller T, Rahmann S, Dandekar T, Wolf M (2004) Accurate and robust phylogeny
625 estimation based on profile distances: a study of the Chlorophyceae
626 (Chlorophyta). BMC Evolutionary Biology 4: 20.
627 Müller T, Philippi N, Dandekar T, Schultz J, Wolf M (2007) Distinguishing species.
628 RNA 13(9): 1469-1472.
629 Nakada T, Nozaki H (2007) Re-evaluation of three *Chlorogonium* (Volvocales,
630 Chlorophyceae) species based on 18S ribosomal RNA gene phylogeny.
631 European Journal of Phycology 42(2): 177-182.
632 Nakada T, Misawa K, Nozaki H (2008a) Molecular systematics of Volvocales
633 (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic
634 analyses. Molecular Phylogenetics and Evolution 48(1): 281-291.
635 Nakada T, Nozaki H, Proeschold T (2008b) Molecular phylogeny, ultrastructure, and
636 taxonomic revision of Chlorogonium (Chlorophyta): Emendation of
637 Chlorogonium and description of Gungnir gen. nov and Rusalka gen. nov.
638 Journal of Phycology 44(3): 751-760.
639 Nakayama T, Watanabe S, Inouye I (1996a) Phylogeny of wall-less green flagellates
640 inferred from 18S rDNA sequence data. Phycological Research 44(3): 151-161.
641 Nakayama T, Watanabe S, Mitsui K, Uchida H, Inouye I (1996b) The phylogenetic
642 relationship between the Chlamydomonadales and Chlorococcales inferred from
643 18S rDNA sequence data. Phycological Research 44(1): 47-55.
644 Nakazawa A, Yamada T, Nozaki H (2004) Taxonomic study of *Asterococcus*
645 (Chlorophyceae) based on comparative morphology and *rbc*L gene sequences.
646 Phycologia 43(6): 711-721.

647 Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants:
648 evaluation of *rbc*L in a multigene tiered approach. Canadian Journal of Botany
649 84(3): 335-341.
650 Nozaki H (2001) Chloroplast multigene phylogeny and systematics of the advanced
651 genera of the Volvocaceae (Chlorophyceae). Phycologia 40(4 Supplement): 37.
652 Nozaki H (2003) Origin and evolution of the genera *Pleodorina* and *Volvox*
653 (Volvocales). Biologia 58(4): 425-431.
654 Nozaki H, Misumi O, Kuroiwa T (2003) Phylogeny of the quadriflagellate Volvocales
655 (Chlorophyceae) based on chloroplast multigene sequences. Molecular
656 Phylogenetics and Evolution 29(1): 58-66.
657 Nozaki H, Ott FD, Coleman AW (2006) Morphology, molecular phylogeny and
658 taxonomy of two new species of *Pleodorina* (Volvoceae, Chlorophyceae).
659 Journal of Phycology 42(5): 1072-1080.
660 Nozaki H, Ohta N, Morita E, Watanabe MM (1998) Toward a natural system of species
661 in *Chlorogonium* (Volvocales, Chlorophyta): A combined analysis of
662 morphological and *rbc*L gene sequence data. Journal of Phycology 34(6): 1024-
663 1037.
664 Nozaki H, Ohta N, Takano H, Watanabe MM (1999) Reexamination of phylogenetic
665 relationships within the colonial Volvocales (Chlorophyta): An analysis of *atp*B
666 and *rbc*L gene sequences. Journal of Phycology 35(1): 104-112.
667 Nozaki H, Ito M, Watanabe MM, Takano H, Kuroiwa T (1997a) Phylogenetic analysis
668 of morphological species of *Carteria* (Volvocales, Chlorophyta) based on *rbc*L
669 gene sequences. Journal of Phycology 33(5): 864-867.
670 Nozaki H, Motomi I, Ryosuke S, Hidenobu U, Makoto MW et al. (1995) Phylogenetic
671 relationships within the colonial Volvocales (Chlorophyta) inferred from *rbc*L
672 gene sequence data. Journal of Phycology 31(6): 970-979.

673 Nozaki H, Misawa K, Kajita T, Kato M, Nohara S et al. (2000) Origin and evolution of
674 the colonial Volvocales (Chlorophyceae) as inferred from multiple, chloroplast
675 gene sequences. Molecular Phylogenetics and Evolution 17(2): 256-268.
676 Nozaki H, Ito M, Sano R, Uchida H, Watanabe MM et al. (1997b) Phylogenetic
677 analysis of *Yamagishiella* and *Platydorina* (Volvocaceae, Chlorophyta) based on
678 *rbc*L gene sequences. Journal of Phycology 33(2): 272-278.
679 Pillmann A, Woolcott GW, Olsen JL, Stam WT, King RJ (1997) Inter- and intraspecific
680 genetic variation in *Caulerpa* (Chlorophyta) based on nuclear rDNA ITS
681 sequences. European Journal of Phycology 32(4): 379-386.
682 Pocock T, Lachance M-A, Pröschold T, Priscu JC, Kim SS et al. (2004) Identification
683 of a psychrophilic green alga from Lake Bonney Antarctica: *Chlamydomonas*
684 *raudensis* Ettl. (UWO 241) Chlorophyceae. Journal of Phycology 40(6): 1138-
685 1148.
686 Pröschold T, Marin B, Schloesser UG, Melkonian M (2001) Molecular phylogeny and
687 taxonomic revision of *Chlamydomonas* (chlorophyta). I. Emendation of
688 *Chlamydomonas* Ehrenberg and *Chloromonas* Gobi, and description of
689 *Oogamochlamys gen. nov*. and *Lobochlamys gen. nov*. Protist 152(4): 265-300.
690 Rahmann S, Müller, T, Dandekar, T, Wolf, M (2006) Efficient and robust analysis of
691 large phylogenetic datasets. . In: Hsu H-H, editor. Advanced Data Mining
692 Technologies in Bioinformatics. Hershey, PA: Idea Group, Inc. pp. 104-117.
693 Rambaut A (2009) FigTree v1.2.3. Program distributed by the author.
694 Sanders ER, Karol KG, McCourt RM (2003) Occurrence of *mat*K in a *trn*K group II
695 intron in Charophyte green algae and phylogeny of the Characeae. American
696 Journal of Botany 90(4): 628-633.
697 Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing
698 the encyclopaedia of life: an introduction to DNA barcoding. Philosophical

699 Transactions of the Royal Society of London B Biological Sciences 360(1462):
700 1805-1811.
701 Schagerl M, Angeler DG, Coleman AW (1999) Infraspecific phylogeny of Pandorina
702 morum (Volvocales, Chlorophyta) inferred from molecular, biochemical and
703 traditional data. European Journal of Phycology 34(1): 87-93.
704 Schultz J, Wolf M (2009) ITS2 sequence-structure analysis in phylogenetics: A how-to
705 manual for molecular systematics. Molecular Phylogenetics and Evolution
706 52(2): 520-523.
707 Schultz J, Maisel S, Gerlach D, Müller T, Wolf M (2005) A common core of secondary
708 structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota.
709 RNA 11(4): 361-364.
710 Schultz J, Müller T, Achtziger M, Seibel PN, Dandekar T et al. (2006) The internal
711 transcribed spacer 2 database - a web server for (not only) low level
712 phylogenetic analyses. Nucleic Acids Research 34: W704-W707.
713 Seberg O, Petersen G (2009) How many loci does it take to DNA barcode a Crocus?
714 PLoS ONE 4(2): e4598.
715 Seibel P, Muller T, Dandekar T, Wolf M (2008) Synchronous visual analysis and
716 editing of RNA sequence and secondary structure alignments using 4SALE.
717 BMC Research Notes 1: 91.
718 Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M (2006) 4SALE - A tool for
719 synchronous RNA sequence and secondary structure alignment and editing.
720 BMC Bioinformatics 7: 498.
721 Selig C, Wolf M, Müller T, Dandekar T, Schultz J (2008) The ITS2 Database II:
722 homology modeling RNA structure for molecular systematics. Nucleic Acids
723 Research 36: D377 - 380.

724 Sluiman HJ, Guihal C, Mudimu O (2008) Assessing phylogenetic affinities and species
725 delimitations in Klebsormidiales (Streptophyta): Nuclear-encoded rDNA
726 phylogenies and its secondary structure models in *Klebsormidium*, *Hormidiella*,
727 and *Entransia*. Journal of Phycology 44(1): 183-195.
728 Smith VS (2005) DNA barcoding: Perspectives from a "Partnerships for Enhancing
729 Expertise in Taxonomy" (PEET) debate. Systematic Biology 54(5): 841-844.
730 Turmel M, Otis C, De Cambiaire J-C, Pombert J-F, Lemieux C (2002) The complete
731 chloroplast DNA sequence of *Chlorokybus atmophyticus*: evidence that
732 charophycean green algae from an early-diverging lineage adapted to terrestrial
733 life. Botany 2002. Madison, WI.
734 van Hannen EJ, Lurling M, van Donk E (2000) Sequence analysis of the ITS-2 region:
735 A tool to identify strains of *Scenedesmus* (Chlorophyceae). Journal of Phycology
736 36(3): 605-607.
737 van Hannen EJ, Fink P, Lurling M (2002) A revised secondary structure model for the
738 internal transcribed spacer 2 of the green algae *Scenedesmus* and *Desmodesmus*
739 and its implication for the phylogeny of these algae. European Journal of
740 Phycology 37(2): 203-208.
741 Verbruggen H, De Clerck O, Schils T, Kooistra WHCF, Coppejans E (2005) Evolution
742 and phylogeography of *Halimeda* section *Halimeda* (Bryopsidales,
743 Chlorophyta). Molecular Phylogenetics and Evolution 37(3): 789-803.
744 Wheeler QD (2005) Losing the plot: DNA "barcodes" and taxonomy. Cladistics 21(4):
745 405-407.
746 White TJ, Bruns T, Lee S, Taylor J (1999) Amplification and direct sequencing of
747 fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH,
748 Sninsky JJ, White TJ, editors. PCR Protocols. San Diego: Academic Press. pp.
749 315-322.

750 Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need
751 for integrative taxonomy. Systematic Biology 54(5).
752 Wolf M, Schultz J (2009) ITS better than its reputation. Science (E-Letter, 10 December
753 2009)
754 Wolf M, Krienitz L, Hepperle D (2002a) Phylogenetic position of *Actinastrum*
755 *hantzschii* Lagerheim 1882 (Chlorophyta, Trebouxiophyceae). Algological
756 Studies 104: 59-67.
757 Wolf M, Hepperle D, Krienitz L (2003a) On the phylogeny of *Radiococcus*,
758 *Planktosphaeria* and *Schizochlamydella* (Radiococcaceae, Chlorophyta).
759 Biologia 58(4): 759-765.
760 Wolf M, Hegewald E, Hepperle D, Krienitz L (2003b) Phylogenetic position of the
761 Golenkiniaceae (Chlorophyta) as inferred from 18S rDNA sequence data.
762 Biologia 58(4): 433-436.
763 Wolf M, Buchheim M, Hegewald E, Krienitz L, Hepperle D (2002b) Phylogenetic
764 position of the Sphaeropleaceae (Chlorophyta). Plant Systematics and Evolution
765 230(3-4): 161-171.
766 Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T (2005) Homology modeling
767 revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary
768 structures. RNA 11(11): 1616-1623.
769 Wolf M, Ruderisch B, Dandekar T, Schultz J, Müller T (2008) ProfDistS: (profile-)
770 distance based phylogeny on sequence-structure alignments. Bioinformatics
771 24(20): 2401-2402.
772 Yamada TK, Miyaji K, Nozaki H (2008) A taxonomic study of *Eudorina unicocca*
773 (Volvocaceae, Chlorophyceae) and related species, based on morphology and
774 molecular phylogeny. European Journal of Phycology 43(3): 317-326.

775 Yao H, Song J, Liu C, Luo K, Han J et al. (2010) Use of ITS2 region as the universal
776 DNA barcode for plants and animals. PLoS One 5(10): 1-9.
777 Zechman FW (2003) Phylogeny of the Dasycladales (Chlorophyta, Ulvophyceae) based
778 on analyses of RUBISCO large subunit (*rbc*L) gene sequences. Journal of
779 Phycology 39(4): 819-827.
780 Zimmer EA, Martin SL, Beverly SM, Kan YW, Wilson AC (1980) Rapid duplication
781 and loss of genes coding for the α chains of hemoglobin. Proceedings of the
782 National Academy of Sciences of the United States of America 77: 2518-2162.
783
784

Figure 1
Click here to download Figure: f01-chloros-RGBrevised.eps

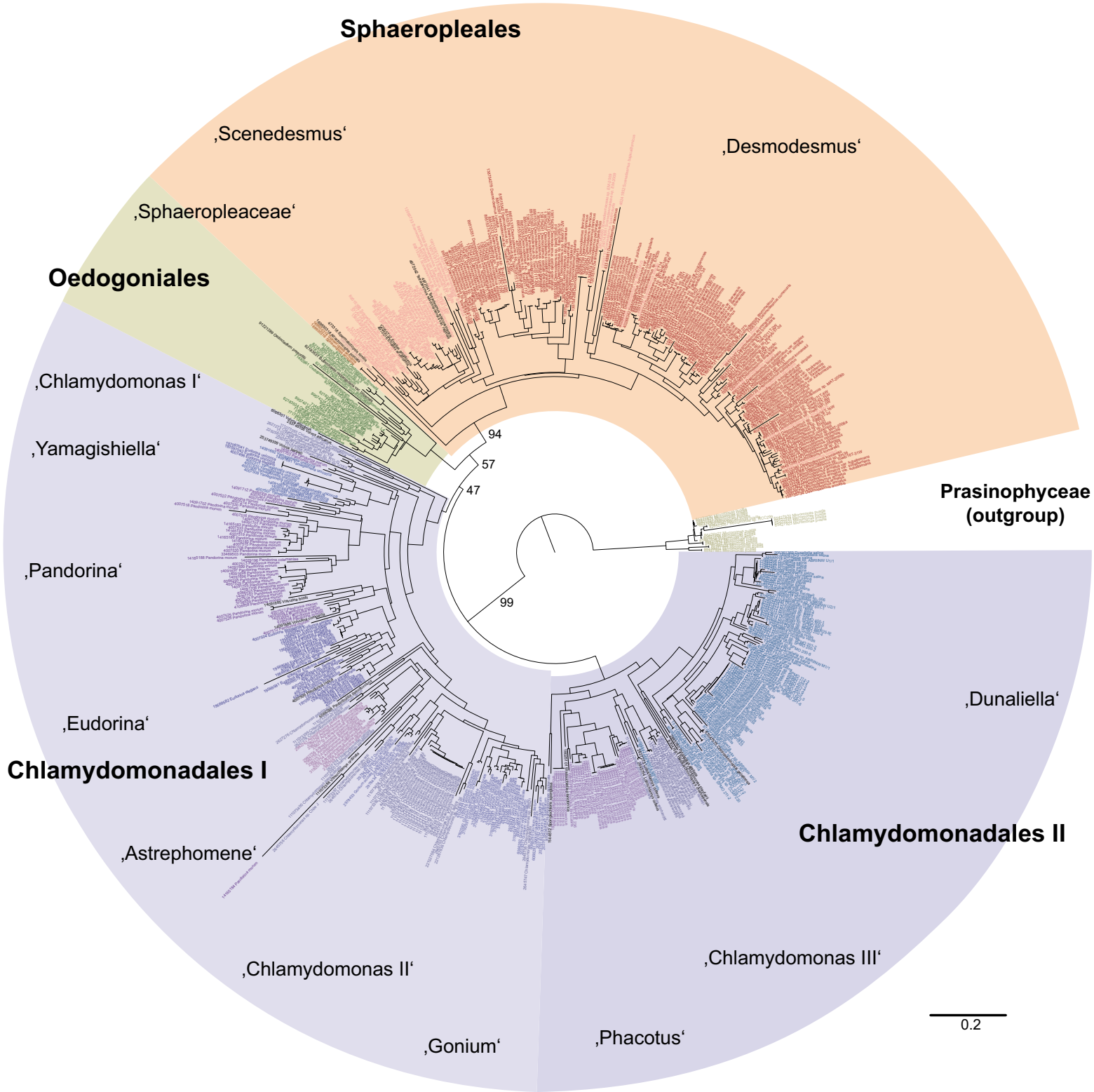


Figure 2
Click here to download Figure: f02-treboux_RGBv3.eps

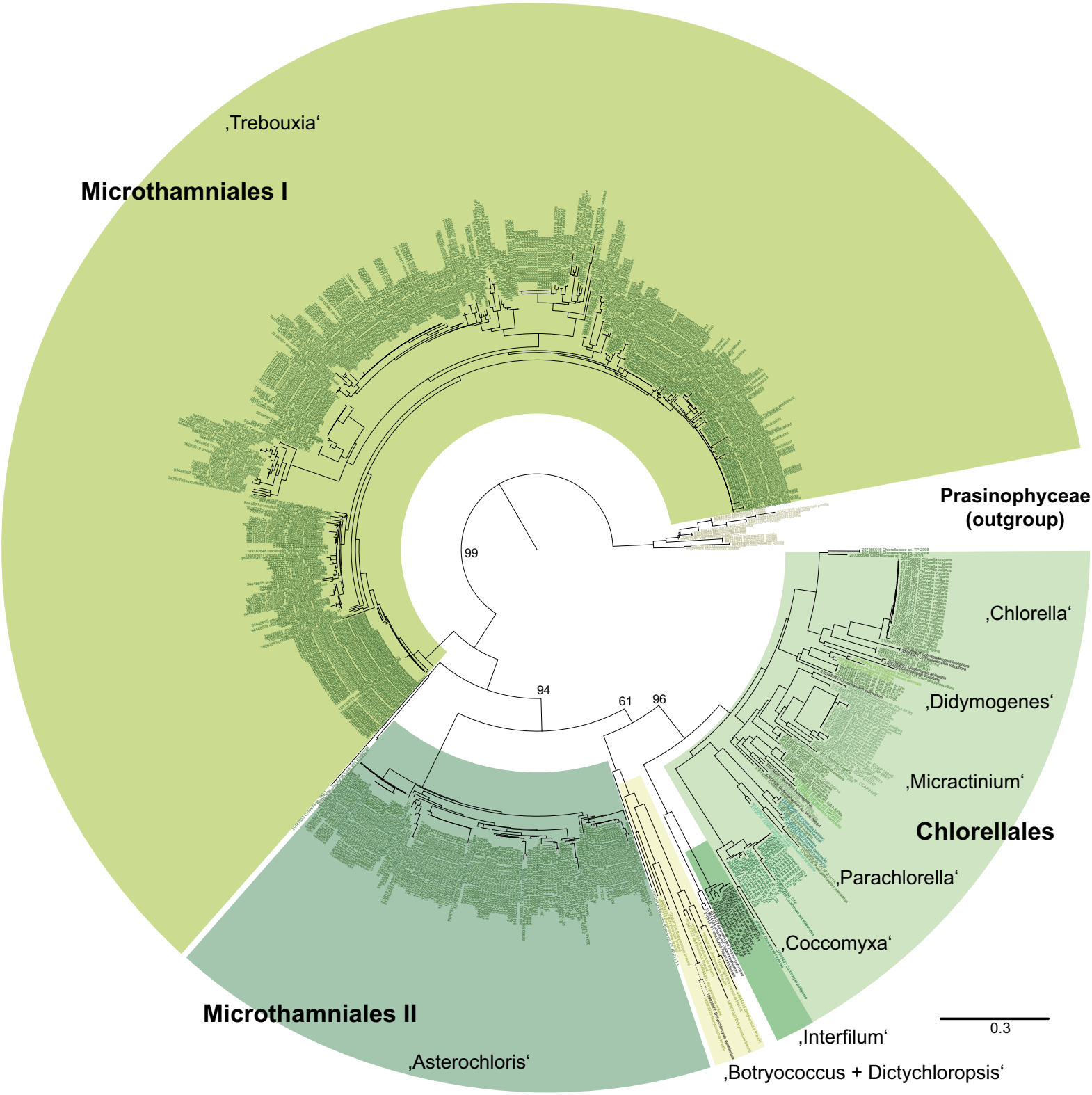


Figure 3
Click here to download Figure: f03-ulva_RGBv3.eps

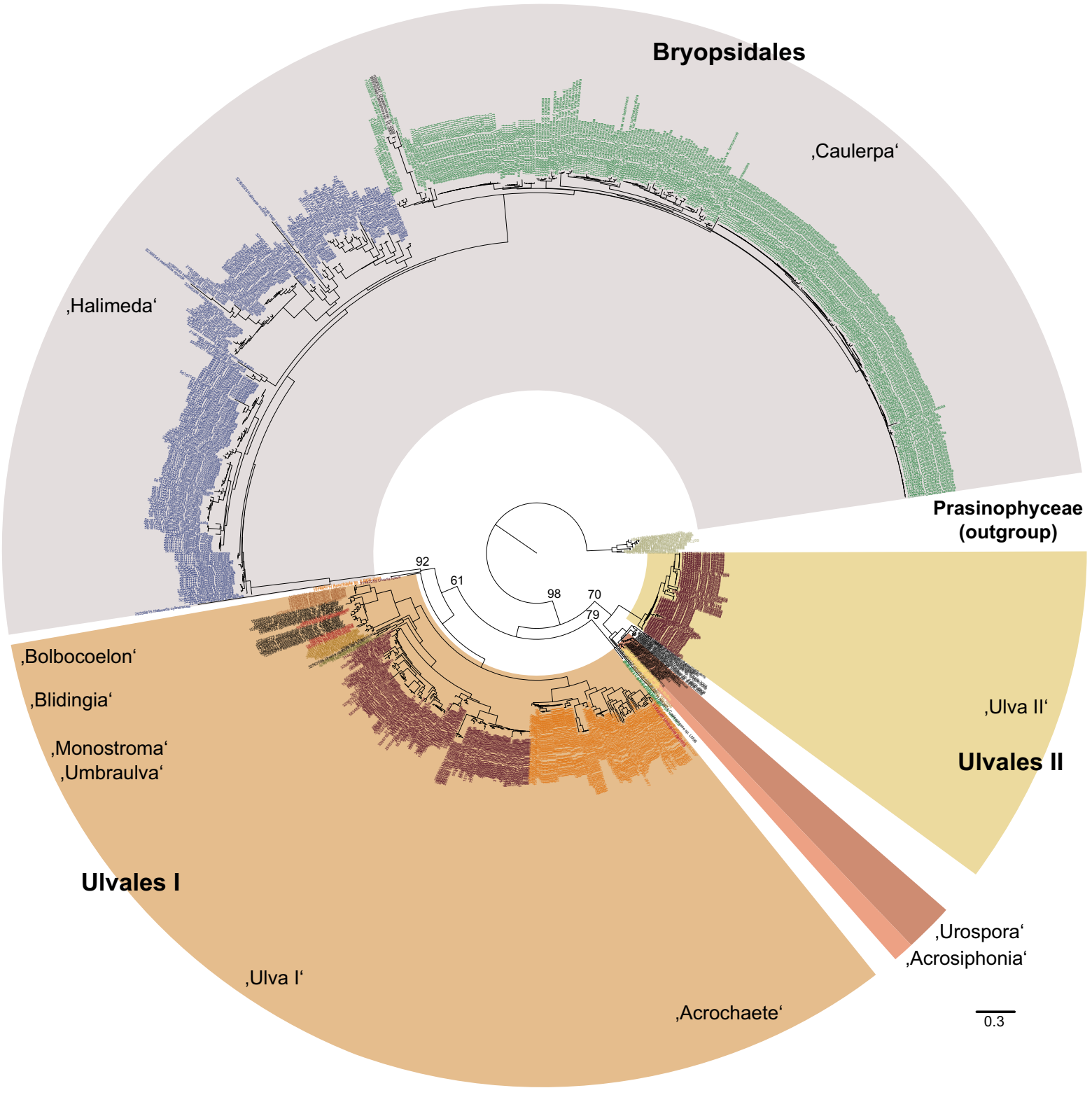
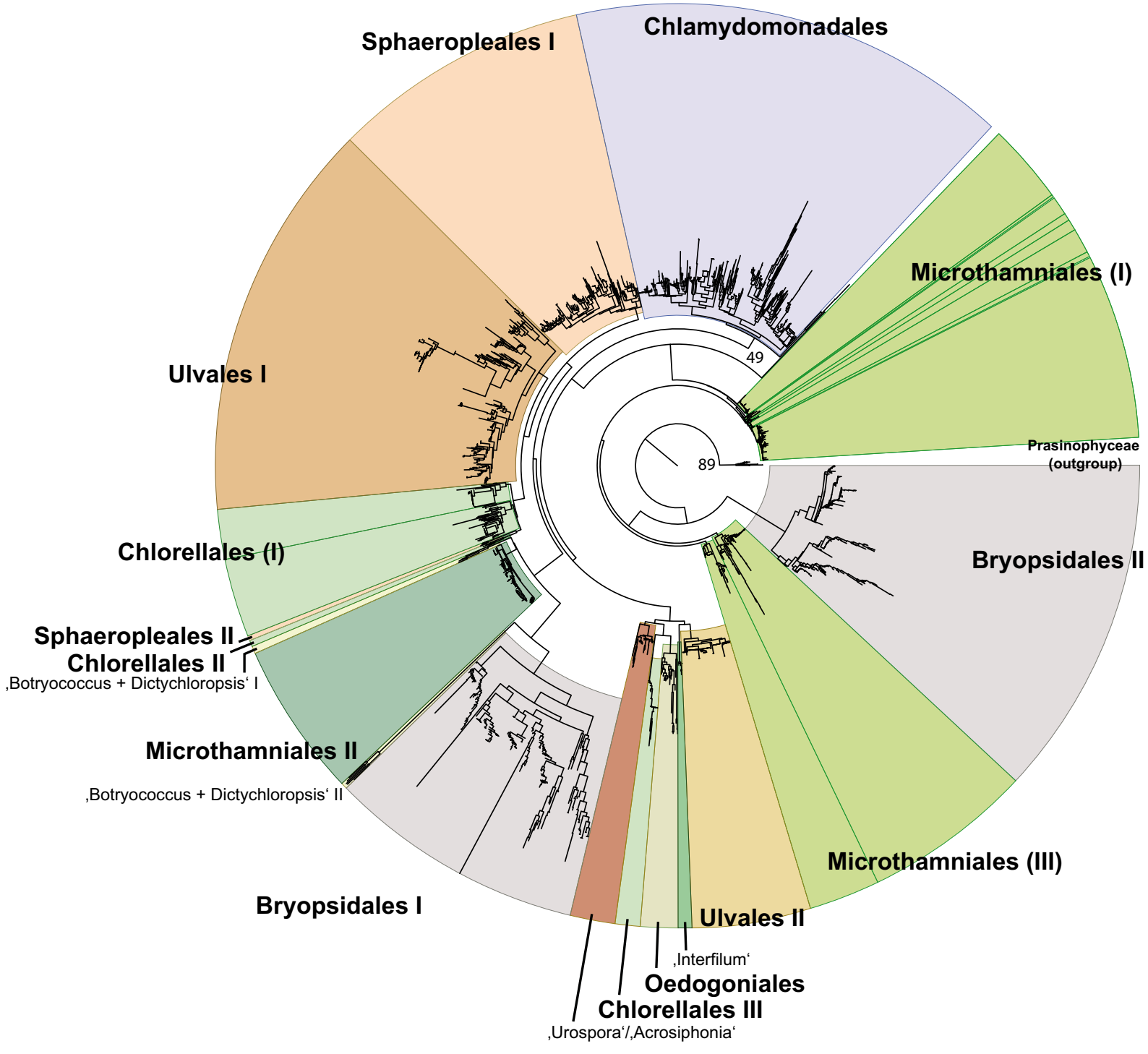


Figure 4

Click here to download Figure: f04-chlorophyta2nonames_RGB.eps



P.7. ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*)

Authors: M. Wiemers, A. Keller*, M. Wolf

published 2009

Journal: *BMC Evolutionary Biology*
Volume: 9, Pages: 300.1–300.27

*Co-author

Author's Contributions:
M. Wiemers conceived and coordinated the study, performed most of the sampling and molecular genetic studies, analyzed data and drafted the manuscript. I performed secondary structure predictions, alignment calculations and phylogenetic reconstructions under supervision of M. Wolf. All authors read and approved the final manuscript.

BMC Evolutionary Biology

Research article

ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*)

Martin Wiemers*[1], Alexander Keller[2] and Matthias Wolf[2]

Address: [1]Department of Animal Biodiversity, Faculty of Life Sciences, University of Vienna, Rennweg 14, 1030 Wien, Austria and [2]Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

Email: Martin Wiemers* - martin.wiemers@univie.ac.at; Alexander Keller - a.keller@biozentrum.uni-wuerzburg.de; Matthias Wolf - matthias.wolf@biozentrum.uni-wuerzburg.de

* Corresponding author

Published: 26 December 2009

BMC Evolutionary Biology 2009, **9**:300 doi:10.1186/1471-2148-9-300

Received: 2 July 2009
Accepted: 26 December 2009

This article is available from: http://www.biomedcentral.com/1471-2148/9/300

Abstract

Background: Current molecular phylogenetic studies of Lepidoptera and most other arthropods are predominantly based on mitochondrial genes and a limited number of nuclear genes. The nuclear genes, however, generally do not provide sufficient information for young radiations. *ITS2* , which has proven to be an excellent nuclear marker for similarly aged radiations in other organisms like fungi and plants, is only rarely used for phylogeny estimation in arthropods, although universal primers exist. This is partly due to difficulties in the alignment of *ITS2* sequences in more distant taxa. The present study uses *ITS2* secondary structure information to elucidate the phylogeny of a species-rich young radiation of arthropods, the butterfly subgenus *Agrodiaetus*. One aim is to evaluate the efficiency of *ITS2* to resolve the phylogeny of the subgenus in comparison with *COI* , the most important mitochondrial marker in arthropods. Furthermore, we assess the use of compensatory base changes in *ITS2* for the delimitation of species and discuss the prospects of *ITS2* as a nuclear marker for barcoding studies.

Results: In the butterfly family Lycaenidae, *ITS2* secondary structure enabled us to successfully align sequences of different subtribes in Polyommatini and produce a Profile Neighbour Joining tree of this tribe, the resolution of which is comparable to phylogenetic trees obtained with *COI+COII* . The subgenus *Agrodiaetus* comprises 6 major clades which are in agreement with *COI* analyses. A dispersal-vicariance analysis (DIVA) traced the origin of most *Agrodiaetus* clades to separate biogeographical areas in the region encompassing Eastern Anatolia, Transcaucasia and Iran.

Conclusions: With the inclusion of secondary structure information, *ITS2* appears to be a suitable nuclear marker to infer the phylogeny of young radiations, as well as more distantly related genera within a diverse arthropod family. Its phylogenetic signal is comparable to the mitochondrial marker *COI* . Compensatory base changes are very rare within Polyommatini and cannot be used for species delimitation. The implementation of secondary structure information into character-based phylogenetic methods is suggested to further improve the versatility of this marker in phylogenetic studies.

Background

Molecular phylogenetic studies aim to reconstruct species trees, e.g. to infer the evolution of morphological characters or life history traits. While in the early days of genetic analyses, the data sets were often confined to single gene fragments, it is now generally acknowledged that analyses should include several genes [1-3]. The use of multiple genes not only provides a greater resolution over different time scales but yields a more accurate estimate of the species tree which may not correspond to a single gene tree, especially in radiations of closely related species [4,5]. Unfortunately, the number of genes which are routinely used for phylogenetic analysis, especially in species rich arthropod assemblages, have remained limited [6]. In the mitochondrial genome, the cytochrome c oxidase subunit I (*COI*) has become the most commonly used marker in molecular phylogenetic studies of arthropods, in part due to it being the focal genetic marker for DNA barcoding studies [7]. This marker is now routinely supplemented by the nuclear marker elongation factor 1 alpha (*ef1*) and sometimes wingless (*wg*) [3,6]. These nuclear markers, however, continue to be of limited use in resolving the phylogeny of young radiations because of their slow evolutionary rate. Recently, novel nuclear genes have been tested in species of Lepidoptera, four of which (*Tektin*, *CAD*, *DDC*, *IDH*) appear promising for such radiations [6,8]. However, experience with these remains limited or lacking.

The internal transcribed spacer 2 (*ITS2*), which separates the nuclear ribosomal genes *5.8S* and *28S* , constitutes a rapidly evolving nuclear DNA fragment and has proved very useful when inferring phylogenetic relationships of closely related species in groups of organisms such as plants and fungi [9]. The highly conserved flanking regions can be used as an anchor for universal primers. However, *ITS2* studies on the phylogeny of metazoans are relatively rare. In arthropods, only 11,927 *ITS2* sequences from 2720 species have been deposited in GenBank [10] as of 02 Feb 2009 compared to 13,347 *ef1* sequences from 7353 species and 375,287 *COI* sequences from 46,385 species in BOLD [11]. This may, in part, be explained by alignment problems which have limited use of *ITS2* in phylogenetic studies of more distantly related taxa. Advances in predicting the secondary structure of *ITS2* enables alignment of *ITS2* data from more distantly related taxa and increases its utility above the genus level [12,13]. In this paper we show that the inclusion of secondary structure information improves phylogeny estimation with *ITS2* in a large radiation of blue butterflies and renders *ITS2* a useful nuclear marker in phylogenetic studies. Furthermore, we suggest that *ITS2* is a promising nuclear candidate for barcode studies, in addition to the mitochondrial marker *COI* .

The Lycaenidae are the second largest family of butterflies with about 6000 species worldwide. Among them is a large radiation of ca 130 Palaearctic species, i.e., the subgenus *Agrodiaetus* . It is extraordinary in Metazoa for its extreme interspecific variation of chromosome numbers, which is present even among closely related species that are often very similar or identical in phenotype [14-17]. Recently, the radiation has become the focus of several molecular phylogenetic studies in order to unravel the evolution of morphological and karyological characters [18-21] and to evaluate the barcoding approach [22]. All these studies employed *COI* as the main genetic marker. Wiemers [18] additionally used *ITS2* as a secondary marker, but phylogenetic resolution without the inclusion of *COI* remained unsatisfactory, and the alignment had to be confined to the subtribe Polyommatina due to alignment problems. Kandul et al. [19] included *ef1* as an additional nuclear marker in a small subset of taxa, but the marker hardly provided any phylogenetic signal and was therefore abandoned in subsequent studies [20,21]. Our aim is to compare and evaluate the phylogenetic trees based on *COI* with independent evidence from the nuclear *ITS2* incorporating sequence, as well as, secondary structure information.

Without doubt, DNA sequence data are an extremely valuable source of information to infer phylogenetic relationships. Another usage of these data has recently come into the focus of both biological scientists and stakeholder groups and attracted much controversy among them: their usage to delimit and identify species [22-33]. Although *COI* has been the marker of choice for the barcoding campaign, *ITS2* is a successful alternative. This is especially true in groups where *COI* fails to work well, e.g. in fungi [34], where it was used in combination with *ITS1* , and, most recently, in diatoms [35]. Furthermore, it has been recently claimed that structural differences in *ITS2* are predictive of species limits. In this view, pairings of CBCs (= compensatory base changes) provide an indication for sexual incompatibility [36], while their absence indicates intercrossing ability [37]. As the investigated taxonomic group provides an interesting and opportune example, a further aim of this study is to test, whether these claims also apply for the large and very recent radiation of the subgenus *Agrodiaetus* with an origin about 2.51-3.85 million years ago [19,21].

Results

Sequencing and alignment results

PCR products amplified successfully from all recently collected ethanol-preserved material, while dried material which had been successfully used for PCR of the mitochondrial cytochrome c oxidase I (*COI*) failed to consistently achieve successful PCR amplification of *ITS2* . Furthermore, in 11% of sequencing reactions, incomplete

sequences were obtained, probably caused by polymerase slippage at positions with highly repetitive motifs. Usually, it was still possible to obtain a complete sequence by sequencing from 5' and 3' ends such that the sequences only rarely remained incomplete after extended sequencing efforts. Incomplete sequences were excluded from the analysis as they may be result from co-amplified pseudogenes or not homogenized *ITS2* copies. No obvious problems with intragenomic sequence variation were encountered in the remaining sequences -- all electropherograms obtained were readable over their entire length. Thus, we assume to have no problems associated with non-homogenized *ITS2* copies, what has been reported in other *ITS* studies [38-41] and is discussed in several reviews [42,43]. Sequence length varied between 450 bp (in *Tarucus theophrastus*) and 602 bp (in *Allotinus portunus* and *Lysandra corydonius*). Sequence length variation in *Agrodiaetus* was between 530 bp (in *A. kurdistanicus*) and 563 bp (in *A. dama*). Nucleotide composition was typical for RNA with a slight overrepresentation of guanine (U : C : A : G = 0.234 : 0.261 : 0.203 : 0.302).

Alignment was successful for all sequences of the tribe Polyommatini (including six subtribes), as well as for the outgroup (Miletini: *Allotinus portunus*). Alignment difficulties were encountered with sequences of three other tribes (Theclini, Eumaeini and Lycaenini) which were therefore excluded from the analysis.

The alignment had 1024 positions of which 419 were variable and 235 were parsimony-informative (with gaps treated as missing data). Within *Agrodiaetus* , 131 positions were variable and 58 were parsimony informative.

Phylogeny of Polyommatus

According to the Profile Neighbour Joining (= PNJ) tree (fig. 1), the genus *Polyommatus* represents a monophyletic unit with the exception of its subgenus *Lysandra* . The subgenus *Lysandra* is clearly monophyletic but its placement within *Plebejus* s.l. is unsupported. Some systematic treatments have united *Lysandra* with *Meleageria* , but the two subgenera appear distinctly distant from each other in our analysis.

The remaining subgenera (*Agrodiaetus*, *Meleageria*, *Polyommatus* s.str., *Neolysandra*) together form a monophyletic group with a bootstrap support of 88%. Regarding these subgenera, the monophyly of the subgenus *Agrodiaetus* is supported with a bootstrap value of 74%. The sister group to *Agrodiaetus* appears to be either the subgenus *Meleageria* or *Polyommatus* s.str. The latter subgenus includes taxa which have sometimes been placed in subgenera *Sublysandra* and *Plebicula* . While the taxa attributed to *Sublysandra* (*P. cornelia*, *P. aedon* and *P. myrrhinus*) appear to form a monophyletic cluster at the base of the remaining species of *Polyommatus* , the subgenus *Plebicula* (in which *P. dorylas*, *P. escheri*, *P. amandus* and *P. thersites* have sometimes been included) does not appear as a monophyletic entity. The taxa of the subgenus *Neolysandra* appear at a basal position relative to the other *Polyommatus* subgenera. The relationships of the remaining Polyommatina genera with each other and with *Polyommatus* are not well supported, except for the monophyly of *Aricia* . Nonetheless, the subtribe Polyommatina received high bootstrap support (95%) and the members of all other Lycaenidae tribes are positioned outside this cluster.

Phylogeny of Agrodiaetus

Agrodiaetus damon (the two sequences from France and Turkey are identical) appears to be the sister taxon to all other *Agrodiaetus* . Unfortunately, the bootstrap support for this position is low. However, a single base-pair substitution is present at position 918 in the alignment that is a further support for the basal position of *A. damon* (although weak). At this position, all other *Agrodiaetus* sequences bear a guanine while *A. damon* and the remaining species of the genus *Polyommatus* bear an adenine base. The following major clades are supported by bootstrap values ≥ 50 among the remaining *Agrodiaetus* species as indicated in fig. 1 (bootstrap values in brackets): *admetus* clade (54%), *dolus* clade (81%), *carmon* clade (50%), *actinides* clade (62%), *iphigenia* clade (59%), *glaucias* clade (56%), *poseidon* clade (79%).

Additionally, there are some minor clades. Most of them are poorly supported and include only two species whose sequences are very similar or identical: *iphidamon* clade (13%, p-distance: 0.006), *erschoffii* clade (57%, p-distance: 0.011), *posthumus* clade (40%, p-distance: 0.002-0.006), *shahrami* clade (9%, p-distance: 0.000), *phyllis* clade (99%, p-distance: 0.000).

The remaining three species cluster with low bootstrap support: *A. valiabadi* as sister to the *admetus* and *dolus* clades (40%), *A. pierceae* as sister to the *carmon* clade (37%), and *A. klausschuriani* as sister to the *poseidon* clade (52%).

The phylogenetic relationships among the clades are usually poorly supported by bootstrap values with the exception of the *admetus* and *dolus* clades which form a clade together with *A. valiabadi* with a bootstrap support of 64%.

A classification based on *Agrodiaetus* clades with bootstrap support ≥ 50% is presented in fig. 1, together with classifications based on previous publications. A comparison of molecular based classifications reveals that 7 major clades are repeatedly found. Their support values are given in table 1.

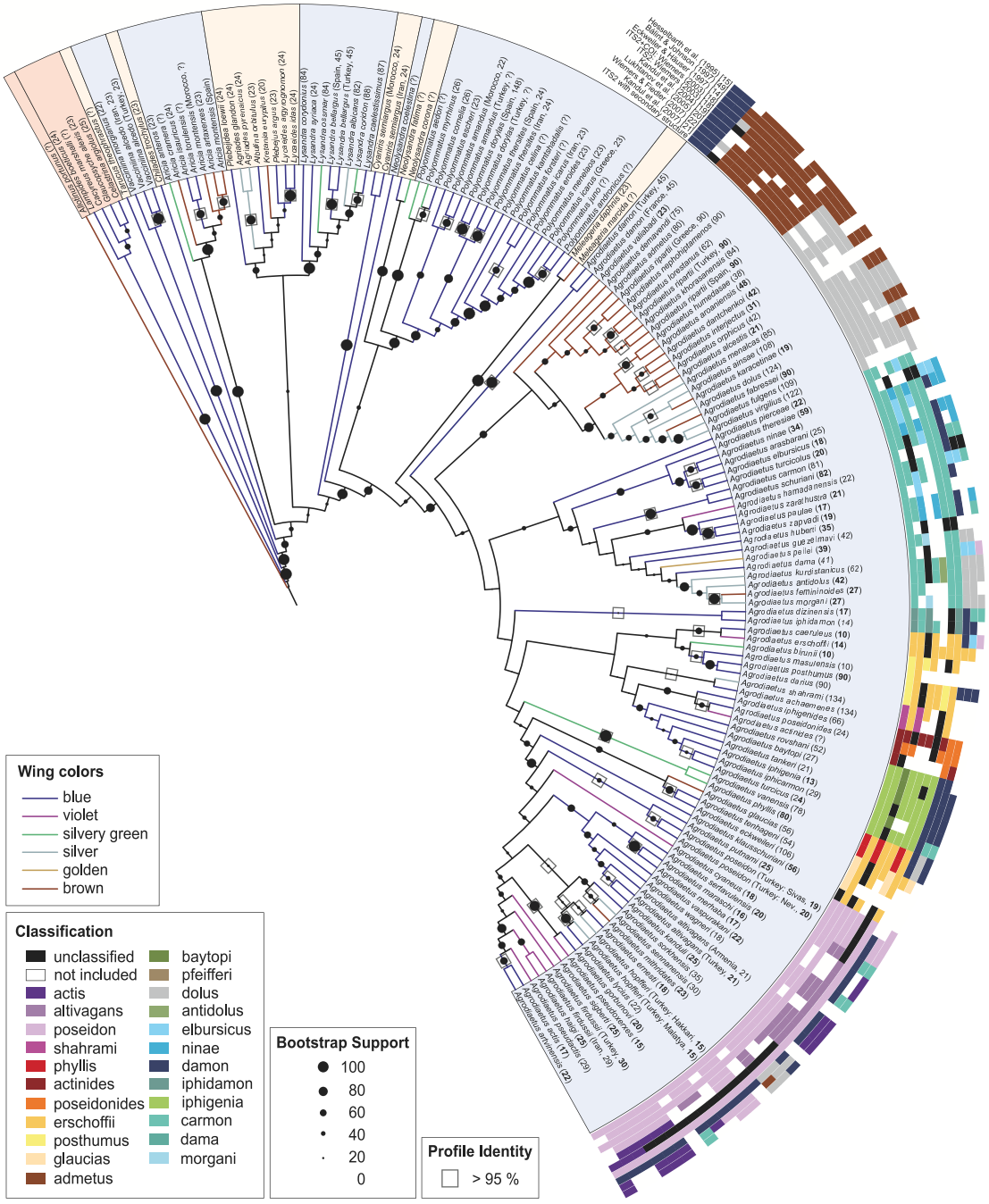


Figure 1

Profile Neighbour-Joining (PNJ) tree of *ITS2*. *ITS2* PNJ tree of 140 Lycaenidae species belonging to the tribe Polyommatini (Polyommatinae) and rooted with *Allotinus portunus* (Miletinae: Miletini) as outgroup. Bootstrap support values and profile identities > 95% are indicated on branches above nodes. Upperside wing colouration of males is indicated by branch colouration, using 6 different classes following Lukhtanov et al. (2005) [20]. Modal chromosome numbers are indicated in brackets after the species name (**bold** = gene sequence and karyotype data obtained from the same specimen; *italics* = sequence and karyotype data of a different individual from the same population [18-21]). Classification schemes of the present and other studies are coded by coloured rings around the tree. References to the corresponding studies are given in square brackets.

Table 1: Support values for major clades in different analyses

Gene(s) & Reference	***ITS2***	***ITS2***	***COI***[18]	***ITS2***[18]	***COI+***	***ITS2***[18]	***COI+COII***[19]		***COI+COII*** [20]		***COI***[22]	***COI+COII*** [21]		
Methods	**PNJ**	**NJ**	**BI***	**BI***	**BI**	**MP**	**BI**	**MP**	**BI**	**ML**	**NJ***	**ML**	**MP**	**BI**
admetus	54	45	100	84	100	100	100	100	100	100	98	100	100	100
dolus	*81*	*64*	100	100	100	100	100	100	100	100	90	100	100	100
carmon	50	0	0	81	100	100	100	73	100	88	9	88	74	100
actinides	62	42	53	<50	56	97	100	97	100	100	0	<50	<50	38
iphigenia	59	57	0	*91*	97	63	98	72	100	84	11	86	75	100
erschoffii	0	0	100	0	100	97	100	0	0	60	45	56	<50	<50
poseidon	79	0	100	*65*	100	98	100	96	100	96	63	97	97	100

Methods: **BI** = Bayesian inference, **ML** = Maximum Likelihood, **MP** = Maximum Parsimony, **NJ** = Neighbour-Joining, **PNJ** = Profile Neighbour Joining; *Support values taken from unpublished data

Biogeographical patterns in Agrodiaetus

According to the dispersal-vicariance model implemented in DIVA, the origin of *Agrodiaetus* remains uncertain, but the ancestral biogeographical areas of most major clades are quite precisely inferred (fig. 2, table 2 &3). An exception is the *admetus* clade whose ancestral area appears to encompass almost the entire range of the subgenus, with the exception of the Central Eurosiberian and Lebanese regions. The reason for this result, however, might be due to the poor taxonomy of this clade. It consists only of monomorphic species which hardly differ in phenotype and possess high chromosome numbers. The precise count of such high chromosome numbers is very difficult with standard karyological techniques [18]. Molecular results (of *ITS2* as well as *COI* [18]) indicate that *A. ripartii* , the most widespread member of this clade, is not monophyletic and consists of several distinct species. The ancestral area of the closely related *dolus* clade also remains ambiguous but is confined either to the Mediterranean, the Central Anatolian, the Armenian, or Kurdistanian region. Most members of the *dolus* clade are also monomorphic or have high chromosome numbers. Therefore its taxonomy is contentious as well and this might have influenced the results. An illustrative example is given in the following section. The ancestral areas of the remaining clades appear to be restricted to four biogeographical regions. The Kurdistanian region is home to the *carmon* clade (as well as to the small Iranian *shahrami* clade) while the *iphigenia* and *poseidon* clades seem to have originated in the neighbouring Armenian region. (The latter clade might also have originated from both.) With the exception of the Turkestanian *actinides* clade, the remaining smaller clades (*erschoffii, posthumus, glaucias*) appear to have originated in the Central Iranian region.

Compensatory base changes (CBCs) in Agrodiaetus

A maximum of only 3 CBCs are found among the 140 investigated species-level taxa of Lycaenidae. One of them occurs between members of the *Agrodiaetus +Polyommatus +Meleageria* clade and the remaining Lycaenidae species (with the exception of *Neolysandra fatima*). In 64% of pairwise species comparisons (and even 99.8% of congeneric comparisons) no CBCs are found. Within *Agrodiaetus* hardly any species is distinguished by a CBC, but some major clades can be delimited by hemi-CBCs such as the *iphigenia* and *dolus* clade. Due to the low number of CBCs and hemi-CBCs, the NJ trees created from CBC or hemi-CBC distance matrices provide little resolution (data not shown).

Although CBCs are uncommon within Polyommatini, most species differ in their *ITS2* sequence. Identical haplotypes were only found in very few sets of taxa (table 4). Most of them concern taxa with questionable species status [18,44]. For example, *A. karacetinae* differs only in karyotype and COI sequence from *A. alcestis* , but not in any morphological characters ("karyospecies"). Its position in fig. 1 (as sister to *A. ainsae*) is an artefact caused by a single missing nucleotide at position 628 in the alignment which causes a change in secondary structure making it similar to *A. ainsae* . The sequence of the latter taxon is most similar to that of *A. fulgens* , and its distant position to this species in fig. 1 can also be explained by several missing nucleotides. According to recent karyological

Table 2: Distribution of *Agrodiaetus* species in biogeographical regions used for DIVA analysis

Species	Distribution	*Species*	Distribution
A. achaemenes	F	*A. karacetinae*	E
A. actinides	K	*A. khorasanensis*	H
A. actis	C	*A. klausschuriani*	H
A. admetus	B C D E	*A. kurdistanicus*	F
A. ainsae	B	*A. lorestanus*	H
A. alcestis	C D E F G	*A. lycius*	D
A. altivagans	E F	*A. maraschi*	C D
A. antidolus	E F	*A. masulensis*	E
A. arasbarani	E	*A. menalcas*	C D E F
A. aroaniensis	B	*A. merhaba*	E
A. artvinensis	E	*A. mithridates*	C D E F
A. baytopi	E F	*A. morgani*	F
A. birunii	H	*A. nephohiptamenos*	B
A. caeruleus	H	*A. ninae*	E
A. carmon	C E F	*A. orphicus*	B
A. cyaneus	E F	*A. paulae*	E
A. dama	D	*A. peilei*	F
A. damon	A B E I	*A. phyllis*	C E F H
A. dantchenkoi	E F	*A. pierceae*	E F
A. darius	H	*A. poseidon*	C D E
A. demavendi	E F H	*A. poseidonides*	K
A. dizinensis	H	*A. posthumus*	H
A. dolus	B	*A. pseudactis*	E
A. eckweileri	H	*A. pseudoxerxes*	H
A. elbursicus	H	*A. putnami*	E
A. ernesti	D	*A. ripartii*	B C D E F I J K
A. erschoffii	H	*A. rovshani*	E
A. fabressei	B	*A. schuriani*	D

Table 2: Distribution of *Agrodiaetus* species in biogeographical regions used for DIVA analysis *(Continued)*

A. femininoides					E					*A. sennanensis*							F		H	
A. firdussii					E	F		H		*A. sertavulensis*				D						
A. fulgens	B									*A. shahrami*							F			
A. glaucias								H		*A. sigberti*			C							
A. gorbunovi					E					*A. sorkhensis*									H	
A. guezelmavi				D						*A. tankeri*						E				
A. haigi					E	F				*A. tenhageni*									H	
A. hamadanensis						F		H		*A. theresiae*				D						
A. hopffferi			C	D	E	F				*A. turcicolus*							F			
A. huberti					E	F				*A. turcicus*						E	F			
A. humedasae	B									*A. valiabadi*									H	
A. interjectus			C							*A. vanensis*			C			E	F		H	
A. iphicarmon				D						*A. vaspurakani*							F			
A. iphidamon								H		*A. virgilius*		B								
A. iphigenia	B	C	D	E	F					*A. wagneri*			C	D		E	F			
A. iphigenides									K	*A. zapvadi*							F			
A. kanduli					E	F				*A. zarathustra*									H	

The abbreviations for the biogeographical regions are: A: Central Eurosiberian, B: Mediterranean, C: Central Anatolian, D: South Anatolian, E: Armenian, F: Kurdistanian, G: Lebanese, H: Central Iranian, I: Turanian, J: Altaian, K: Turkestanian

research, *A. ainsae* appears to be conspecific with *A. fulgens* and the name *A. ainsae* was therefore synonymised with *A. fulgens* [45].

Discussion

Secondary structure information improves phylogenetic signal in ITS2

Wiemers [18] used a mostly comparable set of taxa for phylogenetic inference from *ITS2* but did not include secondary structure information. Although most major clades recovered in our analysis were also found in the Bayesian analysis by Wiemers [18], none of our major clades were recovered with bootstrap support values ≥ 50% in the Maximum Parsimony (MP) analysis of Wiemers [18]. The *poseidon* clade was also not recovered in the Bayesian 80% consensus tree presented. (This clade - with the exclusion of *A. putnami* - only received a Bayesian support of 0.65, Wiemers unpubl., table 1). In a Neighbour Joining (NJ) analysis calculated without secondary structure information only two of the major clades recovered in the PNJ analysis received bootstrap values ≥ 50% while two clades received lower bootstrap values and the remaining two were not recovered at all (table 1). Thus, in a direct comparison of two NJ algorithms (with vs. without secondary structure, table 1), secondary structure information apparently amplifies the phylogenetic information in the data set. Further improvement in phylogeny estimation is to be expected if secondary structure information can be incorporated in Maximum Likelihood (ML) or Bayesian inference (BI) methods, because these character-based methods can be superior compared to distance based methods which discard character-state information.

One disadvantage of using secondary structure information appears to be its sensitivity to missing data in stem regions. Even small amounts of missing data can cause artefacts in phylogeny estimation of closely related taxa

Table 3: Ancestral distributions according to DIVA analysis

Node	Regions included in alternative distributions											Alternative distributions
1	A	B	C	D	E	F		H	I	J	K	ABCDEFHIJK
2	A											A
3		B	C	D	E	F		H	I	J	K	BCDEFHIJK
4		B	C	D	E	F		H	I	J	K	BCDEHIJK, BCDEFHIJK
5		B	C	D	E	F		H	I	J	K	BCDEIJK, BCDEFIJK, BCDEHIJK, BCDEFHIJK
6		B	C	D	E	F		H	I	J	K	more than 10 distributions
7		B	C	D	E	F		H	I	J	K	more than 10 distributions
8		B	C	D	E	F		H	I	J	K	more than 10 distributions
9		B	C	D	E	F			I	J	K	more than 10 distributions
10		B	C	D	E	F		H	I	J	K	more than 10 distributions
11		B	C	D	E	F		H	I	J	K	more than 10 distributions
12		B	C	D	E	F		H	I	J	K	more than 10 distributions
13		B	C		E	F						B, C, E, F
14		B	C		E	F						B, BC, BE, BF
15		B										B
16		B	C	D	E	F	G					more than 10 distributions
17			C		E	F						CE, CF, CEF
18		B	C	D	E	F	G					more than 10 distributions
19		B	C	D	E	F						more than 10 distributions
20		B			E							B, BE
21		B			E							BE
22		B										B
23		B										B
24		B										B
25						F		H				FH
26						F						F
27						F						F
28				D	E	F		H				DF, DEF, DFH, DEFH

Table 3: Ancestral distributions according to DIVA analysis *(Continued)*

29			E	F	H		EF, EFH
30			E				E
31				F	H		FH
32				F			F
33	C	D	E	F			DF, CDF, DEF, CDEF
34				F			F
35				F			F
36			E	F	H		FH, EFH
37			E	F			EF
38				F			F
39		D		F			DF
40				F			F
41		D		F			DF
42				F			F
43				F			F
44			E	F			EF
45					H		H
46					H		H
47					H		H
48			E	F	H	K	EH, FH, EFH, HK, EHK, FHK, EFHK
49					H		H
50					H		H
51					H		H
52			E		H		EH
53					H		H
54			E	F		K	EF, FK, EFK
55				F			F
56			E			K	EK
57						K	K

Table 3: Ancestral distributions according to DIVA analysis *(Continued)*

58							K	K
59			E					E
60			E					E
61			E					E
62			E					E
63		D	E	F				DE, DEF
64					H			H
65					H			H
66					H			H
67					H			H
68					H			H
69			E	F	H			EH, FH, EFH
70			E	F				E, EF
71			E					E
72			E					E
73		D	E	F				E, DE, F, EF, DEF
74		D	E	F	H			D, E, DE, F, EF, DEF, DEH, EFH, DEFH
75	C	D	E	F				E, DE, DF, EF, CEF, DEF, CDEF
76	C	D	E					D, DE, CDE
77	C	D	E					CE, DE, CDE
78			E	F				F, EF
79			E	F				E, F
80			E	F				E, F
81			E	F				E, F
82		D	E	F	H			E, DE, EF, DEF, EH, DEH, EFH, DEFH
83	C	D	E	F	H			more than 10 distributions
84	C	D	E	F	H			more than 10 distributions
85	C	D	E	F	H			more than 10 distributions
86		D						D

Table 3: Ancestral distributions according to DIVA analysis *(Continued)*

87		D		D
88		D		D
89		E		E
90		E	H	EH
91	C	E		CE
92		E		E
93		E		E
94		E		E
95		E		E
96	C	E		CE

The abbreviations for the biogeographical regions are: A: Central Eurosiberian, B: Mediterranean, C: Central Anatolian, D: South Anatolian, E: Armenian, F: Kurdistanian, G: Lebanese, H: Central Iranian, I: Turanian, J: Altaian, K: Turkestanian

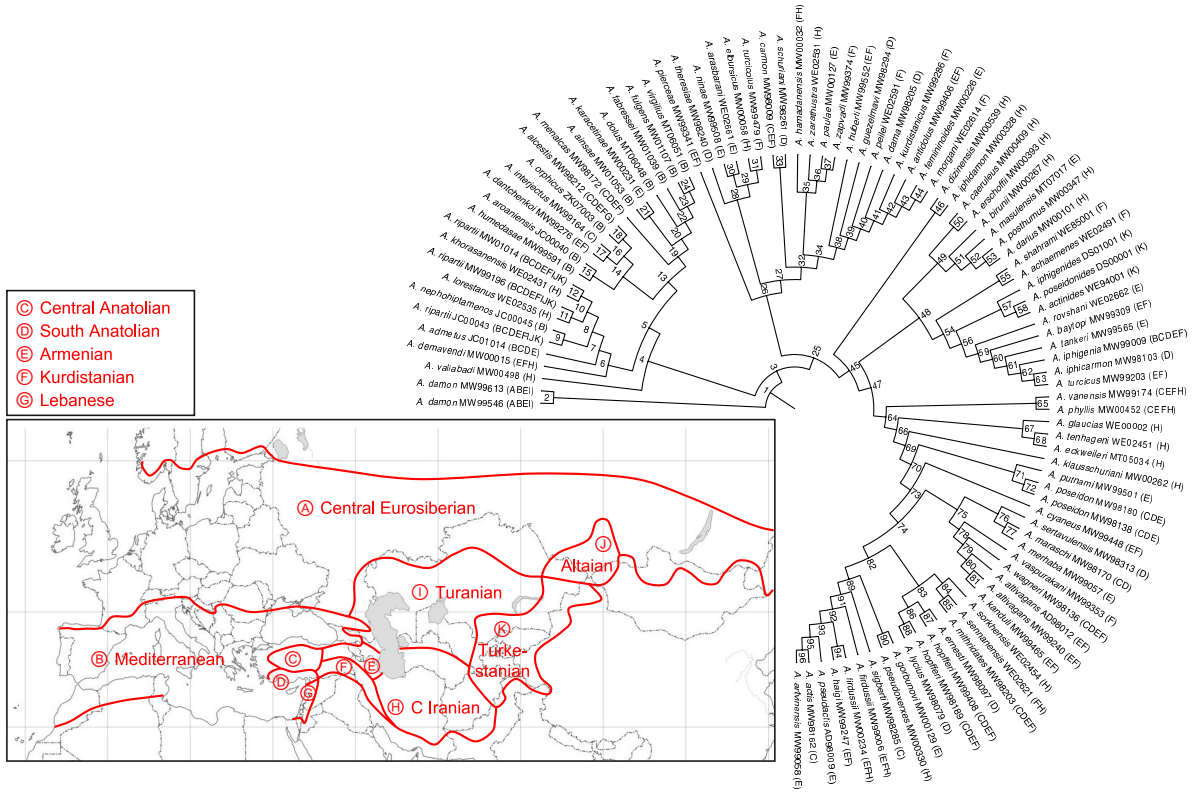


Figure 2
PNJ tree of *ITS2* and biogeographical regions. *ITS2* PNJ tree of 90 *Agrodiaetus* species and a map of biogeographical regions used for DIVA analysis. Occurrences in biogeographical regions are indicated by letters (A-K) after the species name and voucher code number according to the labels used in the map. Internal nodes in the tree are numbered consecutively.

Table 4: List of identical *ITS2* haplotypes in different taxa

Aricia artaxerxes/*A. montensis* (Spain)
Lysandra albicans/*L. coridon*
Polyommatus eroides/*P. menelaos*
Polyommatus icarus (Greece)/*P. andronicus*
Agrodiaetus ripartii (Greece)/*A. nephohiptamenos*
Agrodiaetus alcestis/*A. karacetinae*
Agrodiaetus femininoides/*A. morgani*
Agrodiaetus shahrami/*A. achaemenes*
Agrodiaetus tankeri/*A. iphigenia*
Agrodiaetus altivagans (Armenia)/*A. kanduli*
Agrodiaetus firdussii (Iran)/*A. haigi*/*A. actis*/*A. artvinensis*

with very similar sequences (viz. *A. alcestis* and *A. karacetinae*).

Phylogenetic signal of ITS2 *is comparable to* COI *in* Agrodiaetus

In agreement with *COI* analyses [18], *ITS2* data support the monophyly of Polyommatina which includes the genera *Chilades, Plebejus* and *Polyommatus* . The monophyly of the genera *Plebejus* and *Polyommatus* , however, is not fully supported. This is due to the placement of the subgenus *Lysandra* within *Plebejus* , which however has no bootstrap support and is probably caused by long-branch attraction. Such a placement is also in conflict with the Bayesian analysis of *COI* which places *Lysandra* within the genus *Polyommatus* [18]. The *ITS2* sequences of subgenus *Lysandra* are peculiar in having several longer inserts with repetitive motifs, e.g. in position 70-133 in the alignment. It is noteworthy, on the one hand, that none of the analyses supports a sister-relationship between *Lysandra* and *Meleageria* , even though members of these genera can hybridize with each other [46-48] and therefore were considered to be very closely related [15]. On the other hand, *Cyaniris* is found within *Plebejus* in the *COI* tree but basal within *Polyommatus* in the *ITS2* tree, both times with low support values. Here, the *COI* analysis appears to be more affected by long-branch attraction.

Within *Agrodiaetus* , the phylogenetic analysis of *ITS2* recovers clades which are mostly congruent to those obtained from an analysis of *COI* + *COII* (= cytochrome c oxidase II). Of particular interest is the confirmation of the sister relationship between *A. damon* and the remaining *Agrodiaetus* species that was not or only very weakly supported in the *COI* analyses. *ITS2* and *COI* also agree in the monophyly and sister relationship of the *admetus* and *dolus* clades, only the position of *A. valiabadi* differs (within the *dolus* clade in *COI* , but sister to *admetus* +*dolus* in *ITS2*). The *carmon* clade is also recovered in the *COI* +*COII* analyses but includes the *iphidamon* clade in the analyses by Lukhtanov et al. [20] and Kandul et al. (2007) [21]. Kandul et al. (2004) [19] split this group into three clades although one of them (clade VII) only appears in the MP analysis and has no bootstrap support. In the *COI* analyses by Wiemers [18] and Wiemers & Fiedler [22], which are based on shorter sequences, the *carmon* group receives no bootstrap support. Similarly, the *iphigenia* clade is only recovered in the mtDNA analyses based on the long 1969 bp section of *COI* +*COII* . The *poseidon* clade is recovered in the *COI* analyses, as well. Kandul et al. [19] split this clade into three subclades but the addition of further taxa revealed that they are not monophyletic and thus should be combined [20,21]. Most interesting is the *actinides* clade in the *ITS2* tree which suggests a close relationship between *A. actinides*, *A. poseidonides* and *A. iphigenides* . Although previous analyses have also suggested a close relationship among these taxa, it was never well supported. The relationships of the remaining clades (*glaucias, erschoffii, posthumus, shahrami, phyllis*) are not well supported in the *ITS2* tree. Previous analyses using *COI* [18-20] have suggested a close relationship of these clades, but their combination into an inclusive *erschoffii* clade was only very weakly supported by the latest *COI* analysis [21], probably due to the inclusion of additional taxa (such as *A. eckweileri*). The only major discrepancy is the placement of *A. klausschuriani* in the *ITS2* analyses (sister to the *poseidon* clade) compared to the *COI* analyses (within the *erschoffii* clade), but both placements are only very weakly supported. The missing support for the relationships between the major clades also applies to the *COI* analyses. Most analyses, however, agree in the basal position of the *admetus* +*dolus* clade and all of them recover the *poseidon* clade at the tip of the tree.

We conclude that the phylogenetic signal of *ITS2* is comparable to the signal of a much longer fragment of *COI* / *COII* . This is surprising since the rate of parsimony-informative characters is lower in *ITS2* than in *COI* [18]. Apparently these characters are, however, less "noisy" than those of *COI* , which are almost completely confined to 3rd codon positions.

ITS2 *confirms weaknesses of morphological classifications*

Fig. 1 reveals little congruence between previous classifications based on morphological characters [14,15,49] and those on molecular data (*COI* or *ITS2*). The main reason for this is the small number of available morphological characters (mostly slight differences in wing colouration)

Figure 3
Male wing vouchers of sister species pairs with different upperside colouration. 1-2: *Agrodiaetus lycius* (MW98079) - *A. hopfferi* (MW98189). 3-4: *Agrodiaetus fulgens* (MW01107) - *A. fabressei* (MW01039). 5-6: *Agrodiaetus caeruleus* (MW00409) - *A. erschoffii* (MW00393). 7-8: *Meleageria daphnis* (MW98029) - *M. marcida* (MW00290). Uppersides are shown on the left and undersides on the right side of each image

which are highly susceptible to homoplasy. Illustrative examples are morphology-based groupings formed by species with discoloured males, in which the iridescent bluish colouration on the wing upperside is replaced by a brown, golden or silvery colour (the *admetus* and *dolus* groups). Discolouration of males is coupled with an expansion of the androconial patches, apparently due to a switch from a visual to a scent-based mate recognition system [18]. Although the molecular analyses also recover a clade containing exclusively discoloured males (the clade formed by the *admetus* and *dolus* sister-clades), the molecular data reveal that single discoloured species or small groups of them are also found in most other clades. Discoloured species also appear in many other subgenera of *Polyommatus* and related genera which usually have bluish males. In the sister species pair, *M. daphnis*/*M. marcida* , the discolouration of the latter taxon (which possibly represents only a conspecific population of the former) is probably an adaptation to the specific climatic conditions (low solar radiation) on the north side of Elburs mountains [50]. Such sister species pairs with differing male upperside colouration are also found in *Agrodiaetus* , e.g. *A. fabressei*/*fulgens*, *A. shahrami*/*achaemenes*, *A. erschoffii*/*caeruleus* and *A. hopfferi*/*lycius* (fig. 3).

In some butterfly groups with similar wing patterns, genitalia provide important features for identification and classification. Unfortunately, they are very similar in all *Agrodiaetus* species, possess only few usable characters and therefore have only rarely been evaluated. The little available evidence, however, appears to be more congruent with molecular data than with wing pattern characters. Coutsis [51] analyzed the genitalia of several *Agrodiaetus* taxa which had previously been regarded as subspecies of *Agrodiaetus iphigenia* due to their similar wing colouration, among them *A. iphidamon* and *A. iphigenides* . He concluded that genitalia differences rule out conspecifity. According to the molecular results these taxa belong to different clades. *A. iphidamon* and *A. dizinensis* have been placed in different groups according to wing pattern characters [49], but they share a synapomorphic character in their genitalia: the shape of the labides is short, pointed and "dagger-like" (Coutsis, pers. comm.). Molecular results also clearly show that they are closely related. The monomorphic *Agrodiaetus* species of the *admetus* and *dolus* clades differ in karyotype but are difficult or impossible to identify based on wing pattern characters. Members of these two clades, however, differ in the length of their valves relative to their body size, those in the *admetus* clade (with the possible exception of *A. admetus*) being shorter than those in the *dolus* clade [52-54]. A comprehensive treatment of the genitalia of *Polyommatina* is currently in preparation (Coutsis, pers. comm.).

Historical biogeography

The results of our DIVA analysis confirm earlier assumptions (e.g. [18]) that Eastern Anatolia, Transcaucasia and Iran are the main centres of *Agrodiaetus* radiation. Although the origin of the subgenus could not be inferred with this method, the ancestral biogeographical areas of most major clades are placed in this region. Most interestingly, the origin of each of these clades seems to be confined to a single region (or possibly two neighbouring regions in one case). These results support the evolutionary significance of the clades obtained from the molecular analyses (*ITS2* as well as *COI*/*COII*).

CBCs as predictors of sexual incompatibility and the utility of ITS2 *to delimit species*

Due to the low number of CBCs (and hemi-CBCs) in Lycaenidae, these structural markers cannot be used to predict species limits in the family. Although this does not preclude the possibility that a CBC is a sufficient condition to distinguish species [36], an absence of CBCs cannot be used to predict intercrossing ability as suggested by Coleman [37].

This deficiency does not mean that *ITS2* sequences cannot be used to delimit species. Even in the young radiation of *Agrodiaetus* , scarcely any two species have identical *ITS2* haplotypes, while the same haplotype may be found in distant populations of the same species, e.g. *Agrodiaetus damon* from France and Turkey. On the other hand, sequence differences among populations and among individuals in a single population do exist [18], and we currently lack sufficient intraspecific *ITS2* sequence data to check for the existence of a barcode gap or diagnostic DNA characters [22,25]. Available intraspecific *ITS2* sequences usually cluster together in the PNJ tree. Exceptions occur in species complexes with disputable species borders (*A. ripartii* and *A. altivagans*) and in *Polyommatus icarus* : the Iranian *P. icarus* sequence does not cluster with conspecific sequences but with the almost identical sequence of *P. forsteri* , and is even identical with that of an Iranian specimen (voucher code ILL071) of *Polyomma-*

tus icadius [44]. The latter is a Central Asian species, whose phenotype is very similar to *P. icarus* , but which is well differentiated in *ITS2* and was only recently discovered in Iran [44]. The phenotype of the Iranian *P. icarus* specimen, however, is typical for *P. icarus* and its *COI* sequence is almost identical to those of *P. icarus* from Greece and Anatolia, where *P. icadius* does not occur [22]. Therefore it is possible that the specimen (MW00412) actually represents a hybrid between *P. icarus* and *P. icadius* . Some evidence for introgressive hybridization between these two taxa comes from the Altai where *P. icarus* and *P. icadius* share identical *COI* haplotypes [55]. Although this complex needs further research it is an example for the importance of analysing a fast nuclear locus in addition to the mitochondrial *COI* .

Conclusions

Our analyses show that *ITS2* can be a suitable phylogenetic marker not only for closely related groups of species, but also for higher taxa. In the family Lycaenidae, secondary structure information enabled the alignment of sequences from different subtribes of the tribe Polyommatini.

In *Agrodiaetus*, six major clades were obtained which are corroborated by independent evidence from mitochondrial DNA, genitalia structure, as well as our biogeographical analysis. These clades, however, do not correspond with traditional classifications, which were mainly based on the very limited set of wing pattern characters.

The use of secondary structure information with Profile Neighbour Joining also increased resolution and bootstrap support in the subgenus *Agrodiaetus* to the extent that *ITS2* phylogenetic trees provide a resolution comparable to *COI* .

In insects, *ITS2* currently appears to be the only available and well tested nuclear DNA marker which is informative enough to resolve the phylogeny of young radiations such as *Agrodiaetus* . Therefore we recommend the use of this marker as an addition to mitochondrial markers (like *COI*) in order to prevent erroneous estimation of species trees caused by introgressive hybridization, incomplete lineage sorting or horizontal gene transfer. Although introgression of mitochondrial DNA (mtDNA) appears to be less common in Lepidoptera than in most other Metazoa due to their female-heterogametic sex chromosome system [56] and Haldane's rule [57], recent work shows that such cases exist (Wiemers unpublished; [58]) and therefore should not be ignored.

We cannot, however, corroborate the use of CBCs to delimit species, because CBCs are very rare even among distantly related species in Lycaenidae and, at least, for this group their absence is not a useful predictor for sexual compatibility as claimed by Coleman et al. [37].

Methods

Material

A total of 156 Lycaenidae *ITS2* sequences were included for our analysis. Of these, 17 were exclusively determined for this study. The remainders were selected from the phylogenetic analysis of the PhD thesis by the first author [18]. Five of these sequences were improved in quality by repeating the sequencing procedure.

Generally, only one sequence per species was retained, except for taxa with a large range or with considerable geographic variation. In the latter case, two sequences representing this variation were retained. Selection criterion was the sequence quality in order to minimize ambiguities. For three species, the only available sequence was of insufficient quality and therefore these taxa were excluded from the analysis (*Agrodiaetus surakovi*, *Aricia eumedon*, *Plebejides pylaon*).

Most sequences belong to *Agrodiaetus* (97), the others to closely related genera of the same subtribe Polyommatina (54) or other subtribes within the tribe Polyommatini (5 sequences). *Allotinus portunus* (Miletinae) was chosen as outgroup because it was the only non-Polyommatini sequence available within Lycaenidae which could successfully be aligned. Alignment of sequences from the tribes Lycaenini, Theclini and Eumaeini failed, despite the fact that they are held to be more closely related to Polyommatini according to the morphology-based classification by Eliot [59].

All sequences have been deposited in GenBank [10] with LinkOuts provided to images of the voucher specimens deposited with MorphBank [60] (table 5). Annotation changes of existing entries after HMM-Annotation were as well submitted to this database. No further complete *ITS2* sequences of Lycaenidae are currently available from GenBank. The voucher specimens and DNA extractions are currently stored by the first author at the Department of Animal Biodiversity, Vienna University, but will eventually be deposited at the Alexander Koenig Research Institute and Museum of Zoology in Bonn (Germany).

In many *Agrodiaetus* species groups, especially among the monomorphic, i.e., "brown" species, karyotypes are important for species identification. Therefore in most of the specimens included in molecular analysis, the karyotypes were studied [18] using squash techniques [61,62].

Upperside wing colouration of males was classified according to the method of Lukhtanov et al. (2005) [20]. One additional colour class ("golden" for golden brown)

was added for *Agrodiaetus peilei* , a species which was not assessed in their study.

Taxonomy

The subgenera of *Polyommatus* and *Plebejus* have often been attributed generic rank in recent literature, and we follow this convention for the purposes of the present paper. The following subgenera are included in these genera: *Polyommatus* : *Cyaniris*, *Polyommatus*, *Meleageria*, *Lysandra*, *Neolysandra*, *Agrodiaetus* ; *Plebejus* : *Plebejus*, *Plebejidea*, *Plebejides*, *Lycaeides*, *Kretania*, *Albulina*, *Agriades*, *Aricia*, *Vacciniina* . The subgeneric treatment follows Hesselbarth et al. [15] with the following two exceptions: *Lysandra* (synonymised with *Meleageria* by Hesselbarth et al. [15]) and *Lycaeides* (synonymised with *Plebejus* by Hesselbarth et al. [15]).

The status of many taxa in the genus *Polyommatus* is questionable, especially in the subgenus *Agrodiaetus* which includes many recently described species, some based on disputable evidence. Taxonomic revisions and further research are needed to clarify the status of these taxa. At present, we have retained most species in order to facilitate comparisons with published studies, although some have been synonymised recently. For example, *Agrodiaetus ainsae* has been synonymised with *A. fulgens* [45] and Vodolazsky et al. [44] treat several *Polyommatus* taxa as subspecies or synonyms of *P. eros* (*P. kamtshadalis*, *P. eroides* and *P. menelaos*) and *P. icarus* (*P. andronicus* and *P. juno*).

Laboratory protocols

DNA was extracted from thorax tissue recently collected and preserved in 100% ethanol using QIAGEN® DNeasy Tissue Kit according to the manufacturer's protocol for mouse tail tissue. Occasionally, only dried material was available and either thorax or legs were used for DNA extraction. Amplification of DNA was conducted using the polymerase chain reaction (PCR). The reaction mixture (for a total reaction volume of 25 μl) included: 1 μl DNA, 16.8 μl ddH20, 2.5 μl 10 × PCR II buffer, 3.2 μl 25 mM MgCl2, 0.5 μl 2 mM dNTP-Mix, 0.25 μl Taq Polymerase and 0.375 μl 20 pm of each primer. The two primers used were ITS3 (5'-GCA TCG ATG AAG AAC GCA GC-3') and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') [63].

PCR was conducted on thermal cyclers from BIOMETRA® (models UNO II or T-GRADIENT) or ABI BIOSYSTEMS® (model GENEAMP® PCR-System 2700) using the following profiles: initial 4 minutes denaturation at 94°C and 35 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at 55°C and 1 minute extension at 72°C.

PCR products were purified using purification kits from PROMEGA® or SIGMA® and checked with agarose gel electrophoresis before and after purification.

Cycle sequencing was carried out on BIOMETRA® T-GRADIENT or ABI BIOSYSTEMS® GENEAMP® PCR-System 2700 thermal cyclers using sequencing kits of MWG BIOTECH® (for LI-COR® automated sequencer) or ABI BIOSYSTEMS® (for ABI® 377 automated sequencer) according to the manufacturers' protocols and with the following cycling times: initial 2 minutes denaturation at 95°C and 35 cycles of 15 seconds denaturation at 95°C, 15 seconds annealing at 49°C and 15 seconds extension at 70°C. Primers used were the same as for the PCR reactions for the ABI (primer 1 for forward and primer 2 for independent reverse sequencing). Electrophoresis of sequencing reaction products was carried out on LI-COR® or ABI® 377 automated sequencers using the manufacturer's protocols. Electropherograms were edited and aligned using the LaserGene® Software SeqMan Pro Version 7.1.0 by DNASTAR®.

Data analysis

Secondary Structure Prediction

Data analysis followed the method described in Schultz & Wolf [64] for secondary structure phylogenetics. All retained *ITS2* sequences were delimited and cropped with the HMM-based annotation tool present at the *ITS2* database ([65]; E-value < 0.001, metazoan HMMs). This tool furthermore integrates a visual check for the 5.8S/28S hybridization as the ITS2 proximal stem. Incorrect folding of this region is a good indication for pseudogenes [66]. All sequences of this study passed this test with a correct folding, so that we are confident to exclude pseudogenes in this study. Furthermore, according to Álvarez & Wendel [42], *ITS* pseudogenes have lowered secondary structure stability and an increase in AT content via deaminations. This was not the case for our complete *ITS2* sequences, since their secondary structures were stable and the GC content of each sequence was clearly above 50%. The proximal stem (25 nucleotides of *5.8S* as well as *28S* rDNA) was included to preserve a conserved margin of the alignment. For several sequences, nucleotides near the 3' end of the proximal stem were ambiguous. For these, nucleotides with more than 95% consensus within the remaining aligned sequences were adopted by the majority rule to preserve the marginal secondary structure of the RNA. The secondary structure of the *ITS2* of *Neolysandra coelestina* (MW99013) was predicted with RNA structure 4.6 [67] and ported to Vienna format with CBCanalyzer 1.0.3 [68] (fig. 4). The structures of the remaining sequences were predicted by custom homology modelling at the *ITS2* database [69-72] with the aforementioned structure as a template and at least 70% helix transfer (identity matrix, gap costs: gap open 15, gap extension 2). We further applied a Nussinov Algorithm (perl script) to each sequence to close additional base-pairs within helices, which were left open by homology modelling. For this procedure, no existing base pairs were removed, no pseudo-knots were allowed and exclusively Watson-Crick pairs were added (see fig. 5 for examples).

Table 5: List of taxa included in this study, their provenance and accession numbers

Species	Country	Locality	Collecting Date	Voucher code	Morph-Bank id	GenBank Accession
Agriades glandon	Italy	Stilfser Joch (2300 m), Bozen-Südtirol	27.07.2008	MW08069		GQ166180
Agriades pyrenaicus	Turkey	Çaglayan (1500 m), Erzincan	05.07.1999	MW99018	65226	AY556659
Agrodiaetus achaemenes	Iran	Gardaneh ye Cheri, W Samsami (2800-3000 m), Bakhtiari	21.07.2002	WE02491		AY556740
Agrodiaetus actinides	Kirgizia	Aram-Kungei valley, Alytyn Dara river (3000 m), West Transalai	11.07.1994	WE94001		AY556753
Agrodiaetus actis	Turkey	Gökpinar (1700 m), Sivas	25.07.1998	MW98162	65049	AY556633
Agrodiaetus admetus	Greece	Mt. Taiyetos (1200-1300 m), Peloponnisos	14.06.2001	JC01014	64205	AY556733
Agrodiaetus ainsae	Spain	Sta. Maria (500 m), Huesca	20.07.2001	MW01053	64811	AY556610
Agrodiaetus alcestis	Turkey	Saimbeyli falls (1500 m), Adana	28.07.1998	MW98212	65098	AY556641
Agrodiaetus altivagans	Armenia	Gnyshik village (1800-2200 m), Transcaucasia	20.07.1998	AD98012	64133	AY556717
Agrodiaetus altivagans	Turkey	Güzeldere Geç. (2500 m), Van	17.07.1999	MW99240	65448	AY556676
Agrodiaetus antidolus	Turkey	Dez Çay (1500 m), Hakkari	22.07.1999	MW99406	65614	AY556692
Agrodiaetus arasbarani	Iran	Mahmutabad, W Kaleybar (2200-2400 m), Azarbayjan-e Sharqi	29.07.2002	WE02661		AY556747
Agrodiaetus aroaniensis	Greece	Mt. Helmos (1350 m), Peloponnisos	04.07.2000	JC00040	64181	AY556725
Agrodiaetus artvinensis	Turkey	Kiliçkaya (1350 m), Artvin	08.07.1999	MW99058	65266	AY556663
Agrodiaetus baytopi	Turkey	Çatak (2000-2200 m), Van	18.07.1999	MW99309	65517	AY556684
Agrodiaetus birunii	Iran	Veresk (1800-1950 m), Mazandaran	18.07.2000	MW00267	64474	AY556578
Agrodiaetus caeruleus	Iran	Hajiabad (2150 m), Golestan	23.07.2000	MW00409	64616	AY556589
Agrodiaetus carmon	Turkey	Karabayir (1400 m), Antalya	11.07.1998	MW98009	64896	AY556622
Agrodiaetus cyaneus	Turkey	Zernek Brj. (1900 m), Van	23.07.1999	MW99448	65656	AY556696
Agrodiaetus dama	Turkey	Gündüzbey (1300 m), Malatya	27.07.1998	MW98205		AY556640

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Agrodiaetus damon	Turkey	Köskköy (1900 m), Erzurum	28.07.1999	MW99546	65753	AY556705
Agrodiaetus damon	France	Col de Tende (1850 m), Alpes Maritimes	17.08.1999	MW99613	65820	AY556714
Agrodiaetus dantchenkoi	Turkey	Kurubaş Geçidi (2200 m), Van	17.07.1999	MW99276	65484	AY556679
Agrodiaetus darius	Iran	Dizin Pass (3000 m), Tehran	12.07.2000	MW00101	64310	AY556560
Agrodiaetus demavendi	Iran	Samqabad (1900-2100 m), Tehran	09.07.2000	MW00015	64224	AY556552
Agrodiaetus dizinensis	Iran	Dizin Pass (3200-3300 m), Tehran	04.08.2000	MW00539	64746	AY556599
Agrodiaetus dolus	France	Auriol, La Roussargue (550 m), Bouches-du-Rhône	19.07.2006	MT06048		GQ166173
Agrodiaetus eckweileri	Iran	Fenjan, Surian (3000 m), Fars	08.07.2005	MT05034		GQ166172
Agrodiaetus elbursicus	Iran	Pul-e Zanguleh (2400 m), Mazandaran	11.07.2000	MW00058	64267	AY556556
Agrodiaetus ernesti	Turkey	Dedegöl Geçidi (1700 m), Isparta	21.07.1998	MW98097	64984	AY556626
Agrodiaetus erschoffii	Iran	Hajiabad (2150 m), Golestan	23.07.2000	MW00393	64600	AY556588
Agrodiaetus fabressei	Spain	Abejar (1100 m), Soria	19.07.2001	MW01039	64797	AY556608
Agrodiaetus femininoides	Iran	Qazayd Dagh (2300 m), Zanjan	16.07.2000	MW00226	64435	AY556573
Agrodiaetus firdussii	Iran	Qazayd Dagh (2300 m), Zanjan	16.07.2000	MW00234	64443	AY556576
Agrodiaetus firdussii	Turkey	Çaglayan (1500 m), Erzincan	05.07.1999	MW99006	65214	AY556655
Agrodiaetus fulgens	Spain	Sta. Coloma de Queralt (700 m), Tarragona	23.07.2001	MW01107	64856	AY556615
Agrodiaetus glaucias	Iran	Voluyeh (1500-1600 m), Mazandaran	24.05.2000	WE00002	65829	AY556736
Agrodiaetus gorbunovi	Iran	Ahar Pass (1800-1850 m), Azarbayjan-e Sharqi	13.07.2000	MW00129	64338	AY556565
Agrodiaetus guezelmavi	Turkey	Taşkent (1450 m), Konya	04.08.1998	MW98294	65180	AY556651
Agrodiaetus haigi	Turkey	Güzeldere Geç. (2500 m), Van	17.07.1999	MW99247	65455	AY556677
Agrodiaetus hamadanensis	Iran	Safedabad (2000 m), Tehran	10.07.2000	MW00032	64241	AY556554
Agrodiaetus hopfferi	Turkey	Gündüzbey (1300 m), Malatya	27.07.1998	MW98189	65076	AY556638

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Agrodiaetus hopfferi	Turkey	Dez Çay (1500 m), Hakkari	22.07.1999	MW99408	65616	AY556694
Agrodiaetus huberti	Turkey	Kop Geçidi (2350 m), Bayburt	29.07.1999	MW99552	65759	AY556707
Agrodiaetus humedasae	Italy	Pondel (900 m), Aosta	14.08.1999	MW99591	65798	AY556710
Agrodiaetus interjectus	Turkey	Çiftlik (1900 m), Erzurum	14.07.1999	MW99164	65372	AY556671
Agrodiaetus iphicarmon	Turkey	Dedegöl Geçidi (1700 m), Isparta	21.07.1998	MW98103	64990	AY556627
Agrodiaetus iphidamon	Iran	Shakuh (2600 m), Golestan	21.07.2000	MW00328	64535	AY556584
Agrodiaetus iphigenia	Turkey	Çaglayan (1500 m), Erzincan	05.07.1999	MW99009	65217	AY556656
Agrodiaetus iphigenides	Uzbekistan	Kitabsky national reserve (1500-2500 m)	08.06.2001	DS01001	64175	AY556722
Agrodiaetus kanduli	Turkey	Çatak (1600-1900 m), Van	24.07.1999	MW99465	65673	AY556697
Agrodiaetus karacetinae	Iran	Qazayd Dagh (2300 m), Zanjan	16.07.2000	MW00231	64440	AY556574
Agrodiaetus khorasanensis	Iran	5 km SW Firizi (1700-1900 m), Khorasan	16.07.2002	WE02431		AY556737
Agrodiaetus klausschuriani	Iran	Veresk (1800-1950 m), Mazandaran	18.07.2000	MW00262	64471	AY556577
Agrodiaetus kurdistanicus	Turkey	Çatak (1600-1900 m), Van	18.07.1999	MW99286	65494	AY556680
Agrodiaetus lorestanus	Iran	30 km W Dorud (2100 m), Lorestan	25.07.2002	WE02535	65837	AY556743
Agrodiaetus lycius	Turkey	Cukurelma (1300 m), Antalya	15.07.1998	MW98079	64966	AY556625
Agrodiaetus maraschi	Turkey	Gökpinar (1700 m), Sivas	25.07.1998	MW98170	65057	AY556634
Agrodiaetus masulensis	Iran	Rudbar S Janat (2600-3000 m), Mazandaran	03.07.2007	MT07017		GQ166175
Agrodiaetus menalcas	Turkey	Gökpinar (1700 m), Sivas	25.07.1998	MW98172	65059	AY556635
Agrodiaetus merhaba	Turkey	Kiliçkaya (1350 m), Artvin	08.07.1999	MW99057	65265	AY556662
Agrodiaetus mithridates	Turkey	Gündüzbey (1300 m), Malatya	27.07.1998	MW98203	65090	AY556639
Agrodiaetus morgani	Iran	40 km SW Saqqez (1800-1900 m), Kordestan	27.07.2002	WE02614		AY556745
Agrodiaetus nephohiptamenos	Greece	Mt. Orvilos (1200-2100 m), Macedonia	07.07.2000	JC00045	64186	AY556728

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Agrodiaetus ninae	Turkey	Ağrı (1800 m), Ağrı	26.07.1999	MW99508	65716	AY556701
Agrodiaetus orphicus	Bulgaria	Stara Planina Mts., Karandila Nature Park (1000 m), Sliven	29.07.2007	ZK07003		GQ166185
Agrodiaetus paulae	Iran	Ahar Pass (1800-1850 m), Azarbayjan-e Sharqi	13.07.2000	MW00127	64336	AY556564
Agrodiaetus peilei	Iran	Qamchiyan, 30 km N Chenareh (1800-1900 m), Kordestan	27.07.2002	WE02591	65839	AY556744
Agrodiaetus phyllis	Iran	Polur (2200 m), Tehran	26.07.2000	MW00452	64659	AY556592
Agrodiaetus pierceae	Turkey	Güzeldere Geç. (2600 m), Van	19.07.1999	MW99341	65549	AY556686
Agrodiaetus poseidon	Turkey	Zelve (1100 m), Nevşehir	22.07.1998	MW98138	65025	AY556630
Agrodiaetus poseidon	Turkey	Gökpinar (1700 m), Sivas	25.07.1998	MW98180	65067	AY556636
Agrodiaetus poseidonides	Tajikistan	Safedou (2500 m), Darvaz Mts.	23.06.2000	DS00001	65845	AY556721
Agrodiaetus posthumus	Iran	Shakuh (2600 m), Golestan	21.07.2000	MW00347	64554	AY556586
Agrodiaetus pseudactis	Armenia	Gnyshik village (1800-2200 m), Transcaucasia	20.07.1998	AD98009	64130	AY556716
Agrodiaetus pseudoxerxes	Iran	Shakuh (2600 m), Golestan	21.07.2000	MW00330	64537	AY556585
Agrodiaetus putnami	Turkey	Ağrı (1800 m), Ağrı	26.07.1999	MW99501	65709	AY556700
Agrodiaetus ripartii	Greece	Mt. Helmos (1350-1500 m), Peloponnisos	21.06.2000	JC00043	64184	AY556727
Agrodiaetus ripartii	Spain	Ubierna (900 m), Burgos	18.07.2001	MW01014	64773	AY556603
Agrodiaetus ripartii	Turkey	Çaglayan (1500 m), Erzincan	15.07.1999	MW99196	65404	AY556673
Agrodiaetus rovshani	Iran	Mahmutabad, W Kaleybar (2200-2400 m), Azarbayjan-e Sharqi	29.07.2002	WE02662		AY556748
Agrodiaetus schuriani	Turkey	Gezbeli Geçidi (1800 m), Kayseri	30.07.1998	MW98261	65147	AY556646
Agrodiaetus sennanensis	Iran	20 km E Mahabad (1900 m), Azarbayjan-e Gharbi	28.07.2002	WE02621		AY556746
Agrodiaetus sertavulensis	Turkey	Yellibeli Geçidi (1800 m), Karaman	06.08.1998	MW98313	65199	AY556652
Agrodiaetus shahrami	Iran	30 km N Chelgerd Pass (3000-3200 m), Bakhtiari	23.07.2002	WE85001		AY556752

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Agrodiaetus sigberti	Turkey	Ala Daglar (2700 m), Kayseri	31.07.1998	MW98285	65171	AY556650
Agrodiaetus sorkhensis	Iran	Kuh-e-Sorkh, Kadkan (2100-2500 m), Khorasan	17.07.2002	WE02454	65833	AY556739
Agrodiaetus tankeri	Turkey	Kop Geçidi (2350 m), Bayburt	29.07.1999	MW99565	65772	AY556709
Agrodiaetus tenhageni	Iran	Kuh-e-Sorkh, Kadkan (2100-2500 m), Khorasan	17.07.2002	WE02451	65831	AY556738
Agrodiaetus theresiae	Turkey	Saimbeyli falls (1200-1500 m), Adana	29.07.1998	MW98240	65126	AY556645
Agrodiaetus turcicolus	Turkey	Erek Dagi (2200 m), Van	25.07.1999	MW99479	65687	AY556699
Agrodiaetus turcicus	Turkey	Çaglayan (1500 m), Erzincan	15.07.1999	MW99203	65411	AY556674
Agrodiaetus valiabadi	Iran	5 km S Valiabad (1900 m), Mazandaran	30.07.2000	MW00498	64705	AY556594
Agrodiaetus vanensis	Turkey	Çaglayan (1500 m), Erzincan	15.07.1999	MW99174	65382	AY556672
Agrodiaetus vaspurakani	Turkey	Güzeldere Geç. (2500 m), Van	19.07.1999	MW99353	65561	AY556687
Agrodiaetus virgilius	Italy	Assergi, Gran Sasso (1000 m), Abruzzo	20.07.2006	MT06051		GQ166174
Agrodiaetus wagneri	Turkey	Zelve (1100 m), Nevşehir	22.07.1998	MW98136	65023	AY556629
Agrodiaetus zapvadi	Turkey	Zernek Brj. (1900 m), Van	20.07.1999	MW99374	65582	AY556689
Agrodiaetus zarathustra	Iran	30 km W Dorud (2100 m), Lorestan	25.07.2002	WE02531	65834	AY556741
Albulina orbitulus	Austria	Mitteralm, Grossglockner (1600 m), Salzburg	04.07.2006	MW06120		GQ166176
Allotinus portunus	Indonesia	Ujung Kulon National Park (0 m), West Java	27.01.2008	MW08003		GQ166177
Aricia anteros	Turkey	Erciyes Dagi (2000 m), Kayseri	30.07.1998	MW98270	65156	AY556648
Aricia artaxerxes	Greece	Mt. Taiyetos (1180-1200 m), Peloponnisos	16.06.2000	JC00055	64193	AY556730
Aricia cramera	Spain	Sta. Maria (500 m), Huesca	20.07.2001	MW01061	64819	AY556612
Aricia isauricus	Turkey	Kagizman (1400 m), Kars	11.07.1999	MW99097	65305	AY556666
Aricia montensis	Spain	Abejar (1100 m), Soria	19.07.2001	MW01048	64806	AY556609

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Aricia montensis	Morocco	Oukaimeden (2700 m), Marrakech	15.07.2002	MW02033	64883	AY556620
Aricia torulensis	Turkey	Torul (1100 m), Gümüshane	04.07.1999	MW99001	65209	AY556654
Cacyreus marshalli	France	Maruéjols-les-Gardons (100 m), Hérault	27.07.2001	MW01120	64864	AY556543
Celastrina argiolus	Morocco	Oukaimeden (2300 m), Marrakech	09.07.2002	MW02008	64872	AY556547
Chilades trochylus	Turkey	Dez Çay (1500 m), Hakkari	22.07.1999	MW99425	65633	GQ166186
Cyaniris semiargus	Iran	Takht-e Suleyman (3500-3700 m), Mazandaran	01.08.2000	MW00525	64732	AY556597
Cyaniris semiargus	Morocco	Oukaimeden (2700 m), Marrakech	15.07.2002	MW02034	64884	AY556621
Glaucopsyche alexis	Turkey	Cukurelma (1300 m), Antalya	13.06.2006	MK06007		GQ166171
Kretania eurypilus	Turkey	Çatak (1600-1900 m), Van	18.07.1999	MW99303	65511	AY556683
Lampides boeticus	Morocco	Tourchte (1400 m), Marrakech	14.07.2002	MW02028	64880	AY556546
Lycaeides argyrognomon	Austria	Wien-Donaustadt (200 m)	19.06.2008	MW08032		GQ166178
Lycaeides idas	Italy	Burgeis (1800-1900 m), Bozen-Südtirol	26.07.2008	MW08065		GQ166179
Lysandra albicans	Spain	Boltana (650 m), Huesca	22.07.2001	MW01092	64842	AY556614
Lysandra bellargus	Spain	Ilarduya (550 m), Alava	17.07.2001	MW01011	64770	AY556602
Lysandra bellargus	Turkey	Dez Çay (1500 m), Hakkari	23.07.1999	MW99446	65654	GQ166183
Lysandra caelestissimus	Spain	Moscardon (1600 m), Teruel	30.07.1996	OK96022	65826	AY556735
Lysandra coridon	Italy	Pondel (900 m), Aosta	14.08.1999	MW99612	65819	AY556713
Lysandra corydonius	Turkey	Gaziler (1800 m), Iğdır	26.07.1999	MW99514	65722	AY556702
Lysandra ossmar	Turkey	Zelve (1100 m), Nevşehir	22.07.1998	MW98155	65042	GQ166181
Lysandra syriaca	Turkey	Saimbeyli falls (1500 m), Adana	28.07.1998	MW98228	65114	AY556643
Meleageria daphnis	Turkey	Gülübeli Geçidi (1500 m), Fethiye	12.07.1998	MW98029	64916	AY556623
Meleageria marcida	Iran	Veresk (1800-1950 m), Mazandaran	18.07.2000	MW00290	64497	AY556580

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Neolysandra coelestina	Turkey	Çaglayan (1500 m), Erzincan	05.07.1999	MW99013	65221	AY556657
Neolysandra corona	Iran	Takht-e Suleyman (3000 m), Mazandaran	31.07.2000	MW00504	64711	AY556595
Neolysandra fatima	Turkey	Çatak (1600-1900 m), Van	18.07.1999	MW99301	65509	AY556682
Plebejidea loewii	Turkey	Saimbeyli falls (1500 m), Adana	28.07.1998	MW98220	65106	AY556642
Plebejus argus	Iran	Shemshak (2900 m), Tehran	12.07.2000	MW00116	64325	AY556563
Polyommatus aedon	Iran	Shakuh (2600 m), Golestan	21.07.2000	MW00326	64533	AY556583
Polyommatus amandus	Morocco	Oukaimeden (2300 m), Marrakech	09.07.2002	MW02001	64865	AY556617
Polyommatus amandus	Turkey	Köskköy (1900 m), Erzurum	07.07.1999	MW99047	65255	AY556661
Polyommatus andronicus	Greece	Mt. Falakro (1650 m), Macedonia	09.07.2000	JC00061	64197	AY556731
Polyommatus celina	Morocco	Oukaimeden (2300 m), Marrakech	09.07.2002	MW02006	64870	AY556618
Polyommatus cornelia	Turkey	Gezbeli Geçidi (1800 m), Kayseri	30.07.1998	MW98264	65150	AY556647
Polyommatus dorylas	Spain	Ubierna (900 m), Burgos	18.07.2001	MW01019	64778	AY556605
Polyommatus dorylas	Turkey	Çaglayan (1500 m), Erzincan	05.07.1999	MW99014	65222	AY556658
Polyommatus eroides	Greece	Rodopi Mts. (1200 m), Macedonia	08.07.2000	JC00042	64183	AY556726
Polyommatus escheri	Greece	Mt. Falakro (1650 m), Macedonia	09.07.2000	JC00039	64180	AY556724
Polyommatus forsteri	Iran	Takht-e Suleyman (3500-3700 m), Mazandaran	01.08.2000	MW00530	64737	AY556598
Polyommatus icarus	Greece	Mt. Falakro (1650 m), Macedonia	09.07.2000	JC00063	64199	AY556732
Polyommatus icarus	Iran	Hajiabad (2150 m), Golestan	23.07.2000	MW00412	64619	AY556590
Polyommatus juno	Israel	Mt. Hermon (2050 m)	05.07.2008	DB08003		GQ166170
Polyommatus kamtshadalis	Russia	Sokol, Magadan, NE Siberia	10.07.2002	RU02003		GQ166184
Polyommatus menelaos	Greece	Mt. Taiyetos (1180-1200 m), Peloponnisos	16.06.2000	JC00029	64178	AY556723
Polyommatus myrrhinus	Turkey	Kop Geçidi (2200 m), Erzurum	29.07.1999	MW99550	65757	AY556706

Table 5: List of taxa included in this study, their provenance and accession numbers *(Continued)*

Polyommatus thersites	Iran	Veresk (1800-1950 m), Mazandaran	18.07.2000	MW00302	64509	AY556581
Polyommatus thersites	Spain	Triste (600 m), Huesca	21.07.2001	MW01083	64835	AY556613
Tarucus theophrastus	Morocco	Tourchte (1400 m), Marrakech	14.07.2002	MW02025	64877	AY556619
Vacciniina alcedo	Iran	Samqabad (1900-2100 m), Tehran	09.07.2000	MW00024	64233	AY556553
Vacciniina alcedo	Turkey	Dez Çay (1500 m), Hakkari	22.07.1999	MW99430	65638	GQ166182
Vacciniina morgianus	Iran	Takht-e Suleyman (3600 m), Mazandaran	31.07.2000	MW00517	64724	AY556596

Follow this link (http://www.morphbank.net/Browse/BySpecimen/) to search Morph Bank numbers mentioned in column 6.

Alignment and Phylogenetic Analyses

Sequences and secondary structures were automatically and synchronously aligned with 4SALE 1.5 [73,74]. 4SALE translates sequence-structure tuple information prior to alignment into pseudo-proteins. Pseudo-proteins were coded such that each of the four nucleotides may be present in three different states: unpaired, opening base-pair and closing base-pair. Thus, an *ITS2* specific 12 × 12~scoring matrix was used for calculation of the alignment [73,74]. Sequence-structure alignment is available at the *ITS2* database supplements page [75].

To determine evolutionary distances between organisms simultaneously on sequences and secondary structures we used Profile Neighbour Joining (PNJ) [76] as implemented in ProfDistS 0.98 [77,78]. The tree reconstructing algorithm works similar to the alignment method on a 12 letter alphabet comprised of the 4 nucleotides in three structural states (unpaired, paired left, paired right). We applied an *ITS2* -specific general time reversible substitution model [73]. Profiles were automatically built for nodes with bootstrap support values (1000 replicates) above 70% or with at least 95% nucleotide identities. A profile is regarded as a sequence, however it is composed of probability distribution vectors instead of characters. PNJ is iterated until no more profiles can be defined according to our settings. The resulting tree was displayed with iTol v1.3.1[79] and further refined with CorelDRAW X3 (Corel Corporation, Ottawa, Canada). We utilized CBCanalyzer 1.1 [73,74] to detect CBCs and hemi-CBCs between sequence-structure pairs and to calculate a CBC tree. We used MEGA 4.0.1 [80] to calculate a matrix of p-distances and TCS 1.21 [81] to detect identical haplotypes. MEGA was also used to calculate the bootstrap support values (1000 replicates) of the NJ tree without secondary structure information using the Tamura-Nei model of nucleotide substitution with heterogeneous pattern among lineages and gamma distributed rates among sites. The appropriate model and the gamma parameter (0.8365) were calculated with MODELTEST 3.7 [82].

Classification procedures

To evaluate the results of our approach we constructed a classification of *Agrodiaetus* based on major clusters with bootstrap values ≥ 50% and compared this classification with those constructed in similar ways from published studies which either used the same marker but without secondary structure information or the mitochondrial marker *COI* or both. The clusters were named after the taxonomically most senior taxon. Classifications from published studies were constructed in the following way:

- A classification for *ITS2* without secondary structure information was constructed using major clusters from the Bayesian analysis conducted by Wiemers [18] with 84 *Agrodiaetus* species. Only groups with Bayesian posterior probabilities ≥ 0.80 were considered.

- From an analysis of 1969 bp *COI* and *COII* sequences from 55 *Agrodiaetus* species, Kandul et al. [19] proposed a classification of 12 major clades using Maximum Parsimony and Bayesian inference most of which have high bootstrap and Bayesian support. One notable exception is clade VII (*carmon* clade) which has no support and should have been combined with clade VI (*antidolus* clade) and clade VIII (*ninae* clade).

- Lukhtanov et al. [20] used an extended set of *COI* +*COII* sequences from 80 *Agrodiaetus* species and proposed 8 major clades based on Maximum Likelihood inference of phylogeny all of which are supported by bootstrap values > 50%.

- Kandul et al. [21] produced a Maximum Likelihood tree of a further extended set of *COI* +*COII* sequences

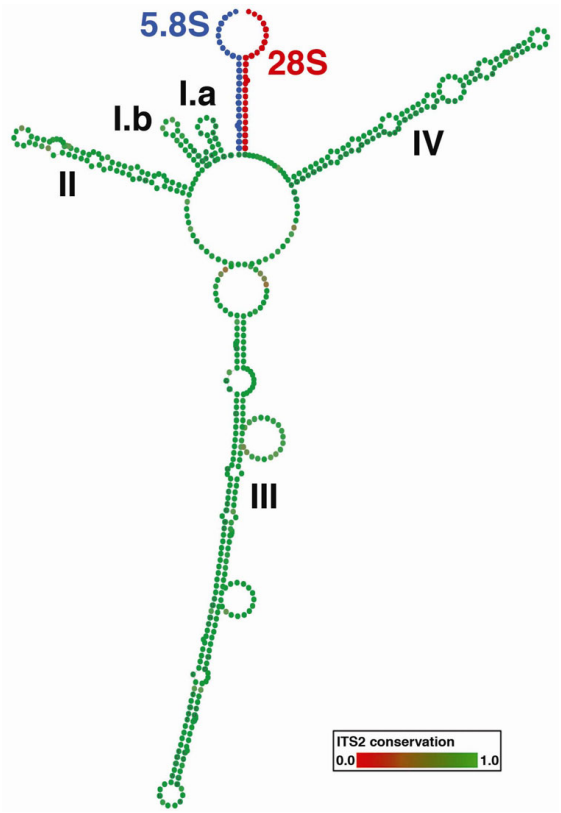


Figure 4
Conserved ITS2 secondary structure of the Polyommatina. The proximal stem of hybridized 5.8S (blue) and 28S (red) rDNA is included. Helices are numbered in Roman numerals. Two small helices are found near the beginning, which are referred to as helices I.a and I.b. The first (basal) internal bulge of helix II with two nucleotides mismatching one nucleotide is the typical U-U mismatch found in the second helix of ITS2 structures throughout the Eukaryota. Degree of conservation is displayed in colour grades from green (conserved) to red (unconserved). The complete structure represents the 51% consensus of aligned structures without gaps.

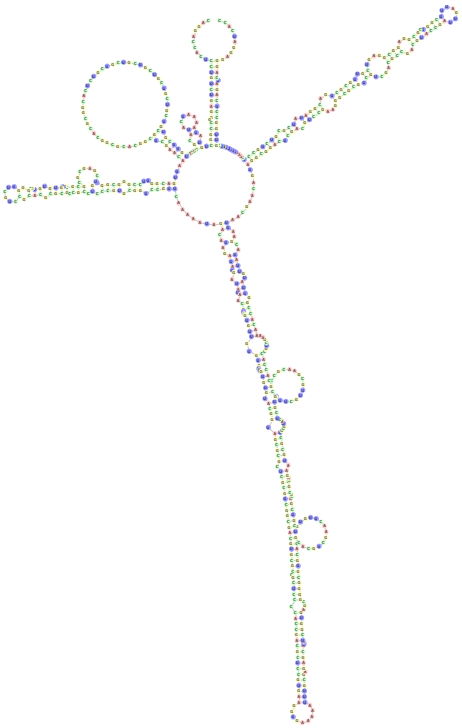


Figure 5
ITS2 secondary structure of *Lysandra syriaca*. In the distal loop of helix I.b an insertion of nucleotides is present in the genus *Lysandra* . Based on homology modelling with a template in which these nucleotides are absent (*Neolysandra*), the nucleotide insertions remain unpaired. This is a distinctive feature for the genus.

from 105 *Agrodiaetus* taxa but did not provide a classification. We inferred one using major clades with support values MP ≥ 50%, ML ≥ 50% or BI ≥ 0.80.

• Wiemers & Fiedler [22] carried out a NJ analysis using a combination of *COI* sequences taken from Wiemers [18] and Lukhtanov et al. [20] which included a total of 116 *Agrodiaetus* species. Major clusters with bootstrap values ≥ 50% were used for the classification.

• A combined analysis of *ITS2* and *COI* sequences of similar length (690 bp) from 88 *Agrodiaetus* species was carried out by Wiemers [18]. He proposed a classification based on clusters obtained with Bayesian inference using a support threshold for posterior probabilities of 0.95.

Biogeographical analysis

A dispersal-vicariance analysis was conducted with the programme DIVA 1.2 [83] to infer the ancestral distributions in the phylogeny of *Agrodiaetus* . Since outgroup relationships of *Agrodiaetus* were not well resolved in previous studies, *A. damon* was used as the outgroup to the remaining *Agrodiaetus* species according to our complete PNJ analysis (Fig. 1). The distribution area of *Agrodiaetus* was divided into 11 biogeographical regions which are based on floral biogeographical regions [84]:

• C Eurosiberian: the Central European region (incl. the Central Siberian subregion) and the Pontic - South Siberian region

• Mediterranean: the Submediterranean and Mediterranean regions excl. the South Anatolian and Palestinian - Lebanese provinces

• C Anatolian: the Central Anatolian province in the Oriental Turanian region

• S Anatolian: the South Anatolian province in the Mediterranean region

• Armenian: the Armenian - NW Iranian province in the Oriental Turanian region

• Kurdistanian: the Kurdistanian - SW Iranian province in the Oriental Turanian region

• Lebanese: the Palestinian - Lebanese province in the Mediterranean region

• C Iranian: the Central Iranian, Hyrcanian, Turkmenian, and Balutchistanian provinces in the Oriental Turanian region

• Turanian: the Turanian subregion in the Oriental Turanian region

• Altaian: the Altaian region

• Turkestanian: the Turkestanian subregion in the Oriental Turanian region

Information on the occurrence of *Agrodiaetus* species in these regions was gathered from published distribution maps and regional faunistic monographs [15,85-95].

FigTree v.1.2.3 [96] was used to draw the tree with labelled internal nodes.

Authors' contributions

The first author conceived and coordinated the study, performed most of the sampling and molecular genetic studies, analyzed data and drafted the manuscript. AK performed secondary structure predictions, alignment calculations and phylogenetic reconstructions under supervision of MWo. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to four anonymous referees for their useful comments on the manuscript. We thank John Plant (University of Vienna) for revising the language. A major part of the sequencing work was carried out by the first author at the molecular lab of the Alexander Koenig Research Institute and Museum of Zoology in Bonn. The first author thanks Clas Naumann (†) and Konrad Fiedler for supervision and assistance in many ways; Bernhard Misof for supervision of the molecular work; Esther Meyer, Ruth Rottscheidt, Meike Thomas, Manuela Brenk, Claudia Huber and Brigitte Gottsberger for assistance in DNA sequencing; Axel Hille, Claudia Etzbauer, Rainer Sonnenberg, Anja Schunke, Oliver Niehuis and Patrick Strutzenberger for general assistance in the lab; Karen Meusemann, Jurate De Prins and Vladimir Lukhtanov for karyological preparations; Wolfgang Eckweiler, Alexandre Dantchenko, John Coutsis, Miguel Munguira, Otakar Kudrna, Zdravko Kolev and Dubi Benyamini for specimen samples. The study was supported by the Deutsche Forschungsgemeinschaft (DFG grant Na 90/14). AK was funded by the BIGSS graduate school (elite network Bavaria).

References

1. Lin C-P, Danforth BN: **How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets.** *Mol Phylogenet Evol* 2004, **30:**686-702.
2. Whitfield JB, Kjer KM: **Ancient rapid radiations of insects: challenges for phylogenetic analysis.** *Annu Rev Entomol* 2008, **53:**449-472.
3. Caterino MS, Cho S, Sperling FAH: **The current state of insect molecular systematics: a thriving tower of Babel.** *Annu Rev Entomol* 2000, **45:**1-54.
4. Nichols R: **Gene trees and species trees are not the same.** *Trends Ecol Evol* 2001, **16(7):**358-364.
5. Seehausen O: **Hybridization and adaptive radiation.** *Trends Ecol Evol* 2004, **19(4):**198-207.
6. Whinnett A, Brower AVZ, Lee MM, Willmott KR, Mallet J: **Phylogenetic utility of *tektin* , a novel region for inferring systematic relationships among Lepidoptera.** *Ann Entomol Soc Am* 2005, **98(6):**873-886.
7. Hebert PD, Ratnasingham S, deWaard JR: **Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species.** *Proc Biol Sci* 2003, **270(Suppl 1):**S96-99.
8. Wahlberg N, Wheat CW: **Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera.** *Syst Biol* 2008, **57(2):**231-242.
9. Coleman AW: **Pan-eukaryote ITS2 homologies revealed by RNA secondary structure.** *Nucleic Acids Res* 2007, **35(10):**3322-3329.
10. National Center for Biotechnology Information: *NCBI* [http://www.ncbi.nlm.nih.gov/].
11. Barcode of Life Data Systems : *(BOLD)* [http://www.boldsystems.org/].
12. Coleman AW: **ITS2 is a double-edged tool for eukaryote evolutionary comparisons.** *Trends Genet* 2003, **19(7):**370-375.
13. Keller A, Schleicher T, Förster F, Ruderisch B, Dandekar T, Müller T, Wolf M: **ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales).** *BMC Evol Biol* 2008, **8:**218.
14. Eckweiler W, Häuser CL: **An illustrated checklist of *Agrodiaetus* Hübner, 1822, a subgenus of *Polyommatus* Latreille, 1804 (Lepidoptera, Lycaenidae).** *Nachr ent Ver Apollo* 1997:113-166.
15. Hesselbarth G, Oorschot Hv, Wagener S: **Die Tagfalter der Türkei unter Berücksichtigung der angrenzenden Länder.** *Bocholt: Author's edition* 1995.
16. Lesse Hd: **Spéciation et variation chromosomiques chez les Lépidoptères Rhopalocères.** *Annls Sci nat, Zool (sér 12)* 1960, **2(1-14):**1-223.
17. Lorković Z: **The butterfly chromosomes and their application in systematics and phylogeny.** In *Butterflies of Europe. Introduction to Lepidopterology Volume 2*. Edited by: Kudrna O. Wiesbaden: Aula; 1990:332-396.
18. Wiemers M: **Chromosome differentiation and the radiation of the butterfly subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*) - a molecular phylogenetic approach.** *Bonn: University of Bonn* 2003 [http://hss.ulb.uni-bonn.de/diss_online/math_nat_fak/2003/wiemers_martin/index.htm].
19. Kandul NP, Lukhtanov VA, Dantchenko AV, Coleman JW, Sekercioglu CH, Haig D, Pierce NE: **Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF1-alpha: karyotype diversification and species radiation.** *Syst Biol* 2004, **53(2):**278-298.
20. Lukhtanov VA, Kandul NP, Plotkin JB, Dantchenko AV, Haig D, Pierce NE: **Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies.** *Nature* 2005, **436(7049):**385-389.
21. Kandul NP, Lukhtanov VA, Pierce NE: **Karyotypic diversity and speciation in *Agrodiaetus* butterflies.** *Evolution* 2007, **61(3):**546-559.

22. Wiemers M, Fiedler K: **Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae).** *Front Zool* 2007, **4:**8.
23. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E: **Defining operational taxonomic units using DNA barcode data.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360(1462):**1935-1943.
24. Brower AVZ: **Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae).** *Syst Biodiv* 2006, **4(2):**127-132.
25. DeSalle R, Egan MG, Siddall M: **The unholy trinity: taxonomy, species delimitation and DNA barcoding.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360(1462**1905-1916 [http://rstb.royalsocietypublishing.org/content/360/1462/1935].
26. Funk DJ, Omland KE: **Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA.** *Annu Rev Ecol Evol Syst* 2003, **34:**397-423.
27. Hebert PD, Cywinska A, Ball SL, deWaard JR: **Biological identifications through DNA barcodes.** *Proc Biol Sci* 2003, **270(1512):**313-321.
28. Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W: **Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*.** *Proc Natl Acad Sci USA* 2004, **101(41):**14812-14817.
29. Meyer CP, Paulay G: **DNA barcoding: error rates based on comprehensive sampling.** *PLoS Biol* 2005, **3(12):**e422.
30. Moritz C, Cicero C: **DNA barcoding: promise and pitfalls.** *PLoS Biol* 2004, **2(10):**e354.
31. Smith VS: **DNA barcoding: perspectives from a "Partnerships for Enhancing Expertise in Taxonomy" (PEET) debate.** *Syst Biol* 2005, **54(5):**841-844.
32. Will KW, Mishler BD, Wheeler QD: **The perils of DNA barcoding and the need for integrative taxonomy.** *Syst Biol* 2005, **54(5):**844-851.
33. Frézal L, Leblois R: **Four years of DNA barcoding: Current advances and prospects.** *Infect Genet Evol* 2008, **8:**727-736.
34. Seifert KA: **Progress towards DNA barcoding of fungi.** *Mol Ecol Resources* 2009, **9(1):**83-89.
35. Moniz MBJ, Kaczmarska I: **Barcoding diatoms: Is there a good marker?** *Mol Ecol Resources* 2009, **9(s1):**65-74.
36. Müller T, Philippi N, Dandekar T, Schultz J, Wolf M: **Distinguishing species.** *RNA* 2007, **13:**1469-1472.
37. Coleman AW: **Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide.** *Mol Phylogenet & Evol* 2009, **50:**197-203.
38. Harris DJ, Crandall KA: **Intragenomic Variation Within ITS1 and ITS2 of Freshwater Crayfishes (Decapoda: Cambaridae): Implications for Phylogenetic and Microsatellite Studies.** *Mol Biol Evol* 2000, **17(2):**284-291.
39. Keller I, Chintauan-Marquier IC, Veltsos P, Nichols RA: **Ribosomal DNA in the Grasshopper *Podisma pedestris* : Escape From Concerted Evolution.** *Genetics* 2006, **174:**863-874.
40. Vollmer SV, Palumbi SR: **Testing the utility of internally transcribed spacer sequences in coral phylogenetics.** *Mol Ecol* 2004, **13:**2763-2772.
41. Alvarez JM, Hoy MA: **Evaluation of the ribosomal ITS2 DNA sequences in separating closely related populations of the parasitoid *Ageniaspis* (Hymenoptera: Encyrtidae).** *Ann Entomol Soc Am* 2002, **95(2):**250-256.
42. Álvarez I, Wendel JF: **Ribosomal ITS sequences and plant phylogenetic inference.** *Mol Phylogenet & Evol* 2003, **29(3):**417-434.
43. Nieto Feliner G, Rosselló JA: **Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants.** *Mol Phylogenet & Evol* 2007, **44(2):**911-919.
44. Vodolazhsky DI, Wiemers M, Stradomsky BV: **A comparative analysis of mitochondrial and nuclear DNA sequences in blue butterflies of the subgenus *Polyommatus* (s. str.) Latreille, 1804 (Lepidoptera: Lycaenidae: *Polyommatus*).** *Kavk Entomol Bjul* 2009, **5(1):**115-120.
45. Lukhtanov VA, Vila R: **Rearrangement of the *Agrodiaetus dolus* species group (Lepidoptera, Lycaenidae) using a new cytological approach and molecular data.** *Insect Syst Evol* 2006, **37:**325-334.
46. Schurian KG: **Bemerkungen zu *Lysandra cormion* Nabokov, 1941 (Lepidoptera: Lycaenidae).** *Nachr ent Ver Apollo, NF* 1989, **10(2):**183-192.
47. Schurian KG: **Nachtrag zu den "Bemerkungen zu "*Lysandra cormion* " (Lepidoptera: Lycaenidae).** *Nachr ent Ver Apollo, NF* 1991, **12(3):**193-195.
48. Schurian KG: **Freilandexemplare des Hybriden *cormion* (= *Polyommatus (Meleageria) coridon x P. (M.) daphnis*) (Lepidoptera: Lycaenidae).** *Nachr ent Ver Apollo, NF* 1997, **18(2/3):**227-230.
49. Bálint Z, Johnson K: **Reformation of the *Polyommatus* section with taxonomic and biogeographic overview (Lepidoptera, Lycaenidae, Polyommatini).** *Neue ent Nachr* 1997, **40:**1-68.
50. Biró LP, Bálint Z, Kertész K, Vértesy Z, Márk GI, Horváth ZE, Balázs J, Méhn D, Kiricsi I, Lousse V, *et al.*: **Role of photonic-crystal type structures in the thermal regulation of a Lycaenid butterfly sister species pair.** *Physical review* 2003, **E67:**. article 021907
51. Coutsis JG: **The blue butterflies of the genus *Agrodiaetus* Hübner (Lep., Lycaenidae): Symptoms of taxonomic confusion.** *Nota lepid* 1986, **9(3-4):**159-169.
52. Coutsis JG, De Prins J: **A new brown *Polyommatus* (*Agrodiaetus*) from northern Greece (Lepidoptera: Lycaenidae).** *Phegea* 2005, **33(4):**129-136.
53. Kolev Z, De Prins W: **A new species of the "brown *Agrodiaetus*" complex from the Crimea.** *Phegea* 1995, **23(2):**119-132.
54. Wakeham-Dawson A, Spurdens P: **Anomalous blue butterflies of the genus *Agrodiaetus* Hübner (Lepidoptera: Lycaenidae) in southern Greece.** *Entomologist's Gaz* 1994, **45(1):**13-20.
55. Lukhtanov VA, Sourakov A, Zakharov EV, Hebert PD: **DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce success of species identification.** *Mol Ecol Resources* 2009, **9(5):**1302-1310.
56. Traut W, Sahara K, Marec F: **Sex chromosomes and sex determination in Lepidoptera.** *Sexual Development* 2007, **1(6):**332-346.
57. Haldane JBS: **Sex ratio and unisexual sterility in hybrid animals.** *J Genet* 1922, **12:**101-109.
58. Zakharov EV, Lobo NF, Nowak C, Hellmann JJ: **Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae).** *Heredity* 2009, **102:**590-599.
59. Eliot JN: **The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement.** *Bulletin of the British Museum (Natural History) Entomology* 1973, **28(6):**371-505.
60. *MorphBank* [http://www.morphbank.net/].
61. Lukhtanov VA, Dantchenko AV: **Principles of the highly ordered arrangement of metaphase I bivalents in spermatocytes of *Agrodiaetus* (Insecta, Lepidoptera).** *Chromosome Res* 2002, **10(1):**5-20.
62. Wiemers M, De Prins J: ***Polyommatus (Agrodiaetus) paulae* sp. nov. (Lepidoptera: Lycaenidae) from Northwest Iran, discovered by means of molecular, karyological and morphological methods.** *Entomol Z* 2004, **114(4):**155-162.
63. White TJ, Bruns T, Lee S, Taylor J: **Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics.** In *PCR Protocols: a Guide to Methods and Applications* Edited by: Innis MA, Gelfand DH, Sninsky JJ, White TJ. New York: Academic Press; 1990:315-322.
64. Schultz J, Wolf M: **ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics.** *Mol Phylogenet Evol* 2009, **52(2):**520-523.
65. Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M: **5.8S-28S rRNA interaction and HMM-based ITS2 annotation.** *Gene* 2009, **430:**50-57.
66. Harpke D, Peterson A: **5.8S motifs for the identification of pseudogenic ITS regions.** *Botany* 2008, **86(3):**300-305.
67. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH: **Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.** *Proceedings of the National Academy of Sciences, USA* 2004, **101:**7287-7292.
68. Wolf M, Friedrich J, Dandekar T, Müller T: **CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures.** In *Silico Biology* 2005, **5:**291-294.
69. Schultz J, Maisel S, Gerlach D, Müller T, Wolf M: **A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota.** *RNA* 2005, **11:**361-364.

70. Schultz J, Müller T, Achtziger M, Seibel PN, Dandekar T, Wolf M: **The internal transcribed spacer 2 database-a web server for (not only) low level phylogenetic analyses.** *Nucl Acids Res* 2006, **34:**704-707.
71. Selig C, Wolf M, Müller T, Dandekar T, Schultz J: **The ITS2 Database II: homology modelling RNA structure for molecular systematics.** *Nucl Acids Res* 2008, **36:**D377-D380.
72. Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T: **Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures.** *RNA* 2005, **11:**1616-1623.
73. Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M: **4SALE - A tool for synchronous RNA sequence and secondary structure alignment and editing.** *BMC Bioinformatics* 2006, **7:**498.
74. Seibel PN, Müller T, Dandekar T, Wolf M: **Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE.** *BMC Res Notes* 2008, **1:**91.
75. ITS2 database Supplements: *ITS2* [http://its2.bioapps.biozentrum.uni-wuerzburg.de/cgi-bin/index.pl?supplements].
76. Müller T, Rahmann S, Dandekar T, Wolf M: **Accurate and robust phylogeny estimation based on profile distances: a study of the Chlorophyceae (Chlorophyta).** *BMC Evol Biol* 2004, **4:**20.
77. Friedrich J, Dandekar T, Wolf M, Müller T: **ProfDist: A tool for the construction of large phylogenetic trees based on profile distances.** *Bioinformatics* 2005, **21:**2108-2109.
78. Wolf M, Ruderisch B, Dandekar T, Schultz J, Müller T: **ProfDistS: (Profile-) Distance based phylogeny on sequence - structure alignments.** *Bioinformatics* 2008, **24:**2401-2402.
79. Letunic I, Bork P: **Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.** *Bioinformatics* 2007, **23:**127-128.
80. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24:**1596-1599.
81. Clement M, Posada D, Crandall KA: **TCS: a computer program to estimate gene genealogies.** *Mol Ecol* 2000, **9(10):**1657-1660.
82. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14(9):**817-818.
83. Ronquist F: **Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography.** *Syst Biol* 1997, **46(1):**195-203.
84. Meusel H, Jäger EJ, eds: **Vergleichende Chorologie der Zentraleuropäischen Flora.** Jena: Gustav Fischer Verlag; 1992.
85. Kudrna O: **The distribution atlas of European butterflies.** *Oedippus* 2002, **20:**1-343.
86. Lukhtanov VA, A L: **Die Tagfalter Nordwestasiens.** *Herbipoliana* 1994, **3(1-440):**.
87. Tshikolovets VV: **The butterflies of Pamir.** In *Bratislava:* F. Slamka; 1997.
88. Tshikolovets VV: **The butterflies of Turkmenistan.** In *Brno:* Konvoj Ltd; 1998.
89. Tshikolovets VV: **The butterflies of Uzbekistan.** In *Kyiv-Brno:* Author's edition; 2000.
90. Tshikolovets VV: **The butterflies of Tajikistan.** In *Kyiv-Brno:* Author's edition; 2004.
91. Tshikolovets VV: **The butterflies of Ladak (N.-W. India).** In *Kyiv-Brno:* Author's edition; 2005.
92. Tshikolovets VV: **The butterflies of Kyrgyzstan.** In *Kyiv-Brno:* Author's edition; 2005.
93. Tshikolovets VV, Bidzilya A, Golovushkin M: **The butterflies of Transbaikal Siberia.** In *Kyiv-Brno:* Author's edition; 2002.
94. Nazari V: **Butterflies of Iran.** In *Stenstrup:* Apollo Books; 2003.
95. Dantchenko A: **Genus *Agrodiaetus*.** In *Guide to the butterflies of Russia and adjacent territories Volume 2*. Edited by: Tuzov VK. Sofia: Pensoft; 2000:196-214.
96. Rambaut A: **FigTree.** *1.2.3* 2006 [http://tree.bio.ed.ac.uk/software/figtree/]. Edinburgh: University of Edinburgh

CHAPTER
NINE

SECONDARY STRUCTURE PHYLOGENETICS IN ECOLOGY

P.8. Ant-flower networks in Hawai'i: native plants are exploited, introduced plants defended

Authors: R.R. Junker, C.C. Daehler, S. Dötterl, A. Keller*, N. Blüthgen

in press 2010

Journal: *Ecological Monographs*
Volume: -, Pages: -

*Co-author

Author's Contributions:
R.R. Junker and N. Blüthgen designed and coordinated the study. R.R. Junker and C.C. Daehler performed the field work experiments and collections. R.R. Junker performed the olfactometer trials. S. Dötterl peformed the scent-analyses. R.R. Junker drafted the manuscript with contributions by me. I performed all bioinformatical steps as well as phylogenetic analyses. All authors read and approved the submitted version of the manuscript.

Ant-flower networks in Hawai'i: nectar-thieving ants prefer undefended native over introduced plants with floral defenses

Robert R. Junker[1], Curtis C. Daehler[2], Stefan Dötterl[3], Alexander Keller[4] and Nico Blüthgen[1, 5]

[1]University of Würzburg, Department of Animal Ecology & Tropical Biology, Biozentrum, Am Hubland, 97074 Würzburg, Germany

[2]University of Hawai'i at Manoa, Botany Department, 3190 Maile Way, Honolulu, HI 96822, USA

[3]University of Bayreuth, Department of Plant Systematics, Universitätsstraße 30, 95447 Bayreuth, Germany

[4]University of Würzburg, Department of Bioinformatics, Biozentrum, Am Hubland, 97074 Würzburg, Germany

[5]Correspondence: bluethgen@biozentrum.uni-wuerzburg.de

Running title: Defensive floral traits explain ant-flower networks

Abstract

Ants are omnipresent in most terrestrial ecosystems, and plants responded to their dominance by evolving traits that either facilitate positive interactions with ants or reduce negative ones. Because ants are generally poor pollinators, plants often protect their floral nectar against ants. Ants were historically absent from the geographically isolated Hawaiian archipelago, which harbors one of the most endemic floras in the world. We hypothesized that native Hawaiian plants lack floral features that exclude ants and therefore would be heavily exploited by introduced, invasive ants. To test this hypothesis, ant-flower interactions involving co-occurring native and introduced plants were observed in ten sites on three Hawaiian Islands. We quantified the residual interaction strength of each pair of ant/plant species as the deviation of the observed interaction frequency from a null-model prediction based on available nectar sugar in a local plant community and local ant activity at sugar baits. As predicted, flowers of plants that are endemic or indigenous to Hawaii were more strongly exploited by ants than flowers of co-occurring introduced plants, which shared an evolutionary history with ants. We also found that the percentage of plant species with ant-visited flowers was much higher in Hawaii than in other continental and island systems, even reaching 100 % in habitats dominated by endemic species. We showed experimentally that the absence of ants on flowers of most introduced and few native plants species was due to morphological barriers, repellent floral scents and, to a lesser extent, unpalatable nectar. Analysis of floral volatiles, however, revealed no consistent ant-repellent "syndrome" attributable to negative responses by ants, probably due to the high chemical variability within the floral scent bouquets. Results from a molecular phylogeny imply that floral defenses against ants were convergently lost in native Hawaiian plants. Exploitation of floral nectar by ants may be an important threat to Hawaiian ecosystems, reducing nectar resources available

40 to native flower visitors and potentially reducing the reproductive success of the endangered
41 endemic flora.

42 **Key-words:** Biological invasions, floral antagonists, Hawaii, ITS2, morphological barriers,
43 nectar thieves, plant defense, olfactometer, repellent floral scents, resource quality.

44

Introduction

Biological invasions, along with other anthropogenic modifications of the environment, are severe threats for ecosystems and biodiversity (Mooney et al. 2005). Low species diversity (Denslow 2003) and functional group diversity (Tilman 1997, Symstad 2000), disharmonic floras and faunas (Denslow 2003) and isolation from source habitats (Lonsdale 1999) in combination with the human capacity to transport biological material over long distances (Mooney 2005) make oceanic islands highly susceptible to invasions. The Hawaiian archipelago, one of the most isolated island groups worldwide, is a paramount example of an island system threatened by biological invasions by non-native plant and animal species, and it features many characteristics that suggest high susceptibility to invasions.

Fifteen percent of the native plant genera and 89 % of the native plant species are endemic to these islands. Today, however, nearly half of the plant species naturally occurring in Hawaii were introduced during the last two centuries (Wagner et al. 1990). Similarly, some insect taxa show high degrees of endemism, e.g. *Drosophila* flies and *Hylaeus* bees (Daly and Magnacca 2003, Magnacca and Danforth 2007) while others had been absent from the islands prior to their human introduction. It is widely accepted that ants are among those previously missing components in the Hawaiian ecosystems (Keeler 1985, Krushelnycky et al. 2005) although it has been suggested that some rather inconspicuous and subterranean ant species could be indigenous to these islands (Wheeler 1934, Medeiros et al. 1986). The vulnerability of the native Hawaiian arthropod fauna to invasive ants (Medeiros et al. 1986, Wetterer 1998, Krushelnycky and Gillespie 2008) suggests, however, that if certain ants had already reached the Hawaiian Islands prior to humans, they were not nearly as ecologically important as ants in most other terrestrial ecosystems.

A number of studies in Hawaii and elsewhere focused on the impact of alien plants on native plants (Stone and Scott 1985, Stone et al. 1992, Allison and Vitousek 2004), or on the

impact of ants on native arthropods (Medeiros et al. 1986, Holway et al. 2002, Krushelnycky et al. 2005, Krushelnycky and Gillespie 2008). Comparably little is known about the interactions between introduced ants and native and / or introduced plants. Hawaii offers a unique opportunity to study those interactions, where plants that shared an evolutionary history with ants (introduced plants) co-occur with plants that had evolved in habitats only recently invaded by ants (native plants). Many plant traits are adaptations to interactions with ants (Heil and McKey 2003, Lach et al. 2010), and Hawaiian plants may be expected to lack many of these traits. Correspondingly, very few endemic Hawaiian plant species possess extrafloral nectaries (EFNs) (Keeler 1985), a trait that is assumed to be ant-related, whereas EFN-bearing plants constitute an important part of tropical floras where ants are common (Blüthgen and Reifenrath 2003).

While the presence of ants on vegetative structures is often beneficial for plants (Rico-Gray and Oliveira 2007), flower visiting ants are – in most cases – detrimental to plant reproduction: they are poor pollinators (Pijl 1955, Beattie et al. 1984, Beattie et al. 1985), nectar thieves (Galen 1983, Galen and Butchart 2003) and negatively interfere with pollinators (Tsuji et al. 2004, Junker et al. 2007); but see e.g. Beattie (2006), Gomez et al. (1992, 1996, 2000) and de Vega et al. (2009). In order to avoid conflicts with ants on their valuable reproductive structures, plants display various mechanisms to reduce or prevent flower visitation by ants (see below). In Hawaii, floral nectar may be an important carbohydrate source since EFNs (Keeler 1985) and ant-attended honeydew-producing hemipterans are uncommon in many habitats (RRJ and NB, personal observation). Accordingly, it was reported that the flowers of a common Hawaiian plant species *Metrosideros polymorpha* (Myrtaceae) are heavily exploited by various introduced ant species (Lach 2005, 2008b, Junker et al. 2010a), suggesting that this resource is not well protected against ants.

From the consumer's (i.e. the ant's) perspective, four distinct but non-exclusive barriers need to be overcome before nectar from a given plant can be consumed (Fig. 1). We regard them as a hierarchical sequence. This conceptual framework, although developed for ants, may be adapted to any type of flower visitor. When ants and nectar-bearing flowers co-occur in space and time (Fig. 1A), floral scents and the flowers' morphology represent the first barriers (B_1 and B_2). Whether morphological barriers or floral scents act first or second may depend on the morphology of the flowers. Floral scents (B_1) are important defensive traits against facultative flower visitors (Junker and Blüthgen 2010b) and have been recently shown to effectively prevent ants from consuming nectar in a wide spectrum of flowering plants (Junker and Blüthgen 2008, Willmer et al. 2009), see also Willmer and Stone (1997) and Ghazoul (2001). Mechanical or morphological barriers (B_2) comprise either narrow nectar tubes (Herrera et al. 1984, Galen 1999, Galen and Cuba 2001, Galen and Geib 2007) or special features like sticky or greasy poles (Harley 1991) or stems or calyxes with dense trichomes that can not be passed by ants and other crawling arthropods (Kerner 1879). Unpalatable or even toxic nectar (C_1) was suggested to be the major reason for the conspicuous absence of ants on flowers observed in many regions of the world (Janzen 1977) resulting from secondary metabolites dissolved in the sugary solution (Adler 2000, Raguso 2004, Kessler and Baldwin 2006). Summarizing several studies that tested the acceptance of nectar offered outside flowers to ants of different species (Feinsinger and Swarm 1978, Guerrant and Fiedler 1981, Haber et al. 1981, Kessler and Baldwin 2006, Junker and Blüthgen 2008), we conclude that unpalatable / toxic nectar has the potential to prevent floral ant visits in a few cases, but its general importance in a large number of plant species is questionable (Junker and Blüthgen 2008). The quality of floral nectar (C_2) may reduce the visitation if ants rate the resource as unfavorable. Blüthgen and Fiedler (2004) support this assumption by showing a strong preference and a more intense recruiting behavior to more concentrated sugar and amino acid solutions by several ant species.

In our study, we observed ant-flower interactions within communities and quantified the interactions between invasive ants and flowers of native and introduced plants, considering both resource quality and the ant species' proportional abundance. We combined the hierarchical framework (Fig. 1), quantitative observations, phylogenetic analysis of the plant species and experimental approaches to test the following hypotheses regarding the visitation pattern found in Hawaii's ant-plant communities: (1) The flowers of plant species that are endemic or indigenous to the Hawaiian Islands are more regularly and strongly exploited by ants than those of introduced plant species after accounting for the nectar quantity and quality. (2) This pattern is due to more effective defensive mechanisms by the introduced plant species. (3) As suggested by Willmer et al. (2009), we expect that flowers possess mainly one type of defense (morphological barriers or repellent scent, Fig. 1) as a result from a trade-off between these. (4) The combination of floral features either allowing or preventing strong nectar exploitation by ants is the result of independent evolutionary processes, which may have been triggered by the absence / presence of nectar thieving ants, respectively. Potential implications regarding the evolution and the conservation of the flora and fauna in Hawaii are critically discussed.

Materials and Methods

Study sites

The study was conducted on the islands of Hawaii, Oahu and Kauai in natural habitats and garden settings. Sites were selected due to their accessibility, the availability of flowers and the presence of ants. Names, location, altitude and number of ant and plant species are given in Table 1. The ten study sites featured a varying degree of endemism of the plants ranging between 0 – 100 % endemic plant species. The study was conducted between March and June 2009.

Ant-flower networks

In each study site, on two consecutive days (6 am – 10 am) all flowers within a small area (0.01 – 0.1 ha) were individually checked for presence of ants. Samples of ants were taken for identification using the "Key to the ants of Hawaii" (Reimer 2006). The total number of ant workers momentarily visiting flowers of a certain species (abundance) was recorded once per plant individual per day. Since ants are social insects, these counts do not represent independent decisions, but provide a suitable estimate of the nectar consumption rate and thus reflect the ants' preferences and aversions. Number of flowers of each species present in the habitat was counted in small plants or estimated in larger shrubs or trees by multiplying the number of inflorescences with mean number of flowers per inflorescence. Nectar samples of 2 – 90 haphazardly chosen flowers per species were taken with micro-capillaries (5 µl) to quantify the amount [µl] and the sugar content [% w/w] using a handheld refractometer (Eclipse, Bellingham + Stanley, UK). Total volume of sugar provided by each plant species (standing crop) was calculated by multiplying number of flowers with mean amount of nectar [µl] and with mean sugar content [% w/w]. Plant species were assigned to the three categories: endemic, indigenous and introduced, following Wagner et al. (1990). Since bird pollination plays an important role on the Hawaiian Islands, plant species were classified as bird-pollinated or non-bird pollinated based on literature reports and/or floral syndrome. A complete list of plant species used in this study and information on their origin (endemic, indigenous or introduced) and their typical pollinators is given in Appendix A.

In order to determine the species pool of ants in the area, underneath every plant or, in dense clusters of plants, every 5 m, pieces of cardboard were laid out baited with sucrose solution (50 % w/w). After approximately one hour, all baits were checked and number and species of ants on each of the baits were recorded. In cases where two or more ant species shared bait, interactions between these species were noted (i.e. which species defended the resource against another species).

Quantification of residual interaction strength

Because we were interested in traits that promote or prevent interactions between ants and flowers, we focused our analysis to the residual interaction strength (i.e. the degree to which ants interact more or less often than expected with flowers from particular plant species) after accounting for a null model. We generated the null model prediction based on two assumptions: (1) In the absence of mechanical or chemical barriers, preferences and constraints, ants distribute themselves proportionally to the sugar supply of the different plant species they encounter in a given habitat, as predicted by optimal foraging theory (Taylor 1977, Bonser et al. 1998) and ideal free distribution (Fretwell and Lucas 1970). (2) Ant species composition on sugar baits reflects their potential composition on flowers. This is supported by a study in an Australian tropical rain forest where nearly the same ant species composition was found on baits (Blüthgen and Fiedler 2004) as on naturally occurring sugar sources like honeydew, EFNs and floral nectar (Blüthgen et al. 2004).

In the interaction matrix, each link defines the interaction between an ant species i and a plant species j, and the total number of potential links in a site is $I \times J$, with I being the total number of ant species and J the total number of plant species. The expected relative proportion E_{ij} of each link between ant species i and plant species j of the total number of interactions would be $E_{ij} = A_i \cdot P_j$, with A_i as the proportional number of workers of species i among all I ant species visiting the sugar baits, and P_j as proportional amount of sugar offered by plant species j of all J plants at the site. Thus, at sites with one ant species i and several plant species j, $E_{ij} = P_j$. The deviation of the observed from the expected proportion of a given interaction was expressed as the residual $R_{ij} = O_{ij} - E_{ij}$, with O_{ij} as observed proportion of ant species i on plant species j of the total number of ants visiting flowers in the focal interaction network. R_{ij} thus ranges from -1 to 1, and $\sum_{i}^{I}\sum_{j}^{J} R_{ij} = 0$. Negative R_{ij} indicate that interactions occurred less frequently than expected, positive R_{ij} unexpectedly frequent interactions.

196 Whether each R_{ij} significantly deviated from zero was tested by Monte Carlo statistics: We
197 randomly assigned the same total number of ant individuals that were actually found on
198 flowers in a given network to all possible links $I \times J$ one million times, with E_{ij} as the
199 probability that each link *ij* is occupied. The randomly assigned values were compared to the
200 observed numbers of ants in each link *ij*. When the observed number of ants in link *ij*
201 overlapped with less than 5 % of the simulated values, it was regarded as significant. For each
202 of the randomizations, we calculated the residuals in the same way as described above.
203 Additionally, we calculated the variance of the observed and randomized residuals $var(R_{ij})$
204 which expresses the overall deviation from the expected distribution of ants on flowers and
205 thus the degree of specialization within the habitat. Commands for R software (R: A language
206 and environment for statistical computing. R Foundation for Statistical Computing, Vienna,
207 Austria) are available in Supplement. In addition to R_{ij}, $R_i = \sum_j^J R_{ij}$ and $R_j = \sum_i^I R_{ij}$ were
208 calculated which are the row and column totals of each ant species *i* and plant species *j*,
209 respectively, and denote the total deviance from the expected contribution of the species
210 within the other species in the same trophic level in each network. We compared R_i of the
211 three different ant subfamilies (Dolichoderinae, Formicinae and Myrmicinae) and R_j of plants
212 that are endemic, indigenous or introduced to the Hawaiian Islands using an ANOVA.
213 According to our first hypothesis, we predict that R_j is higher for native than for introduced
214 plant species. Hence, we expect that native plants receive more ant visits than expected based
215 on the amount of nectar sugar, while introduced plants receive fewer visits. Note that R_j
216 (unlike R_i) only depends on the sugar quantity and relative visitation rate on plant *j* and is thus
217 independent of the ant species' identities and their responses to sugar baits.
218 Prior to statistical analysis, values were transformed to meet requirement of normality: R'_i
219 $= s_i \cdot \log(|R_i|+1)$ where s_i maintains the original sign of R_i, thus $s_i = +1$ if $R_i > 0$ and $s_i = -1$ if
220 $R_i < 0$. The same applies to R_j. We performed the ANOVA including data only from networks

with at least two subfamilies of ants (for R_i) or plants with at least two different origins (for R_j). Furthermore, we tested the influence of the "pollination syndrome" (bird vs. insect pollination) on the residuals R'_j of plant species with a *t*-test. Note that flowers assigned as bird-pollinated are additionally visited and potentially pollinated by insects.

Comparison to other oceanic islands and continents

The proportion of ant-visited flowering plant species within each of the Hawaiian networks was compared to flower-visitor networks elsewhere that included ants. Additional to published networks known to the authors, further datasets were found online using appropriate search terms or were provided by colleagues. For each network, the proportion of plant species that were visited by ants was quantified. Studies without ants where omitted from the analysis. Prior to statistical analysis, proportional data were arcsin-sqrt-transformed.

Olfactometer trials

Ants' responses to floral scents were examined in a mobile olfactometer which allowed behavioral assays in the field with unpicked flowers and free living ants (Fig. 2, Junker et al. 2010b). A battery driven electric pump (Thomas Gardner Denver, G 24/08 30W) produced an airstream of filtered air that was cleaned and humidified in charcoal and distilled water. The airstream supplied four flowmeters (Analyt-MTC, 112-08SA) that led the airstream to spiral Teflon tubes (PKMSA, CH) and regulated it to 100 ml min^{-1}. Flower stems were swathed with Teflon tape (PTFE) and one side of an oven bag (Toppits, Melitta Haushaltsprodukte GmbH & Co. KG, Minden, Germany) was tightly affixed at the Teflon tape using masking tape. The other open end of the oven bag was then pushed through a cut top-part of a washing flask and a Teflon washing flask topping was pressed into the overlapping oven bag resulting in a tight connection between the Teflon tubes coming from the flowmeter and the oven bag which thereupon inflated itself. The whole assemblage was held in place by a post and a laboratory clamp. Another Teflon tube attached to the washing flask topping supplied a four-

field arena with scented air. Usually, two separate flowers / inflorescences of an individual plant where used as the scent source for the two scented air-fields within the arena. In exceptional cases the scent of one flower / inflorescence was split into two Teflon tubes. Scented air was pumped in two opponent fields of the arena, the remaining two fields were supplied with neutral, unscented air. The four-pointed star-shaped arena (Fig. 2) was modified after Petterson (1970) and Vet (1983) and was similarly used by Junker et al. (2008, 2010b). The arena allowed creation of four distinct odor fields and was manufactured from a single Teflon block. Air left the arena at a central hole. The whole olfactometer setup was fitted in a wheeled aluminium box for its application in the field (for more information see Appendix B). For the tests, six ants were caught on sugar baits and were placed in the arena. After 60, 90, 120 and 150 s number of ants in scented and neutral fields were counted. 150 s intervals were repeated twelve times with each ant / plant combination and with different sets of ants and data from each interval were condensed to a mean number of ants in the scented fields. These values were used to calculate a response index $Q_{ij} = \frac{2(N_{obs} - N_{\exp})}{N_{total}}$, with N_{obs} = number of ants in scented fields; N_{exp} = expected number of ants in each field assuming random choices, i.e. 50 % of tested animals; and N_{total} = total number of ants tested. Like R_{ij}, Q_{ij} varies between -1 (repellence) and 1 (attraction). Scented and neutral fields were reversed after each 150 s interval to compensate for potential side preferences. All parts of the olfactometer that had contact to floral scents and ants were thoroughly cleaned with hexane and acetone. Oven bags were used only once. Olfactometer tests were performed for selected ant-plant pairs (*ij*), including the most common ant- and plant species. For several plant species, the response of two or more ant species was examined. For statistical analysis of the ants' responses, Q_{ij} values were transformed in the same way as described above. In order to compare responses to plants of different origins, the mean value of response indices Q_{ij} of different ant species to each plant species was used in the ANOVA. In cases where ant

271 species *i* encountered plant species *j* in two or more different communities, we tested the
272 interaction only once but used Q_{ij} in all communities for analysis. In four different habitats,
273 we tested the responses of *Linepithema humile* and *Pheidole megacephala* to the floral scent
274 of *Metrosideros polymorpha* in order to compare different populations. The responses Q_{ij}
275 were similar and did not change signs: -0.16 ± 0.06 for *L. humile* (mean ± standard error,
276 ANOVA: $F_{3,56} = 1.4$, $p = 0.26$) and -0.07 ± 0.01 for *P. megacephala* ($F_{3,56} = 0.09$, $p = 0.96$).

277 *Volatile collection and analysis*

278 Scent samples were taken from the same flower individuals as used in the olfactometer
279 trials. Additionally, further plant species that were not included in the ant-flower networks
280 were used for additional olfactometer trials with *Pheidole megacephala* workers and scent
281 sampling. After each olfactometer trial, the oven bag was closed with masking tape and scent
282 was either immediately sucked through a volatile trap or scent first accumulated in the oven
283 bag and was than sucked through a volatile trap using a battery driven pump (Method and
284 sampling time is given in Appendix C). Scent traps consisted of microvials (Varian,
285 Darmstadt, Germany) from which the bottoms were removed and which were filled with a
286 mixture of 1.5 mg Tenax-TA (mesh 60-80) and 1.5 mg Carbotrap (mesh 20-40). Microvials
287 with trapped scents were frozen at -20°C as soon as possible and stored in glass vials until
288 further use.

289 Scent samples were analyzed using a Varian 3800 gas chromatography fitted with a 1079
290 injector and a ZB-5 column (5% phenyl polysiloxane; length, 60 m; inner diameter, 0.25 mm;
291 film thickness, 0.25 µm; Phenomenex) and a Varian Saturn 2000 mass spectrometer. Scent
292 traps were placed into the injector port of the GC by means of the ChromatoProbe kit (Amirav
293 and Dagan 1997, Dötterl et al. 2005). The injector split vent was opened, and the injector was
294 heated at 40°C to flush any air from the system. After 2 min the split vent was closed and the
295 injector heated at 200 °C min^{-1}, then held at 200 °C for 4.2 min, after which the split vent was

296 opened (1/20) and the injector cooled down. Electronic flow control was used to maintain a
297 constant helium carrier gas flow rate (1.8 ml min^{-1}). The GC oven temperature was held for
298 7 min at 40 °C, then increased by 6 °C min^{-1} to 260 °C and held for 1 min at this temperature.
299 The mass spectra were taken at 70 eV with a scanning speed of 1 scan s^{-1} from m/z 30 to 350.
300 To identify the floral scent compounds of the GC-MS spectra, the data bases NIST 08, Wiley
301 7, Adams (2007), and MassFinder 3 were used, and identifications were confirmed by
302 comparison of retention times with published data (Adams 2007). Identification of some
303 compounds was also confirmed by comparison of mass spectra and retention times with those
304 of authentic standards. We estimated total scent emission (absolute amount) by injecting
305 known amounts of monoterpenoids, benzenoids, and fatty acid derivatives. The mean
306 response of these compounds (mean peak area) was used to determine the total amount of
307 each compound available in the samples (Dötterl et al. 2009).
308 For statistical analysis, mean amounts of individual substances were taken in cases of
309 repeated scent sampling of plant species and emission was standardized to one hour. We
310 tested three alternative hypotheses in search for patterns explaining ant-repellence. (a) Firstly,
311 we tested whether the total hourly emission of the flowers was correlated to the mean values
312 $\overline{Q}_{ij}$ across several ant species, or to Q_{ij} of *Pheidole megacephala* ants alone that were used for
313 tests with most plant species. (b) We secondly tested whether response index Q_{ij} correlates
314 with the amount [ng h^{-1}] of individual substances within floral scent bouquets. For
315 correlations we used individual substances only if they were emitted by at least seven plant
316 species. For responses, we either used the mean values $\overline{Q}_{ij}$ of several ant species or Q_{ij} of
317 *Pheidole megacephala* ants. (c) We thirdly tested whether the floral scent composition of
318 plant species that share certain features are separated from groups of plants with different
319 features. The features we tested included the significant repellence against at least one ant
320 species, the presence of mechanical barriers and the origin of the plant species, i.e. whether
321 they are endemic, indigenous or introduced to the Hawaiian Islands. Most individual

substances were emitted by one or few plant species only, thus we grouped the compounds according to their biosynthetic pathways and the presence or absence of a functional group: benzenoids (B), fatty acid derivates (FAD), monoterpenes (MT), oxidized monoterpenes (MTO), sesquiterpenes (ST), oxidized sesquiterpenes (STO) and others. We performed two non-metric multidimensional scalings (NMDS) based on Bray-Curtis distances, the first with quantitative data and the second with proportional data. Environmental vectors were fitted into the plots indicating the most rapid change in the indicated scent group (direction of vector) and strength of the gradient (length of vector). Environmental vectors were only included if the significance was $p < 0.1$.

Nectar accessibility and palatability

Using a micrometer, we measured the width of the nectar holder tube from three to 15 flowers per plant species and the width of the head capsules of 10 individuals of each ant species to the nearest 0.01 mm. The mean width of each flower was compared to the mean width of the head capsules of the ant species present in the respective habitat in order to assess the accessibility of nectar for the ants. In the di- or polymorphic species *Pheidole megacephala* and *Solenopsis* spp. we measured the width of the head capsule of the smallest caste. We never observed cases of nectar robbing, i.e. cases where ants bit holes in the perianth in order to access the nectar. In addition to narrow nectar tubes, we also checked for further mechanical barriers that could prevent the nectar consumption by ants. Furthermore, the nectar of some plant species was extracted with micro capillaries and small amounts were offered to the ants next to the sugar baits in order to test the palatability.

Phylogenetic Analysis

For the phylogenetic analysis of the plant species encountered in our study, we used internal transcribed spacer 2 (ITS2) sequences of the ribosomal cistron. Secondary structures of the ITS2 were included in the analysis to receive support for a broad range of taxonomic

relationships (Keller et al. 2010). Sequences were obtained from GenBank (Benson et al. 2009) and delimited at the ITS2 database (Keller et al. 2009). For several plant species no ITS2 sequences were obtainable from GenBank. For these, we chose representatives of close relatives for a complete taxon sampling. The amount of sequences per species was dependent of the availability of complete sequences at the database. GenBank accession numbers and representative taxa are listed in Appendix D. Data analysis followed the method described in Schultz and Wolf (2009) and Keller et al. (2008) for secondary structure phylogenetics with the ITS2.

The ITS2 secondary structures of *Armeria villosa* (Caryophyllales), *Musa velutina* (Liliopsida), *Myoporum parvifolium* (Asterids) and *Sida fallax* (Rosids) were predicted with RNA structure 4.6 (Mathews et al. 2004). These structures served as templates for homology modelling for the remaining sequences of the respective taxonomic groups at the ITS2 database (Koetschan et al. 2009). Thus, each sequence in the data set was complemented with an individual secondary structure.

Sequences and secondary structures were automatically and synchronously aligned with 4SALE 1.5 (Seibel et al. 2008). 4SALE translates the individual pairs of sequences and structures prior to alignment into a pseudo-protein code. Pseudo-proteins were coded such that each of the four nucleotides may be present in three different states: unpaired, opening base pair and closing base pair. Thus, an ITS2 specific 12x12 scoring matrix was used for calculation of the alignment (Seibel et al. 2008).

To determine evolutionary distances between plant species simultaneously on sequences and secondary structures we used Profile Neighbor Joining (PNJ) as implemented in ProfDistS 0.98 (Wolf et al. 2008). The tree-reconstructing algorithm works similarly to the alignment method on a 12-letter alphabet with an ITS2-specific general time reversible substitution model. Profiles were automatically built for nodes with bootstrap support values (1000 replicates) above 70% or with at least 95% nucleotide identities. A profile is regarded

as a sequence, although it is composed of probability distribution vectors instead of characters. PNJ was iterated until no more profiles can be defined. The resulting tree was displayed with FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and further refined with Adobe Illustrator CS4 (Adobe Corporation, San Jose, CA).

The resulting Neighbor-Joining distance matrix was compared with distances of species-specific mean nectar volume [µl], mean nectar sugar concentrations [% w/w], nectar holder tube width and mean residual R_j values. Distances based on a single variable were standardized between 0 and 1 as $D = |x_i - x_j| / (x_{max} - x_{min})$, where x_i and x_j represent the value for species *i* and *j*, x_{max} and x_{min} the maximum and minimum value of all species, respectively. Mean evolutionary distances were used for taxa with multiple sequences in the analysis. Matrices were compared using a Mantel test (Spearman rank correlation, 10000 randomizations). Distances were additionally calculated for the ants' responses to floral scents and amount of total hourly emission per dry weight [ng h^{-1} g^{-1}]. Bray-Curtis distances were calculated for floral scent composition both with quantitative data and proportional data. Since this information was not available for all plant species, we compared this matrix with a subset of the evolutionary distance matrix including only those plant species used for olfactometer trials.

Results

Ant-flower networks and residuals

In total, we screened 21,940 flowers of 39 species in ten habitats for ant visits. The flowers of 24 species were visited by a total of 1,635 ants from 12 species; on the remaining 15 plant species we never observed any ant. Five additional ant species visited sugar baits, but we did not observe them on flowers. Twelve plant species were endemic to the Hawaiian Islands (10 of them visited by ants), ten were indigenous (7 visited by ants) and 17 were introduced (7 visited by ants). The introduced species are native to continental regions where

398 they have shared an evolutionary history with rich ant faunas. Forty-four of 194 potential
399 interactions between plants and ants were recorded (Appendix E).

400 The proportion of plant species with ant-visited flowers varied between 33.3 and 100 % in
401 the ten communities and was strongly and positively correlated with the proportion of
402 endemic plant species in a habitat (see below). The total deviation from the expected
403 visitation pattern expressed by the variance of the residuals *var* (R_{ij}) was less pronounced in
404 habitats with a high proportion of endemic plant species (exponential regression: $R^2 = 0.79$, df
405 $= 9$, $p < 0.01$) indicating that ants distributed more disproportional to the available resources
406 in habitats dominated by introduced plant species. The variance of the randomized residuals,
407 however, was independent of the plant species composition in the habitats ($R^2 < 0.001$, $df = 9$,
408 $p = 0.49$). Thirty-three percent of all residuals R_{ij} deviated significantly from zero (i.e. number
409 of observed ants in link ij overlapped with less than 5% of the simulated ones): 14.4 % of all
410 potential interactions occurred significantly more frequently than expected, 18.6 % occurred
411 less frequently (Appendix E). In two of the observed sites, ants were distributed proportional
412 to the resources offered by the plant species, i.e. all residuals R_{ij} did not significantly deviate
413 from zero. In both sites only *Linepithema humile* ants and plants native to the Hawaiian
414 Islands were present (Tab. 1, #3 and #5). On average, R_j values of endemic and indigenous
415 plants were positive, while those of introduced plants were negative in sites where plants of at
416 least two origins were present (ANOVA: $F_{2,43} = 3.7$, $p = 0.03$, Fig. 3), indicating that flowers
417 of endemic and indigenous plants are favored over flowers of introduced plants. This result
418 based on residuals R_j is confirmed by the average number of ants per flower, which was
419 highest in endemic plants and lowest in introduced plant species (Tab. 2). Across ant
420 subfamilies, R_i values of Dolichoderinae were positive (mean ± SE $R_i = 0.38 \pm 0.12$), while
421 those of Formicinae (-0.01 ± 0.12) and Myrmicinae (-0.18 ± 0.07) were negative (ANOVA:
422 $F_{2,25} = 7.3$, $p = 0.003$), i.e. Dolichoderines used floral nectar as resource more intensely than
423 expected by their relative abundance in each habitat.

424 The "pollination syndrome", i.e. bird-pollinated plants (13 species, R_j values = mean ±
425 *SE*: -0.09 ± 0.06) or insect-pollinated plants (26, 0.05 ± 0.05), did not significantly influence
426 R_j values (Welch two sample *t*-test: $t_{36.3} = 1.6$, $p = 0.11$).
427 We rarely observed interactions between ant species although baits were occasionally
428 shared by two ant species. *Solenopsis geminata* displayed aggression against *Ochetellus*
429 *glaber*, *Paratrechina vaga* against *Technomyrmex albipes*, *Tetramorium tonganum* against *P.*
430 *vaga* and *Pheidole megacephala* against *Plagiolepis alluaudi*. However, we never saw similar
431 interactions on flowers. Thus it is unlikely that inter-specific aggressions on flowers
432 influenced immediate foraging decisions, but they may still have long-term effects on the
433 network pattern.

434 *Comparison to other oceanic islands and continents*
435 We compared the Hawaiian ant-flower networks to ten flower-visitor networks from
436 continents and 15 from other islands (Appendix F). On average, the proportion of ant-visited
437 flowering plant species within each network was lower on continents and other islands than in
438 Hawaii (ANOVA: $F_{2,32} = 6.6$, $p < 0.01$, Fig. 4). Within the Hawaiian networks, those with no
439 or only a low proportion of endemic plant species had a similarly low proportion of ant-
440 visited flowering plants as networks on islands in other parts of the world (Fig. 4). The
441 proportion of ant-visited plants increased linearly with the proportion of endemic plant
442 species occurring in the networks (Pearson's $R^2 = 0.93$, $df = 8$, $p < 0.001$, Fig. 4). The
443 proportion of ant-visited flowering plants was independent on the size of the network, i.e.,
444 product of ant and plant species (Pearson's $R^2 = 0.06$, $df = 33$, $p = 0.17$)

445 *Olfactometer trials*
446 In total, we performed 46 olfactometer trials ($n = 771$ individual trials with 6 ants each)
447 where we tested the response of nine ant species to floral scents of nine endemic, eight
448 indigenous and eight introduced plant species (Appendix G). Ants showed 34 negative (10 of

449 them significantly) and 12 positive (none of them significantly) responses (Q_{ij}). On average,
450 introduced plant species emitted volatiles that were stronger ant repellent than those of
451 endemic and indigenous species (ANOVA: $F_{2,21} = 4.0$, $p = 0.034$, Fig. 5). *Pheidole*
452 *megacephala* was the most abundant ant species observed; hence we tested their responses to
453 16 plant species. In these trials, we found the same distinction between plant origins albeit
454 only marginally significant (ANOVA: $F_{2,13} = 3.1$, $p = 0.08$, Fig. 5). Residuals R_{ij} were
455 positively correlated with the response index Q_{ij} (Pearson's $R^2 = 0.13$, $df = 53$, $p < 0.01$),
456 suggesting that olfactory cues influence foraging decisions of ants.

457 *Volatile collection and analysis*

458 In 29 floral scent bouquets analyzed, we found a total of 222 different substances. Most
459 substances were confined to a single plant species (median = 1), whereas 30 substances
460 occurred in seven or more floral scent bouquets. We did not find a consistent pattern that
461 explained the qualitatively different responses towards scents by ants: (a) Total amount of
462 hourly emission neither influenced the mean responses by all ant species $\overline{Q}_{ij}$ nor the responses
463 Q_{ij} of *Pheidole megacephala* alone (Pearsons $R^2 \leq 0.05$, $df = 27, 22$, $p \geq 0.27$). (b) We did not
464 find an individual floral scent compound that explained the variance of Q_{ij} in *P. megacephala*
465 and only one out of 30 that was correlated to the mean ant species' responses $\overline{Q}_{ij}$: the amount
466 of an unidentified sesquiterpene was negatively correlated to $\overline{Q}_{ij}$ (Pearsons $R^2 = 0.70$, $df = 6$, p
467 $= 0.0097$, but note the Bonferroni-corrected $\alpha = 0.05/30 = 0.0017$). (c) After grouping the
468 scent compounds by their biochemical pathway and the presence or absence of functional
469 groups, the compositions of floral scent bouquets of plants that repelled ants were not
470 different from those that did not. The same is true for plant species with mechanical barriers
471 and for plants endemic, indigenous and introduced to the Hawaiian Islands: non-metric
472 multidimensional scaling revealed no distinction between these groups, neither for the

quantitative data nor for the proportional data (Appendix C). Total hourly emission and emission of substances from different classes of compounds are given in Appendix C.

Nectar accessibility and palatability

Narrow nectar tubes prevented nectar access in 32.2 % of all possible interactions, i.e. the head capsules were broader than the tubular width. While most flowers of endemic (78.9 %) and indigenous (72.6 %) plant species granted access to ants by broad nectar tubes, the floral nectar of only 42.2 % of introduced plant species was available to ants that occurred in the same habitats ($\chi^2 = 18.5$, $df = 2$, $p < 0.001$, Fig 6, Appendix G). Apart from narrow nectar tubes, we found three cases of rather unusual mechanical barriers: (1) The calyx of *Plumbago zeylanica* (Plumbaginaceae) possessed very sticky glandular hairs that effectively function as barrier for crawling insects. (2) The calyx of *Abutilon eremitopetalum* (Malvaceae) was covered with dense, fine hairs that prevented ants from reaching the nectaries of these flowers. However, stamens, stigmas and petals of these flowers were often connected with leaves of the same plant or other parts of the surrounding vegetation, resulting in ant visits and associated nectar theft. (3) The inflorescence stalk of *Russelia equisetiformis* (Scrophulariaceae) deterred / repelled ants: *Pheidole megacephala* ants avoided walking on these stems (median = 0 ants min^{-1}) while they readily climbed control sticks (median = 1 ants min^{-1}) in a bioassay (Wilcoxon signed rank test: $V = 21$, $n = 7$, $p = 0.02$). The presence of mechanical barriers (including narrow nectar tubes and the three mechanisms described) effectively suppressed ant visits to floral nectar. Correspondingly, on average, links between ants and flowers where the head capsules were broader than the width of the nectar tube received negative residuals R_{ij} (-0.035 ± 0.009, mean ± se) while the others received positive ones (0.017 ± 0.016, *t*-test: $t_{175} = 3.1$, $p < 0.01$). In five cases, however, one or few ants were recorded on flowers despite a predicted mechanical barrier but it remains unclear whether they reached the nectaries.

Endemic, indigenous and introduced plant species strongly varied in volume and sugar quality of the floral nectar and in the average number of ants per flower (Tab. 2). However, the difference in ant visits on flowers could not be explained by any of the nectar features (Tab. 2; Spearman rank correlations: $R \leq 0.23$, $n = 55$, $p \geq 0.1$). In contrast to olfactory and mechanical mechanisms that may effectively exclude ants from flowers, unpalatable nectar explained only in five out of 43 cases tested negative residuals R_{ij}. The nectar of one endemic (*Gardenia brighamii*, Rubiaceae), two indigenous (*Myoporum sandwicense*, Scrophulariaceae and *Osteomeles anthyllidifolia*, Rosaceae) and one introduced (*Saraca asoca*, Fabaceae) plant species was not consumed by the ant species the nectar was offered to (Appendix G).

Trade off between repellent floral scents and morphological barriers

We found patterns consistent with a trade-off between repellent floral scents and morphological barriers across introduced plant species (logistic regression: $R^2 = 0.45$, $df = 13$, $p < 0.01$, Fig. 7), i.e. many flowers possess either one or the other defensive mechanism. Among flowers of indigenous plants, we did not find such a trade-off ($R^2 = 0.014$, $df = 13$, $p = 0.34$, Fig. 7). For endemic species, we even found a highly significant opposite trend ($R^2 = 0.56$, $df = 17$, $p < 0.001$, Fig. 7). However, this result was strongly influenced by *Hibiscus brackenridgei* subsp. *brackenridgei* (Malvaceae) with both repellent scent and morphological barrier. Apart from this species, only *Nama sandwicensis* (Hydrophyllaceae) possessed morphological barriers among the endemic plants in our study.

Phylogenetic Analysis

The phylogenetic analyses resulted in a Neighbor-Joining tree (Fig. 8) that is comparable to the current classification presented at the NCBI taxonomy database (Sayers et al. 2009). Major clades were clearly separated, supported by high bootstrap values. Evolutionary relationships between orders were in some cases not well resolved so that inter-order

relationships should be regarded with caution (e.g. the clustering of the three clades Ericales, Asterales and Lamiales).

The phylogenetic analysis shows that the endemic, indigenous and introduced plant species in our study are polyphyletic groups, which was also true for the floral features that may or may not promote ant visits (Fig. 8). Distances of mean sugar concentrations in the nectar [% w/w] and nectar volume per flower correlated with the evolutionary distances of the plant species (Mantel statistic $R \geq 0.12$; $p \leq 0.04$). Traits related to protection against ants featured by the plants included in our study (including R_j values, nectar holder tube width and responses to floral scents Q_{ij}) were independent of the evolutionary signal ($R \leq 0.05$; $p \geq 0.23$). Total hourly floral scent emission and the distances of the scent composition (quantitatively and proportionally) did not correlate with phylogenetic distances either ($R \leq 0.1$; $p \geq 0.54$).

Greater rates of ant visitation in endemic species than in introduced species was observed within the Fabaceae, the only plant family were representatives of both endemic and introduced species were available in sufficient replication: The residuals R_{ij} of endemic Fabaceae-species were 0.00 ± 0.01 (mean ± se, n = 18), those of introduced species -0.06 ± 0.02 (n = 11, *t*-test: $t = 3.0$, $p < 0.01$).

Discussion

Our four hypotheses about ant-flower interactions in Hawaii and the underlying mechanisms were confirmed: (1) Ants visited flowers of plant species endemic or indigenous to the Hawaiian Islands more frequently than those of introduced plant species. This was evident on the link and community level: Introduced plants were visited by no or few ants per flower and had negative residuals R_j, while flowers of indigenous and endemic plants were visited by more ants and therefore had positive residuals R_j. Furthermore, in communities with a higher proportion of endemic species, a higher proportion of all plants within the

547 community had ant- visited flowers and the variance of the residuals *var*(R_{ij}) was close to zero
548 in these communities, where ants were distributed across the flowers as expected by the
549 amount of nectar and the ants' abundances. The proportion of plant species with ant-visited
550 flowers in communities that were dominated by endemic plant species was not only
551 exceptionally high compared to other Hawaiian habitats with few or no endemic plant species
552 but also compared to other ecosystems on other oceanic islands or on continents. (2) The poor
553 visitation of introduced plants was the result of more efficient defense mechanisms. Repellent
554 floral scents, morphological barriers and, to a lesser extent, unpalatable nectar each explained
555 negative residuals R_{ij} of ant-flower interactions. (3) We confirmed a trade-off between
556 different floral defense mechanisms as suggested by Willmer et al. (2009) for introduced plant
557 species that often prevented ants from visiting their flowers either by repellent floral scents or
558 by morphological barriers. We did not find such a relationship in endemic and indigenous
559 plant species, which overall showed little evidence of floral defenses. (4) The distribution of
560 plants whose flowers were heavily exploited by ants (positive R_{ij}) among the taxa was found
561 to be independent of the phylogenetic classifications. Therefore, the different susceptibility to
562 floral ant visits of native and introduced plant species was not a result of an inadvertent
563 selection of a phylogenetically narrow or isolated group of study species in this study. Floral
564 defenses against ants are more likely convergently lost in response to prior absence of ants in
565 native Hawaiian ecosystems. In contrast, nectar features (volume and sugar concentration)
566 correlated with the phylogenetic signal.

567 Ants are dominant components of many ecosystems and interact with other organisms of
568 all trophic levels with varying net effects (Lach et al. 2010). Many of those interaction
569 partners adapted to the ecological importance of ants, either by evolving traits that intensify
570 mutualistic interactions (Heil and McKey 2003) or by traits that reduce or even prevent
571 interactions where ants have negative effects (Rico-Gray and Oliveira 2007). Some
572 myrmecophytic Acacias for example have an ambivalent relationship with ants: they benefit

from ants patrolling on their foliage, flower buds and fruits but also profit from keeping ants away from flowers (Willmer and Stone 1997). They succeed in both tasks by offering food and housing to ants (Heil and McKey 2003) and by repelling them from flowers during anthesis (Willmer and Stone 1997, Ghazoul 2001, Willmer et al. 2009). Ant repellent floral scents were documented, or at least suggested, for many plant species from many different regions on nearly all continents (Willmer and Stone 1997, Ghazoul 2001, Junker et al. 2007, Junker and Blüthgen 2008, Willmer et al. 2009). The examples involve different plant life forms, not only myrmecophytes or other plants with a tight relationship to ants. Our result that ants heavily exploit nectar of Hawaiian plant species, while introduced plants that share an evolutionary history with ants are not as affected by these antagonists, suggests that the presence of ants had selected for floral traits that protect this valuable resource (including morphology, scent and nectar features). Accordingly, the ants' selective influence on floral morphology has been suggested in several studies (Herrera et al. 1984, Galen 1999, Galen and Cuba 2001, Galen and Geib 2007). Galen and Cuba (2001), for example, showed that flowers of *Polemonium viscosum* in populations with high densities of nectar-thieving ants are morphologically better defended against these antagonists than flowers in populations with low densities of ants. A similar relationship was suggested for scent-morphs of the same plant species (Galen 1983).

The historic lack of ants in Hawaii is thus likely to be a possible evolutionary cause of ant accessible flowers. However, the fauna of Hawaii lacked – next to ants – also other social hymenopterans and other groups of insects that are common flower visitors in other parts of the world and their absence may have also contributed to the lack of certain floral features. Furthermore, the vacant functional niche of insect-pollinators was often filled by nectarivorous birds (Lammers and Freeman 1986, Gardener and Daehler 2006). A shift from insect to bird pollination may result in changes of floral traits such as scent (Raguso and Pichersky 1995) or morphology (Wilson et al. 2004). In our study, however, bird pollinated

599 flowers were not more heavily exploited by ants than insect pollinated flowers, suggesting
600 that the unusual commonness of bird-pollination in Hawaii did not bias our conclusions.
601 Our study clearly confirms the findings of Junker and Blüthgen (2008) and Willmer et al.
602 (2009) that floral odors often repel ants. However, our methodology using unpicked flowers
603 as scent source, a controlled and constant air stream and the distribution of ants in scented and
604 neutral fields within an olfactometer-arena to measure repellence instead of aggression against
605 manually confronted odor sources (see Willmer et al. 2009) may be even better suited to
606 unequivocally reveal the repellent effect of naturally emitted floral scents. We furthermore
607 demonstrated the ecological significance of the defensive function of floral scents by
608 combining the olfactometer results in a community network analysis. We found, however, no
609 evidence for ant attraction by floral scents. Despite the ants' qualitatively broad spectrum of
610 responses to floral scents, we were unable to depict features of floral scent that are shared by
611 ant-repellent bouquets: ant-repellence could neither be attributed to the total hourly emission,
612 nor the presence of individual volatiles, nor the composition of substances deriving from
613 different biochemical pathways with and without functional groups. We did not find a
614 consistent ant-repellent "syndrome": the composition of ant-repellent floral scent bouquets
615 did not stand out against non-repellent scents, suggesting that the composition is not crucial
616 for defensive functions. Similarly, the attractive function of floral scent compositions is often
617 elicited by one or few substances within complex bouquets. Learned responses are often
618 based on "key odorants" that are required to recognize a reinforced multi-component signal
619 (Laloi et al. 2000, Dötterl et al. 2006, Riffell et al. 2009, Reinhard et al. 2010). However, note
620 that naïve responses may be more pronounced towards blends of volatiles instead of
621 individual substances (Stringer et al. 2008). Several studies demonstrated that several
622 individual substances strongly repel ants (Cane 1986, Kessler and Baldwin 2006, Junker and
623 Blüthgen 2008, Junker and Blüthgen 2010a). Thus, the presence of one ant-repellent
624 substance in a relevant concentration within a complex bouquet may determine the ants'

625 responses. These specific compounds may be notoriously difficult to determine within highly
626 diverse compositions in a multivariate or correlative approach where most substances are
627 emitted by one of few plants only. Potential additive or synergistic effects (Junker and
628 Blüthgen 2008) may even further complicate the detection of clear patterns.

629 Altogether, this study emphasizes that protection is an important function of floral traits
630 (Kerner 1879, Irwin et al. 2004, Junker and Blüthgen 2008, Gomez et al. 2009, Hanley et al.
631 2009, Junker and Blüthgen 2010b, Junker et al. 2010b), and that floral traits operate as filters
632 allowing only a selection of floral visitors to access the rewards (Johnson et al. 2006, Stang et
633 al. 2006, 2007, Raguso 2008a, 2008b). Defenses may also involve energetic costs of
634 synthesizing chemical substances and forming specific floral structures, or costs in terms of
635 loosing potential pollinators that are also negatively affected by these traits. Thus, the trade-
636 off between chemical and mechanical defense mechanisms observed in introduced plant
637 species but not in endemic and indigenous species is also consistent with the hypothesis that
638 ants are selective forces on floral traits.

639 It has been shown that introduced plants and flower visiting animals are well integrated in
640 native interaction networks and often even outnumber the native competitors (Kato et al.
641 1999, Memmott and Waser 2002, Morales and Aizen 2006, Lopezaraiza-Mikel et al. 2007,
642 Vila et al. 2009). The consideration of interactions between native and introduced flowering
643 plants and antagonists (ants) that have not been present prior to their recent introduction
644 provides novel insights in invasional processes. Flower visiting invasive ants can have
645 devastating effects on the reproduction of native plants and their pollinators (Holway et al.
646 2002, Lach 2005, 2007, 2008b, a) suggesting that plants endemic or indigenous to the
647 Hawaiian Islands are negatively affected by nectar feeding ants (especially in pollen limited
648 plants), while introduced plants remain largely unaffected. The success of an introduced
649 species often depends on other non-indigenous species that promote their establishment
650 (Simberloff and Von Holle 1999, Ricciardi 2005), e.g. introduced plants often rely on

introduced pollinators (Richardson et al. 2000). In the Hawaiian scenario, the introduced plants may be indirectly facilitated by introduced ants due to their negative impact on the reproduction of native plants. Thus, the defensive traits featured by the flowers of introduced plants along with the introduction of ants in the same habitat may be disadvantageous for the heavily exploited natives. While the detrimental effects of ants on the Hawaiian arthropod community are well documented (Medeiros et al. 1986, Krushelnycky and Gillespie 2008), their effect on plant communities and their pollinators still needs to be assessed. However, the studies of Lach (2005, 2008b) in combination with our results on the visitation pattern of ants on flowers imply that ant impacts on the Hawaiian flora may be similarly detrimental.

Acknowledgement

We thank Richard Bleil for assistance in the field, Bernice Akamine, Robert P. Borris, Peter Van Dyke, Melody Euaparadorn, Oliver Geissler, Lori Lach, Rhonda Loh, David H. Lorence, Sierra McDaniel, Florian Menzel, Michelle Montgomery, Robert W. Peck, Linda Pratt, Heather Sahli, Norbert Schneider and Karen Shigematsu for organizational, logistical, and technical support. Linda Pratt, Sierra McDaniel and Karen Shigematsu helped with the identification of plants. We thank the United States Department of the Interior (National Park Service), the Bishop Museum Hawaii, the National Tropical Botanical Garden and the University of Hawaii for permits to conduct our study in the Hawaii Volcanoes National Park, the Amy B.H. Greenwell Ethnobotanical Garden, the McBryde Garden and in the Lyon Arboretum, respectively. Natacha Chacoff, Christopher Kaiser-Bunbury and Diego Vázquez kindly provided additional data on ant-flower networks. The project was supported by the Deutsche Forschungsgemeinschaft (DFG, BL 960/1-1), the Evangelisches Studienwerk e.V. Villigst, and the BIGSS graduate school.

676 **Literature**

677 Adams, R. P. 2007. Identification of essential oil components by gas chromatography / mass
678 spectrometry. 4 edition. Allured Publishing Corporation, Carol Stream, Illinois.
679 Adler, L. S. 2000. The ecological significance of toxic nectar. Oikos **91**:409–420.
680 Allison, S. D. and P. M. Vitousek. 2004. Rapid nutrient cycling in leaf litter from invasive
681 plants in Hawai'i. Oecologia **141**:612-619.
682 Amirav, A. and S. Dagan. 1997. A direct sample introduction device for mass spectrometry
683 studies and gas chromatography mass spectrometry analyses. European Mass
684 Spectrometry **3**:105-111
685 Beattie, A. J. 2006. The evolution of ant pollination systems. Botanische Jahrbücher für
686 Systematik **127**:43-55.
687 Beattie, A. J., C. Turnbull, T. Hough, S. Jobson, and R. B. Knox. 1985. The vulnerability of
688 pollen and fungal spores to ant secretions: evidence and some evolutionary
689 implications. American Journal of Botany **72**:606-614.
690 Beattie, A. J., C. Turnbull, R. B. Knox, and E. G. Williams. 1984. Ant inhibition of pollen
691 function: a possible reason why ant pollination is rare. American Journal of Botany
692 **71**:421-426.
693 Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2009. GenBank.
694 Nucleic Acids Research **37**:D26-D31.
695 Blüthgen, N. and K. Fiedler. 2004. Preferences for sugars and amino acids and their
696 conditionality in a diverse nectar-feeding ant community. Journal of Animal Ecology
697 **73**:155-166.
698 Blüthgen, N. and K. Reifenrath. 2003. Extrafloral nectaries in an Australian rainforest:
699 structure and distribution. Australian Journal of Botany **51**:515-527.

Blüthgen, N., N. E. Stork, and K. Fiedler. 2004. Bottom-up control and co-occurrence in complex communities: honeydew and nectar determine a rainforest ant mosaic. Oikos **106**:344-358.

Bonser, R., P. J. Wright, S. Bament, and U. O. Chukwu. 1998. Optimal patch use by foraging workers of *Lasius fuliginosus*, *L. niger* and *Myrmica ruginodis*. Ecological Entomology **23**:15-21.

Cane, J. H. 1986. Predator deterrence by mandibular gland secretions of bees (Hymenoptera, Apoidea). Journal of Chemical Ecology **12**:1295-1309.

Daly, H. V. and K. N. Magnacca. 2003. Insects of Hawaii. University of Hawaii Press, Honolulu.

de Vega, C., M. Arista, P. L. Ortiz, C. M. Herrera, and S. Talavera. 2009. The ant-pollination system of *Cytinus hypocistis* (Cytinaceae), a Mediterranean root holoparasite. Annals of Botany **103**:1065-1075.

Denslow, J. S. 2003. Weeds in paradise: thoughts on the invasibility of tropical islands. Annals of the Missouri Botanical Garden **90**:119-127.

Dötterl, S., A. Jürgens, K. Seifert, T. Laube, B. Weißbecker, and S. Schütz. 2006. Nursery pollination by a moth in *Silene latifolia*: the role of odours in eliciting antennal and behavioural responses. New Phytologist **169**:707-718.

Dötterl, S., A. Jürgens, L. Wolfe, and A. Biere. 2009. Disease status and population origin effects on floral scent: potential consequences for oviposition and fruit predation in a complex interaction between a plant, fungus, and noctuid moth. Journal of Chemical Ecology **35**:307-319.

Dötterl, S., L. M. Wolfe, and A. Jurgens. 2005. Qualitative and quantitative analyses of flower scent in *Silene latifolia*. Phytochemistry **66**:203-213.

Feinsinger, P. and L. A. Swarm. 1978. How common are ant-repellent nectars? Biotropica **10**:238-239.

726 Fretwell, S. and H. L. Lucas. 1970. On territorial behavior and other factors influencing
727 habitat distribution in birds. I. Theoretical Development. Acta Biotheoretica **19**:16-36.
728 Galen, C. 1983. The effects of nectar thieving ants on seedset in floral scent morphs of
729 *Polemonium viscosum*. Oikos **41**:245-249.
730 Galen, C. 1999. Flowers and enemies: predation by nectar-thieving ants in relation to
731 variation in floral form of an alpine wildflower, *Polemonium viscosum*. Oikos **85**:426-
732 434.
733 Galen, C. and B. Butchart. 2003. Ants in your plants: effects of nectar-thieves and pollen
734 fertility and seed-siring capacity in the alpine wildflower, *Polemonium viscosum*.
735 Oikos **101**:521 - 528.
736 Galen, C. and J. Cuba. 2001. Down the tube: pollinators, predators, and the evolution of
737 flower shape in the Alpine Skypilot, *Polemonium viscosum*. Evolution **55**:1963-1971.
738 Galen, C. and J. C. Geib. 2007. Density-dependent effects of ants on selection for bumble bee
739 pollination in *Polemonium viscosum*. Ecology **88**:1202–1209.
740 Gardener, M. C. and C. C. Daehler. 2006. Documenting floral visitors to rare Hawaiian plants
741 using automated video recordings. Pacific Conservation Biology **12**:189-194.
742 Ghazoul, J. 2001. Can floral repellents pre-empt potential ant-plant conflicts? Ecology Letters
743 **4**:295 - 299.
744 Gomez, J. M. 2000. Effectiveness of ants as pollinators of *Lobularia maritima*: effects on
745 main sequential fitness components of the host plant. Oecologia **122**:90-97.
746 Gomez, J. M., F. Perfecti, J. Bosch, and J. P. M. Camacho. 2009. A geographic selection
747 mosaic in a generalized plant–pollinator–herbivore system. Ecological Monographs
748 **79**:245-263.
749 Gomez, J. M. and R. Zamora. 1992. Pollination by ants: consequences of quantitative effects
750 on a mutualistic system. Oecologia **91**:410-418.

751 Gomez, J. M., R. Zamora, J. A. Hodar, and D. Garcia. 1996. Experimental study of
752 pollination by ants in mediterranean high mountain and arid habitats. Oecologia
753 **105**:236-242.
754 Guerrant, E. O., Jr and P. G. Fiedler. 1981. Flower defenses against nectar-pilferage by ants.
755 Biotropica **13**:25-33.
756 Haber, W. A., G. W. Frankie, H. G. Baker, I. Baker, and S. Koptur. 1981. Ants like flower
757 nectar. Biotropica **13**:211-214.
758 Hanley, M. E., B. B. Lamont, and W. S. Armbruster. 2009. Pollination and plant defence
759 traits co-vary in Western Australian Hakeas. New Phytologist **182**:251-260.
760 Harley, R. 1991. The greasy pole syndrome. Pages 430-433 *in* C. R. Huxley and D. F. Cutler,
761 editors. Ant-Plant Interactions. Oxford University Press, Oxford.
762 Heil, M. and D. McKey. 2003. Protective ant-plant interactions as model systems in
763 ecological and evolutionary research. Annual Review of Ecology Evolution and
764 Systematics **34**:425-453.
765 Herrera, C. M., J. Herrera, and X. Espadaler. 1984. Nectar thievery by ants from southern
766 Spanish insect-pollinated flowers. Insectes Sociaux **31**:142-154.
767 Holway, D. A., L. Lach, A. V. Suarez, N. D. Tsutsui, and T. J. Case. 2002. The causes and
768 consequences of ant invasions. Annual Review of Ecology and Systematics **33**:181-
769 233.
770 Irwin, R. E., L. S. Adler, and A. K. Brody. 2004. The dual role of floral traits: Pollinator
771 attraction and plant defense. Ecology **85**:1503-1511.
772 Janzen, D. H. 1977. Why don't ants visit flowers? Biotropica **9**:252.
773 Johnson, S. D., A. L. Hargreaves, and M. Brown. 2006. Dark, bitter-tasting nectar functions
774 as a filter of flower visitors in a bird-pollinated plant. Ecology **87**:2709–2716.

775 Junker, R., A. Y. C. Chung, and N. Blüthgen. 2007. Interaction between flowers, ants and
776 pollinators: additional evidence for floral repellence against ants. Ecological Research
777 **22**:665–670.
778 Junker, R. R., R. Bleil, C. C. Daehler, and N. Blüthgen. 2010a. Intrafloral resource
779 partitioning between endemic and invasive flower visitors: consequences for
780 pollinator effectiveness. Ecological Entomology, in press.
781 Junker, R. R. and N. Blüthgen. 2008. Floral scents repel potentially nectar-thieving ants.
782 Evolutionary Ecology Research **10**:295-308.
783 Junker, R. R. and N. Blüthgen. 2010a. Dependency on floral resources determines the
784 animals' responses to floral scents. Plant Signaling and Behavior **5**:1014-1016.
785 Junker, R. R. and N. Blüthgen. 2010b. Floral scents repel facultative flower visitors, but
786 attract obligate ones. Annals of Botany **105**:777-782.
787 Junker, R. R., N. Höcherl, and N. Blüthgen. 2010b. Responses to olfactory signals reflect
788 network structure of flower-visitor interactions. Journal of Animal Ecology **79**:818-
789 823.
790 Kato, M., A. Shibata, T. Yasui, and H. Nagamasu. 1999. Impact of introduced honeybees,
791 *Apis mellifera*, upon native bee communities in the Bonin (Ogasawara) Islands.
792 Researches on Population Ecology **41**:217-228.
793 Keeler, K. H. 1985. Extrafloral nectaries on plants in communities without ants: Hawaii.
794 Oikos **44**:407-414.
795 Keller, A., F. Forster, T. Muller, T. Dandekar, J. Schultz, and M. Wolf. 2010. Including RNA
796 secondary structures improves accuracy and robustness in reconstruction of
797 phylogenetic trees. Biology Direct **5**:4.
798 Keller, A., T. Schleicher, F. Forster, B. Ruderisch, T. Dandekar, T. Muller, and M. Wolf.
799 2008. ITS2 data corroborate a monophyletic chlorophycean DO-group
800 (Sphaeropleales). BMC Evolutionary Biology **8**:218.

Keller, A., T. Schleicher, J. Schultz, T. Muller, T. Dandekar, and M. Wolf. 2009. 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. Gene **430**:50-57.

Kerner, A. 1879. Die Schutzmittel der Blüthen gegen unberufene Gäste. Verlag der Wagner'schen Universitäts-Buchhandlung, Innsbruck.

Kessler, D. and I. T. Baldwin. 2006. Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata*. The Plant Journal **49**:840-854.

Koetschan, C., F. Förster, A. Keller, T. Schleicher, B. Ruderisch, R. Schwarz, T. Müller, M. Wolf, and J. Schultz. 2009. The ITS2 Database III—sequences and structures for phylogeny. Nucleic Acids Research.

Krushelnycky, P. D. and R. G. Gillespie. 2008. Compositional and functional stability of arthropod communities in the face of ant invasions. Ecological Applications **18**:1547-1562.

Krushelnycky, P. D., L. L. Loope, and N. J. Reimer. 2005. The ecology, policy, and management of ants in Hawaii. Proceedings of the Hawaiian Entomological Society. **37**:1-25.

Lach, L. 2005. Interference and exploitation competition of three nectar-thieving invasive ant species. Insectes Sociaux **52**:257-262.

Lach, L. 2007. A mutualism with a native membracid facilitates pollinator displacement by Argentine ants. Ecology **88**:1994-2004.

Lach, L. 2008a. Argentine ants displace floral arthropods in a biodiversity hotspot. Diversity and Distributions **14**:281-290.

Lach, L. 2008b. Floral visitation patterns of two invasive ant species and their effects on other hymenopteran visitors. Ecological Entomology **33**:155-160.

Lach, L., C. Parr, and K. Abbott, editors. 2010. Ant ecology. Oxford University Press, Oxford.

827 Laloi, D., O. Bailez, M. M. Blight, B. Roger, M.-H. Pham-Delegue, and L. J. Wadhams.
828 2000. Recognition of complex odors by restrained and free-flying honeybees, *Apis*
829 *mellifera*. Journal of Chemical Ecology **26**:2307-2319.
830 Lammers, T. G. and C. E. Freeman. 1986. Ornithophily among the Hawaiian *Lobelioideae*
831 (Campanulaceae) - evidence from floral nectar sugar compositions. American Journal
832 of Botany **73**:1613-1619.
833 Lonsdale, W. M. 1999. Global patterns of plant invasions and the concept of invasibility.
834 Ecology **80**:1522-1536.
835 Lopezaraiza-Mikel, M. E., R. B. Hayes, M. R. Whalley, and J. Memmott. 2007. The impact of
836 an alien plant on a native plant-pollinator network: an experimental approach. Ecology
837 Letters **10**:539-550.
838 Magnacca, K. N. and B. N. Danforth. 2007. Low nuclear DNA variation supports a recent
839 origin of Hawaiian *Hylaeus* bees (Hymenoptera : Colletidae). Molecular
840 Phylogenetics and Evolution **43**:908-915.
841 Mathews, D. H., M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner.
842 2004. Incorporating chemical modification constraints into a dynamic programming
843 algorithm for prediction of RNA secondary structure. Proceedings of the National
844 Academy of Sciences of the United States of America **101**:7287-7292.
845 Medeiros, A. C., L. L. Loope, and F. R. Cole. 1986. Distribution of ants and their effects on
846 endemic biota of Haleakala and Hawaii Volcanoes National Park: a preliminary
847 assessment. Pages 39-52 *in* Proceedings 6th Conference in natural sciences, Hawaii
848 Volcanoes National Park.
849 Memmott, J. and N. M. Waser. 2002. Integration of alien plants into a native flower-pollinator
850 visitation web. Proceedings of the Royal Society of London Series B-Biological
851 Sciences **269**:2395-2399.

852 Mooney, H. A. 2005. Invasive alien species: the nature of the problem.*in* H. A. Mooney, R.
853 N. Mack, J. A. McNeely, L. E. Neville, P. J. Schei, and J. K. Waage, editors. Invasive
854 alien species. Island Press, Washington, Covelo, London.
855 Mooney, H. A., R. N. Mack, J. A. McNeely, L. E. Neville, P. J. Schei, and J. K. Waage,
856 editors. 2005. Invasive alien species. Island Press, Washington, Covelo, London.
857 Morales, C. L. and M. A. Aizen. 2006. Invasive mutualisms and the structure of plant-
858 pollinator interactions in the temperate forests of north-west Patagonia, Argentina.
859 Journal of Ecology **94**:171-180.
860 Pettersson, J. 1970. An aphid sex attractant. I. Biological studies. Entomologica Scandinavica
861 **1**:63-73.
862 Pijl, L. v. d. 1955. Some remarks on myrmecophytes. Phytomorphology **5**:190-200.
863 Raguso, R. A. 2004. Why are some floral nectars scented? Ecology **85**:1486-1494.
864 Raguso, R. A. 2008a. Start making scents: the challenge of integrating chemistry into
865 pollination ecology. Entomologia Experimentalis et Applicata **128**:196-207.
866 Raguso, R. A. 2008b. Wake up and smell the roses: The ecology and evolution of floral scent.
867 Annual Review of Ecology, Evolution and Systematics **39**:549-569.
868 Raguso, R. A. and E. Pichersky. 1995. Floral volatiles from *Clarkia breweri* and *C. concinna*
869 (Onagraceae): recent evolution of floral scent and moth pollination. Plant Systematics
870 and Evolution **194**:55-67.
871 Reimer, N. J. 2006. Key to the ants of Hawaii. Plant Pest Control Branch, Hawaii Department
872 of Agriculture.
873 Reinhard, J., M. Sinclair, M. V. Srinivasan, and C. Claudianos. 2010. Honeybees learn odour
874 mixtures via a selection of key odorants. Plos One **5:**e9110.
875 Ricciardi, A. 2005. Facilitation and synergistic interactions between introduced aquatic
876 species. Pages 162-178 *in* H. A. Mooney, R. N. Mack, J. A. McNeely, L. E. Neville,

P. J. Schei, and J. K. Waage, editors. Invasive Alien Species. Island press, Washington, Covelo, London.

Richardson, D. M., N. Allsopp, C. M. D'Antonio, S. J. Milton, and M. Rejmanek. 2000. Plant invasions - the role of mutualisms. Biological Reviews **75**:65-93.

Rico-Gray, V. and P. S. Oliveira. 2007. The ecology and evolution of ant-plant interactions. The University of Chicago Press, Chicago.

Riffell, J. A., H. Lei, T. A. Christensen, and J. G. Hildebrand. 2009. Characterization and coding of behaviorally significant odor mixtures. Current Biology **19**:335-340.

Sayers, E. W., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, E. Yaschenko, and J. Ye. 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research **37**:D5-D15.

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

Seibel, P., T. Müller, T. Dandekar, and M. Wolf. 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Research Notes **1**:91.

Simberloff, D. and B. Von Holle. 1999. Positive interactions of nonindigenous species: invasional meltdown? Biological Invasions **1**:21-32.

Stang, M., P. G. L. Klinkhamer, and E. v. d. Meijden. 2006. Size constraints and flower abundance determine the number of interactions in a plant-flower visitor web. Oikos **112**:111-121.

903 Stang, M., P. G. L. Klinkhamer, and E. v. d. Meijden. 2007. Asymmetric specialization and
904 extinction risk in plant–flower visitor webs: a matter of morphology or abundance?
905 Oecologia **151**:442-453.
906 Stone, C. P. and J. M. Scott, editors. 1985. Hawaii's terrestrial ecosystems: Preservation and
907 management. University of Hawaii, Honolulu.
908 Stone, C. P., C. W. Smith, and J. T. Tunison, editors. 1992. Alien plant invasions in native
909 ecosystems of Hawaii: Management and research. University of Hawaii Press,
910 Honolulu.
911 Stringer, L. D., A. M. El-Sayed, L. M. Cole, L. A. M. Manning, and D. M. Suckling. 2008.
912 Floral attractants for the female soybean looper, *Thysanoplusia orichalcea*
913 (Lepidoptera: Noctuidae). Pest Management Science **64**:1218-1221.
914 Symstad, A. J. 2000. A test of the effects of functional group richness and composition on
915 grassland invasibility. Ecology **81**:99-109.
916 Taylor, F. 1977. Foraging behavior of ants - experiments with 2 species of Myrmecine ants.
917 Behavioral Ecology and Sociobiology **2**:147-167.
918 Tilman, D. 1997. Community invasibility, recruitment limitation, and grassland biodiversity.
919 Ecology **78**:81-92.
920 Tsuji, K., A. Hasyim, H. Nakamura, and K. Nakamura. 2004. Asian weaver ants, *Oecophylla*
921 *smaragdina*, and their repelling of pollinators. Ecological Research **19**:669-673.
922 Vet, L. E. M., J. C. Van Lenteren, M. Heymans, and E. Meelis. 1983. An airflow olfactometer
923 for measuring olfactory responses of hymenopterous parasitoids an other small
924 insects. Physiological Entomology **8**:97-106.
925 Vila, M., I. Bartomeus, A. C. Dietzsch, T. Petanidou, I. Steffan-Dewenter, J. C. Stout, and T.
926 Tscheulin. 2009. Invasive plant integration into native plant-pollinator networks
927 across Europe. Proceedings of the Royal Society B-Biological Sciences **276**:3887-
928 3893.

929 Wagner, W. L., D. R. Herbst, and S. H. Sohmer. 1990. Manual of the flowering plants of
930 Hawai'i. Bishop Museum, Honolulu.
931 Wetterer, J. K. 1998. Nonindigenous ants associated with geothermal and human disturbance
932 in Hawai'i Volcanoes National Park. Pacific Science **52**:40-50.
933 Wheeler, W. M. 1934. Revised list of Hawaiian ants. Bishop Museum Occasional Papers
934 **10**:1-21.
935 Willmer, P. G., C. V. Nuttman, N. E. Raine, G. N. Stone, J. G. Pattrick, K. Henson, P.
936 Stillman, L. McIlroy, S. G. Potts, and J. T. Knudsen. 2009. Floral volatiles controlling
937 ant behaviour. Functional Ecology **23**:888-900.
938 Willmer, P. G. and G. N. Stone. 1997. How aggressive ant-guards assist seed-set in Acacia
939 flowers. Nature **388**:165-167.
940 Wilson, P., M. C. Castellanos, J. N. Hogue, J. D. Thomson, and W. S. Armbruster. 2004. A
941 multivariate search for pollination syndromes among penstemons. Oikos **104**:345-361.
942 Wolf, M., B. Ruderisch, T. Dandekar, J. Schultz, and T. Muller. 2008. ProfDistS: (profile-)
943 distance based phylogeny on sequence-structure alignments. Bioinformatics **24**:2401-
944 2402.
945
946
947

APPENDIX A

Complete list of plant species used in this study including information on their origin (endemic, indigenous or introduced) and their typical pollinators.

APPENDIX B

Mobile olfactometer - technical details and pictures.

APPENDIX C

Floral scent sampling and results of NMDS.

APPENDIX D

GenBank accession numbers and representative taxa for phylogenetic analysis.

APPENDIX E

Ant-flower networks of all 10 study sites.

APPENDIX F

Sources for the flower-visitor networks from other islands and continents.

APPENDIX G

Potential links encountered in the ten habitats, nectar tube width, head capsule width, olfactometer results and nectar palatability.

SUPPLEMENT

Commands for R (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) for Monte-Carlo statistics for the calculation of significance levels of the deviation of the Residuals R_{ij} from zero.

968 **Tables**

969 **Table 1** Study sites on the Hawaiian Islands. Names and locations of the study sites, the altitude [m above sea level] and number of ant and plant
970 species are given.

#	Island	Location	Geographic coordinates	Altitude	Ant species	Plant species
1	Big Island	Amy B.H. Greenwell Ethnobotanical Garden	N19°29.5 W155°54.7	461	7	10
2	Big Island	Hawaii Volcanoes National Park	N19°17.5 W155°08.7	96	4	3
3	Big Island	Hawaii Volcanoes National Park	N19°26.2 W155°17.9	1240	1	2
4	Big Island	Hawaii Volcanoes National Park	N19°20.7 W155°12.7	901	3	2
5	Big Island	Hawaii Volcanoes National Park	N19°19.9 W155°16.7	901	1	3
6	Big Island	Hawaii Volcanoes National Park	N19°17.6 W155°05.9	17	2	5
7	Kauai	McBryde Garden	N21°54.3 W159°30.5	61	4	9
8	Oahu	Sandy Beach	N21°17.5 W157°39.7	20	5	6
9	Oahu	University of Hawaii at Manoa	N21°18.1 W157°48.9	71	1	10
10	Oahu	Lyon Arboretum	N21°19.9 W157°48.2	177	3	6

971

Table 2 Nectar features and average number of ants per flower of endemic, indigenous and introduced plant species. Median (interquartile range) and results of Kruskal – Wallis ANOVA are shown. Significant values are bold.

	Volume [µl]	**Mass % w/w**	**Volume * Mass %**	**Ants flower^{-1}**
Endemic	8.3 (1.2 – 15.2)	17.2 (10.7 – 19.5)	1.1 (0.1 – 2.7)	0.12 (0.03 – 1.1)
Indigenous	0.6 (0.2 – 1.1)	23.6 (13.8 – 36.5)	0.1 (0.04 – 0.22)	0.03 (0.0 – 0.1)
Introduced	2.0 (0.5 – 5.6)	18.2 (14.3 – 22.5)	0.3 (0.09 – 1.2)	0.0 (0.0 – 0.3)
χ^2	13.5	5.6	8.4	6.9
p	**< 0.01**	0.06	**0.015**	**0.032**

977 **Figure Legends**

978 **Fig. 1** Hierarchical framework that summarizes prerequisites and potential barriers for
979 exploitation of floral nectar by ants. Steps A – C are sequentially encountered by ants
980 approaching floral nectar. The order of B_1 and B_2 depends on the morphology of the flower,
981 thus either the solid or the dashed path can be followed from A to C. C_1 and C_2 operate
982 simultaneously.

983 **Fig. 2** Mobile olfactometer with flowmeters, Teflon tubes directing the scented air from the
984 flowers towards the arena and the inflated oven bags covering the flowers (*Ipomoea pes-*
985 *caprae*). Inlet shows shape and dimensions of the arena.

986 **Fig. 3** Residuals R_j of plant species that are endemic, indigenous and introduced to the
987 Hawaiian Islands. Mean and 95 % confidence intervals are shown. Positive residuals R_j denote
988 interactions that occurred more frequently than expected by the null model, negative R_j less
989 frequent ones than expected. Black bars denote R_j from plant species that were sympatric to at
990 least one species of a different origin (i.e. endemic, indigenous or introduced). Letters indicate
991 significant differences according to pairwise *t*-tests. Difference between indigenous and
992 introduced plants remained significant after Bonferroni-correction. White bars denote R_j from
993 all plants in our study regardless the plant community. Sample size of each group is given in
994 bars.

995 **Fig. 4** Proportion (mean and 95% confidence intervals) of plant species with ant-visited
996 flowers in Hawaii and other on islands or continents. Letters indicate significant differences
997 according to pair wise *t*-tests based on arcsin-sqrt transformed data. Differences remained
998 significant after Bonferroni-correction. Additionally, proportion of plant species with ant-
999 visited flowers as a function of the proportion of endemic plant species within the Hawaiian

networks is shown. Closed circles denote to two overlapping points. Sample size is given above bars.

Fig. 5 Response indices Q_{ij} of ants toward the floral scent of plant species that are endemic, indigenous and introduced to the Hawaiian Islands. Shown are mean and 95 % confidence intervals. Black bars denote response indices from olfactometer trials with various ant species. In cases where plant species were tested with two ore more ant species, the mean value of Q_{ij} was taken (total number of olfactometer trials: 46). Letters indicate significant differences according to pair wise *t*-tests. Differences between indigenous and introduced plants remained significant after Bonferroni-correction. White bars denote response indices from trials with *Pheidole megacephala* only. Sample size of each group is given in bars.

Fig. 6 Proportion of links between ants and flowers of plants that are endemic, indigenous or introduced to the Hawaiian Islands that are prevented by mechanical barriers. Letters indicate significant differences according to pair wise *Chi-square* tests.

Fig. 7 Trade off between repellent floral scents and morphological barriers. Negative Q_{ij} values indicate repellence, positive attraction. Morphological barriers are either present (1) or absent (0) from the flowers. Trait combination for endemic (open circles), indigenous (open triangles) and introduced plant species (filled circles) and logistic regression for introduced plant species is shown.

Fig. 8 Evolutionary relationship of plants species encountered in the habitats studied as revealed by Profile Neighbor Joining with ITS2-RNA sequences and structures. Given are bootstrap values (1000 replicates), order (according to NCBI taxonomy, Sayers et al. 2009), family[1] and origin (endemic, indigenous and introduced) of each plant species. Additionally, mean residuals R_j are given. Plant species that were substituted by representatives for the

1023 phylogenetic analysis are marked with asterisks (see Appendix D). Species with more than
1024 one sequence in the analysis were collapsed in the tree for clarity.
1025 Footnote: [1] ACA = Acanthaceae; CAM = Campanulaceae; COM = Combretaceae; CON =
1026 Convolvulaceae; ERI = Ericaceae; FAB = Fabaceae; GOO = Goodeniaceae; HYD =
1027 Hydrophyllaceae; LAM = Lamiaceae; MAL = Malvaceae; MUS = Musaceae; MYO =
1028 Myoporaceae; MYR = Myrtaceae; NYC = Nyctaginaceae; PLU = Plumbaginaceae; ROS =
1029 Rosaceae; RUB = Rubiaceae; SCR = Scrophulariaceae; TUR = Turneraceae; VER =
1030 Verbenaceae.

Figures

Fig. 1

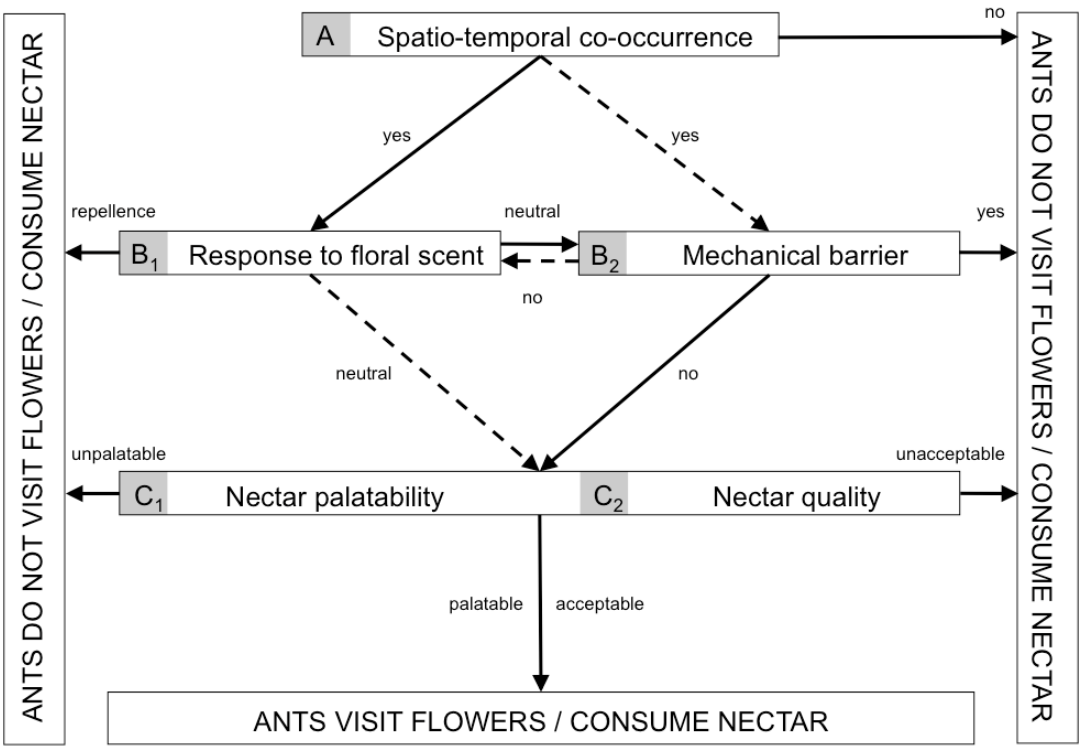


Fig. 2

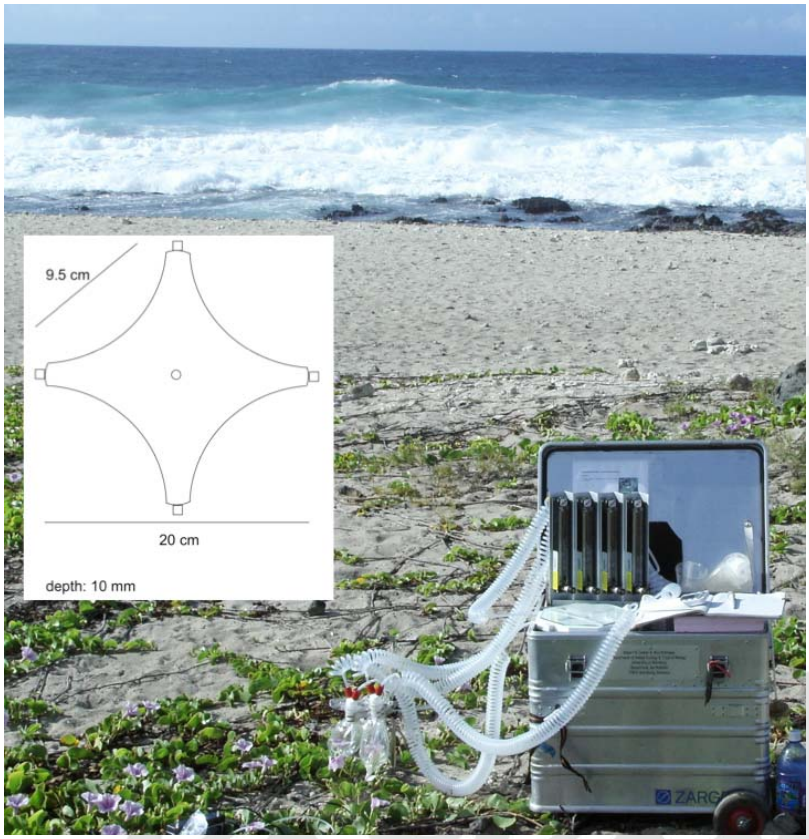


Fig. 3

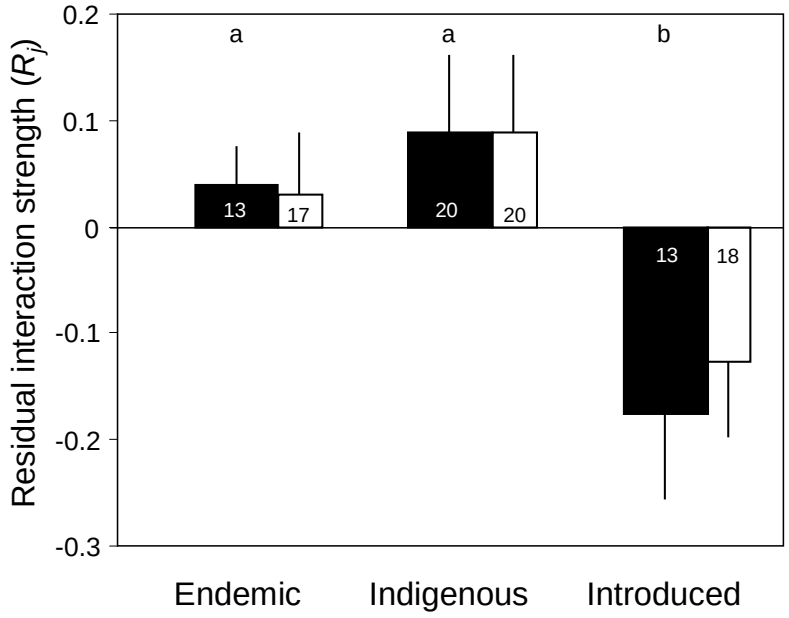


Fig. 4

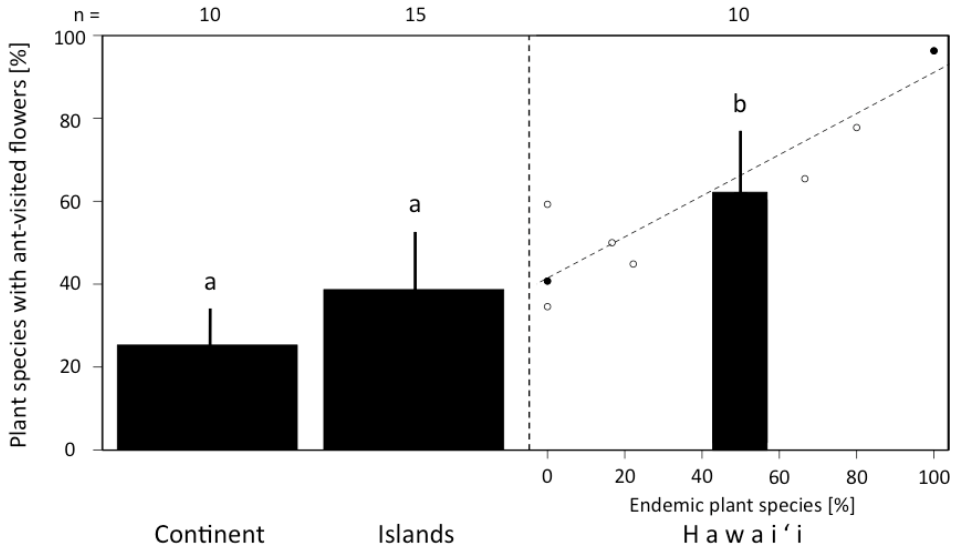


Fig. 5

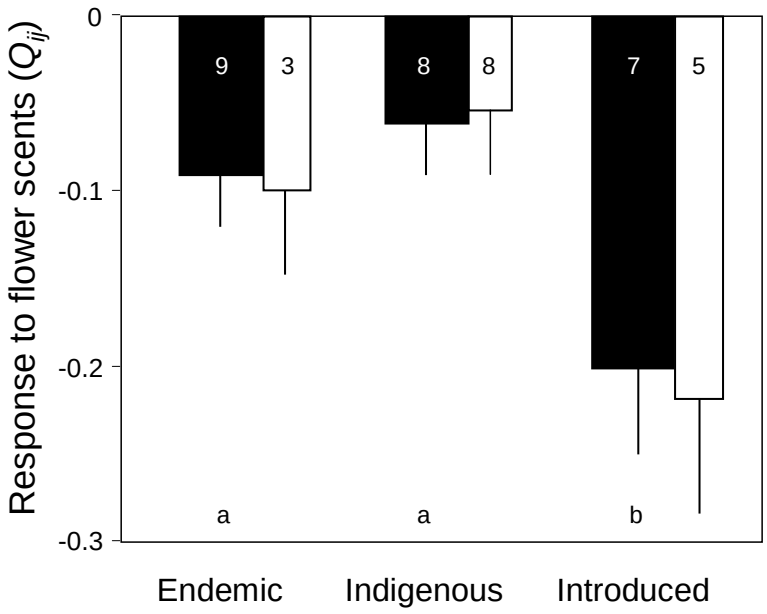


Fig. 6

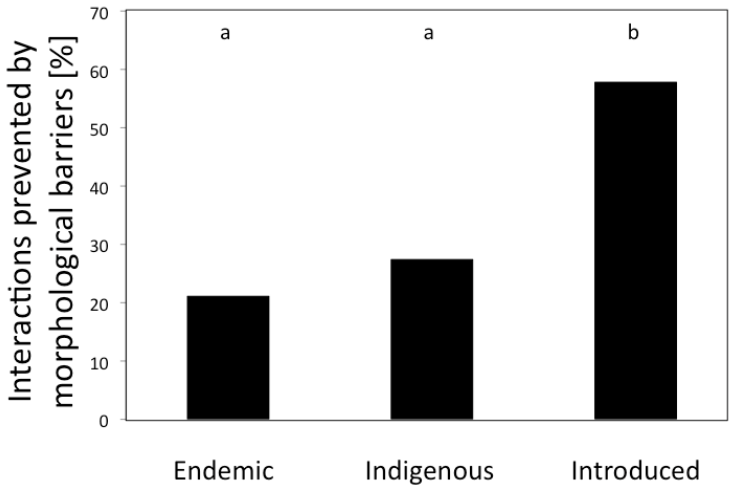


Fig. 7

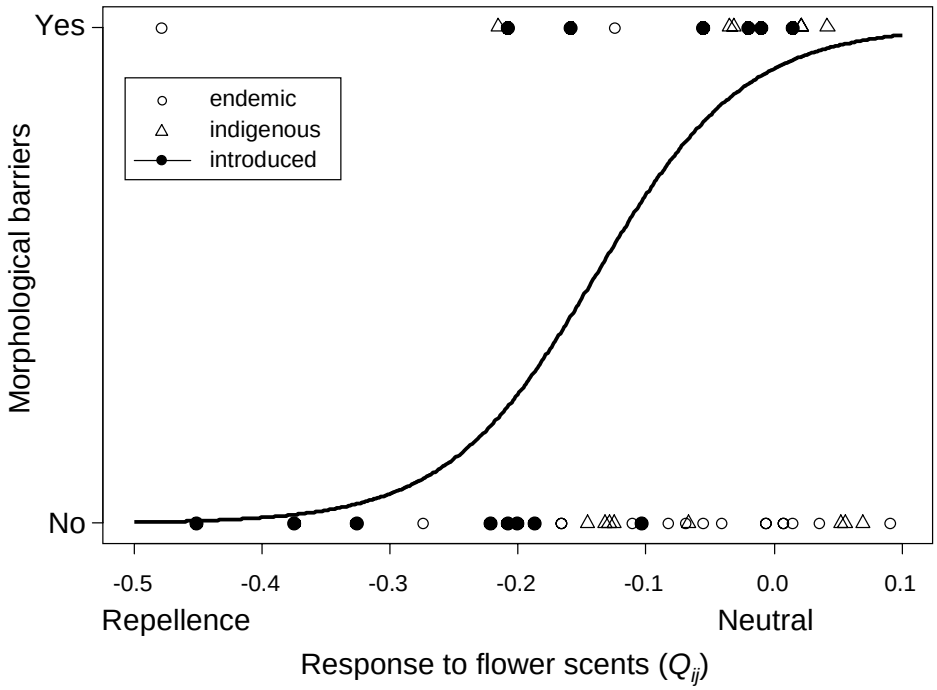
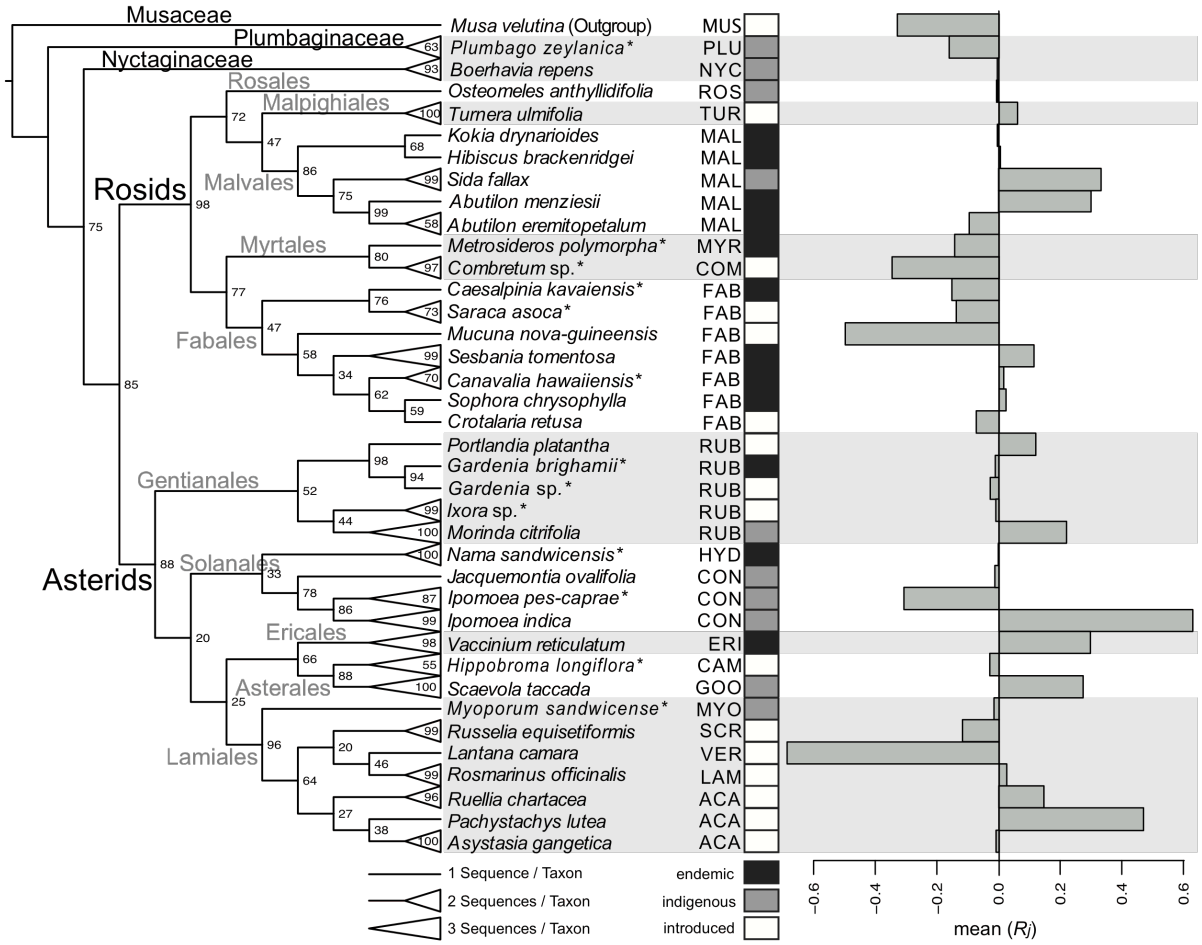


Fig. 8


Musaceae
Plumbaginaceae
Nyctaginaceae
Rosales
Malpighiales
Rosids
Malvales
Myrtales
Fabales
Gentianales
Asterids
Solanales
Ericales
Asterales
Lamiales
Musa velutina (Outgroup) MUS
Plumbago zeylanica* PLU
Boerhavia repens NYC
Osteomeles anthyllidifolia ROS
Turnera ulmifolia TUR
Kokia drynarioides MAL
Hibiscus brackenridgei MAL
Sida fallax MAL
Abutilon menziesii MAL
Abutilon eremitopetalum MAL
Metrosideros polymorpha* MYR
Combretum sp.* COM
Caesalpinia kavaiensis* FAB
Saraca asoca* FAB
Mucuna nova-guineensis FAB
Sesbania tomentosa FAB
Canavalia hawaiiensis* FAB
Sophora chrysophylla FAB
Crotalaria retusa FAB
Portlandia platantha RUB
Gardenia brighamii* RUB
Gardenia sp.* RUB
Ixora sp.* RUB
Morinda citrifolia RUB
Nama sandwicensis* HYD
Jacquemontia ovalifolia CON
Ipomoea pes-caprae* CON
Ipomoea indica CON
Vaccinium reticulatum ERI
Hippobroma longiflora* CAM
Scaevola taccada GOO
Myoporum sandwicense* MYO
Russelia equisetiformis SCR
Lantana camara VER
Rosmarinus officinalis LAM
Ruellia chartacea ACA
Pachystachys lutea ACA
Asystasia gangetica ACA
1 Sequence / Taxon
2 Sequences / Taxon
3 Sequences / Taxon
endemic
indigenous
introduced
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
mean (R_i)

P.9. Composition of epiphytic bacterial communities differs on flowers and leaves

Authors: R.R. Junker, C. Loewel, R. Gross, S. Dötterl, A. Keller*, N. Blüthgen

to be submitted 2010

*Co-author

Author's Contributions:
R.R. Junker, N. Blüthgen, R. Gross and S. Dötterl designed the study. C. Loewel and R.R. Junker collected samples in the field. C. Loewel determined the new sequences in the laboratory and performed the Agar diffusion assays. S. Dötterl peformed the scent-analyses. R.R. Junker made the statistical analyses and drafted the manuscript with contributions by me and C. Loewel. I performed the bioinformatical processing of the sequences and structures as well as the phylogenetic analyses.

Composition of epiphytic bacterial communities differs on flowers and leaves

Robert R. Junker[1], Christina Loewel[1], Roy Gross[2], Stefan Dötterl[3], Alexander Keller[4] and Nico Blüthgen[1,5]

[1]University of Würzburg, Department of Animal Ecology & Tropical Biology, Biozentrum, Am Hubland, 97074 Würzburg, Germany

[2]University of Würzburg, Department of Microbiology, Biozentrum, Am Hubland, 97074 Würzburg, Germany

[3]University of Bayreuth, Department of Plant Systematics, Universitätsstraße 30, 95447 Bayreuth, Germany

[4]University of Würzburg, Department of Bioinformatics, Biozentrum, Am Hubland, 97074 Würzburg, Germany

[5]Correspondence: bluethgen@biozentrum.uni-wuerzburg.de

Abstract

The epiphytic bacterial communities colonizing roots and leaves have been described for many plant species. The ephemeral floral surfaces of naturally growing plants have rarely been considered by microbiologists. We identified bacteria isolated from petals and leaves of two plant species, *Saponaria officinalis* (Caryophyllaceae) and *Lotus corniculatus* (Fabaceae). The bacterial diversity was much lower on flowers than on leaves and the compositions on the plant organs were different: while Pseudomonadaceae and Microbacteriaceae were the most abundant families on leaves, Enterobacteriaceae dominated floral communities. We hypothesize that antibacterial floral volatiles trigger the low diversity on petals, which is supported by agar diffusion assays using substances emitted by flowers and leaves of *S.*

officinalis. These results suggest that bacteria should be included in the interpretation of floral traits and bacterial effects on pollination biology are proposed and discussed.

Introduction

Above ground plant surfaces provide diverse habitats for bacterial colonists. Environmental factors and specific features of the plant organs determine the character of these surfaces and thus may affect the composition of the bacterial communities (Andrews and Harris 2000). The establishment and the growth of bacteria strongly depends on the availability of nutrients that may be variable on a macroscopic level (different plant parts) and on a microscopic level where nutrients are heterogeneously distributed within small areas (Andrews and Harris 2000, Mercier and Lindow 2000). Next to nutrients, the emission or secretion of secondary metabolites that either inhibit or facilitate bacterial growth may have an impact on the distribution of bacteria on different plant parts (Bednarek and Osbourn 2009). This notion is supported by numerous studies that investigated the antibacterial properties of essential oils (Harrewijn et al. 1995, Lokvam and Braddock 1999, Velickovic et al. 2003, Gershenzon and Dudareva 2007, Tomczykowa et al. 2008).

Besides roots that may be the best examined plant part regarding its associated bacteria (Andrews and Harris 2000), leaves were often the target of microbiologists that isolated and identified the microbial taxa dwelling on them. The most common bacteria found on leaves are representatives of the families Enterobacteriaceae, Pseudomonadaceae and Microbacteriaceae (Ercolani 1991, Thompson et al. 1993, Lindow and Brandl 2003, Krimm et al. 2005) that build diverse communities. Several studies have focused on the distribution of specific groups or species of bacteria across plant species (Corpe and Rheem 1989, Brighigna et al. 1992). These studies revealed non species-specific distribution of the investigated taxa (but see Yang et al. 2001) leading to the conclusion that these bacteria may be well adapted to the phyllosphere irrespective species-specific properties of leaf surfaces (Hirano and Upper

2000). A recent study by Östman *et al.* (2010) also indicates that habitat-specific microbial communities have a high degree of similarity across sites within a large spatial scale.

Similar to leaves, petals offer colonizable surfaces, but received much less attention. Due to the severe economic and social impacts caused by pathogenic microorganisms, previous work on flower dwelling bacteria focused on crop diseases (e.g. Windels 2000) such as the bacterium *Erwinia amylovora* that causes fire blight (Buban et al. 2003). Much less is known about bacteria growing on flowers of uncultured plants or about those with no obvious detrimental effect on the plants' reproduction. However, nectar and exudates of stigma and pollen offer excellent growing media for microorganisms (Brysch-Herzberg 2004, Stockwell 2005) and the visitation by pollinators or other dissemination mechanisms of pollen provide ideal dispersal conditions for microorganisms (Giles et al. 2005). Nonetheless, a study by Krimm *et al.* (2005) indicates that the diversity of bacteria is lower on flowers than on leaves.

In this study we compared the bacterial communities on flowers and leaves of two naturally growing plants species. Within the flowers we excluded stigmas, nectar and pollen from our investigation and restricted it to petals in order to ensure a better comparability to leaves. Additionally, we examined the role of plant volatiles in structuring the bacterial communities.

Material and Methods

Isolation and identification of epiphytic bacteria

At different sites in Würzburg and Reichenberg, Germany, we sampled young leaves and flowers from *Lotus corniculatus* (Fabaceae) and *Saponaria officinalis* (Caryophyllaceae) from spatially separated patches. Several leaves and flowers per sample were placed in 30 ml phosphate buffered saline (PBS) and were sonificated for 7 min to separate bacteria from plant material. 100 µl of different dilutions of PBS were plated on LB agar plates. After incubation at 30 °C for 48 h colony forming units (cfu) were counted and density of bacteria

on plant parts were estimated by calculation of the surface area of all leaves and flowers in each sample [cfu cm^{-2}]. Three colonies per distinct morphotype were cultivated on a separate LB agar plate at the same conditions as described above.

From isolated bacterial strains one colony was picked and DNA was isolated as template for polymerase chain reaction using the primer pair 27f and 1492r targeting the 16S rRNA gene. Purified DNA was sent to SeqLab (Sequence Laboratories, Göttingen, Germany) for sequencing. For methodological details see Appendix methods.

Sequences were matched with sequences at GenBank nucleotide database (accessed 23. March 2010) (Benson et al. 2009). We decided to integrate ribosomal secondary structure information additionally to sequence information into the phylogenetic reconstructions, as a recent simulation study confirmed the benefit regarding accuracy and robustness (Keller et al. 2010). Thus, we used according to the workflow published for ITS2 sequence and structure phylogenetics 4SALE alignments (Seibel et al. 2008) and Profile Neighbor-Joining (Wolf et al. 2008) for our 16S data with Jukes Cantor distances and 100 bootstrap replicates. Sequences of the genus *Deinococcus* (GI:219846824, GI:222083990 and GI:110277976) were used as outgroup. The resulting tree was displayed with iToL (Letunic and Bork 2007). Taxonomy information was added according to the most often occurring taxonomic annotation (genus and family) within the best 25 BLAST hits with minimal manual corrections for recently split genera. For methodological details see Appendix methods.

Volatile collection

Scents of leaves and flowers of both plant species were sampled in mixture (1:1) of Tenax-TA (mesh 60-80) and Carbotrap (mesh 20-40) and samples were analysed in a Varian Saturn 2000 system that was equipped with a ChromatoProbe kit. For further details see Dötterl et al. (2005) and Appendix Methods.

Agar diffusion assay

In the agar diffusion assay, the potential effect of volatile compounds on the growth of bacteria that were isolated from leaves or flowers of *S. officinalis* was examined. We used two volatiles that were predominately emitted by leaves and three that were predominately emitted by flowers (see Fig. 2 and Appendix Results). Bacterial strains used for the tests were either isolated from leaves or flowers of *S. officinalis* (see Fig. 2 and Appendix Methods). 100 µl of a bacterial suspension was mixed with 5ml top agar and poured upon dried LB agar plates. 0.06 or 0.04 mMol of the substances dissolved in acetone were applied on sterile cellulose discs (Ø 6 mm, Oxoid, Hampshire, United Kingdom) and cellulose discs were placed on agar plates with bacterial suspensions. Pure acetone was used to control for potential growth inhibitory effects of the solvent. The control did never inhibit growth of any bacterial strain and was thus removed from statistical analysis. After incubation for 48 h, the diameter of inhibition zones was measured.

Statistical analysis

We used random forest, a machine-learning algorithm (Breiman 2001), to assign individual bacterial communities and scent compositions to specific groups (leaves and flowers of *L. corniculatus* and *S. officinalis*) and to estimate the variable importance (bacterial genus and scent compound) for the correctness of the assignment. Recently, this statistical classification tool was established for the interpretation of ecological multivariate data and its utility and advantages (i.e. it calculates the importance of each variable for a right classification independently of the others but also considers multivariate interactions with the others) were demonstrated (Prasad et al. 2006, Ranganathan and Borges 2009). For each analysis n_{tree} = 100,000 bootstrap samples were drawn with m_{try} = 2 variables randomly selected at each node. For each bacterial family or scent compound with a variable importance > 0, we used a

t-test for bacteria or an ANOVA for scent compounds as post-hoc test to validate the results of random forest.

Results

Bacterial communities

In total, we identified 130 bacterial strains from 10 families and 25 genera (Fig. 1). Density of bacteria on plant surfaces [bacteria cm^{-2}] and diversity of bacteria families and genera differed between flowers and leaves of *S. officinalis* and *L. corniculatus* (Tab. 1). In general, diversity of bacteria colonizing flowers was much lower than those colonizing leaves, both on family (Wilcoxon rank sum test: $W = 1$, $n = 10$, $p < 0.001$) and genus level ($W = 17$, $n = 10$, $p = 0.012$). We did neither find differences between the communities colonizing flowers of *S. officinalis* and *L. corniculatus* nor between the leaves of these species as the flowers and the leaves, regardless of the plant species, were each assigned to one group only by random forest analysis (result not shown). Thus, we repeated the random forest analysis considering only the plant part, not the plant species. On the family level, flower-communities were all correctly assigned to flowers, 9 out of 10 leaf-communities to leaves (Tab. 2a). On the genus level, flower communities were correctly assigned, but half of the leaf-communities were also assigned to flowers (Tab. 2b). Bacterial communities on flowers were dominated by representatives of genera belonging to the Enterobacteriaceae, but *Pseudomonas* was the most common bacterial genus colonizing leaves (Fig. 2 a and b).

Volatile compositions

Scent compositions from flowers and leaves of *S. officinalis* and *L. corniculatus* were distinct from each other, except for one floral scent composition of *S. officinalis* which was assigned to leaf scents of the same species (Tab. Rf scent in Appendix Scent). Leaves and flowers of *L. corniculatus* emitted the same volatiles but in different proportions. Leaves and flowers of *S.*

officinalis shared some compounds but some were exclusively emitted by flowers or in much higher amounts (Appendix scent).

Agar diffusion assay

In total, we performed 450 agar diffusions assays with five bacterial strains; two scent compounds that were predominantly emitted by leaves of *S. officinalis* and three floral volatiles of the same species. The diameter of the inhibition zones was affected by the bacterial strain, the scent compound used and interaction of both (multiple ANOVA: bacterial strain: $F_4 = 43.5$, $p < 0.001$; scent: $F_4 = 131.5$, $p < 0.001$; bacterial strain · scent: $F_{16} = 9.9$, $p < 0.001$; residuals = 425). Benzyl nitrile and 2-Phenylethylalcohol, both floral scent compounds, had the strongest growth-inhibitory effect on most bacterial strains, while Methyl-benzoate and the green leaf volatiles only slightly affected the growth of the bacteria (Fig. 2). *Serratia* sp. (Enterobacteriaceae, strain: SR1-2-f) was least inhibited in its growth by the floral scents compounds (Fig. 2). The different concentrations of substances applied in the assay (0.06 and 0.04 mMol) did not affect the diameter of the inhibition zones (Welch corrected *t*-test: $t_{422.02} = 1.58$, $p = 0.11$). Both concentrations are well beyond the daily emission [ng d^{-1} dry weight $[g]^{-1}$] of the substances (e.g. 50 times more in the case of Methyl benzoate) suggesting that the maximal inhibition is reached. Thus, the extent of inhibition may not reflect natural conditions but the comparison between substances remains valid.

Discussion

On the family and genus level, we found the bacteria that colonized leaves of *Lotus corniculatus* and *Saponaria officinalis* to be consistent with those found on leaves of other plant species (*cf*. Ercolani 1991, Thompson et al. 1993, Krimm et al. 2005). The bacteria that colonized the flowers of these plant-species where generally from the same families as those found on leaves but their composition was fundamentally different. The communities on

flowers were less diverse than those on leaves – as also suggested by data from Krimm et al. (2005) – and were dominated by bacteria of the family Enterobacteriaceae. Overall, the bacterial compositions were differed between the plant parts but not between the plant-species, which suggests that flowers and leaves to a certain extend have their distinct communities. In the agar diffusion assay we explored one out of several causes that may be responsible for the flower-specific and taxonomically restricted bacterial communities. The antimicrobial function of substances that are also frequently produced by flowers including terpenoids (Velickovic et al. 2003, Gershenzon and Dudareva 2007) and benzenoids (Karapinar and Aktug 1987) is well known. Correspondingly, with the exception of *Serratia* sp. (Enterobacteriaceae, isolated from *S. officinalis* flowers) scent compounds emitted by flowers of *S. officinalis* had a stronger antibacterial effect on most bacteria tested than those emitted by leaves. Thus, floral scents may contribute to the relatively low diversity of bacteria colonizing petals. These results may suggest that floral volatiles serve as defenses against microorganisms that potentially could be pathogenetic or otherwise detrimental for the reproduction of the plants. This hypothesis may contribute to the recent discussion about alternative functions of floral scents besides pollinator attraction (Raguso 2008) and the notion that defensive properties of floral volatiles are crucial for the fitness of plants (Junker and Blüthgen 2010).

In pollination biology, the presence of microorganisms and their potential impact on floral signals, rewards, and consequently on pollinator behavior and plants' reproduction was mostly neglected. Exceptions from this gap are the interactions between yeast and nectar (Kevan et al. 1988, Herrera et al. 2008), fungi altering flower traits or induce pseudo-flowers (Raguso and Roy 1998, Dötterl et al. 2009) and floral pathogens (Johnson and Stockwell 1998). The omnipresence of bacteria and their virtually endless biochemical abilities as well as insights into floral pathogens presume that bacteria may have additional profound impacts on ecological processes related to flowers and pollination. These potential bacterial impacts

may include effects on floral rewards and signals and the bacterial communities may in turn be affected by the visitation pattern of flower visiting insects. (1) Bacteria colonizing flower surfaces may spoil nectar or pollen e.g. by the activity of pollinators, which may severely affect the nutritional composition, an effect that was recently demonstrated for yeasts dwelling in nectar (Herrera et al. 2008). Next to the alteration of resources, bacterial metabolites such as ethanol may accumulate in nectar and thereby make it toxic to pollinators (Ehlers and Olesen 1997). (2) The scents emitted by bacteria include many of those that are also emitted by flowers (Knudsen et al. 2006, Schulz and Dickschat 2007) but also include unknown substances (Kai et al. 2008). Floral scents mediate several mutualistic and antagonistic interactions (Junker and Blüthgen 2010, Junker et al. 2010) and the complementation of floral volatile compositions by bacterial odors may interfere with those interactions. For instance, alternations of the original bouquet (e.g. due to bacteria) may lead to an reproductive isolation of flowers with modified scents (Waelti et al. 2008). (3) The ability of different bee species to spread antagonistic bacteria of plant pathogens has been demonstrated in several studies (Johnson et al. 1993, Maccagnani et al. 2009) suggesting that naturally occurring bacteria may be dispersed similarly. Therefore, the taxonomically relatively restricted visitor spectrum of flowers (Blüthgen et al. 2007) may contribute to the establishment of floral bacterial communities.

Acknowledgement

We thank Kristina Keidel and Norman Mauder for practical help and helpful comments. The project was financially supported by the Deutsche Forschungsgemeinschaft (DFG, BL 960/1-1). AK appreciates financial help by the BIGSS graduate school.

References

Adler, L. S. 2000. The ecological significance of toxic nectar. Oikos **91**:409–420.

Andrews, J. H. and R. F. Harris. 2000. The ecology and biogeography of microorganisms on plant surfaces. Annual Review of Phytopathology **38**:145-180.

Bednarek, P. and A. Osbourn. 2009. Plant-Microbe Interactions: Chemical Diversity in Plant Defense. Science **324**:746-748.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 2009. GenBank. Nucleic Acids Research **37**:D26-D31.

Blüthgen, N., F. Menzel, T. Hovestadt, B. Fiala, and N. Blüthgen. 2007. Specialization, constraints, and conflicting interests in mutualistic networks. Current Biology **17**:341-346.

Breiman, L. 2001. Random forests. Machine Learning **45**:5-32.

Brighigna, L., P. Montaini, F. Favilli, and A. C. Trejo. 1992. Role of the Nitrogen-Fixing Bacterial Microflora in the Epiphytism of Tillandsia (Bromeliaceae). American Journal of Botany **79**:723-727.

Brysch-Herzberg, M. 2004. Ecology of yeasts in plant–bumblebee mutualism in Central Europe. FEMS Microbiology Ecology **50**:87-100.

Buban, T., Z. Orosz-Kovacs, and A. Farkas. 2003. The nectary as the primary site of infection by *Erwinia amylovora* (Burr.) Winslow et al.: a mini review. Plant Systematics and Evolution **238**:183-194.

Carter, C. and R. W. Thornburg. 2004. Is the nectar redox cycle a floral defense against microbial attack? Trends in Plant Science **9**:320-324.

Corpe, W. A. and S. Rheem. 1989. Ecology of the Methylotrophic Bacteria on Living Leaf Surfaces. FEMS Microbiology Ecology **62**:243-249.

Dötterl, S., A. Jürgens, L. Wolfe, and A. Biere. 2009. Disease status and population origin effects on floral scent: potential consequences for oviposition and fruit predation in a complex interaction between a plant, fungus, and noctuid moth. Journal of Chemical Ecology **35**:307-319.

Dötterl, S., L. M. Wolfe, and A. Jurgens. 2005. Qualitative and quantitative analyses of flower scent in Silene latifolia. Phytochemistry **66**:203-213.

Ehlers, B. K. and J. M. Olesen. 1997. The fruit-wasp route to toxic nectar in Epipactis orchids? Flora **192**:223-229.

Ercolani, G. L. 1991. Distribution of Epiphytic Bacteria on Olive Leaves and the Influence of Leaf Age and Sampling Time. Microbial Ecology **21**:35-48.

Gershenzon, J. and N. Dudareva. 2007. The function of terpene natural products in the natural world. Nature Chemical Biology **3**:408-414.

Giles, B. E., T. M. Pettersson, U. Carlsson-Granér, and P. K. Ingvarsson. 2005. Natural selection on floral traits of female *Silene dioica* by a sexually transmitted disease. New Phytologist **169**:729-739.

Harrewijn, P., A. K. Minks, and C. Mollema. 1995. Evolution of plant volatiles production in insect-plant relationships. Chemoecology **5**:55-73.

Herrera, C. M., I. M. Garcia, and R. Perez. 2008. Invisible floral larcenies: Microbial communities degrade floral nectar of bumble bee-pollinated plants. Ecology **89**:2369-2376.

Hirano, S. S. and C. D. Upper. 2000. Bacteria in the leaf ecosystem with emphasis on Pseudomonas syringae - a pathogen, ice nucleus, and epiphyte. Microbiology and Molecular Biology Reviews **64**:624-+.

Johnson, K. B. and V. O. Stockwell. 1998. Management of fire blight: a case study in microbial ecology. Annual Review of Phytopathology **36**:227-248.

Johnson, K. B., V. O. Stockwell, D. M. Burgett, D. Sugar, and J. E. Loper. 1993. Dispersal of Erwinia-Amylovora and Pseudomonas-Fluorescens by Honey-Bees from Hives to Apple and Pear Blossoms. Phytopathology **83**:478-484.

Junker, R. R. and N. Blüthgen. 2010. Floral scents repel facultative flower visitors, but attract obligate ones. Annals of Botany **105**:777-782.

Junker, R. R., N. Höcherl, and N. Blüthgen. 2010. Responses to olfactory signals reflect network structure of flower-visitor interactions. Journal of Animal Ecology **doi: 10.1111/j.1365-2656.2010.01698.x**.

Kai, M., M. Haustein, F. Molina, A. Petri, B. Scholz, and B. Piechulla. 2008. Bacterial volatiles and their action potential. Applied and microbiological biotechnology **DOI 10.1007/s00253-008-1760-3**.

Karapinar, M. and S. E. Aktug. 1987. Inhibition of Foodborne Pathogens by Thymol, Eugenol, Menthol and Anethole. International Journal of Food Microbiology **4**:161-166.

Keller, A., F. Forster, T. Muller, T. Dandekar, J. Schultz, and M. Wolf. 2010. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. Biol Direct **5**:4.

Kevan, P. G., S. Eisikowitch, S. Fowle, and K. Thomas. 1988. Yeast-contaminated nectar and its effects on bee foraging. Journal of Apicultural Research **27**:26-29.

Knudsen, J. T., R. Eriksson, J. Gershenzon, and B. Stahl. 2006. Diversity and distribution of floral scent. Botanical Review **72**:1-120.

Krimm, U., D. Abanda-Nkpwatt, W. Schwab, and L. Schreiber. 2005. Epiphytic microorganisms on strawberry plants (*Fragaria ananassa* cv. Elsanta): identification of bacterial isolates and analysis of their interaction with leaf surfaces. FEMS Microbiology Ecology **53**:483–492.

Letunic, I. and P. Bork. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics **23**:127-128.

Lindow, S. E. and M. T. Brandl. 2003. Microbiology of the Phyllosphere. Applied and environmental microbiology **69**:1875–1883.

Lokvam, J. and J. F. Braddock. 1999. Anti-bacterial function in the sexually dimorphic pollinator rewards of Clusia grandiflora (Clusiaceae). Oecologia **119**:534-540.

Maccagnani, B., F. Giacomello, M. Fanti, D. Gobbin, S. Maini, and G. Angeli. 2009. Apis mellifera and Osmia cornuta as carriers for the secondary spread of Bacillus subtilis on apple flowers. Biocontrol **54**:123-133.

Mercier, J. and S. E. Lindow. 2000. Role of leaf surface sugars in colonization of plants by bacterial epiphytes. Applied and environmental microbiology **66**:369–374.

Östman, Ö., S. Drakare, E. S. Kritzberg, S. Langenheder, J. B. Logue, and E. S. Lindström. 2010. Regional invariance among microbial communities. Ecology Letters **13**:118-127.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems **9**:181-199.

Raguso, R. A. 2008. Wake up and smell the roses: The ecology and evolution of floral scent. Annual Review of Ecology, Evolution and Systematics **39**:549–569.

Raguso, R. A. and B. A. Roy. 1998. 'Floral' scent production by Puccinia rust fungi that mimic flowers. Molecular Ecology **7**:1127-1136.

Ranganathan, Y. and R. M. Borges. 2009. Reducing the babel in plant volatile communication: using the forest to see the trees. Plant Biology **10.1111/j.1438-8677.2009.00278.x**.

Schulz, S. and J. Dickschat. 2007. Bacterial volatiles: the smell of small organisms. Natural Product Report **24**:814–842.
Seibel, P., T. Müller, T. Dandekar, and M. Wolf. 2008. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Research Notes **1**:91.
Stockwell, V. O. 2005. Flowers: diverse and mutable microbial habitats. Phytopathology **95**:128.
Sundin, G. W. and J. L. Jacobs. 1999. Ultraviolet radiation (UVR) sensitivity analysis and UVR survival strategies of a bacterial community from the phyllosphere of field-grown peanut (Arachis hypogeae L.). Microbial Ecology **38**:27-38.
Thompson, I. P., M. J. Bailey, J. S. Fenlon, T. R. Fermor, A. K. Lilley, J. M. Lynch, P. J. Mccormack, M. P. Mcquilken, K. J. Purdy, P. B. Rainey, and J. M. Whipps. 1993. Quantitative and Qualitative Seasonal-Changes in the Microbial Community from the Phyllosphere of Sugar-Beet (Beta-Vulgaris). Plant and Soil **150**:177-191.
Tomczykowa, M., M. Tomczyk, P. Jakoniuk, and E. Tryniszewska. 2008. Antimicrobial and antifungal activities of the extracts and essential oils of *Bidens tripartita.* Folia histochemica et cytobiologica **46**:389 - 393.
Velickovic, D. T., N. V. Randjelovic, M. S. Ristic, A. S. Velickovic, and A. A. Smelcerovic. 2003. Chemical constituents and antimicrobial activity of the ethanol extracts obtained from the flower, leaf and stem of *Salvia officinalis* L. Journal of the Serbian Chemical Society **68**:17-24.
Waelti, M. O., J. K. Muhlemann, A. Widmer, and F. P. Schiestl. 2008. Floral odour and reproductive isolation in two species of *Silene*. Journal of Evolutionary Biology **21**:111-121.
Windels, C. E. 2000. Economic and social impacts of Fusarium head blight: changing farms and rural communities in the Northern Great Plains. Phytopathology **90**:17-21.
Wolf, M., B. Ruderisch, T. Dandekar, J. Schultz, and T. Muller. 2008. ProfDistS: (profile-) distance based phylogeny on sequence-structure alignments. Bioinformatics **24**:2401-2402.
Yang, C.-H., D. E. Crowley, J. Borneman, and N. T. Keen. 2001. Microbial phyllosphere populations are more complex than previously realized. PNAS **98**:3889–3894.

Tables

Tab. 1 Density and diversity of bacteria colonizing flowers and leaves of *Saponaria officinalis* and *Lotus corniculatus*. Shown are Median and interquartile range.

	Saponaria officinalis		*Lotus corniculatus*	
	fower	leaf	fower	leaf
N	3	3	7	7
Bacteria cm^{-2}	3759 (3192 - 7999)	7166 (6998 - 7297)	146247 (43823 - 449895)	700 (454 - 1360)
Diversity (family)	1.02 (1.01 - 1.14)	1.47 (1.28 - 2.18)	1.00 (1.00 - 1.06)	1.74 (1.68 - 2.39)
Diversity (genus)	1.67 (1.47 - 1.71)	3.54 (2.65 - 4.07)	1.17 (1.09 - 1.65)	1.74 (1.69 - 2.39)

Tab. 2 Classification of the bacterial communities colonizing flowers and leaves of *Saponaria officinalis* and *Lotus corniculatus* based on bacterial families (a) or genera (b) using random forest. Confusion matrix shows number of correctly assigned communities and proportional class error. Families (a) or genera (b) that were important in the classification (i.e. variable importance $E > 0$) are listed in decreasing order. Additionally, number of samples from which each family or genus was isolated is given and in parenthesis the proportion of colony forming units that belong to it in the samples were the family or genus occurred. Flower and leaf samples were compared with *t*-tests, asterisks indicate significance level with *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

a)

Confusion matrix

	Flower	Leaf	Class error
Flower	10	0	0
Leaf	1	9	0.1

Variable importance

Family	*E*	Flower	Leaf	*t*
Enterobacteriaceae	75.78	10 (0.98 ± 0.01)	8 (0.29 ± 0.07)	9.66 ***
Pseudomonadaceae	64.34	3 (0.03 ± 0.00)	9 (0.40 ± 0.11)	3.42 **
Microbacteriaceae	34.34	2 (0.06 ± 0.05)	8 (0.37 ± 0.13)	2.53 *
Burkholderiaceae	19.20	0	3 (0.06 ± 0.04)	1.44
Xanthomonadaceae	14.85	0	2 (0.15 ± 0.04)	1.44
Rhizobiaceae	10.61	0	2 (0.08 ± 0.07)	1.09

b)

Confusion matrix

	Flower	Leaf	Class error
Flower	10	0	0
Leaf	5	5	0.5

Variable importance

Genus	*E*	Flower	Leaf	*t*
Pseudomonas	53.76	3 (0.03 ± 0.00)	9 (0.40 ± 0.11)	3.42 **
Serratia	28.69	6 (0.58 ± 0.12)	1 (0.37)	2.50 *
Yersinia	22.68	6 (0.72 ± 0.15)	3 (0.37 ± 0.20)	1.94
Plantibacter	20.69	0	3 (0.13 ± 0.06)	1.56
Ralstonia	20.51	0	3 (0.06 ± 0.04)	1.44
Microbacterium	20.16	0	3 (0.09 ± 0.08)	1.12
Frigoribacterium	19.96	1 (0.01)	5 (0.24 ± 0.14)	1.58
Rathayibacter	16.27	0	2 (0.20 ± 0.05)	1.46
Stenotrophomonas	13.45	0	2 (0.15 ± 0.04)	1.44
Curtobacterium	8.90	0	2 (0.03 ± 0.01)	1.31
Rahnella	2.21	2 (0.46 ± 0.42)	1 (0.01)	1.04

Figures legends

Fig. 1 Phylogenetic Profile Neighbor Joining tree representing evolutionary relationships between all sampled specimen. Bootstrap values were determined with 1000 pseudoreplicates. Specimens were assigned to genera according to the majority of the first 20 BLAST hits against the GenBank database. Voucher identifiers are displayed in parenthesis. Sample communities are indicated by the inner ring, whereas the outer ring represents current family classifications. Three Deinococcus species were added as the outgroup. Strains used for the agar diffusion assay are highlighted in gray.

Fig. 2 Results of agar diffusion assay. Mean and 95% confidence intervals are given; significant differences in the growth inhibition are indicated if confidence intervals do not overlap. (Z)-3-Hexen-ol and (Z)-3-Hexen-1-ol acetate were predominantly emitted by leaves, the others predominantly from the flowers of *Saponaria officinalis*.

Figures

Fig. 1

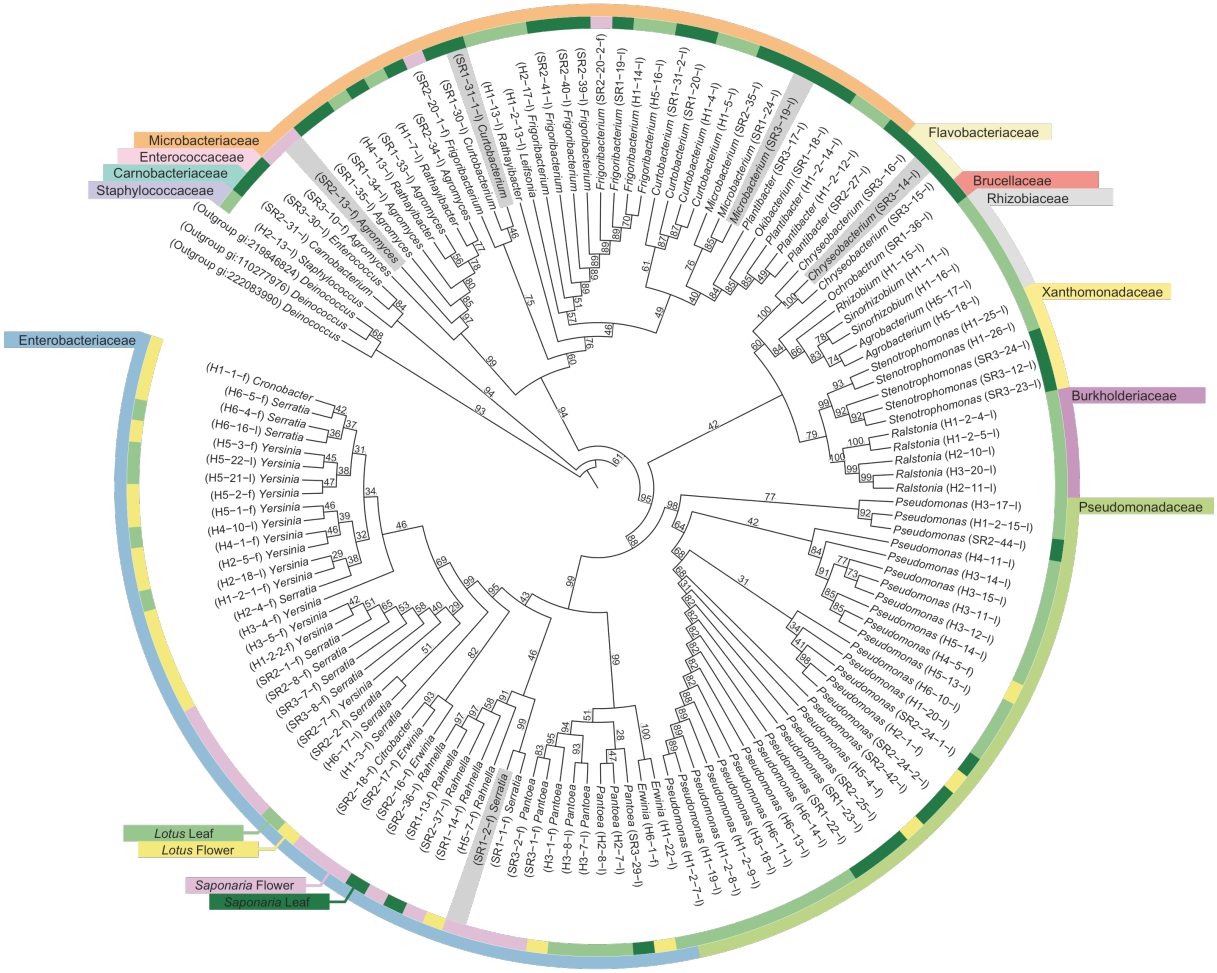
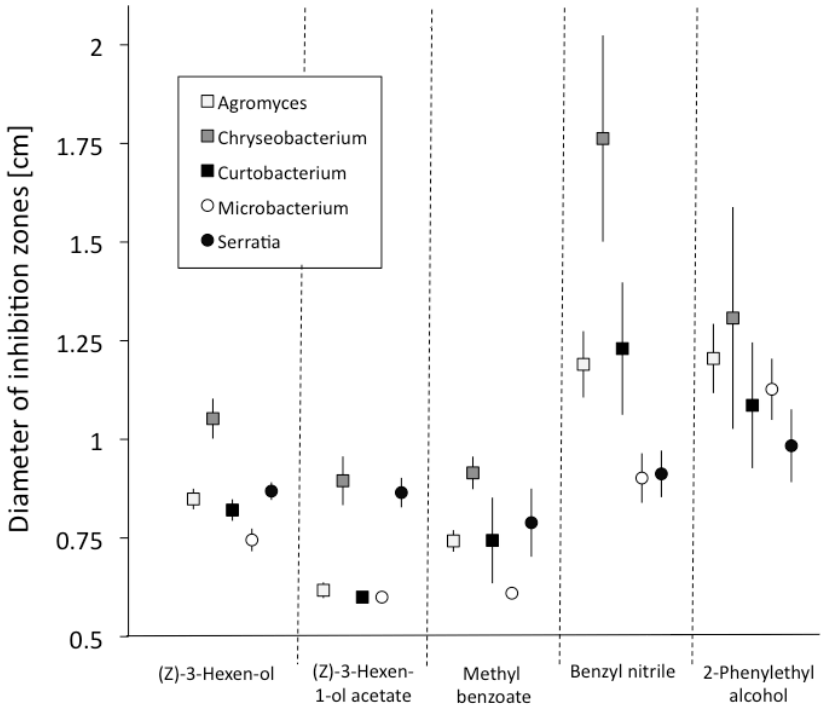


Fig. 2


Diameter of inhibition zones [cm]
2
1.75
1.5
1.25
1
0.75
0.5
Agromyces
Chryseobacterium
Curtobacterium
Microbacterium
Serratia
(Z)-3-Hexen-ol
(Z)-3-Hexen-1-ol acetate
Methyl benzoate
Benzyl nitrile
2-Phenylethyl alcohol

CHAPTER
TEN

FUTURE PROSPECTS OF STRUCTURAL PHYLOGENETICS

P.10. Ribosomal RNA phylogenetics: the third dimension

Authors: A. Keller*, M. Wolf, T. Dandekar

published 2010

Journal: *Biologia*
Volume: 65, Pages: 388–391

*First author

Author's Contributions:
All authors designed the study together. I folded tertiary structure models and performed all analytical procedures. I drafted the manuscript. All authors contributed to the final version of the manuscript and approved it.

VERSITA

Biologia **65**/3: 388—391, 2010
Section Cellular and Molecular Biology
DOI: 10.2478/s11756-010-0045-3

Ribosomal RNA phylogenetics: the third dimension

Alexander Keller, Matthias Wolf* & Thomas Dandekar

Department of Bioinformatics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany; e-mail: matthias.wolf@biozentrum.uni-wuerzburg.de

Abstract: With integration of secondary structures, ribosomal genes have once again become very popular for phylogenetic analyses. This additional source of information to the nucleotide sequence provides a massive boost for taxonomic inferences. Herein, we propose that in the near future a further benefit for phylogenetics with such genes will be very likely by inclusion of the third dimension. For the first time, we determined the tertiary structure of the ribosomal internal transcribed spacer 2 for *Chlamydomonas rheinhardtii* by application of two different *in silico* prediction algorithms. We compared these methods with focus on phylogenetic usability. Further, we determined the tertiary structures for closely related green algae to provide a small phylogenetic example. The results suggest that the tertiary structure inherits evolutionary information observable neither within the sequence nor in the secondary structure.

Key words: molecular systematics; internal transcribed spacer 2; ITS2; non-coding RNA; secondary structure; tertiary structure.

Abbreviations: ITS2, internal transcribed spacer 2; PDB, Protein Data Bank; RMSD, root mean square deviation.

Introduction

The use of nucleic acid sequence information of the four nucleotides adenine, cytosine, guanine and thymine or uracil has been a major breakthrough for investigations of species relationships. With this first dimension of molecular data, traditional morphological systematics has been augmented by novel molecular phylogenetics.

However, regarding ribosomal RNA a second dimension has been approached for this purpose in recent years: the additional information from RNA secondary structure. Several very different methods have been applied to infer phylogenies with the help of secondary structure. They either use substitution matrices for encoded pseudoproteins (Wolf et al. 2008), substitution matrices based on dinucleotide interactions (Biffin et al. 2007) or morphometrical matrices of characteristics of the secondary structure (Grajales et al. 2007). Independent of the method applied, inclusion of secondary structure information improves the phylogeny in contrast to sequence only information (Biffin et al. 2007, Grajales et al. 2007, Keller et al. 2008). Furthermore, a recent simulation study confirms the benefit of secondary structures in phylogenetics (Keller et al. 2010). Secondary structure phylogenetics with ribosomal RNA in particular may use an additional source of information for reconstructions, as it is a way of incorporating analysis methods and models of evolution that purportedly more adequately represent the generated nucleotide data.

Inclusion of secondary structures is becoming a more and more accepted and sophisticated procedure in the phylogenetic community to improve phylogenies (Schultz & Wolf 2009). In this context, secondary structure predictions with bioinformatics tools from scratch or via homology modeling are efficient high-throughput approaches (Jossinet et al. 2007).

Several bioinformatics tools have been developed which allow three-dimensional structure predictions of RNA molecules (Shapiro et al. 2007). Similar to secondary structure predictions, they are based on data obtained from experimental structure investigations in the laboratory. However, the advantage of computational calculations for large-scale comparisons of RNA structures is obvious. The achieved rapid gain of three-dimensional information is of major importance for comparative studies and identification of homologous characteristics of a molecule. In our opinion, this is still the case after the loss of accuracy, in comparison with wet laboratory verified structures, is taken into account.

In this short study we want to demonstrate that bioinformatics predictions of the third dimension of ribosomal RNA structure may be usable for phylogenetic studies in the near future.

Material and methods

For phylogenetic studies it has been shown that subsidiary secondary structures are particularly a major advantage in cases where secondary structures are very conserved,

* Corresponding author

Springer

yet mutations of nucleotides occur frequently (Keller et al. 2010). As a case in point the internal transcribed spacer 2 (ITS2) of the ribosomal cistron benefits from the inclusion of secondary structures (Keller et al. 2010). Its secondary structure is evolutionarily maintained as it is of importance in ribogenesis (Côté et al. 2002; Venema & Tollervey 1999). By contrast, the evolutionary rate of its sequence is relatively high and it is not present in the mature ribosome. There is long experience with this marker that allows testing for consistency of predictions.

We applied two bioinformatics methods to determine the ITS2 (including 25 nucleotides of each *5.8S* and *28S* ribosomal RNA as a proximal stem) three-dimensional structure for the model organism *Chlamydomonas rheinhardtii*. The first tool was ASSEMBLE as part of the S2S platform developed by Jossinet & Westhof (2005). Tertiary structure models are generated by splitting paired and unpaired regions in separate building blocks. Helical properties are calculated so that stem regions result in a double helix, whereas bulges and loops result in single stranded helical regions. An integrated motif database can be applied to selections so that the topologies are adapted according to structural motives present at the Protein Data Bank (PDB) (Henrick et al. 2008). During or after such processing, the building blocks may be stacked to a single three-dimensional model of the complete molecule. Furthermore, the software allows alignment and homology modeling of homologous molecules.

As a second tool, we used RNA2D3D (Martinez et al. 2008), which is a more automated attempt for three-dimensional model prediction of a complete molecule. Unpaired regions are simple estimations of a planar topology and thus no further manipulation is necessary to receive a continuous structure. However, further manipulations are possible if the knowledge is present for the molecule of interest (Martinez et al. 2008). In a comparison with laboratory-verified structures within this publication, it is further described that models are not too far from reality and thus good initial estimations. Several features of the models of a hammerhead ribozyme were nearly identical with the X-ray resolved structures.

For a comparison between tertiary structures of different organisms as a small phylogenetic example we determined the ITS2 tertiary structure (without additional the *5.8S/28S* hybridized proximal stem) of *Chlamydomonas*

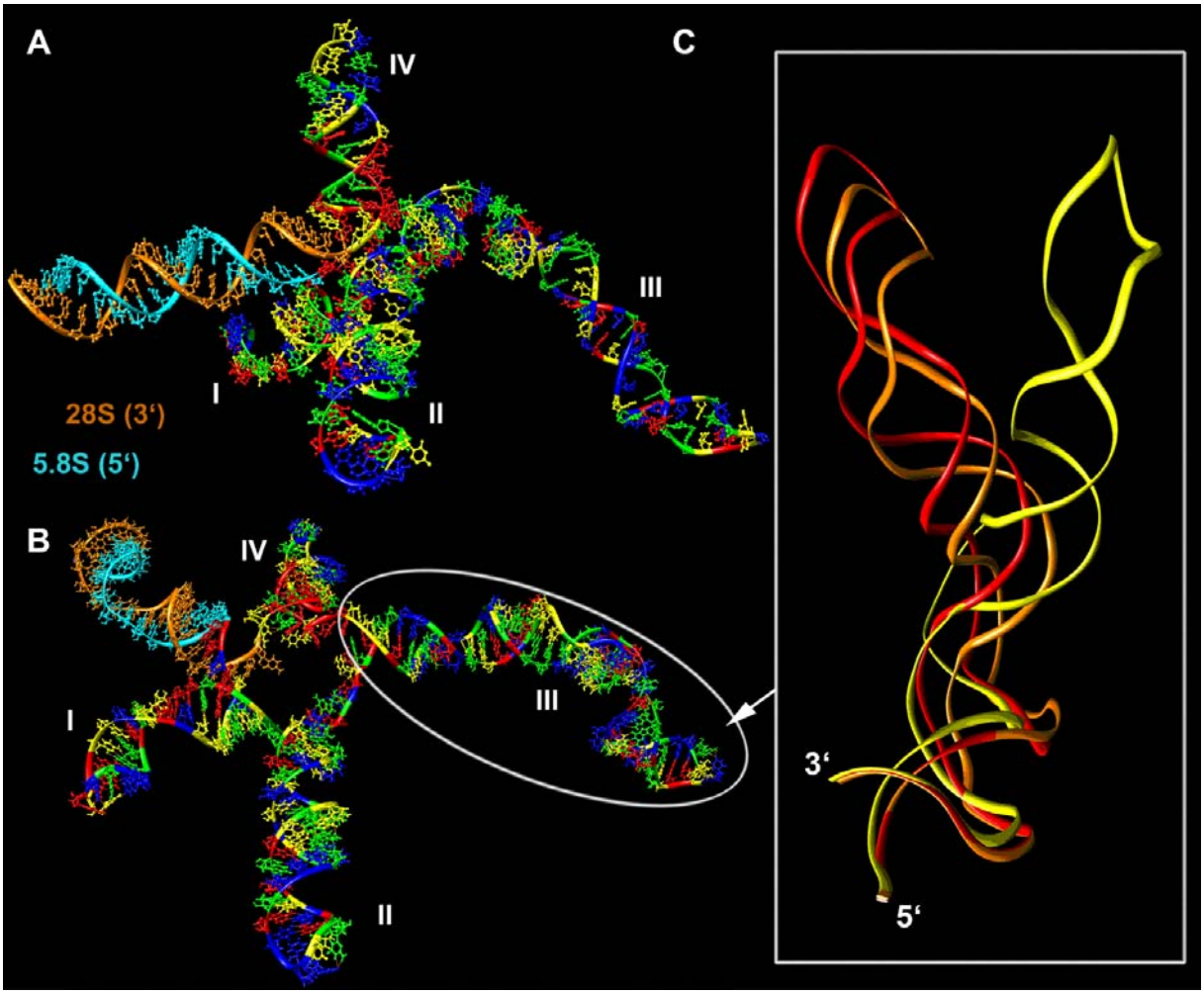


Fig. 1. Comparison of three-dimensional structure predictions by (A) ASSEMBLE and (B) RNA2D3D. Numbering denotes helices I-IV of the ITS2 of *Chlamydomonas rheinhardtii*. Full atomic models are shown excluding hydrogens. Nucleotides coloration: adenine, red; cytosine, green; guanine, yellow; and uracil, blue. (C) An example of genus-specific differences is displayed for the isolated third helix. Complete structures were determined by RNA2D3D and aligned with the Smith-Waterman algorithm according to RMSD distance using UCSF Chimera (Pettersen et al. 2004). Structures are from the following species: red, *Chlamydomonas rheinhardtii*; orange, *Chlamydomonas debaryana*; and yellow, *Gonium pectorale*.

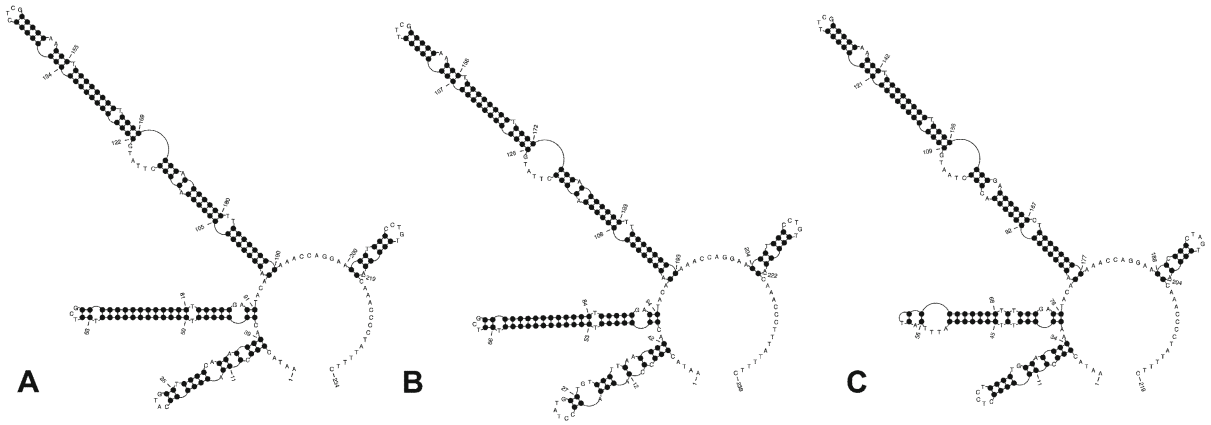


Fig. 2. ITS2 secondary structure of the three algae organisms compared in Figure 1. (A) *Chlamydomonas rheinhardtii*, (B) *Chlamydomonas debaryana* and (C) *Gonium pectorale*. Displayed with PseudoViewer 3.0 (Byun & Han 2009).

rheinhardtii and two close relatives (*Chlamydomonas debaryana* and *Gonium pectorale*) with RNA2D3D. Structures were aligned with the Smith-Waterman algorithm according to root mean square deviation (RMSD) distance using the software UCSF Chimera (Pettersen et al. 2004). PDB files with the modeled atom coordinates of all resulting tertiary structures of this study are available from the authors.

Results and discussion

Both algorithms were able to determine a tertiary structure for ITS2 (Fig. 1a,b). However, using ASSEMBLE, a lot of manipulations have to be performed by hand. This increases the proportion of user-related errors. Furthermore, tertiary structure prediction is almost not reproducible and very time-consuming. Its advantage is in the precise manipulation power for refinement studies of molecules, e.g., where electron density maps or structural information from homologous molecules are present. RNA2D3D is more likely a tool usable for automated approaches, as e.g. large databases and comparisons of homologous molecules in phylogenetic studies. For example, the ITS2 database stores approximately 200,000 secondary structures of ITS2 sequences (database accessed: 16$^{\text{th}}$ October 2009; Koetschan et al. 2010). An automated prediction of tertiary structures in such large-scale databases would be a pleasant addition. The major difference between the two applied methods in the resulting tertiary structure is that unpaired regions in RNA2D3D are planar whereas ASSEMBLE is able to apply user-defined motifs of tertiary structure as well to unpaired nucleotides. Thus, in these regions the latter is likely closer to the "real" structure. However, in a comparison between homologous molecules for phylogenetics this is only a minor drawback for RNA2D3D in contrast to the major benefit of automated and fast predictions.

In our comparison between the different green algae organisms, we see a genus specific difference in the region of the third helix. Figure 1c illustrates the similarity between the two *Chlamydomonas* organisms. Only small changes occur in the basal region, which are, however, compensated in the further alignment so that the resulting coordination of the distal part is mostly similar between the two structures. By contrast, the three-dimensional model of *Gonium pectorale* indicates a major shift in the complete coordination of the third helix. This is caused by substitutions in the proximal region near to the central core. The resulting torsion angles within the medial bending region are largely affected by these substitutions. This is a first indication that phylogenetic signals might be well observable and usable with tertiary structures that are less pronounced in sequences or secondary structures (Fig. 2). However, as this study constitutes a prospect to future works and thus only covers three individual ITS2 tertiary structures, future studies with a more extensive sampling effort are necessary to demonstrate the general capability of three-dimensional RNA phylogenetics using this marker. Further, comparisons with other markers in their effectiveness to resolve phylogenies using the tertiary structure will be very appreciated.

The direct use of three-dimensional models in phylogenetic studies is, however, still a major issue. Currently no phylogenetic method is present that is able to automatically reconstruct trees including information due to RNA tertiary structure. Hence structural insights into phylogeny are currently only shown here by superposition of calculated structures, e.g., with UCSF Chimera (Pettersen et al. 2004) and its tools for three-dimensional analysis. However, with more and more available tools for tertiary structure predictions of RNA molecules and the lessons we learned from secondary structures, we are confident that this is only a matter of time.

The use of three-dimensional comparisons between RNA or protein molecules is not new in the field of functional molecular biology. Just to mention a few studies, e.g., Kroemer et al. (1998), Alon et al. (1995) and Pley et al. (1994) have successfully applied such comparisons to find structural motifs and differences between species, isoforms or during temporal changes. With this paper, we intend to arise interest in the systematic community to enter the third dimension and to apply these methods to answer phylogenetic questions, i.e. to investigate species relationships. First studies

may be based on morphometrical matrices, since simple characteristics of the molecules, as e.g. torsion angles, distances between helices as well as nucleotides or coordination patterns and geometrical features may be easily extracted from three-dimensional structure models (Carugo & Pongor 2002). It is furthermore worthwhile to invest in the development of alignment methods (Brown et al. 2009; Hasegawa & Holm 2009) and substitution matrices that include three-dimensional interactions or distance measurements beyond RMSD, which comprise RNA-specific substitution matrices and characteristics as for example proposed by Parisien et al (2009).

Acknowledgements

We greatly appreciate financial support by the graduate school BIGSS for AK and land Bavaria.

References

Alon R.N., Mirny L., Sussman J.L. & Gutnick D.L. 1995. Detection of α/β-hydrolase fold in the cell surface esterases of *Acinetobacter* species using an analysis of 3D profiles. FEBS Lett. **371:** 231–235.

Biffin E., Harrington M.G., Crisp M.D., Craven L.A. & Gadek P.A. 2007. Structural partitioning, paired-sites models and evolution of the ITS transcript in *Syzygium* and *Myrtaceae*. Mol. Phylogenet. Evol. **43:** 124–139.

Brown J.W., Birmingham A., Griffiths P.E., Jossinet F., Kachouri-Lafond R., Knight R., Lang B.F., Leontis N., Steger G., Stombaugh J. & Westhof E. 2009. The RNA structure alignment ontology. RNA **15:** 1623–1631.

Byun Y. & Han K. 2009. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. Bioinformatics **25:** 1435–1437.

Carugo O. & Pongor S. 2002. Recent progress in protein 3D structure comparison. Curr. Protein Pept. Sci. **3:** 441–449.

Côté C., Greer C. & Peculis B.A. 2002. Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast. RNA **8:** 786–797.

Grajales A., Aguilar C. & Sanchez J.A. 2007. Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals. BMC Evol. Biol. **7:** 90.

Jossinet F., Ludwig T.E. & Westhof E. 2007. RNA structure: bioinformatic analysis. Curr. Opin. Microbiol. **10:** 279–285.

Jossinet F. & Westhof E. 2005. Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. Bioinformatics **21:** 3320–3321.

Hasegawa H. & Holm L. 2009. Advances and pitfalls of protein structural alignment. Curr. Opin. Struct. Biol. **19:** 341–348.

Henrick K., Feng Z., Bluhm W.F., Dimitropoulos D., Doreleijers J.F., Dutta S., Flippen-Anderson J.L., Ionides J., Kamada C., Krissinel E., Lawson C.L., Markley J.L., Nakamura H., Newman R., Shimizu Y., Swaminathan J., Velankar S., Ory J., Ulrich E.L., Vranken W., Westbrook J., Yamashita R., Yang H., Young J., Yousufuddin M. & Berman H.M. 2008. Remediation of the protein data bank archive. Nucleic Acids Res.**36 (Database Issue):** D426–D433.

Keller A., Forster F., Muller T., Dandekar T., Schultz J. & Wolf M. 2010. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. Biol. Direct **5:** 4.

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 Evol. Biol. **8:** 218.

Koetschan C., Förster F., Keller A., Schleicher T., Ruderisch B., Schwarz R., Müller T., Wolf M. & Schultz J. 2010. The ITS2 Database III – sequences and structures for phylogeny. Nucleic Acids Res. **38 (Database Issue):** D275–D279**.**

Kroemer R.T., Kröncke R., Gerdes J. & Richards W.G. 1998. Comparison of the 3D models of four different human IL-7 isoforms with human and murine IL-7. Protein Eng. **11:** 31–40.

Martinez H.M., Maizel J.V. Jr. & Shapiro B.A. 2008. RNA2D3D: a program for generating, viewing, and comparing 3–dimensional models of RNA. J. Biomol. Struct. Dyn. **25:** 669–683.

Parisien M., Cruz J.A., Westhof E. & Major F. 2009. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. RNA **15:** 1875–1885.

Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C. & Ferrin T.E. 2004. UCSF Chimera – a visualization system for exploratory research and analysis. J. Comput. Chem. **25:** 1605–1612.

Pley H.W., Flaherty K.M. & McKay D.B. 2002. Three-dimensional structure of a hammerhead ribozyme. Nature **372:** 68–74.

Schultz J. & Wolf M. 2009. ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics. Mol. Phylogenet. Evol. **52:** 520–523.

Shapiro B.A., Yingling Y.G., Kasprzak W. & Bindewald E. 2007. Bridging the gap in RNA structure prediction. Curr. Opin. Struct. Biol. **17:** 157–165.

Venema J. & Tollervey D. 1999. Ribosome synthesis in *Saccharomyces cerevisiae*. Annu. Rev. Genet. **33:** 261–311.

Wolf M., Ruderisch B., Dandekar T., Schultz J. & Müller T. 2008. ProfDistS: (Profile-) distance based phylogeny on sequence-structure alignments. Bioinformatics **24:** 2401–2402.

Received January 27, 2010
Accepted March 9, 2010

Part IV.

General Discussion and Conclusions

Secondary Structure Phylogenetics

The use of genetic data, i.e. nucleic acid sequence information of the four nucleotides A, C, G and T/U, has been one of the major advantages for biodiversity research and investigations of species relationships. With this first dimension of molecular data, traditional morphological systematics has been augmented by novel molecular phylogenetics (Felsenstein 2004) and species delineation techniques, as e.g. DNA barcoding (Casiraghi et al. 2010; Hebert et al. 2003b).

But, to go one step further, biological observations indicated that ribosomal phylogenetic markers amended through secondary structures yield enhanced phylogenetic trees especially on higher taxonomic levels (Alvarez and Wendel 2003; Coleman 2003; Hillis and Dixon 1991; Mai and Coleman 1997; Soltis et al. 1998; Wheeler and Honeycutt 1988). This hypothetical idea with early, small scale and manually performed studies was consecutively transformed into a functional and sophisticated methodological pipeline (Schultz and Wolf 2009) ranging from secondary structure prediction (Mathews et al. 2004; Wolf et al. 2005b) over alignment (Seibel et al. 2006, 2008; Thompson et al. 1994) to phylogenetic tree reconstructions (Friedrich et al. 2005; Wolf et al. 2008). Several drawbacks currently exist with this pipeline, e.g. the reconstruction step is currently restricted to NJ. Yet, these software tools are pioneering and will likely be expanded in the near future for further functionality, so that other methods like e.g. MP or ML are incorporated.

When applying this pipeline, one of the most crucial tasks is to work with reliable data. We were able to show for the ribosomal marker ITS2 that misannotated sequences, with only few basepairs surplus or less at one of the ends, are likely to yield unwanted results during secondary structure predictions and, with that, all following steps (Publication P.1, Keller et al. 2009a). Thus, we developed an annotation method that is independent of GenBank (Benson et al. 2008) annotation data and based on HMMs (Publication P.1, Eddy 1998; Keller et al. 2009a). This method, contrasted with other methods, provides high quality annotations and reliable sequence fragments for phylogenetic analyses. The method has been made available as an online tool for ITS2 annotation at the ITS2-database (Koetschan et al. 2010; Schultz et al. 2006; Selig et al. 2008). There sequencing results or otherwise retained sequence data may be easily analysed for occuring ITS2 regions. Beside that, valuable ITS2 data is now easily extractable from sequences with unknown taxonomic origin as e.g. environmental samples in metagenomic studies. It will be fortunate to expand this tool in the future to the remaining ribosomal genes so that the complete ribosomal cistron is annotatable with this method for taxonomic and other comparative studies.

Not only a methodological pipeline exists that is able to cope with ITS2 secondary structures. Furthermore, a database has been established that automatically collects ITS2 data, as well as folds individual secondary structures and deposits them permanently (Koetschan et al. 2010; Schultz et al. 2006; Selig et al. 2008). This database, namely the ITS2-database is a good starting point for any phylogenetic analysis with secondary structures and this ribosomal gene, as all the data is freely available through its web-frontend. As the HMM-annotation procedure turned out to be a very profitable method, it was included into the automatic procedure of data allocation of the ITS2-database (Publication P.2, Koetschan et al. 2010). During that process, the complete GenBank (Benson et al. 2008) nucleotide data is scanned independently of the entries' annotations. This increased the number as well as the quality of the retained sequences largely in comparison to the previously used string pattern searches.

This fundamental framework of a new pipeline for phylogenetics and corresponding data stored in a database was as part of this thesis subject of an extensive evaluation through simulation experiments (Publication P.3, Keller et al. 2010a). The results indicate that secondary

structures of ITS2 provide a valuable gain of information content that is useful for phylogenetics. Both, the robustness and accuracy of tree reconstructions are largely improved. Furthermore, we found that the range of the optimum level of sequence divergence for phylogenetic studies is (1) broadend and (2) shifted to higher divergences when secondary structures are included. The optimum level is defined with the upper limits saturated in substitutions, whereas the lower boundary is the lack thereof (Yang 1998).

As a conclusion from our simulation experiments, organisms that are more distantly related can be included into phylogenies with secondary structures of the ITS2, however the functionality and benefits still account for low-level phylogenies. This makes the toolbox of secondary structure phylogenetics a valuable mean for applied studies in biodiversity and evolutionary biology research. Confinements of this study are that we were only able to test for improvements for the ITS2 marker, as unfortunately today there is no comparable amount of data available concerning secondary structures of other RNAs, as provided through the ITS2-database. From a theoretical point of view, benefits may be equally for markers that share the ambiguity of a common secondary structure, but almost minimal conservation on the nucleotide level. Furthermore, evaluation of more algorithms to reconstruct phylogenies as e.g. MP and ML will be very interesting future tasks. Yet, sophisticated phylogenetic methods to apply these on the secondary structure level still lack.

Alternate tools that are able to include secondary structures have also been proposed in the last years. For example, RNAsalsa (Stocsits et al. 2009) aligns RNA by concurrent simulation of secondary structures from a template structure and the actual sequence alignment. The software PHASE (Jow et al. 2003) allows Bayesian inferences of phylogenetic trees by differing between RNA stem regions and unbound nucleotides. This different treatment of unpaired and paired nucleotides is appreciable and is very close to the initial idea of substitution models with RNA structures (Schöniger and von Haeseler 1994). Both tools however, are not able to integrate an individual secondary structure for each of the sequences of interest, but rather a consensus structure. This may be useful for investigations with strongly conserved markers (e.g. 18S), however even for such genes extensive and sophisticated evaluations of these methods are to my knowledge still missing. If any, the benefit may be negligible with markers that have high substitution rates as for example the ITS2, where strict consensus constraints may also lead to inconsistencies and ambiguities in the alignment. Further, no comprehensive pipeline from structure prediction over alignment to phylogenetic inferences has been proposed as these tools are not designed for interplay with each other. A further interesting alignment tool is RSmatch (Liu et al. 2005) and accordingly the corresponding RADAR web interface (Khaladkar et al. 2007), which is – like 4SALE – capable of using individual secondary structures, but uses two scoring schemes, i.e. for unpaired positions and positions with H-bonds. This poses a true alternate in the alignment procedure, however currently the software is only designed for small datasets and comparative studies are still lacking.

Applied Case Studies

In our case studies, we kept close to the published review on how to perform secondary structure phylogenetics (Schultz and Wolf 2009). Some of the tools were just in a state of development so that not all of the methods were applicable in the first manuscripts (e.g. the annotation method and ProfDistS in an alpha developmental state). Throughout the last three years, these methods became more comfortable and co-operative in their workflow, what is reflected by several aspects of the performed case studies, e.g. the number of taxa included and the increase of automation in the analyses.

Our case studies started with an investigation of a rather well known group, namely the

Sphaeropleales of the Chlorophyta, and a small taxon sampling size (Publication P.4, Keller et al. 2008c). Most authors agree to a monophyly of this group (Deason et al. 1991), due to the direct opposite orientation of the flagellar apparatus as a distinct morphological feature (Deason et al. 1991). However no molecular study was able to support this with high statistical support (Buchheim et al. 2001; Cáceres and Robinson 1980; Lewis and McCourt 2004; Müller et al. 2004; Pröschold and Leliaert 2007; Wolf et al. 2002b). From a methodological point of view, we were interested to provide a first biological comparison between secondary structure and sequence only phylogenetics. We were able to show that analyses that included secondary structures performed better in supporting the taxonomic hypothesis that the Sphaeropleales form a monophyletic cluster within the green algae.

Furthermore, we were able to pinpoint a structural feature to an evolutionary event: within the Sphaeropleales, most genera show an untypical Y-branched first helix in the ITS2 (Buchheim et al. 2005; Hegewald and Wolf 2003; van Hannen et al. 2002). We were able to show with this study that this feature evolved past the "Sphaeroplea" clade and thus is not a feature of the complete order Sphaeropleales, but of a specific subgroup. The sequences of these organisms are however now, two years after the study, listed in the ITS2-database (Koetschan et al. 2010) without this Y-branch feature. It thus seems that this Y-branch is only one of the possible options of folding of this first helix. Energetically it seems to be the most optimal structure, however given a possible modeling imperfection of folding software and the heterogeneious environment of a living cell, we have to consider existing fluctuations between the optimal and suboptimal structures. It is possible that one of the suboptimal structures states the actually functional structure for the living organisms, what conforms to the general model of conservation of a four helical unbranched ITS2 (An et al. 2008; Joseph et al. 1999; Schultz et al. 2005).

A following and corroborating phylogenetic study went more into detail regarding the taxonomic status of subgroups of the Sphaeropleales (Publication P.5, Hegewald et al. 2010). Most interesting was the status of the family Coelastraceae including the genera *Asteracys*, *Coelastrella*, *Coelastrum* and *Hariotina*. We were able to show that the Coelastraceae are not a sister group to the monophyletic family of the Scenedesmaceae, but rather included as a subfamily. Furthermore, we were able to erect the new genera *Comasiella* and *Pectinodesmus* with this study with support of morphological features determined by scanning electron microscopy. This approach especially shows that the mutation rates of the ITS2 is high enough to resolve phylogenies and delineate monophyletic groups on the very low level of genera and species.

As determined by the simulation experiments, the benefit of secondary structures becomes in a theoretical point of view more prominent in studies on a larger scale. Thus, we were very interested in a biological example to show the massive range of secondary structure phylogenetics. For this, we chose the same organismal group for which we already performed studies on a small scale (low phylogenetic level and small to moderate taxon sample size) to perform a study on a very large scale (high phylogenetic level and very large taxon sample size): the complete major division Chlorophyta of the green algae (Publication P.6, Buchheim et al. 2010b). With a completly automated pipeline, we determined phylogenetic trees for the complete Chlorphyta, and for each of its inherent classes Chlorophyceae, Trebouxiophyceae and Ulvophyceae.

In general, the complete Chlorophyta tree shows several clusterings, which correspond to monophyletic groups mostly on a family level. However, the order and arrangement of these groups are mostly dissatisfying, since the three classes are not represented as monophyla. However, the individual trees of these classes determined in parallel are very promising for large scale phylogenetics with the ITS2. Despite being a quite short marker with high

substitution rates, we received robust broad phylogenetic trees from our analyses, which are remarkably congruent with analyses of the much longer and mostly very conserved markers typically used in plants that are 18S rRNA, 26S rRNA, rbcL and atpB (e.g. Buchheim et al. 1990, 2010a; Friedl et al. 2009; Leliaert et al. 2003; Mei et al. 2007; Nozaki 2001; Nozaki et al. 2003; Zechman 2003). We conclude that in green algae, the upper limit of the ITS2 is with the given methods between the level of class and division, whereas the lower limits are probably below or approximately at the species level. This is a remarkable range of applicability for phylogenetic studies in plants and biologically corroborates the observations made with the simulation experiments.

Moreover this study showed that taxonomic classification on a species level with the ITS2 is possible even under the condition that the taxonomic affiliation is not known. Thus it shows high practicability as a marker for species identification. With that, and its general features as e.g. the conserved priming regions and its short length, it is a valuable DNA barcoding candidate. Despite some notable exceptions (Chen et al. 2010; Gao et al. 2010; jie Zhu et al. 2010; Luo et al. 2010; Wolf and Schultz 2009), the ITS2 gene has largely been shunned by those investigators that are designing or promoting DNA barcodes for the land plants (Chase and Fay 2009a,b; Chase et al. 2003). Concern about the confounding impact of pseudogenes and the potential presence of intraspecific or even intra-individual variation were cited as reasons for relegating ITS2 to, at best, a supporting role in DNA barcoding for the land plants (Chase and Fay 2009a,b). However, virtually all of the other candidate genomic targets for DNA barcoding in the Chlorophyta exhibit one or more serious deficiencies. Our results and that of others (Chen et al. 2010; Gao et al. 2010; Luo et al. 2010) illustrate that ITS2 data from unknown chlorophytan organisms can be plugged into a high resolution tool for taxonomic assessment and that the ITS2 gene can serve as a powerful plant DNA barcode.

Despite this discussion about barcoding, ITS2 has always been an accepted and appreciated marker in the botanical part of molecular phylogenetics. This also accounts for studies in fungi (Lumbsch 2002; Mullineux and Hausner 2009; Seifert 2009; White et al. 1990), however and in contrast to this, the potential of ITS2 was mostly overlooked in the animal kingdom. When looking at sequence lengths of the marker in animals, it falls into place that ITS2 sequences are typically longer in comparison to those of plants and fungi and even more variable. This is obviously a drawback for phylogenetics, where alignment procedures are likely to yield unwanted results when this occurs simultanuously with high substitution rates of the nucleotides. This may be on of the reasons why this markers has been of less significance in studies of animals and also other organismical groups (e.g. Rhizaria and Excavata). In a study about the blue butterfly genus *Agrodiaetus*, we were very interested whether the inclusion of secondary structure may be a possibility to counter this deficiency in the animal kingdom (Publication P.7, Wiemers et al. 2009). Although the size of the determined sequences was by far larger than those we usualy encountered, the structure was remarkably similar with its four helices, although each of these were largely elongated. This enabled us to reliably align the sequences and retain a phylogenetic tree of high quality. It was comparable to phylogenetic trees obtained with a concatenated Cytochrome c oxidase (CO) subunit I + CO subunit II alignment, whereas it outrivals each of these markers solely. According to our analyses, the subgenus *Agrodiaetus* comprises 6 major clades which are in agreement with COI analyses (Kandul et al. 2004; Wiemers 2003).

When regarding the age of this taxonomic group, it is remarkable that the ITS2 appears to be a suitable nuclear marker to infer the phylogeny of young radiations of animals. Furthermore, we were able to trace the biogeographical distribution of species for this radiation with a dispersal-vicariance analysis, which is in accordance with earlier assumptions (Wiemers

2003). Most *Agrodiaetus* clades seem to origin of biogeographical areas in the region encompassing Eastern Anatolia, Transcaucasia and Iran.

Further recent studies confirm that the ITS2 is also usable for taxonomic inferences within the animal kingdom (Aguilar and Reimer 2010; Gebiola et al. 2010; Oh et al. 2009; Ruhl et al. 2010; Schill et al. 2010). The applicability for phylogenetics as well as barcoding approaches has been demonstrated as well for Diatoms (Moniz and Kaczmarska 2009a,b; Sorhannus et al. 2009). These results are promising that the marker is valuable in the remaining and often unheeded groups of eukaryotes.

Unfortunately, we were not able to support the hypothesis that CBCs can be considered to detect species boundaries for blue butterflies. In this young radiation, the elapsed evolutionary time was likely not enough for compensatory changes, so that we were only able to trace hCBCs within the genus *Agrodiaetus*, which separate the most important clusters of species. Between the different genera, we found several CBCs, suggesting the resolution of this characteristic is approximately at the genus level for this group of organisms. This is in contrast to our and others results within the green algae (Publication P.4, Coleman 2003, 2009; Keller et al. 2008c), tardigrades (Schill et al. 2010) and diatoms (Sorhannus et al. 2009) and as well the large scale approach regarding plants and fungi (Müller et al. 2007), where it seems to be a valuable characteristic to distinguish between species.

Ecological Questions

Knowledge about the evolutionary relationship between organisms becomes very important in explaining or refusing hypotheses in other biological disciplines. For example, several ecological considerations may only be investigated under the assumption of a specific evolutionary tree.

Such a investigation states our comparison of flower-ant interaction networks in Hawai'i (Publication P.8, Junker et al. 2010a). Historically, ants were absent from the geographically isolated hawai'ian archipelago. This group of islands harbors one of the most endemic floras in the world. Most likely due to anthropogenic traffic and trade to and fro (Krushelnycky and Gillespie 2008; Medeiros et al. 1986; Wetterer 1998), ants and plants from other parts of the world invaded the islands (Keeler 1985; Krushelnycky et al. 2005; Wagner et al. 1990). We hypothesised that invasive plants that shared an evolutionary history with ants are better primed against nectar robbing by such than the hawai'ian native plants. Under this assumption, the latter should be more frequently visited by ants due to less pronounced resistances. Furthermore and where phylogeny becomes important, we hypothesized that this pattern is not only restricted to one taxonomic group where specific defensive characters were developed – or lost – only once, but as multiple convergent events in different plant lineages during hawai'ian plant evolution. As the sampled species origin from very different taxonomic lineages, application of a genetic marker was required that shows good resolution capabilities on the small scale as well as for large scale inferences. The secondary structure approach with ITS2 data turned out to be very useful for this investigation.

These hypotheses were confirmed in our study; endemic or indigenous flowers were more frequently exploited by ants than introduced species, as a result of less efficient defense mechanisms. This pattern was independent of the phylogeny, so that the different susceptibility to floral ant visits of native and introduced plant species. Floral defenses against ants are thus likely convergently lost in response to prior absence of ants in native hawai'ian ecosystems. In contrast, nectar features (volume and sugar concentration) correlated with the phylogenetic signal. Flower visiting invasive ants can have devastating effects on the reproduction of native plants and their pollinators (Holway et al. 2002; Lach 2005, 2007, 2008a,b). This suggests that plants endemic or indigenous to the hawai'ian islands are neg-

atively affected by nectar feeding ants, while introduced plants remain largely unaffected. The resulting severe thread to the indigenous ecological system of Hawai'i stated by introduced ants is obvious.

Another example, where secondary structure phylogenetics have been very useful to answer ecological community questions has been an investigation of epiphytic bacterial communities on flowers and leaves (Publication P.9, Junker et al. 2010b). Only little is known about bacteria growing on flowers of uncultured plants or about those with no obvious detrimental effect on the plants' reproduction. However, nectar and exudates of stigma and pollen offer excellent growing media for microorganisms (Brysch-Herzberg 2004; Stockwell 2005). Pollinators or other dissemination mechanisms of pollen provide ideal dispersal conditions for microorganisms (Giles et al. 2006). Nonetheless, a community study (Krimm et al. 2005) indicated that the diversity of bacteria is lower on flowers than on leaves. In this study we compared the bacterial communities on flowers and leaves of two naturally growing plants species to explore the differences in their communities in detail. Genetic sequences of bacteria originating from petals or leaves of *Saponaria officinalis* and *Lotus corniculatus* were characterized at the genus level by means of secondary structure phylogenetics. As bacteria lack the eukaryotic ITS2 region, we transfered the approach to a ribosomal marker present in these organisms, namely 16S rRNA.

The bacteria that colonized the flowers of these plant species were generally from the same families as those found on leaves. Yet, their composition was fundamentally different. The communities on flowers were less diverse than those on leaves and were dominated by bacteria of the family Enterobacteriaceae. This suggests that flowers and leaves have – to a certain extend – distinct communities. Assays showed that floral scents may contribute to the relatively low diversity of bacteria colonizing petals, as an adaptation against microorganisms that potentially could be pathogenetic or otherwise detrimental for the reproduction of the plants.

The approach of secondary structure phylogenetics was flawlessly transferable to the region of 16S rRNA, although the length of the marker was by far longer. The resulting tree was very robust and convincingly fitted to the results obtained by the other experiments of the study. With that, our results indicate that the benefits of the method are not only obtainable with the ITS2 region, but the complete ribosomal cistron.

Future Aspects

In the nearby future, not only the secondary, but as well the tertiary structure of rRNA will likely be usable for phylogenies (Publication P.10, Keller et al. 2010b). However and equivalently to secondary structure, tertiary structure prediction by laboratory means will be too money and time expensive. Further, given the length of phylogenetic markers, which is in most cases longer than the maximum of 95 nt $\approx$ 30 kDa proposed for nuclear magnetic resonance (NMR) spectroscopy (P. J. Lukavsky 2007), the typical treatment methods of RNA in the laboratory (Lukavsky and Puglisi 2004) and the large amount of examined individuals (Lukavsky and Puglisi 2004) render such experiments quite complex up to impossible with the current techniques. To step further into the next dimension of rRNA-phylogenetics, *ab initio* bioinformatical inferences pose once again a time-efficient and inexpensive solution (Shapiro et al. 2007).

First software tools have been developed in the last years, which enable tertiary structure prediction of RNA molecules (Jossinet and Westhof 2010; Jossinet et al. 2007; Martinez et al. 2008). These pioneering software tools lay the fundament for any analyses with the third dimension, however are not technically mature. For example, RNA-specific alignment procedures that incorporate tertiary features are requested by the community of structural

biologists and scientists in nucleic acids research (Parisien et al. 2009). However, even with the current rudimentary methods we were able to show that tertiary structures may be helpful for the investigation of phylogenetic and biodiversity questions (Publication P.10, Keller et al. 2010b). Any advancements achieved by the structural science community will thus be also very fortunate for studies in molecular phylogenetics. Perhaps even comparable features to CBCs may be found that ease the process of species identification and delineation.

Conclusions

The major part of this thesis concentrates on questions in evolutionary biology and biodiversity research. These are complemented with investigations in the closely related field of community ecology. However, the data used to answer these questions originates from molecular biology laboratories, whereas the analytical methods are of bioinformatical nature. Lack of tools in the bioinformatical pipeline stated an opportunity to develop new software by the means of informatics, which has been amended by database development and management. Mathematical i.e. statistical simulation experiments have been used to evaluate these methods. Further, aspects of structural biology and nucleic acids research are integrated into these methods to increase their effectiveness.

To combine these very different disciplines and techniques into a harmonical and integrative thesis has been a challenge. Yet, this combination shows that to look beyond the theoretical and technical horizon of a specific scientific discipline often states a striking opportunity for new ideas and advancements in the way to perform research. The current trends and developments as e.g. the rising of high throughput sequencing technologies and the massive increase of available data are new prospects for ecological questions amongst others, but they will require new multidisciplinary approaches and analytical methods developed by other branches of biology and even research areas beyond the life sciences. In many cases and also this thesis, bioinformatics states the central important connective link between these disciplines.

Summarizing this thesis, it presents an anchor point for new ideas and an example for the integrative use of bioinformatics as a tie between biodiversity research and other biological disciplines. With that, I am very delighted at being able to make my personal contribution to two of the general aims of the "International Year of Biodiversity": to learn about biodiversity and to share this knowledge with other people.

Part V.

Bibliography and additional Information

BIBLIOGRAPHY

Adams, R. P. (2007). *Identification of Essential Oil Components by Gas Chromatography / Mass Spectrometry*. 4th ed. IL USA: Allured Publishing Corporation. (Cit. on p. 154).

Adler, L. S. (2000). "The ecological significance of toxic nectar". In: *Oikos* 91, 409–420. (Cit. on pp. 154, 210).

Agardh, C. (1824). *Systema algarum XXV*. Lund Sweden: Literis Berlingianis: Societatis Physiographicae Lundensis. (Cit. on p. 60).

Aguilar, C. and J. Reimer (2010). "Molecular phylogenetic hypotheses of *Zoanthus* species (Anthozoa:Hexacorallia) using RNA secondary structure of the internal transcribed spacer 2 (ITS2)". In: *Marine Biodiversity* 40 (3), pp. 195–204. (Cit. on p. 241).

Alfaro, M. E., S. Zoller, and F. Lutzoni (2003). "Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence". In: *Molecular Biology and Evolution* 20.2, p. 255. (Cit. on pp. 5, 46).

Allison, S. D. and P. M. Vitousek (2004). "Rapid nutrient cycling in leaf litter from invasive plants in Hawai'i". In: *Oecologia* 141.4, pp. 612–619. (Cit. on p. 154).

Alon, R. N., L. Mirny, J. L. Sussman, and D. L. Gutnick (1995). "Detection of α/β -hydrolase fold in the cell surface esterases of *Acinetobacter* species using an analysis of 3D profiles". In: *FEBS Letters* 371.3, pp. 231–235. (Cit. on p. 230).

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". In: *Nucleic Acids Research* 25.17, p. 3389. (Cit. on p. 38).

Alvarez, I. and J. F. Wendel (2003). "Ribosomal ITS sequences and plant phylogenetic inference". In: *Molecular Phylogenetics and Evolution* 29.3, pp. 417–434. (Cit. on pp. 6, 7, 28, 124, 237).

Alvarez, J. M. and M. A. Hoy (2002). "Evaluation of the ribosomal ITS2 DNA sequences in separating closely related populations of the parasitoid *Ageniaspis* (Hymenoptera: Encyrtidae)". In: *Annals of the Entomological Society of America* 95.2, pp. 250–256. (Cit. on p. 124).

Amirav, A. and S. Dagan (1997). "A direct sample introduction device for mass spectrometry studies and gas chromatography mass spectrometry analyses". In: *European Mass Spectrometry* 3, pp. 105–111. (Cit. on p. 154).

An, S. S., T. Friedl, and E. Hegewald (2008). "Phylogenetic relationships of *Scenedesmus* and *Scenedesmus*-like coccoid green algae as inferred from ITS-2 rDNA sequence comparisons". In: *Plant Biology* 1.4, pp. 418–428. (Cit. on pp. 60, 74, 86, 239).

Andrews, J. H. and R. F. Harris (2000). "The ecology and biogeography of microorganisms on plant surfaces". In: *Annual Review of Phytopathology* 38, pp. 145–180. (Cit. on p. 210).

Andronescu, M., V. Bereg, H. H. Hoos, and A. Condon (2008). "RNA STRAND: the RNA secondary structure and statistical analysis database". In: *BMC Bioinformatics* 9.1, p. 340. (Cit. on p. 12).

Angeler, D. G., M. Schagerl, and A. W. Coleman (1998). "Does the infraspecific disjunct occurrence of the carotenoid loroxanthin reflect phylogenetical relationships within syngens of *Pandorina morum* (Volvocales, Chlorophyta)?" In: *Biologia* 53, pp. 567–576. (Cit. on p. 86).

– (1999). "Phylogenetic relationships among isolates of *Eudorina* species (Volvocales, Chlorophyta) inferred from molecular and biochemical data". In: *Journal of Phycology* 35, pp. 815–823. (Cit. on p. 86).

Bailey, T. L. and C. Elkan (1994). "Fitting a mixture model by expectation maximization to

discover motifs in biopolymers". In: *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*. Vol. 2. 1553-0833, pp. 28–36. (Cit. on p. 38).

Bakker, F. T., J. L. Olsen, and W. T. Stam (1995). "Evolution of nuclear rDNA ITS sequences in the *Cladophora albida/sericea* clade (Chlorophyta)". In: *Journal of Molecular Evolution* 40.6, pp. 640–651. (Cit. on p. 86).

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, and M. J. Donoghue (1995). "The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny". In: *Annals of the Missouri Botanical Garden*, pp. 247–277. (Cit. on p. 6).

Balint, Z. and K. Johnson (1997). "Reformation of the *Polyommatus* section with taxonomic and biogeographic overview (Lepidoptera, Lycaenidae, Polyommatini)". In: *Neue entomologische Nachrichten* 40, pp. 1–68. (Cit. on p. 124).

Barcode of Life Data Systems (2010). *BOLD*. URL: http://www.boldsystems.org. (Cit. on p. 124).

Beattie, A. J. (2006). "The evolution of ant pollination systems". In: *Botanische Jahrbücher für Systematik* 127, pp. 43–55. (Cit. on p. 154).

Beattie, A. J., C. Turnbull, R. B. Knox, and E. G Williams (1984). "Ant inhibition of pollen function: A possible reason why ant pollination is rare". In: *American Journal of Botany* 71.3, pp. 421–426. (Cit. on p. 154).

Beattie, A. J., C. Turnbull, T. Hough, S. Jobson, and R. B. Knox (1985). "The vulnerability of pollen and fungal spores to ant secretions: evidence and some evolutionary implications". In: *American Journal of Botany* 72, pp. 606–614. (Cit. on p. 154).

Bednarek, P. and A. Osbourn (2009). "Plant-microbe interactions: chemical diversity in plant defense". In: *Science* 324.5928, p. 746. (Cit. on p. 210).

Ben-David, T., S. Melamed, U. Gerson, and S. Morin (2007). "ITS2 sequences as barcodes for identifying and analyzing spider mites (Acari: Tetranychidae)". In: *Experimental and Applied Acarology* 41.3, pp. 169–181. (Cit. on pp. 28, 38).

Benson, D. A., M. S. Boguski, D. J. Lipman, J. Ostell, B. F. Ouellette, B. A. Rapp, and D. L. Wheeler (1999). "GenBank". In: *Nucleic Acids Research* 27.1, p. 12. (Cit. on pp. 12, 13, 23).

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler (2008). "GenBank". In: *Nucleic Acids Research* 36, pp. D25–D30. (Cit. on pp. 28, 38, 74, 154, 210, 237).

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. Ingram, and I. Das (2007). "Cryptic species as a window on diversity and conservation". In: *Trends in Ecology & Evolution* 22.3, pp. 148–155. (Cit. on p. 4).

Biffin, E., M. G. Harrington, M. D. Crisp, L. A. Craven, and P. A. Gadek (2007). "Structural partitioning, paired-sites models and evolution of the ITS transcript in *Syzygium* and Myrtaceae". In: *Molecular Phylogenetics and Evolution* 43.1, pp. 124–139. (Cit. on pp. 28, 46, 230).

Biro, L. P., Z. Balint, K. Kertesz, Z. Vertesy, G. I. Mark, Z. E. Horvath, J. Balazs, D. Mehn, I. Kiricsi, and V. Lousse (2003). "Role of photonic-crystal type structures in the thermal regulation of a Lycaenid butterfly sister species pair". In: *Physical Review* E67, p. 021907. (Cit. on p. 124).

Blaxter, M. L. (2004). "The promise of a DNA taxonomy". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359.1444, p. 669. (Cit. on pp. 3, 86).

Blaxter, M. L., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe (2005). "Defining operational taxonomic units using DNA barcode data". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 360.1462, pp. 1935–1943. (Cit. on pp. 86, 124).

Blüthgen, N. and K. Fiedler (2004). "Preferences for sugars and amino acids and their conditionality in a diverse nectar-feeding ant community". In: *Journal of Animal Ecology* 73, pp. 155–166. (Cit. on p. 154).

Blüthgen, N., N. E. Stork, and K. Fiedler (2004). "Bottom-up control and co-occurrence in complex communities: honeydew and nectar determine a rainforest ant mosaic". In: *Oikos* 106, pp. 344–358. (Cit. on p. 154).

Blüthgen, N., F. Menzel, T. Hovestadt, B. Fiala, and N. Blüthgen (2007). "Specialization, constraints, and conflicting interests in mutualistic networks". In: *Current Biology* 17.4, pp. 341–346. (Cit. on p. 210).

Bonser, R., P. J. Wright, S. Bament, and U. O. Chukwu (1998). "Optimal patch use by foraging workers of Lasius fuliginosus, L. niger and Myrmica ruginodis". In: *Ecological Entomology* 23.1, pp. 15–21. (Cit. on p. 154).

Booton, G. C., G. L. Floyd, and P. A. Fuerst (1998a). "Origins and affinities of the filamentous green algal orders Chaetophorales and Oedogoniales based on 18S rRNA gene sequences". In: *Journal of Phycology* 34.2, pp. 312–318. (Cit. on p. 60).

– (1998b). “Polyphyly of tetrasporalean green algae inferred from nuclear small-subunit ribosomal DNA”. In: *Journal of Phycology* 34, pp. 306–311. (Cit. on p. 60).
Bourrelly, P. (1990). *Les algues d'eau douce: initiation a la systematique*. Paris France: Societe nouvelle des Editions Boubee. (Cit. on p. 74).
Breiman, L. (2001). “Random forests”. In: *Machine Learning* 45.1, pp. 5–32. (Cit. on p. 210).
Bremer, B., R. K. Jansen, B. Oxelman, M. Backlund, H. Lantz, and K. J. Kim (1999). “More characters or more taxa for a robust phylogeny-case study from the coffee family (Rubiaceae)”. In: *Systematic Biology* 48.3, pp. 413–435. (Cit. on p. 46).
Brighigna, L., P. Montaini, F. Favilli, and A. C. Trejo (1992). “Role of the nitrogen-fixing bacterial microflora in the epiphytism of *Tillandsia* (Bromeliaceae)”. In: *American Journal of Botany* 79.7, pp. 723–727. (Cit. on p. 210).
Brower, A. V. Z. (2006). “Problems with DNA barcodes for species delimitation: ten species of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae)”. In: *Systematics and Biodiversity* 4.2, pp. 127–132. (Cit. on p. 124).
Brysch-Herzberg, M. (2004). “Ecology of yeasts in plant–bumblebee mutualism in Central Europe”. In: *FEMS Microbiology Ecology* 50.2, pp. 87–100. (Cit. on pp. 210, 242).
Buban, T. and Z. Orosz-Kovacs (2003). “The nectary as the primary site of infection by *Erwinia amylovora* (Burr.) Winslow *et al.*: a mini review”. In: *Plant Systematics and Evolution* 238.1, pp. 183–194. (Cit. on p. 210).
Buchheim, M. A. and R. L. Chapman (1991). “Phylogeny of the colonial green flagellates: a study of 18S and 26S rRNA sequence data”. In: *BioSystems* 25.1-2, pp. 85–100. (Cit. on p. 86).
– (1992). “Phylogeny of *Carteria* (Chlorophyceae) inferred from molecular and organismal data”. In: *Journal of Phycology* 28.3, pp. 362–374. (Cit. on p. 86).
Buchheim, M. A. and L. R. Hoffman (1986). “Ultrastructure of male gametes of *Sphaeroplea robusta* (Chlorophyceae)”. In: *Journal of Phycology* 22.2, pp. 176–185. (Cit. on p. 60).
Buchheim, M. A., M. Turmel, E. A. Zimmer, and R. L. Chapman (1990). “Phylogeny of *Chlamydomonas* (Chlorophyta) based on cladistic analysis of nuclear 18S rRNA sequence data”. In: *Journal of Phycology* 26.4, pp. 689–699. (Cit. on pp. 86, 240).
Buchheim, M. A., E. A. Michalopulos, and J. A. Buchheim (2001). “Phylogeny of the Chlorophyceae with special reference to the Sphaeropleales: a study of 18S and 26S rDNA data”. In: *Journal of Phycology* 37.5, pp. 819–835. (Cit. on pp. 60, 86, 239).
Buchheim, M. A., J. A. Buchheim, T. Carlson, and P. Kugrens (2002). “Phylogeny of *Lobocharacium* (Chlorophyceae) and allies: a study of 18S and 26S rDNA data”. In: *Journal of Phycology* 38.2, pp. 376–383. (Cit. on p. 86).
Buchheim, M. A., J. A. Buchheim, T. Carlson, A. Braband, D. Hepperle, L. Krienitz, M. Wolf, and E. Hegewald (2005). “Phylogeny of the Hydrodictyaceae (Chlorophyceae): inferences from rDNA data”. In: *Journal of Phycology* 41.5, pp. 1039–1054. (Cit. on pp. 60, 86, 239).
Buchheim, M. A., A. Kirkwood, J. A. Buchheim, B. Verghese, and W. J. Henley (2010a). “Hypersaline soil supports a diverse community of *Dunaliella* (Chlorophyceae)”. In: *Journal of Phycology* in press. (Cit. on pp. 86, 240).
Buchheim, M. A., A. Keller, C. Koetschan, F. Förster, B. Merget, and M. Wolf (2010b). “Internal transcribed spacer 2 (nu ITS2 rRNA) sequence-structure phylogenetics: towards an automated reconstruction of the green algal tree of life”. In: submitted. (Cit. on pp. 86, 239).
Byun, Y. and K. Han (2006). “PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures”. In: *Nucleic Acids Research* 34.Web Server issue, W416. (Cit. on p. 60).
– (2009). “PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots”. In: *Bioinformatics* 25.11, p. 1435. (Cit. on pp. 11, 15, 38, 230).
Cáceres, E. J. and D. G. Robinson (1980). “Ultrastructural studies on *Sphaeroplea annulina* (Chlorophyceae). 1. Vegetative structure and mitosis”. In: *Journal of Phycology* 16.3, pp. 313–320. (Cit. on pp. 60, 239).
– (1981). “Ultrastructural studies on *Sphaeroplea annulina* (Chlorophyceae). 2. Spermatogenesis and male gamete structure”. In: *Journal of Phycology* 17. (Cit. on p. 60).
Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden (2009). “BLAST+: architecture and applications”. In: *BMC Bioinformatics* 10.1, p. 421. (Cit. on p. 11).
Camin, J. H. and R. R. Sokal (1965). “A method for deducing branching sequences in phylogeny”. In: *Evolution* 19.3, pp. 311–326. (Cit. on pp. 5, 17, 18, 60).
Cane, J. H. (1986). “Predator deterrence by mandibular gland secretions of bees (Hymenoptera, Apoidea)”. In: *Journal of Chemical Ecology* 12.6, pp. 1295–1309. (Cit. on p. 154).

Cangelosi, G. A., A. M. Hamlin, R. Marin 3rd, and C. A. Scholin (1997). "Detection of stable pre-rRNA in toxigenic *Pseudo-nitzschia* species". In: *Applied and Environmental Microbiology* 63.12, p. 4859. (Cit. on p. 28).

Carter, C. and R. W. Thornburg (2004). "Is the nectar redox cycle a floral defense against microbial attack?" In: *Trends in Plant Science* 9.7, pp. 320–324. (Cit. on p. 210).

Carugo, O. and S. Pongor (2002). "Recent progress in protein 3D structure comparison". In: *Current Protein and Peptide Science* 3.4, pp. 441–449. (Cit. on p. 230).

Casiraghi, M., M. Labra, E. Ferri, A. Galimberti, and F. De Mattia (2010). "DNA barcoding: a six-question tour to improve users' awareness about the method". In: *Briefings in Bioinformatics* 11.4, pp. 440–453. (Cit. on pp. 4, 237).

Caterino, M. S., S. Cho, and F. A. H. Sperling (2000). "The current state of insect molecular systematics: a thriving tower of Babel". In: *Annual Reviews of Entomology* 45, pp. 1–54. (Cit. on p. 124).

Chan, K. (1973). "A review of the genus *Coelastrum* (Chlorophyceae)". In: *Journal of the Chinese University of Hong Kong* 1, pp. 275–281. (Cit. on p. 74).

Chapman, R. L., M. A. Buchheim, C. F. Delwiche, T. Friedl, V. A. R. Huss, K. G. Karol, L. A. Lewis, J. Manhart, R. M. McCourt, J. L. Olsen, and D. A. Waters (1998). *Molecular systematics of the green algae*. Köln Germany: Kluwer Academic Publishing, pp. 508–540. (Cit. on p. 60).

Chase, M. W. and M. F. Fay (2009a). "Barcoding of plants and fungi". In: *Science* 325.5941, p. 682. (Cit. on pp. 86, 240).

– (2009b). "Barcoding of plants and fungi: Response to J. Schultz and M. Wolf's E-Letter". In: *Science Online E-letters* 325.5941, p. 682. (Cit. on pp. 86, 240).

Chase, M. W., S. Knapp, A. V. Cox, J. J. Clarkson, Y. Butsko, J. Joseph, V. Savolainen, and A. S. Parokonny (2003). "Molecular systematics, GISH and the origin of hybrid taxa in Nicotiana (Solanaceae)". In: *Annals of Botany* 92.1, p. 107. (Cit. on pp. 86, 240).

Chase, M. W., N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, and V. Savolainen (2005). "Land plants and DNA barcodes: short-term and long-term goals". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 360.1462, p. 1889. (Cit. on p. 86).

Chase, M. W., R. S. Cowan, P. M. Hollingsworth, C. van den Berg, S. Madriñán, G. Petersen, O. Seberg, T. Jorgsensen, K. M. Cameron, and M. Carine (2007). "A proposal for a standardised protocol to barcode all land plants". In: *Taxon* 56.2, pp. 295–299. (Cit. on p. 86).

Chen, S., H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao, and X. Pang (2010). "Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species". In: *PLoS ONE* 5, p. 8613. (Cit. on pp. 86, 240).

Cifuentes, A. S., M. A. González, I. Inostroza, and A. Aguilera (2001). "Reappraisal of physiological attributes of nine strains of *Dunaliella* (Chlorophyceae): growth and pigment content across a salinity gradient". In: *Journal of Phycology* 37.2, pp. 334–344. (Cit. on p. 86).

Clement, M., D. Posada, and K. A. Crandall (2000). "TCS: a computer program to estimate gene genealogies". In: *Molecular Ecology* 9.10, pp. 1657–1660. (Cit. on p. 124).

Coat, G., P. Dion, M. C. Noailles, B. de Reviers, J. M. Fontaine, Y. Berger-Perrot, and S. Goér (1998). "Ulva armoricana (Ulvales, Chlorophyta) from the coasts of Brittany (France). II. Nuclear rDNA ITS sequence analysis". In: *European Journal of Phycology* 33.1, pp. 81–86. (Cit. on p. 86).

Coleman, A. W. (1999). "Phylogenetic analysis of "Volvocacae" for comparative genetic studies". In: *Proceedings of the National Academy of Sciences of the USA* 96.24, p. 13892. (Cit. on p. 86).

– (2000). "The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence." In: *Protist* 151.1, p. 1. (Cit. on pp. 6, 28).

– (2001). "Biogeography and speciation in the *Pandorina/Volvulina* (Chlorophyta) superclade". In: *Journal of Phycology* 37.5, pp. 836–851. (Cit. on p. 86).

– (2003). "ITS2 is a double-edged tool for eukaryote evolutionary comparisons". In: *Trends in Genetics* 19.7, pp. 370–375. (Cit. on pp. 6, 7, 28, 38, 46, 60, 74, 86, 124, 237, 241).

– (2007). "Pan-eukaryote ITS2 homologies revealed by RNA secondary structure". In: *Nucleic Acids Research* 35.10, p. 3322. (Cit. on pp. 6, 28, 38, 46, 60, 86, 124).

– (2009). "Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide". In: *Molecular Phylogenetics and Evolution* 50.1, pp. 197–203. (Cit. on pp. 7, 74, 86, 124, 241).

Coleman, A. W. and J. C. Mai (1997). "Ribosomal DNA and ITS-2 sequence comparisons as a tool for predicting genetic relatedness". In: *Journal of Molecular Evolution* 45.2, pp. 168–177. (Cit. on p. 86).

Coleman, A. W. and V. D. Vacquier (2002). "Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (*Haliotis*)". In: *Journal of Molecular Evolution* 54.2, pp. 246–257. (Cit. on pp. 6, 28, 60).

Comas, A. (1989). "Taxonomische Übersicht der zönobialen Chlorokokkalalgen von Kuba. II. Fam. Coelastraceae". In: *Archiv für Hydrobiologie. Supplementband, Algological studies* 82.3, pp. 347–364. (Cit. on p. 74).

– (1996). "Las Chlorococcales dulciacuicolas de Cuba". In: *Bibliotheca Phycologica* 99, pp. 1–192. (Cit. on p. 74).

Corpe, W. A. and S. Rheem (1989). "Ecology of the methylotrophic bacteria on living leaf surfaces". In: *FEMS Microbiology Letters* 62.4, pp. 243–249. (Cit. on p. 210).

Costello, M. J., M. Coll, R. Danovaro, P. Halpin, H. Ojaveer, and P. Miloslavich (2010). "A census of marine biodiversity knowledge, resources, and future challenges". In: (cit. on p. 3).

Côté, C. A. and B. A. Peculis (2001). "Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast". In: *Nucleic Acids Research* 29.10, pp. 2106–2116. (Cit. on pp. 28, 60).

Côté, C. A., C. L. Greer, and B. A. Peculis (2002). "Dynamic conformational model for the role of ITS2 in pre-rRNA processing in yeast". In: *RNA* 8.06, pp. 786–797. (Cit. on pp. 6, 7, 28, 60, 230).

Couté, A. (1984). "Premieres observations au MET et au MEB sur la cytologie de *Ducellieria chodatii* (Ducel., Xanthophyceae, Mischococcales, Chlorobotrydaceae). First observations with TEE and SEM on the cytology". In: *Nova Hedwigia* 39.3-4, pp. 651–669. (Cit. on p. 74).

Coutsis, J. G. (1986). "The blue butterflies of the genus *Agrodiaetus* Hubner (Lep., Lycaenidae): symptoms of taxonomic confusion". In: *Nota Lepidopterologica* 9.3-4, pp. 159–169. (Cit. on p. 124).

Coutsis, J. G. and J. de Prins (2005). "A new brown *Polyommatus* (*Agrodiaetus*) from northern Greece (Lepidoptera: Lycaenidae)". In: *Phegea* 33.4, pp. 129–136. (Cit. on p. 124).

Cracraft, J. (1989). *Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation*. MA USA: Sinauer Associates, pp. 28–59. (Cit. on p. 3).

Cronn, R. C., R. L. Small, T. Haselkorn, and J. F. Wendel (2002). "Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes". In: *American Journal of Botany* 89.4, p. 707. (Cit. on p. 6).

Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner (2004). "WebLogo: a sequence logo generator". In: *Genome Research* 14.6, p. 1188. (Cit. on p. 28).

Cyberinfrastructure for phylogenetic research (2010). *Cipres Portal*. URL: http://www.phylo.org/sub_sections/portal. (Cit. on p. 18).

Daly, H. V. and K. N. Magnacca (2003). *Insects of Hawaii*. Vol. Hawaiian Hylaeus (Nesoprospis) bees (Hymenoptera: Apoidea). HI USA: University of Hawaii Press. (Cit. on p. 154).

Dangeard, P. A. (1889). "Memoire sur les Algues". In: *Le Botaniste* 1, pp. 127–174. (Cit. on p. 74).

Dantchenko, A. (2000). "Genus *Agrodiaetus*". In: *Guide to the butterflies of Russia and adjacent territories*. Vol. 2. Sofia, Bulgaria: Pensoft, pp. 196–214. (Cit. on p. 124).

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. NY USA: D. Appleton New York. (Cit. on pp. 3, 4).

Daugbjerg, N., O. Moestrup, and P. Arctander (1994). "Phylogeny of the genus *Pyramimonas* (Prasinophyceae, Chlorophyta) inferred from the rbcL gene". In: *Journal of Phycology* 30.6, pp. 991–999. (Cit. on p. 86).

– (1995). "Phylogeny of genera of Prasinophyceae and Pedinophyceae (Chlorophyta) deduced from molecular analysis of the rbcL gene". In: *Phycological Research* 43.4, pp. 203–213. (Cit. on p. 86).

de Lesse, H. (1960). "Speciation et variation chromosomiques chez les Lepidopteres Rhopaloceres". In: *Annales des Sciences Naturelles comprenant la zoologie* 2.1, pp. 1–223. (Cit. on p. 124).

de Salle, R., M. G. Egan, and M. Siddall (2005). "The unholy trinity: taxonomy, species delimitation and DNA barcoding". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 360.1462, pp. 1905–1916. URL: http://rstb.royalsocietypublishing.org/content/360/1462/1935. (Cit. on pp. 86, 124).

de Vega, C., M. Arista, P. L. Ortiz, C. M. Herrera, and S. Talavera (2009). "The ant-pollination system of Cytinus hypocistis (Cytinaceae), a Mediterranean root holoparasite". In: *Annals of Botany* 103.7, pp. 1065–1075. (Cit. on p. 154).

Deason, T. R., P. C. Silva, S. Watanabe, and G. L. Floyd (1991). "Taxonomic status of the species of the green algal genus *Neochloris*". In: *Plant*

Systematics and Evolution 177.3, pp. 213–219. (Cit. on pp. 60, 239).

Denslow, J. S. (2003). "Weeds in paradise: Thoughts on the invasibility of tropical islands". In: *Annals of the Missouri Botanical Garden* 90.1, pp. 119–127. (Cit. on p. 154).

Dixon, M. T. and D. M. Hillis (1993). "Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis". In: *Molecular Biology and Evolution* 10.1, p. 256. (Cit. on p. 7).

Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, G. W. Rouse, M. Obst, G. D. Edgecombe, et al. (2008). "Broad phylogenomic sampling improves resolution of the animal tree of life". In: *Nature* 452.7188, pp. 745–749. (Cit. on p. 6).

Durand, C., M. Manuel, C. F. Boudouresque, A. Meinesz, M. Verlaque, and Y. le Parco (2002). "Molecular data suggest a hybrid origin for the invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta) in the Mediterranean Sea". In: *Journal of Evolutionary Biology* 15.1, pp. 122–133. (Cit. on p. 86).

Dötterl, S., L. M. Wolfe, and A. Jürgens (2005). "Qualitative and quantitative analyses of flower scent in Silene latifolia". In: *Phytochemistry* 66.2, pp. 203–213. (Cit. on pp. 154, 210).

Dötterl, S., A. Jürgens, K. Seifert, T. Laube, B. Weißbecker, and S. Schütz (2006). "Nursery pollination by a moth in Silene latifolia: the role of odours in eliciting antennal and behavioural responses". In: *New Phytologist* 169, pp. 707–718. (Cit. on p. 154).

Dötterl, S., A. Jürgens, L. Wolfe, and A. Biere (2009). "Disease status and population origin effects on floral scent: potential consequences for oviposition and fruit predation in a complex interaction between a plant, fungus, and noctuid moth". In: *Journal of Chemical Ecology* 35.3, pp. 307–319. (Cit. on pp. 154, 210).

Ebach, M. C. and C. Holdrege (2005). "More taxonomy, not DNA barcoding". In: *BioScience* 55.10, pp. 823–824. (Cit. on p. 86).

Eckweiler, W. and C. L. Hauser (1997). "An illustrated checklist of *Agrodiaetus* Hubner, 1822, a subgenus of *Polyommatus* Latreille, 1804 (Lepidoptera, Lycaenidae)". In: *Nachrichten Entomologischer Verein Apollo* Supplement 16, pp. 113–166. (Cit. on p. 124).

Eddy, S. R. (1998). "Profile hidden Markov models". In: *Bioinformatics* 14.9, p. 755. (Cit. on pp. 11, 13, 28, 38, 237).

Edlind, T. D., C. Sharetzky, and M. E. Cha (1990). "Ribosomal RNA of the primitive eukaryote *Giardia lamblia*: large subunit domain I and potential processing signals". In: *Gene* 96.2, pp. 289–293. (Cit. on pp. 28, 86).

Egger, B., S. KoblMüller, C. Sturmbauer, and K. M. Sefc (2007). "Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus* ". In: *BMC Evolutionary Biology* 7.1, p. 137. (Cit. on p. 46).

Ehlers, B. K. and J. M. Olesen (1997). "The fruit-wasp route to toxic nectar in *Epipactis orchids*". In: *Flora* 192, pp. 223–229. (Cit. on p. 210).

Eliot, J. N. (1973). "The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement". In: *Bulletin of the British Museum (Natural History) Entomology* 28.6, pp. 371–505. (Cit. on p. 124).

Engelmann, J. C., S. Rahmann, M. Wolf, J. Schultz, E. Fritzilas, S. Kneitz, T. Dandekar, and T. Müller (2008). "Modelling cross-hybridization on phylogenetic DNA microarrays increases the detection power of closely related species". In: *Molecular Ecology Resources* 9.1, pp. 83–93. (Cit. on pp. 28, 38, 86).

Engler, A. and K. Prantl (1897). *Die natürlichen Pflanzenfamilien I*. Leipzig Germany: Verlag von Wilhelm Engelmann. (Cit. on p. 60).

Ercolani, G. L. (1991). "Distribution of epiphytic bacteria on olive leaves and the influence of leaf age and sampling time". In: *Microbial Ecology* 21.1, pp. 35–48. (Cit. on p. 210).

Erixon, P., B. Svennblad, T. Britton, and B. Oxelman (2003). "Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics". In: *Systematic Biology* 52.5, pp. 665–673. (Cit. on p. 46).

Erwin, T. L. (2002). "Tropical forests: their richness in Coleoptera and other arthropod species". In: *Foundations of tropical forest biology: classic papers with commentaries* 36.1, p. 438. (Cit. on p. 3).

Ettl, H. (1978). "Xanthophyceae, 1. Teil". In: *Süßwasserflora von Mitteleuropa* 3. (Cit. on p. 74).

European Distributed Institute of Taxonomy (2010). *EDIT*. URL: http://www.e-taxonomy.eu. (Cit. on p. 3).

Fabry, S., A. Köhler, and A. W. Coleman (1999). "Intraspecies analysis: comparison of ITS sequence data and gene intron sequence data with breeding data for a worldwide collection of *Gonium pectorale*". In: *Journal of Molecular Evolution* 48.1, pp. 94–101. (Cit. on p. 86).

Famá, P., J. L. Olsen, W. T. Stam, and G. Procaccini (2000). "High levels of intra-and inter-individual polymorphism in the rDNA ITS1 of *Caulerpa racemosa* (Chlorophyta)". In: *European*

Journal of Phycology 35.4, pp. 349–356. (Cit. on p. 86).

Fawley, M. W., K. P. Fawley, and H. A. Owen (2005). "Diversity and ecology of small coccoid green algae from Lake Itasca, Minnesota, USA, including *Meyerella planktonica*, gen. et sp. nov". In: *Phycologia* 44.1. (Cit. on p. 86).

Fazekas, A. J., K. S. Burgess, P. R. Kesanakurthi, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, M. Hajibabaei, and S. C. H. Barrett (2008). "Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well". In: *PLoS One* 3.7. (Cit. on p. 86).

Feinberg, J. (2009). *Wordle*. URL: http://www.wordle.net. (Cit. on pp. 11, 285).

Feinsinger, P. and L. A. Swarm (1978). "How common are ant-repellent nectars?" In: *Biotropica* 10, pp. 238–239. (Cit. on p. 154).

Feliner, G. N. and J. A. Rosselló (2007). "Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants". In: *Molecular Phylogenetics and Evolution* 44.2, pp. 911–919. (Cit. on p. 6).

Felsenstein, J. (1981). "Evolutionary trees from DNA sequences: a maximum likelihood approach". In: *Journal of Molecular Evolution* 17.6, pp. 368–376. (Cit. on pp. 5, 17, 18, 60).

– (1985). "Confidence limits on phylogenies: an approach using the bootstrap". In: *Evolution* 39.4, pp. 783–791. (Cit. on pp. 18, 46, 60).

– (1989). "PHYLIP-phylogeny inference package (version 3.2)". In: *Cladistics* 5.1, pp. 164–166. (Cit. on pp. 11, 18, 24, 46).

– (2004). "Inferring phylogenies". In: *Sunderland (Massachusetts): Sinauer Associates*. (Cit. on pp. 4, 5, 237).

Felsenstein, J., J. Archie, W. H. E. Day, W. Maddinson, C. Meacham, F. J. Rohlf, and D. L. Swofford (1986). "The Newick tree format". In: *Meeting of the Society for the Study of Evolution*. IL USA: American Society of Naturalists. (Cit. on p. 18).

Fenwick, M. G. (1962). "Some interesting algae from Lake Huron". In: *Transactions of the American Microscopical Society* 81.1, pp. 72–76. (Cit. on p. 74).

– (1968). "Review of the status of some green algae in the genus *Coelastrum*". In: *Michigan Botanist* 7.1, pp. 129–131. (Cit. on p. 74).

Fenwick, M. G., L. O. Hansen, and D. L. Lynch (1966). "Polymorphic forms of *Coelastrum proboscideum* Bohn". In: *Transactions of the American Microscopical Society* 85.4, pp. 579–581. (Cit. on p. 74).

Frezal, L. and R. Leblois (2008). "Four years of DNA barcoding: current advances and prospects". In: *Infection, Genetics and Evolution* 8, pp. 727–736. (Cit. on p. 124).

Friedl, T. (1996). "Evolution of the polyphyletic genus *Pleurastrum* (Chlorophyta): inferences from nuclear-encoded ribosomal DNA sequences and motile cell ultrastructure". In: *Phycologia* 35.5, pp. 456–469. (Cit. on p. 86).

Friedl, T., T. Pröschold, L. A. Lewis, and M. R. Letsch (2009). "Symbiosis in green algae: origin and diversity of a successful life style". In: *Phycologia* 48, pp. 31–32. (Cit. on pp. 86, 240).

Friedrich, J., T. Dandekar, M. Wolf, and T. Müller (2005). "ProfDist: a tool for the construction of large phylogenetic trees based on profile distances". In: *Bioinformatics* 21.9, p. 2108. (Cit. on pp. 11, 18, 21, 46, 60, 74, 86, 124, 237).

Fritsch, F. E. (1918). "A first report on the freshwater algae mostly from the Cape Peninsula in the herbarium of the South African Museum". In: *Annales of the South African Museum* 9.1, 483–611. (Cit. on p. 74).

Funk, D. J. and K. E. Omland (2003). "Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA". In: *Annual Review of Ecology, Evolution, and Systematics* 34, pp. 397–423. (Cit. on p. 124).

Galen, C. (1983). "The effects of nectar thieving ants on seedset in floral scent morphs of Polemonium viscosum". In: *Oikos* 41, pp. 245–249. (Cit. on p. 154).

– (1999). "Flowers and enemies: predation by nectar-thieving ants in relation to variation in floral form of an alpine wildflower, Polemonium viscosum". In: *Oikos* 85.3, pp. 426–434. (Cit. on p. 154).

Galen, C. and B. Butchart (2003). "Ants in your plants: effects of nectar-thieves and pollen fertility and seed-siring capacity in the alpine wildflower, Polemonium viscosum". In: *Oikos* 101, pp. 521 –528. (Cit. on p. 154).

Galen, C. and J. Cuba (2001). "Down the tube: pollinators, predators, and the evolution of flower shape in the Alpine Skypilot, Polemonium viscosum". In: *Evolution* 55, pp. 1963–1971. (Cit. on p. 154).

Galen, C. and J. C. Geib (2007). "Density-dependent effects of ants on selection for bumble bee pollination in Polemonium viscosum". In: *Ecology* 88.5, 1202–1209. (Cit. on p. 154).

Gao, T., H. Yao, J. Song, C. Liu, Y. Zhu, X. Ma, X. Pang, H. Xu, and S. Chen (2010). "Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2". In: *Jour-*

nal of Ethnopharmacology 130.1, pp. 116 –121. (Cit. on p. 240).

Gardener, M. C. and C. C. Daehler (2006). "Documenting floral visitors to rare Hawaiian plants using automated video recordings". In: *Pacific Conservation Biology* 12, pp. 189–194. (Cit. on p. 154).

Gascuel, O. (1997). "BioNJ: an improved version of the NJ algorithm based on a simple model of sequence data". In: *Molecular Biology and Evolution* 14.7, p. 685. (Cit. on p. 18).

Gebiola, M., U. Bernardo, and R. A. Burks (2010). "A reevaluation of the generic limits of *Pnigalio* Schrank (Hymenoptera: Eulophidae) based on molecular and morphological evidence". In: *Zootaxa* 2484, pp. 35–44. (Cit. on p. 241).

Gershenzon, J. and N. Dudareva (2007). "The function of terpene natural products in the natural world". In: *Nature Chemical Biology* 3.7, pp. 408–414. (Cit. on p. 210).

Gesell, T. and A. von Haeseler (2006). "In silico sequence evolution with site-specific interactions along phylogenetic trees". In: *Bioinformatics* 22.6, p. 716. (Cit. on pp. 11, 21, 46).

Ghazoul, J. (2001). "Can floral repellents preempt potential ant-plant conflicts?" In: *Ecology Letters* 4, pp. 295 –299. (Cit. on p. 154).

Giles, B. E., T. M. Pettersson, U. Carlsson-Graner, and P. K. Ingvarsson (2006). "Natural selection on floral traits of female *Silene dioica* by a sexually transmitted disease". In: *New Phytologist* 169.4, pp. 729–739. (Cit. on pp. 210, 242).

Gillespie, J. J., J. S. Johnston, J. J. Cannone, and R. R. Gutell (2006). "Characteristics of the nuclear (18S, 5.8 S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements". In: *Insect Molecular Biology* 15.5, p. 657. (Cit. on pp. 13, 28).

Global Taxonomy Initiative (2010). *GTI*. URL: http://www.cbd.int/gti/. (Cit. on p. 3).

Gomez, J. M. and R. Zamora (1992). "Pollination by ants: consequences of quantitative effects on a mutualistic system". In: *Oecologia* 91, pp. 410–418. (Cit. on p. 154).

Gomez, J. M., R. Zamora, J. A. Hodar, and D. Garcia (1996). "Experimental study of pollination by ants in mediterranean high mountain and arid habitats". In: *Oecologia* 105, pp. 236–242. (Cit. on p. 154).

Gomez, J. M., F. Perfecti, J. Bosch, and J. P. M. Camacho (2009). "A geographic selection mosaic in a generalized plant–pollinator–herbivore system". In: *Ecological Monographs* 79, pp. 245–263. (Cit. on p. 154).

Gonzalez, I. L., C. Chambers, J. L. Gorski, D. Stambolian, R. D. Schmickel, and J. E. Sylvester (1990). "Sequence and structure correlation of human ribosomal transcribed spacers." In: *Journal of Molecular Biology* 212.1, p. 27. (Cit. on p. 28).

González, M. A., A. W. Coleman, P. I. Gómez, and R. Montoya (2001). "Phylogenetic relationship among various strains of *Dunaliella* (Chlorophyceae) based on nuclear ITS rDNA sequences". In: *Journal of Phycology* 37.4, pp. 604–611. (Cit. on p. 86).

Google Inc. (2010). *Google Scholar™*. URL: http://scholar.google.com. (Cit. on p. 11).

Gorodkin, J., L. J. Heyer, S. Brunak, and G. D. Storomo (1997). "Displaying the in formation contents of structural RNA alignments: the structure logos". In: *Bioinformatics* 13.6, p. 583. (Cit. on p. 38).

Gouy, M. (1995). *NJplot*. URL: http://pbil.univ-lyon1.fr/software/njplot.html. (Cit. on pp. 11, 18).

Gowri-Shankar, V. and M. Rattray (2006). "On the correlation between composition and site-specific evolutionary rate: implications for phylogenetic inference". In: *Molecular Biology and Evolution* 23.2, p. 352. (Cit. on p. 6).

Grajales, A., C. Aguilar, and J. Sánchez (2007). "Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS 2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals". In: *BMC Evolutionary Biology* 7.1, p. 90. (Cit. on pp. 46, 60, 230).

Graybeal, A. (1998). "Is it better to add taxa or characters to a difficult phylogenetic problem?" In: *Systematic Biology* 47.1, pp. 9–17. (Cit. on p. 46).

Greeff, R. (1873). "Über Radiolarien und radiolarienartige Rhizopoden des süßen Wassers." In: *Sitzungsberichte der Gesellschaft zur Beförderung der gesammten Naturwissenschaften zu Marburg* 5, 47–64. (Cit. on p. 74).

Griekspoor, A. (2010). *Papers*. URL: http://mekentosj.com/papers. (Cit. on p. 11).

Griffiths-Jones, S., A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy (2003). "Rfam: An RNA family database". In: *Nucleic Acids Research* 31.1, p. 439. (Cit. on pp. 12, 13, 28).

Guerrant, E. O. and P. G. Fiedler (1981). "Flower defenses against nectar-pilferage by ants". In: *Biotropica* 13, pp. 25–33. (Cit. on p. 154).

Guiry, M. D. "AlgaeBase—listing the world's algae". In: *The Irish Scientist 2005 Yearbook*, pp. 74–75. (Cit. on p. 60).

Gunnison, D. and M. Alexander (1975). "Basis for the resistance of several algae to microbial decomposition". In: *Applied and Environmental Microbiology* 29.6, p. 729. (Cit. on p. 74).
Gutell, R. R., N. Larsen, and C. R. Woese (1994). "Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective." In: *Microbiology and Molecular Biology Reviews* 58.1, p. 10. (Cit. on pp. 7, 28).
Haber, W. A., G. W. Frankie, H. G. Baker, I. Baker, and S. Koptur (1981). "Ants like flower nectar". In: *Biotropica* 13, pp. 211–214. (Cit. on p. 154).
Hajdu, L., H. E., and C. G. (1976). "Beiträge zur Taxonomie der Gattung *Coelastrum*(Chlorophyta, Chlorococcales". In: *Annales Historico Naturales Musei Nationalis Hungarici* 68, 31–37. (Cit. on p. 74).
Haldane, J. B. S. (1922). "Sex ratio and unisexual sterility in hybrid animals". In: *Journal of Genetics* 12, pp. 101–109. (Cit. on p. 124).
Hanagata, N. (1998). "Phylogeny of the subfamily Scotiellocystoideae (Chlorophyceae, Chlorophyta) and related taxa inferred from 18S ribosomal RNA gene sequence data". In: *Journal of Phycology* 34.6, pp. 1049–1054. (Cit. on p. 74).
– (2001). "New species of *Coelastrella* and *Scenedesmus* (Chlorophyceae, Chlorophyta)". In: *Journal of Japanese Botany* 76.3, pp. 129–136. (Cit. on p. 74).
Hanley, M. E., B. B. Lamont, and W. S. Armbruster (2009). "Pollination and plant defence traits co-vary in Western Australian Hakeas". In: *New Phytologist* 182.1, pp. 251–260. (Cit. on p. 154).
Harley, R. (1991). "The greasy pole syndrome". In: *Ant-Plant Interactions*. Ed. by C. R. Huxley and D. F. Cutler. Oxford UK: Oxford University Press, pp. 430–433. (Cit. on p. 154).
Harpke, D. and A. Peterson (2006). "Non-concerted ITS evolution in *Mammillaria* (Cactaceae)". In: *Molecular Phylogenetics and Evolution* 41.3, pp. 579–593. (Cit. on p. 86).
– (2008a). "5.8S motifs for the identification of pseudogenic ITS regions". In: *Botany* 86.3, pp. 300–305. (Cit. on pp. 28, 86, 124).
– (2008b). "Extensive 5.8S nrDNA polymorphism in *Mammillaria* (Cactaceae) with special reference to the identification of pseudogenic internal transcribed spacer regions". In: *Journal of Plant Research* 121.3, pp. 261–270. (Cit. on p. 28).
Harrewijn, P., A. K. Minks, and C. Mollema (1994). "Evolution of plant volatile production in insect-plant relationships". In: *Chemoecology* 5.2, pp. 55–73. (Cit. on p. 210).
Harris, D. J. and K. A. Crandall (2000). "Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies". In: *Molecular Biology and Evolution* 17.2, pp. 284–291. (Cit. on p. 124).
Hasegawa, H. and L. Holm (2009). "Advances and pitfalls of protein structural alignment". In: *Current Opinion in Structural Biology*, pp. 341–348. (Cit. on p. 230).
Havgaard, J. H., R. B. Lyngso, G. D. Stormo, and J. Gorodkin (2005). "Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%". In: *Bioinformatics* 21.9, p. 1815. (Cit. on p. 6).
Hebert, P. D., S. Ratnasingham, and J. R. de Waard (2003a). "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species". In: *Proceedings of the Royal Society of London. Series B: Biological sciences* 270.Suppl 1, S96–99. (Cit. on p. 124).
Hebert, P. D., A. Cywinska, S. L. Ball, and J. R. de Waard (2003b). "Biological identifications through DNA barcodes". In: *Proceedings of the Royal Society of London. Series B: Biological sciences* 270.1512, pp. 313–321. (Cit. on pp. 124, 237).
Hebert, P. D., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs (2004). "Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* ". In: *Proceedings of the National Academy of Sciences of the USA* 101.41, pp. 14812–14817. (Cit. on pp. 3, 4, 124).
Hebert, P. D. N. and T. R. Gregory (2005). "The promise of DNA barcoding for taxonomy". In: *Systematic Biology* 54.5, p. 852. (Cit. on p. 86).
Hegewald, E. (1989). "The *Scenedesmus* strains of the culture collection of the University of Texas at Austin (UTEX)". In: *Algological Studies* 55, pp. 153–189. (Cit. on p. 74).
– (1997). "Taxonomy and phylogeny of *Scenedesmus*". In: *Algae* 12, pp. 235–246. (Cit. on p. 74).
Hegewald, E. and T. R. Deason (1989). "*Pseudodidymocystis*, a new genus of Scenedesmaceae (Chlorophyceae)". In: *Archiv für Hydrobiologie. Supplementband, Algological studies* 82, pp. 119–127. (Cit. on p. 74).
Hegewald, E. and N. Hanagata (1992). "*Asterarcys* Comas, eine weit verbreitete tropische Grünalgengattung". In: *Archiv für Hydrobiologie. Supplementband, Algological studies* 95, 25–30. (Cit. on p. 74).
– (2000). "Phylogenetic studies on Scenedesmaceae (Chlorophyta)". In: *Archiv für Hydro-*

biologie. Supplementband, Algological studies 136, pp. 29–49. (Cit. on pp. 74, 86).

Hegewald, E. and M. Wolf (2003). "Phylogenetic relationships of *Scenedesmus* and *Acutodesmus* (Chlorophyta, Chlorophyceae) as inferred from 18S rDNA and ITS-2 sequence comparisons". In: *Plant Systematics and Evolution* 241.3, pp. 185–191. (Cit. on pp. 28, 74, 86, 239).

Hegewald, E., D. Hepperle, M. Wolf, and L. Krienitz (2001). "Phylogenetic placement of *Chlorotetraedron incus*, *C. polymorphum* and *Polyedriopsis spinulosa* (Neochloridaceae, Chlorophyta)". In: *Phycologia* 40.5, pp. 399–402. (Cit. on p. 60).

Hegewald, E., P. F. M. Coesel, and P. Hegewald (2002). "A phytoplankton collection from Bali, with the description of a new *Desmodesmus* species (Chlorophyta, Scenedesmaceae)". In: *Algological Studies* 105, pp. 51–78. (Cit. on p. 74).

Hegewald, E., A. Schmidt, A. Braband, and P. M. Tsarenko (2005). "Revision of the Desmodesmus (Sphaeropleales, Scenedesmaceae) species with lateral spines. 2. The multi-spined to spineless taxa". In: *Algological Studies* 116.1, pp. 1–38. (Cit. on p. 74).

Hegewald, E., M. Wolf, A. Keller, T. Friedl, and L. Krienitz (2010). "ITS2 sequence-structure phylogeny in the Scenedesmaceae with special reference to *Coelastrum* (Chlorophyta, Chlorophyceae), including the new genera *Comasiella* and *Pectinodesmus*". In: *Phycologia* 49, pp. 325–335. (Cit. on pp. 74, 86, 239).

Heil, M. and D. McKey (2003). "Protective ant-plant interactions as model systems in ecological and evolutionary research". In: *Annual Review of Ecology Evolution and Systematics* 34, pp. 425–453. (Cit. on p. 154).

Hennig, W. (1966). *Phylogenetic systematics*. IL USA: University of Illinois Press. (Cit. on p. 3).

Henrick, K., Z. Feng, W. F. Bluhm, D. Dimitropoulos, J. F. Doreleijers, S. Dutta, J. L. Flippen-Anderson, J. Ionides, C. Kamada, and E. Krissinel (2008). "Remediation of the protein data bank archive". In: *Nucleic Acids Research* 36.Database issue, p. D426. (Cit. on pp. 12, 16, 230).

Hepperle, D., L. Krienitz, and T. Hollibaugh (2001a). "Molecular phylogeny of picoplanktonic chlorophytes and the discovery of a new evolutionary lineage". In: *Phycologia* 40, p. 36. (Cit. on p. 86).

Hepperle, D., E. Hegewald, and L. Krienitz (2001b). "Phylogenetic position of the Oocystaceae (Chlorophyta)". In: *Journal of Phycology* 36.3, pp. 590–595. (Cit. on pp. 74, 86).

Herrera, C. M., J. Herrera, and X. Espadaler (1984). "Nectar thievery by ants from southern Spanish insect-pollinated flowers". In: *Insectes Sociaux* 31, pp. 142–154. (Cit. on p. 154).

Herrera, C. M., I. Garcia, and R. Perez (2008). "Invisible floral larcenies: microbial communities degrade floral nectar of bumble bee-pollinated plants". In: *Ecology* 89.9, pp. 2369–2376. (Cit. on p. 210).

Hershkovitz, M. A. and L. A. Lewis (1996). "Deep-level diagnostic value of the rDNA-ITS region". In: *Molecular Biology and Evolution* 13.9, p. 1276. (Cit. on pp. 5, 28, 86).

Hesse, M., E. Kusel-Fetzmann, and K. Carniel (1989). "Life cycle and ultrastructure of *Ducellieria chodati* (Oomycetes)". In: *Plant Systematics and Evolution* 165.1, pp. 1–15. (Cit. on p. 74).

Hesselbarth, G., von Oorschot, H., and S. Wagener (1995). *Die Tagfalter der Türkei unter Berücksichtigung der angrenzenden Länder*. Author's edition. (Cit. on p. 124).

Hillis, D. M. and J. J. Bull (1993). "An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis". In: *Systematic Biology* 42.2, p. 182. (Cit. on pp. 5, 46).

Hillis, D. M. and M. T. Dixon (1991). "Ribosomal DNA: molecular evolution and phylogenetic inference". In: *The quarterly Review of Biology* 66.4. (Cit. on pp. 6, 7, 237).

Hillis, D. M., J. P. Huelsenbeck, and C. W. Cunningham (1994). "Application and accuracy of molecular phylogenies". In: *Science* 264.5159, p. 671. (Cit. on p. 46).

Hirano, S. and C. Upper (2000). "Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*—a pathogen, ice nucleus, and epiphyte". In: *Microbiology and Molecular Biology Reviews* 64.3, p. 624. (Cit. on p. 210).

Hofacker, I. L. (2003). "Vienna RNA secondary structure server". In: *Nucleic Acids Research* 31.13, p. 3429. (Cit. on p. 11).

Hoffman, L. R. (1983). "*Atractomorpha echinata* gen. et sp. nov., a new anisogamous member of the Sphaeropleaceae (Chlorophyceae)". In: *Journal of Phycology* 19, pp. 76–86. (Cit. on p. 60).

– (1984). "*Atractomorpha porcata* sp. nov., new member of the Sphaeropleaceae (Chlorophyceae) from California". In: *Journal of Phycology* 20.2, pp. 225–236. (Cit. on p. 60).

Holdrege, C. and M. C. Ebach (2006). "Response from Holdrege and Ebach: What about Taxa?" In: *BioScience* 56.2, pp. 94–94. (Cit. on p. 86).

Hollingsworth, M. L., A. A. Clark, L. L. Forrest, J. Richardson, R. T. Pennington, D. G. Long, R. S. Cowan, M. W. Chase, M. Gaudeul, and P. M. Hollingsworth (2009a). "Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants". In: *Molecular Ecology Resources* 9.2, pp. 439–457. (Cit. on p. 86).

Hollingsworth, P. M., L. L. Forrest, J. L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M. W. Chase, R. S. Cowan, D. L. Erickson, and A. J. Fazekas (2009b). "A DNA barcode for land plants". In: *Proceedings of the National Academy of Sciences of the USA* 106.31, p. 12794. (Cit. on p. 86).

Holway, D. A., L. Lach, A. V. Suarez, N. D. Tsutsui, and T. J. Case (2002). "The causes and consequences of ant invasions". In: *Annual Review of Ecology and Systematics* 33, pp. 181–233. (Cit. on pp. 154, 241).

Huelsenbeck, J. P. and F. Ronquist (2001). "MrBayes: a program for the Bayesian inference of phylogeny". In: *Bioinformatics* 17, pp. 754–755. (Cit. on pp. 5, 18, 60).

Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham (1996). "Combining data in phylogenetic analysis". In: *Trends in Ecology and Evolution* 11.4, pp. 152–158. (Cit. on pp. 6, 28).

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback (2001). "Bayesian inference of phylogeny and its impact on evolutionary biology". In: *Science* 294.5550, p. 2310. (Cit. on pp. 5, 17, 18).

Irwin, R. E., L. S. Adler, and A. K. Brody (2004). "The dual role of floral traits: Pollinator attraction and plant defense". In: *Ecology* 85.6, pp. 1503–1511. (Cit. on p. 154).

ITS2 database Supplements (2009). "ITS2 database Supplements". In: URL: `http://its2.bioapps.biozentrum.uni-wuerzburg.de/cgi-bin/index.pl?supplements`. (Cit. on p. 124).

Jakupciak, J. P. and R. R. Colwell (2009). "Biological agent detection technologies". In: *Molecular Ecology Resources* 9, pp. 51–57. (Cit. on p. 86).

Janzen, D. H. (1977). "Why don't ants visit flowers?" In: *Biotropica* 9, p. 252. (Cit. on p. 154).

Jeon, S. L. and E. Hegewald (2006). "A revision of the species *Desmodesmus perforatus* and *D. tropicus* (Scenedesmaceae, Chlorophyceae, Chlorophyta)". In: *Phycologia* 45.5. (Cit. on p. 74).

Jie Zhu, Y., C. Shi-lin, Y. Hui, T. Rui, S. Jing-yuan, K. Luo, and L. Jing (2010). "DNA barcoding the medicinal plants of the genus *Paris*". In: *Acta Pharmaceutica Sinica*, p. 03. (Cit. on p. 240).

John, D. M., B. A. Whitton, and A. J. Brook (2002). *The freshwater algal flora of the British Isles: an identification guide to freshwater and terrestrial algae*. Cambridge UK: Cambridge University Press. (Cit. on p. 74).

Johnson, J. L., M. W. Fawley, and K. P. Fawley (2007). "The diversity of *Scenedesmus* and *Desmodesmus* (Chlorophyceae) in Itasca State Park, Minnesota, USA". In: *Phycologia* 46.2. (Cit. on p. 74).

Johnson, K. B. and V. O. Stockwell (1998). "Management of fire blight: A case study in microbial ecology". In: *Annual Review of Phytopathology* 36.1, pp. 227–248. (Cit. on p. 210).

Johnson, K. B., V. O. Stockwell, D. M. Burgett, D. Sugar, and J. E. Loper (1993). "Dispersal of *Erwinia amylovora* and *Pseudomonas fluorescens* by honey bees from hives to apple and pear blossoms". In: *Phytopathology* 83.5, pp. 478–484. (Cit. on p. 210).

Johnson, S. D., A. L. Hargreaves, and M. Brown (2006). "Dark, bitter-tasting nectar functions as a filter of flower visitors in a bird-pollinated plant". In: *Ecology* 87, 2709–2716. (Cit. on p. 154).

Jones, G. and S. M. van Parijs (1993). "Bimodal echolocation in pipistrelle bats: are cryptic species present?" In: *Proceedings of the Royal Society B: Biological Sciences* 251.1331, pp. 119–125. (Cit. on p. 3).

Joseph, N., E. Krauskopf, M. I. Vera, and B. Michot (1999). "Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast". In: *Nucleic Acids Research* 27.23, p. 4533. (Cit. on pp. 6, 28, 239).

Jossinet, F. and E. Westhof (2005). "Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure". In: *Bioinformatics* 21.15, p. 3320. (Cit. on pp. 16, 230).

– (2010). *Assemble*. URL: `http://www.bioinformatics.org/assemble`. (Cit. on pp. 11, 16, 242).

Jossinet, F., T. E. Ludwig, and E. Westhof (2007). "RNA structure: bioinformatic analysis". In: *Current Opinion in Microbiology* 10.3, pp. 279–285. (Cit. on pp. 230, 242).

Jow, H., C. Hudelot, M. Rattray, and P. G. Higgs (2002). "Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution". In: *Molecular Biology and Evolution* 19.9, p. 1591. (Cit. on pp. 6, 46).

Jow, H., V. Gowri-Shankar, and B. Guillard (2003). *PHASE: a software package for phylogenetics and*

sequence evolution. Manchester UK: University of Manchester. (Cit. on p. 238).

Junker, R. R. and N. Blüthgen (2008). "Floral scents repel potentially nectar-thieving ants". In: *Evolutionary Ecology Research* 10, pp. 295–308. (Cit. on p. 154).

– (2010a). "Dependency on floral resources determines the animals' responses to floral scents". In: *Plant Signaling and Behavior* 5.8, in press. (Cit. on p. 154).

– (2010b). "Floral scents repel facultative flower visitors, but attract obligate ones". In: *Annals of Botany* 105, pp. 777–782. (Cit. on p. 154).

Junker, R. R., A. Y. C. Chung, and N. Blüthgen (2007). "Interaction between flowers, ants and pollinators: additional evidence for floral repellence against ants". In: *Ecological Research* 22, 665–670. (Cit. on p. 154).

Junker, R. R., C. C. Daehler, S. Dötterl, A. Keller, and N. Blüthgen (2010a). "Ant-flower networks in Hawaii: native plants are exploited, introduced plants defended". In: *Ecological Monographs*, in press. (Cit. on pp. 154, 241).

Junker, R. R., C. Loewel, R. Gross, S. Dötterl, A. Keller, and N. Blüthgen (2010b). "Flower- and leaf- specificity of epiphytic bacterial communities – implications for pollination ecology". In: to be submitted. (Cit. on p. 242).

Junker, R. R., N. Höcherl, and N. Blüthgen (2010c). "Responses to olfactory signals reflect network structure of flower-visitor interactions". In: *Journal of Animal Ecology* 79, pp. 818–823. (Cit. on p. 154).

Kai, M., M. Haustein, F. Molina, A. Petri, B. Scholz, and B. Piechulla (2009). "Bacterial volatiles and their action potential". In: *Applied Microbiology and Biotechnology* 81.6, pp. 1001–1012. (Cit. on p. 210).

Kandul, N. P., V. A. Lukhtanov, A. V. Dantchenko, J. W. Coleman, C. H. Sekercioglu, D. Haig, and N. E. Pierce (2004). "Phylogeny of *Agrodiaetus* Hubner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF1-α: karyotype diversification and species radiation". In: *Systematic Biology* 53.2, pp. 278–298. (Cit. on pp. 124, 240).

Kandul, N. P., V. A. Lukhtanov, and N. E. Pierce (2007). "Karyotypic diversity and speciation in *Agrodiaetus* butterflies". In: *Evolution* 61.3, pp. 546–559. (Cit. on p. 124).

Kapraun, D. F. (1993). "Karyology of marine green algae". In: *Phycologia* 32.1, pp. 1–21. (Cit. on p. 86).

– (1994). "Cytophotometric estimation of nuclear DNA contents in thirteen species of the Caulerpales (Chlorophyta)". In: *Cryptogamic Botany* 4, pp. 410–410. (Cit. on p. 86).

Kapraun, D. F. and J. R. Buratti (1998). "Evolution of genome size in the Dasycladales (Chlorophyta) as determined by DAPI cytophotometry". In: *Phycologia* 37.3, pp. 176–183. (Cit. on p. 86).

Karapinar, M. et al. (1987). "Inhibition of foodborne pathogens by thymol, eugenol, menthol and anethole". In: *International Journal of Food Microbiology* 4.2, pp. 161–166. (Cit. on p. 210).

Kato, M., A. Shibata, T. Yasui, and H. Nagamasu (1999). "Impact of introduced honeybees, Apis mellifera, upon native bee communities in the Bonin (Ogasawara) Islands". In: *Researches on Population Ecology* 41.2, pp. 217–228. (Cit. on p. 154).

Keeler, K. H. (1985). "Extrafloral nectaries on plants in communities without ants: Hawaii". In: *Oikos* 44, pp. 407–414. (Cit. on pp. 154, 241).

Keller, A., T. Schleicher, F. Förster, B. Ruderisch, T. Dandekar, T. Müller, and M. Wolf (2008c). "ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales)". In: *BMC Evolutionary Biology* 8.1, p. 218. (Cit. on pp. 18, 28, 46, 60, 86, 124, 154, 230, 239, 241).

Keller, A., T. Schleicher, J. Schultz, T. Müller, T. Dandekar, and M. Wolf (2009a). "5.8S-28S rRNA interaction and HMM-based ITS2 annotation". In: *Gene* 430.1-2, pp. 50–57. (Cit. on pp. 15, 28, 38, 46, 74, 124, 154, 237).

Keller, A., F. Förster, T. Müller, T. Dandekar, J. Schultz, and M. Wolf (2010a). "Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees." In: *Biology Direct* 5.1, p. 4. (Cit. on pp. 46, 86, 154, 210, 230, 237).

Keller, A., M. Wolf, and T. Dandekar (2010b). "Ribosomal RNA phylogenetics: the third dimension". In: *Biologia* 65.3, pp. 388–391. (Cit. on pp. 230, 242, 243).

Keller, I., I. C. Chintauan-Marquier, P. Veltsos, and R. A. Nichols (2006). "Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution". In: *Genetics* 174, pp. 863–874. (Cit. on p. 124).

Kerner, A. (1879). *Die Schutzmittel der Blüthen gegen unberufene Gäste*. Innsbruck Austria: Verlag der Wagner'schen Universitäts-Buchhandlung. (Cit. on p. 154).

Kessler, D. and I. T. Baldwin (2006). "Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of

Nicotiana attenuata". In: *The Plant Journal* 49, 840–854. (Cit. on p. 154).
Kevan, P. G., D. Eisikowitch, S. Fowle, and K. Thomas (1988). "Yeast-contaminated nectar and its effects on bee foraging." In: *Journal of Apicultural Research* 27.1, pp. 26–29. (Cit. on p. 210).
Khaladkar, M., V. Bellofatto, J. T. L. Wang, B. Tian, and B. A. Shapiro (2007). "RADAR: a web server for RNA data analysis and research". In: *Nucleic Acids Research* 35.Web Server issue, W300. (Cit. on p. 238).
Knudsen, J. T., R. Eriksson, J. Gershenzon, and B. Stahl (2006). "Diversity and distribution of floral scent". In: *The Botanical Review* 72.1, pp. 1–120. (Cit. on p. 210).
Koetschan, C., F. Förster, A. Keller, T. Schleicher, B. Ruderisch, R. Schwarz, T. Müller, M. Wolf, and J. Schultz (2010). "The ITS2 Database III–sequences and structures for phylogeny". In: *Nucleic Acids Research* 38.Database issue, p. D275. (Cit. on pp. 12, 15, 17, 38, 86, 230, 237, 239).
Kolev, Z. and W. de Prins (1995). "A new species of the "brown *Agrodiaetus*" complex from the Crimea". In: *Phegea* 23.2, pp. 119–132. (Cit. on p. 124).
Komárek, J. and B. Fott (1983). *Chlorophyceae (Grünalgen): Ordnung, Chlorococcales*. Stuttgart Germany: E. Schweizerbart. (Cit. on p. 74).
Koonin, E. V., A. R. Mushegian, and P. Bork (1996). "Non-orthologous gene replacement". In: *Trends in Genetics* 12, pp. 334–336. (Cit. on p. 28).
Kress, W. J. and D. L. Erickson (2007). "A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region". In: *PLoS One* 2.6. (Cit. on p. 86).
Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen (2005). "Use of DNA barcodes to identify flowering plants". In: *Proceedings of the National Academy of Sciences of the USA* 102.23, p. 8369. (Cit. on p. 86).
Krienitz, L., I. Ustinova, T. Friedl, and A. R. van Huss (2001). "Traditional generic concepts versus 18S rRNA gene phylogeny in the green algal family Selenastraceae (Chlorophyceae, Chlorophyta)". In: *Journal of Phycology* 37.5, pp. 852–865. (Cit. on p. 86).
Krienitz, L., E. Hegewald, D. Hepperle, and A. Wolf (2003). "The systematics of coccoid green algae: 18S rRNA gene sequence data versus morphology". In: *Biologia* 58.4, pp. 437–446. (Cit. on pp. 60, 74, 86).
Krienitz, L., E. Hegewald, D. Hepperle, V. A. R. Huss, T. Rohr, and M. Wolf (2004). "Phylogenetic relationship of *Chlorella* and *Parachlorella* gen. nov.(Chlorophyta, Trebouxiophyceae)". In: *Phycologia* 43.5, pp. 529–542. (Cit. on pp. 74, 86).
Krimm, U., D. Abanda-Nkpwatt, W. Schwab, and L. Schreiber (2005). "Epiphytic microorganisms on strawberry plants (*Fragaria ananassa* cv. *Elsanta*): identification of bacterial isolates and analysis of their interaction with leaf surfaces". In: *FEMS Microbiology Ecology* 53.3, pp. 483–492. (Cit. on pp. 210, 242).
Kroemer, R. T., R. Kroncke, J. Gerdes, and W. G. Richards (1998). "Comparison of the 3D models of four different human IL-7 isoforms with human and murine IL-7". In: *Protein Engineering Design and Selection* 11.1, p. 31. (Cit. on p. 230).
Krushelnycky, P. D. and R. G. Gillespie (2008). "Compositional and functional stability of arthropod communities in the face of ant invasions". In: *Ecological Applications* 18, 1547–1562. (Cit. on pp. 154, 241).
Krushelnycky, P. D., L. L. Loope, and N. J. Reimer (2005). "The ecology, policy, and management of ants in Hawaii". In: *Proceedings of the Hawaiian Entomological Society* 37, pp. 1–25. (Cit. on pp. 154, 241).
Kudrna, O. (2002). *The distribution atlas of European butterflies*. Vol. 20. Germany: Oedippus, pp. 1–343. (Cit. on p. 124).
Kusel-Fetzmann, E. and H. Nouak (1981). "*Ducellieria chodati* – Alge oder Pilz?" In: *Plant Systematics and Evolution* 138.3, pp. 199–207. (Cit. on p. 74).
Kützing, F (1849). *Species Algarum*. Leipzig Germany: Brockhaus Verlag. (Cit. on p. 60).
Kyrpides, N. C. and C. A. Ouzounis (1999). "Whole-genome sequence annotation:'Going wrong with confidence'". In: *Molecular Microbiology* 32.4, pp. 886–887. (Cit. on p. 28).
Lach, L. (2005). "Interference and exploitation competition of three nectar-thieving invasive ant species". In: *Insectes Sociaux* 52, pp. 257–262. (Cit. on pp. 154, 241).
– (2007). "A mutualism with a native membracid facilitates pollinator displacement by Argentine ants". In: *Ecology* 88.8, pp. 1994–2004. (Cit. on pp. 154, 241).
– (2008a). "Argentine ants displace floral arthropods in a biodiversity hotspot". In: *Diversity and Distributions* 14.2, pp. 281–290. (Cit. on pp. 154, 241).
– (2008b). "Floral visitation patterns of two invasive ant species and their effects on other hy-

menopteran visitors". In: *Ecological Entomology* 33, 155–160. (Cit. on pp. 154, 241).

Lach, L., C. Parr, and K. Abbott (2010). *Ant ecology*. Oxford UK: Oxford University Press. (Cit. on p. 154).

Lafontaine, D. L. J. and D. Tollervey (2001). "The function and synthesis of ribosomes". In: *Nature Reviews Molecular Cell Biology* 2.7, pp. 514–520. (Cit. on pp. 6, 28).

Laloi, D., O. Bailez, M. M. Blight, B. Roger, M. H. Pham-Delegue, and L. J. Wadhams (2000). "Recognition of complex odors by restrained and free-flying honeybees, Apis mellifera". In: *Journal of Chemical Ecology* 26.10, pp. 2307–2319. (Cit. on p. 154).

Lammers, T. G. and C. E. Freeman (1986). "Ornithophily among the Hawaiian Lobelioideae (Campanulaceae) - Evidence from Floral Nectar Sugar Compositions". In: *American Journal of Botany* 73.11, pp. 1613–1619. (Cit. on p. 154).

Lanave, C., G. Preparata, C. Sacone, and G. Serio (1984). "A new method for calculating evolutionary substitution rates". In: *Journal of Molecular Evolution* 20.1, pp. 86–93. (Cit. on pp. 5, 60).

Landis, F. C. and A. Gargas (2007). "Using ITS2 secondary structure to create species-specific oligonucleotide probes for fungi". In: *Mycologia* 99.5, p. 681. (Cit. on pp. 28, 38).

Latex Project Team (2010). *LaTeX – A document preparation system*. URL: http://www.latex-project.org. (Cit. on p. 11).

Leaché, A. D. and M. K. Fujita (2010). "Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*)". In: *Proceedings of the Royal Society B: Biological Sciences*. (Cit. on p. 3).

Ledford, H. (2008). "Botanical identities". In: *Nature* 451.7179, p. 616. (Cit. on p. 86).

Leliaert, F., F. Rousseau, B. de Reviers, and E. Coppejans (2003). "Phylogeny of the Cladophorophyceae (Chlorophyta) inferred from partial LSU rRNA gene sequences: is the recognition of a separate order Siphonocladales justified?" In: *European Journal of Phycology* 38.3, pp. 233–246. (Cit. on pp. 86, 240).

Lemieux, C., C. Otis, and M. Turmel (2000). "Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution". In: *Nature* 403.6770, pp. 649–652. (Cit. on p. 86).

Lemmermann, E. (1910). "Beitrage zur Kenntnis der Planktonalgen". In: *Archiv für Hydrobiologie und Planktonkunde* 5, pp. 291–338. (Cit. on p. 74).

LEO GmbH (2010). *Leo*. URL: http://dict.leo.org. (Cit. on p. 11).

Leontis, N. B., R. B. Altman, H. M. Berman, S. E. Brenner, J. W. Brown, D. R. Engelke, S. C. Harvey, S. R. Holbrook, F. Jossinet, and S. E. Lewis (2006). "The RNA Ontology Consortium: an open invitation to the RNA community". In: *RNA* 12.4, p. 533. (Cit. on p. 230).

Letunic, I. and P. Bork (2007). "Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation". In: *Bioinformatics* 23.1, p. 127. (Cit. on pp. 11, 18, 74, 124, 210).

Lewis, L. A. (2008). "Diversity and phylogenetic placement of *Bracteacoccus tereg* (Chlorophyceae, Chlorophyta) based on 18S ribosomal RNA gene sequence data". In: *Journal of Phycology* 33.2, pp. 279–285. (Cit. on p. 60).

Lewis, L. A. and V. R. Flechtner (2004). "Cryptic species of *Scenedesmus* (Chlorophyta) from desert soil communities of western North America". In: *Journal of Phycology* 40.6, pp. 1127–1137. (Cit. on pp. 74, 86).

Lewis, L. A. and R. M. McCourt (2004). "Green algae and the origin of land plants". In: *American Journal of Botany* 91.10, p. 1535. (Cit. on pp. 60, 239).

Lin, C.-P. and B. N. Danforth (2004). "How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets". In: *Molecular Phylogenetics and Evolution* 30, pp. 686–702. (Cit. on p. 124).

Lin, Y. H., B. C. H. Chang, P. W. Chiang, and S. L. Tang (2008). "Questionable 16S ribosomal RNA gene annotations are frequent in completed microbial genomes". In: *Gene* 416.1-2, pp. 44–47. (Cit. on p. 28).

Lindow, S. E. and M. T. Brandl (2003). "Microbiology of the phyllosphere". In: *Applied and Environmental Microbiology* 69.4, p. 1875. (Cit. on p. 210).

Liu, J., J. T. L. Wang, J. Hu, and B. Tian (2005). "A method for aligning RNA secondary structures and its application to RNA motif detection". In: *BMC Bioinformatics* 6.1, p. 89. (Cit. on p. 238).

Liu, J. S. and C. L. Schardl (1994). "A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes". In: *Plant Molecular Biology* 26.2, pp. 775–778. (Cit. on pp. 6, 7, 60).

Lokvam, J. and J. F. Braddock (1999). "Antibacterial function in the sexually dimorphic pollinator rewards of *Clusia grandiflora* (Clusiaceae)". In: *Oecologia* 119.4, pp. 534–540. (Cit. on p. 210).

Lonsdale, W. M. (1999). "Global patterns of plant invasions and the concept of invasibility". In: *Ecology* 80.5, pp. 1522–1536. (Cit. on p. 154).

Lopezaraiza-Mikel, M. E., R. B. Hayes, M. R. Whalley, and J. Memmott (2007). "The impact of an alien plant on a native plant-pollinator network: an experimental approach". In: *Ecology Letters* 10.7, pp. 539–550. (Cit. on p. 154).

Lorkovic, Z. (1990). "The butterfly chromosomes and their application in systematics and phylogeny". In: *Butterflies of Europe. Introduction to Lepidopterology*. Vol. 2. Wiesbaden Germany: Aula, pp. 332–396. (Cit. on p. 124).

Loughnane, C. J., L. M. McIvor, F. Rindi, D. B. Stengel, and M. D. Guiry (2008). "Morphology, rbcL phylogeny and distribution of distromatic *Ulva* (Ulvophyceae, Chlorophyta) in Ireland and southern Britain". In: *Phycologia* 47.4, pp. 416–429. (Cit. on p. 86).

Lukavsky, J. (2006). "*Coelastrum pascheri* sp. n., a new green alga from lakes of the Bohemian Forest". In: *Biologia* 61, pp. 485–490. (Cit. on p. 74).

Lukavsky, P. J. (2007). "Basic Principles of RNA NMR Spectroscopy". In: *Structure and Biophysics – New Technologies for Current Challenges in Biology and Beyond*, pp. 65–80. (Cit. on p. 242).

Lukavsky, P. J. and J. D. Puglisi (2004). "Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides." In: *RNA* 10.5, p. 889. (Cit. on p. 242).

Lukhtanov, V. A. and A. V. Dantchenko (2002). "Principles of the highly ordered arrangement of metaphase I bivalents in spermatocytes of *Agrodiaetus* (Insecta, Lepidoptera)". In: *Chromosome Research* 10.1, pp. 5–20. (Cit. on p. 124).

Lukhtanov, V. A. and A. G. Lukhtanov (1994). *Die Tagfalter Nordwestasiens*. Vol. 3. Marktleuthen, Germany: Herbipoliana, pp. 1–440. (Cit. on p. 124).

Lukhtanov, V. A. and R. Vila (2006). "Rearrangement of the *Agrodiaetus dolus* species group (Lepidoptera, Lycaenidae) using a new cytological approach and molecular data". In: *Insect Systematics and Evolution* 37, pp. 325–334. (Cit. on p. 124).

Lukhtanov, V. A., N. P. Kandul, J. B. Plotkin, A. V. Dantchenko, D. Haig, and N. E. Pierce (2005). "Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies". In: *Nature* 436.7049, pp. 385–389. (Cit. on p. 124).

Lukhtanov, V. A., A. Sourakov, E. V. Zakharov, and P. D. Hebert (2009). "DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce success of species identification". In: *Molecular Ecology Resources* 9.5, pp. 1302–1310. (Cit. on p. 124).

Lumbsch, H. T. (2002). "How objective are genera in euascomycetes?" In: *Perspectives in Plant Ecology, Evolution and Systematics* 5.2, pp. 91–101. (Cit. on p. 240).

Luo, K., S. Chen, K. Chen, J. Y. Song, H. Yao, X. Ma, Y. J. Zhu, X. H. Pang, H. Yu, X. W. Li, and Z. Liu (2010). "Assessment of candidate plant DNA barcodes using the Rutaceae family". In: *Science China Life Sciences* 53 (6), pp. 701–708. (Cit. on p. 240).

Maccagnani, B., F. Giacomello, M. Fanti, D. Gobbin, S. Maini, and G. Angeli (2009). "*Apis mellifera* and *Osmia cornuta* as carriers for the secondary spread of *Bacillus subtilis* on apple flowers". In: *Biocontrol* 54.1, pp. 123–133. (Cit. on p. 210).

Macromates (2010). *Textmate*. URL: `http://macromates.com`. (Cit. on p. 11).

Magnacca, K. N. and B. N. Danforth (2007). "Low nuclear DNA variation supports a recent origin of Hawaiian *Hylaeus* bees (Hymenoptera: Colletidae)". In: *Molecular Phylogenetics and Evolution* 43.3, pp. 908–915. (Cit. on p. 154).

Mahner, M. and M. Bunge (1997). *Foundations of biophilosophy*. Heidelberg Germany: Springer. (Cit. on p. 3).

Mai, J. C. and A. W. Coleman (1997). "The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants". In: *Journal of Molecular Evolution* 44.3, pp. 258–271. (Cit. on pp. 6, 7, 38, 60, 86, 237).

Mailund, T. and C. N. S. Pedersen (2004). "QDist–quartet distance between evolutionary trees". In: *Bioinformatics*, p. 971. (Cit. on pp. 24, 46).

Markham, N. R. and M. Zuker (2008). "UNAFold: software for nucleic acid folding and hybridization." In: *Methods in Molecular Biology* 453, p. 3. (Cit. on p. 38).

Martinez, H. M., J. V. Maizel Jr, and B. A. Shapiro (2008). "RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA." In: *Journal of Biomolecular Structure and Dynamics* 25.6, p. 669. (Cit. on pp. 11, 16, 230, 242).

Mathews, D. H., M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner (2004). "Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure". In: *Proceedings of the National Academy*

of Sciences of the USA 101.19, p. 7287. (Cit. on pp. 11, 15, 28, 60, 74, 124, 154, 237).

Mattox, K. R. and K. D. Stewart (1984). "Classification of the green algae: a concept based on comparative cytology". In: *Systematics of the green algae*, pp. 29–72. (Cit. on p. 60).

May, R. M. (1988). "How many species are there on earth?" In: *Science* 241.4872, pp. 1441–1441. (Cit. on p. 3).

Mayr, E. (1970). *Populations, species, and evolution: an abridgment of animal species and evolution*. MA USA: Belknap Press Cambridge. (Cit. on p. 3).

McManus, H. A. and L. A. Lewis (2005). "Molecular phylogenetics, morphological variation and colony-form evolution in the family Hydrodictyaceae (Sphaeropleales, Chlorophyta)". In: *Phycologia* 44.6, pp. 582–595. (Cit. on p. 86).

Medeiros, A. C., L. L. Loope, and F. R. Cole (1986). "Distribution of ants and their effects on endemic biota of Haleakala and Hawaii Volcanoes National Park: a preliminary assessment". In: *Proceedings 6th Conference in natural sciences*. HI USA: Hawaii Volcanoes National Park, 39–52. (Cit. on pp. 154, 241).

Mei, H., G.-X. Liu, and Z.-Y. Hu (2007). "Phylogenetic studies of Oedogoniales (Chlorophyceae, Chlorophyta) based on 28S rDNA sequences". In: *Acta Hydrobiologica Sinica*, p. 4. (Cit. on pp. 86, 240).

Melkonian, M. (1982). "Structural and evolutionary aspects of the flagellar apparatus in green algae and land plants". In: *Taxon* 31.2, pp. 255–265. (Cit. on p. 60).

Melkonian, M. and B. Surek (1995). "Phylogeny of the Chlorophyta: congruence between ultrastructural and molecular evidence". In: *Bulletin de la Societe zoologique de France* 120.2, pp. 191–208. (Cit. on p. 60).

Memmott, J. and N. M. Waser (2002). "Integration of alien plants into a native flower-pollinator visitation web". In: *Proceedings of the Royal Society of London Series B-Biological Sciences* 269.1508, pp. 2395–2399. (Cit. on p. 154).

Mercier, J. and S. E. Lindow (2000). "Role of leaf surface sugars in colonization of plants by bacterial epiphytes". In: *Applied and Environmental Microbiology* 66.1, p. 369. (Cit. on p. 210).

Meusel, H. and E. J. Jager (1992). *Vergleichende Chorologie der zentraleuropaischen Flora*. Jena Germany: Gustav Fischer Verlag. (Cit. on p. 124).

Meyen, F. J. (1829). "Beobachtungen über einige niedere Algenformen". In: *Nova acta physico-medica Academiae Caesareae Leopoldino-Carolinae Naturae Curiosorum* 14, pp. 769–778. (Cit. on p. 74).

Meyer, C. P. and G. Paulay (2005). "DNA barcoding: error rates based on comprehensive sampling". In: *PLoS Biology* 3.12, e422. (Cit. on p. 124).

Meyer, S. and A. von Haeseler (2003). "Identifying site specific substitution rates". In: *Molecular Biology and Evolution*, p. 191. (Cit. on p. 46).

Mitchell, P., E. Petfalski, and D. Tollervey (1996). "The 3'end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism." In: *Genes and Development* 10.4, p. 502. (Cit. on pp. 6, 28).

Moniz, M. B. J. and I. Kaczmarska (2009a). "Barcoding diatoms: Is there a good marker?" In: *Molecular Ecology Resources* 9, pp. 65–74. (Cit. on p. 241).

– (2009b). "Barcoding of diatoms: nuclear encoded ITS revisited". In: *Protist*. (Cit. on pp. 38, 124, 241).

Mooney, H. A. (2005). "The nature of the problem". In: *Invasive alien species*. Ed. by H. A. Mooney, R. N. Mack, J. A. McNeely, L. E. Neville, P. J. Schei, and J. K. Waage. London UK: Island Press. (Cit. on p. 154).

Mooney, H. A., R. N. Mack, J. A. McNeely, L. E. Neville, P. J. Schei, and J. K. Waage (2005). *Invasive alien species*. Scope. London UK: Island press. (Cit. on p. 154).

Morales, C. L. and M. A. Aizen (2006). "Invasive mutualisms and the structure of plant-pollinator interactions in the temperate forests of north-west Patagonia, Argentina". In: *Journal of Ecology* 94.1, pp. 171–180. (Cit. on p. 154).

Moritz, C. and C. Cicero (2004). "DNA barcoding: promise and pitfalls". In: *PLoS Biology* 2.10, e354. (Cit. on p. 124).

MorphBank (2009). *MorphBank*. URL: `http://www.morphbank.net`. (Cit. on p. 124).

Mortensen, A. (2010). *Fugu*. URL: `http://rsug.itd.umich.edu/software/fugu`. (Cit. on p. 11).

Mullineux, T. and G. Hausner (2009). "Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of rhe fungal genera *Grosmannia* and *Leptographium* ". In: *Fungal Genetics and Biology* 46.11, pp. 855–867. (Cit. on p. 240).

Müller, T. and M. Vingron (2000). "Modeling amino acid replacement". In: *Journal of Computational Biology* 7.6, pp. 761–776. (Cit. on p. 46).

Müller, T., S. Rahmann, T. Dandekar, and M. Wolf (2004). "Accurate and robust phylogeny estimation based on profile distances: a study of the Chlorophyceae (Chlorophyta)". In: *BMC Evolutionary Biology* 4.1, p. 20. (Cit. on pp. 60, 86, 124, 239).

Müller, T., N. Philippi, T. Dandekar, J. Schultz, and M. Wolf (2007). "Distinguishing species". In: *RNA* 13.9, p. 1469. (Cit. on pp. 7, 18, 38, 74, 86, 124, 241).

Nakada, T. and H. Nozaki (2007). "Re-evaluation of three *Chlorogonium* (Volvocales, Chlorophyceae) species based on 18S ribosomal RNA gene phylogeny". In: *European Journal of Phycology* 42.2, pp. 177–182. (Cit. on p. 86).

Nakada, T., H. Nozaki, and T. Pröschold (2008a). "Molecular phylogeny, ultrastructure, and taxonomic revision of *Chlorogonium* (Chlorophyta): emendation of *Chlorogonium* and description of *Gungnir* gen. nov and *Rusalka* gen. nov." In: *Journal of Phycology* 44.3, pp. 751–760. (Cit. on p. 86).

Nakada, T., K. Misawa, and H. Nozaki (2008b). "Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses". In: *Molecular Phylogenetics and Evolution* 48.1, pp. 281–291. (Cit. on p. 86).

Nakayama, T., S. Watanabe, and I. Inouye (1996a). "Phylogeny of wall-less green flagellates inferred from 18SrDNA sequence data". In: *Phycological Research* 44.3, pp. 151–161. (Cit. on p. 86).

Nakayama, T., S. Watanabe, K. Mitsui, H. Uchida, and I. Inouye (1996b). "The phylogenetic relationship between the Chlamydomonadales and Chlorococcales inferred from 18SrDNA sequence data". In: *Phycological Research* 44.1, pp. 47–55. (Cit. on p. 86).

Nakazawa, A., T. K. Yamada, and H. Nozaki (2004). "Taxonomic study of *Asterococcus* (Chlorophyceae) based on comparative morphology and rbcL gene sequences". In: *Phycologia* 43.6, pp. 711–721. (Cit. on p. 86).

National Center for Biotechnology Information (2010). *NCBI*. URL: http://www.ncbi.nlm.nih.gov. (Cit. on p. 124).

Nazari, V. (2003). *Butterflies of Iran*. Stenstrup Denmark: Apollo Books. (Cit. on p. 124).

NCBI (2010). *PubMed*. URL: http://www.ncbi.nlm.nih.gov/pubmed. (Cit. on p. 11).

Newmaster, S. G., A. J. Fazekas, and S. Ragupathy (2006). "DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach". In: *Botany* 84.3, pp. 335–341. (Cit. on p. 86).

Nichols, R. (2001). "Gene trees and species trees are not the same". In: *Trends in Ecology and Evolution* 16.7, pp. 358–364. (Cit. on p. 124).

Nieto-Feliner, G. and J. A. Rossello (2007). "Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants". In: *Molecular Phylogenetics and Evolution* 44.2, pp. 911–919. (Cit. on p. 124).

Nilsson, R. H., M. Ryberg, E. Kristiansson, K. Abarenkov, K. H. Larsson, and U. Koljalg (2006). "Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective". In: *PLoS One* 1.1. (Cit. on p. 28).

Nilsson, R. H., E. Kristiansson, M. Ryberg, N. Hallenberg, and K. H. Larsson (2008). "Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification". In: *Evolutionary Bioinformatics* 4, p. 193. (Cit. on p. 28).

Nozaki, H. (2001). "Chloroplast multigene phylogeny and systematics of the advanced genera of the Volvocaceae (Chlorophyceae)". In: *Phycologia* 40, p. 37. (Cit. on pp. 86, 240).

– (2003). "Origin and evolution of the genera *Pleodorina* and *Volvox* (Volvocales)". In: *Biologia* 58.4, pp. 425–432. (Cit. on p. 86).

Nozaki, H., M. Itoh, R. Sano, H. Uchida, M. M. Watanabe, and T. Kuroiwa (1995). "Phylogenetic relationships within the colonial Volvocales (Chlorophyta) inferred from rbcL gene sequence data". In: *Journal of Phycology* 31.6, pp. 970–979. (Cit. on p. 86).

Nozaki, H., M. Itoh, R. Sano, H. Uchida, M. M. Watanabe, H. Takahashi, and T. Kuroiwa (1997a). "Phylogenetic analysis of *Yamagishiella* and *Platydorina* (Volvocaceae, Chlorophyta) based on rbcL gene sequences". In: *Journal of Phycology* 33.2, pp. 272–278. (Cit. on p. 86).

Nozaki, H., M. Itoh, M. M. Watanabe, H. Takano, and T. Kuroiwa (1997b). "Phylogenetic analysis of morphological species of *Carteria* (Volvocales, Chlorophyta) based on rbcL gene sequences". In: *Journal of Phycology* 33.5, pp. 864–867. (Cit. on p. 86).

Nozaki, H., N. Ohta, E. Morita, and M. M. Watanabe (1998). "Toward a natural system of species in *Chlorogonium* (Volvocales, chlorophyta): a combined analysis of morphological and rbcL gene sequence data". In: *Journal of Phycology* 34, pp. 1024–1037. (Cit. on p. 86).

Nozaki, H., N. Ohta, H. Takano, and M. M. Watanabe (1999). "Reexamination of phylogenetic relationships within the colonial Volvocales (Chlorophyta): an analysis of atpB and rbcL gene sequences". In: *Journal of Phycology* 35, pp. 104–112. (Cit. on p. 86).

Nozaki, H., K. Misawa, T. Kajita, M. Kato, S. Nohara, and M. M. Watanabe (2000). "Origin and evolution of the colonial Volvocales (Chlorophyceae) as inferred from multiple, chloro-

plast gene sequences". In: *Molecular Phylogenetics and Evolution* 17.2, pp. 256–268. (Cit. on p. 86).

Nozaki, H., O. Misumi, and T. Kuroiwa (2003). "Phylogeny of the quadriflagellate Volvocales (Chlorophyceae) based on chloroplast multigene sequences". In: *Molecular Phylogenetics and Evolution* 29.1, pp. 58–66. (Cit. on pp. 86, 240).

Nozaki, H., F. D. Ott, and A. W. Coleman (2006). "Morphology, molecular phylogeny and taxonomy of two new species of *Pleodorina* (Volvoceae, Chlorophyceae)". In: *Journal of Phycology* 42.5, pp. 1072–1080. (Cit. on p. 86).

Oh, H., H. Yoon, M. Kim, H. Jeong, S. Kim, J. Hwang, C. Bae, and I. Kim (2009). "ITS2 ribosomal DNA sequence variation of the bumblebee *Bombus ardens* (Hymenoptera: Apidae)". In: *Genes & Genomics* 31 (4), pp. 293–303. (Cit. on p. 241).

Oltmanns, F. (1829). *Morphologie und Biologie der Algen, Spezieller Teil*. Jena Germany: Fischer. (Cit. on p. 74).

Östman, O., S. Drakare, E. S. Kritzberg, S. Langenheder, J. B. Logue, and E. S. Lindström (2010). "Regional invariance among microbial communities". In: *Ecology Letters* 13.1, pp. 118–127. (Cit. on p. 210).

Parisien, M., J. A. Cruz, E. Westhof, and F. Major (2009). "New metrics for comparing and assessing discrepancies between RNA 3D structures and models". In: *RNA* 15.10, p. 1875. (Cit. on pp. 230, 243).

Park, M., C. Sim, J. Baek, and G. Min (2007). "Identification of genes suitable for DNA barcoding of morphologically indistinguishable korean Halichondriidae sponges". In: *Molecules and Cells* 23.2, p. 220. (Cit. on p. 28).

Pascher, A. (1931). "Systematische Übersicht über die mit Flagellaten in Zusammenhang stehenden Algenreihen und Versuch einer Einreihung dieser Algenstämme in die Stämme des Pflanzenreiches". In: *Beihefte zum Botanischen Centralblatt* 48, pp. 317–332. (Cit. on p. 60).

– (1939). "Über geisselbewegliche Eier, mehrköpfige Schwärmer und vollständigen Schwärmverlust bei *Sphaeroplea*". In: *Beihefte zum Botanischen Centralblatt* 59, pp. 188–213. (Cit. on p. 60).

Paschma, R. and E. Hegewald (1986). "DNA base composition within the genus *Scenedesmus* (Chlorophyta)". In: *Plant Systematics and Evolution* 153.3, pp. 171–180. (Cit. on p. 74).

Pearson, W. R. and D. J. Lipman (1988). "Improved tools for biological sequence comparison". In: *Proceedings of the National Academy of Sciences of the USA* 85.8, p. 2444. (Cit. on p. 28).

Peculis, B. A. and C. L. Greer (2002). "The structure of the ITS2-proximal stem is required for pre-rRNA processing in yeast". In: *RNA* 4.12, pp. 1610–1622. (Cit. on pp. 6, 28).

Perl Foundation (2010). *Perl*. URL: http://www.perlfoundation.org. (Cit. on pp. 11, 13).

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin (2004). "UCSF Chimera – a visualization system for exploratory research and analysis". In: *Journal of Computational Chemistry* 25.13, pp. 1605–1612. (Cit. on pp. 11, 230).

Pettersson, J. (1970). "An aphid sex attractant. I. Biological studies". In: *Entomologica Scandinavica* 1, pp. 63–73. (Cit. on p. 154).

Pillmann, A., G. W. Woolcott, J. L. Olsen, W. T. Stam, and R. J. King (1997). "Inter- and intraspecific genetic variation in *Caulerpa* (Chlorophyta) based on nuclear rDNA ITS sequences". In: *European Journal of Phycology* 32.4, pp. 379–386. (Cit. on p. 86).

Pley, H. W., K. M. Flaherty, and D. B. McKay (1994). "Three-dimensional structure of a hammerhead ribozyme". In: *Nature* 372.6501, pp. 68–74. (Cit. on p. 230).

Pocock, T., M. A. Lachance, T. Pröschold, J. C. Priscu, S. S. Kim, and N. P. A. Huner (2004). "Identification of a psychrophilic green alga from Lake Bonney Antarctica: *Chlamydomonas raudensis* Ettl. (UWO 241) Chlorophyceae". In: *Journal of Phycology* 40.6, pp. 1138–1148. (Cit. on p. 86).

Posada, D. and K. A. Crandall (1998). "Modeltest: testing the model of DNA substitution". In: *Bioinformatics* 14.9, p. 817. (Cit. on pp. 17, 60, 124).

– (2002). "The effect of recombination on the accuracy of phylogeny estimation". In: *Journal of Molecular Evolution* 54.3, pp. 396–402. (Cit. on p. 46).

Poulton, E. B. (1903). "What is a species?" In: *Transactions of the Entomological Society of London*, pp. 77–116. (Cit. on p. 3).

Prasad, A. M., L. R. Iverson, and A. Liaw (2006). "Newer classification and regression tree techniques: bagging and random forests for ecological prediction". In: *Ecosystems* 9.2, pp. 181–199. (Cit. on p. 210).

Printz, H. (1927). "Chlorophyceae (nebst Conjugatae, Heterocontae und Charophyta)". In: *Die natürlichen Pflanzenfamilien nebst ihren Gattungen und wichtiger Arten insbesondere der Nutzpflanzen* 3, p. 463. (Cit. on p. 74).

Protist Information Server (2010). *Protist Information Server*. URL: http://protist.i.hosei.ac.jp. (Cit. on p. 74).
Pröschold, T. and F. Leliaert (2007). "Systematics of the green algae: conflict of classic and modern approaches". In: *Unravelling the algae: the past, present, and future of algal systematics*, p. 123. (Cit. on pp. 60, 239).
Pröschold, T., B. Marin, U. Schlösser, and M. Melkonian (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." In: *Protist* 152.4, p. 265. (Cit. on pp. 60, 86).
Pröschold, T., C. Bock, W. Luo, and L. Krienitz (2010). "Polyphyletic distribution of bristle formation in Chlorellaceae: *Micractinium*, *Diacanthos*, *Didymogenes* and *Hegewaldia* gen. nov.(Trebouxiophyceae, Chlorophyta)". In: *Phycological Research* 58.1, pp. 1–8. (Cit. on p. 74).
Pun-Cocharova, M. and T. Kalina (1994). "Taxonomy of the genus *Scotiellopsis* Vinatzer (Chlorococcales, Chlorophyta)". In: *Archiv für Hydrobiologie: Organ der Internationalen Vereinigung für Theoretische und Angewandte Limnologie* 60, p. 119. (Cit. on p. 74).
R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org. (Cit. on pp. 11, 24, 46).
Raguso, R. A. (2004). "Why are some floral nectars scented?" In: *Ecology* 85.6, pp. 1486–1494. (Cit. on p. 154).
– (2008a). "Start making scents: the challenge of integrating chemistry into pollination ecology". In: *Entomologia Experimentalis Et Applicata* 128.1, pp. 196–207. (Cit. on p. 154).
– (2008b). "Wake up and smell the roses: The ecology and evolution of floral scent". In: *Annual Review of Ecology, Evolution and Systematics* 39, 549–69. (Cit. on pp. 154, 210).
Raguso, R. A. and E. Pichersky (1995). "Floral volatiles from Clarkia breweri and C. concinna (Onagraceae): recent evolution of floral scent and moth pollination". In: *Plant Systematics and Evolution* 194, pp. 55–67. (Cit. on p. 154).
Raguso, R. A. and B. A. Roy (1998). "Floral scent production by *Puccinia rust* fungi that mimic flowers". In: *Molecular Ecology* 7.9, pp. 1127–1136. (Cit. on p. 210).
Rahmann, S., T. Müller, T. Dandekar, and M. Wolf (2006). "Efficient and robust analysis of large phylogenetic datasets". In: *Advanced Data Mining Techonolgies in Bioinformatics*. Hershey USA: Idea Group Publising, pp. 104–117. (Cit. on p. 86).
Rambaut, A. (2007). *FigTree*. URL: http://tree.bio.ed.ac.uk/software/figtree. (Cit. on pp. 11, 18, 86, 124).
Ranganathan, Y. and R. M. Borges. "Reducing the babel in plant volatile communication: using the forest to see the trees". In: *Plant Biology*. (Cit. on p. 210).
Rayss, T. (1915). "Le *Coelastrum proboscideum*Bohl. Etude de planctologie experimentale suivide d'une revision des *Coelastrum* de la Suisse." PhD thesis. Universite Geneve, Institute de Botanique. (Cit. on p. 74).
Reinhard, J., M. Sinclair, M. V. Srinivasan, and C. Claudianos (2010). "Honeybees Learn Odour Mixtures via a Selection of Key Odorants". In: *Plos One* 5.2, pp. –. (Cit. on p. 154).
Reinsch, P. F. (1875). *Contributiones ad algologiam et fungologiam*. Leipzig Germany: T. O. Weigel. (Cit. on p. 74).
– (1877). "On freshwater algae from the Cape of Good Hope". In: *Journal of the Linnean Society London, Botany* 16, pp. 232–248. (Cit. on p. 74).
Reymond, O. (1975). "La paroi cellulaire de *Coelastrum* (Chlorophycees)". In: *Archiv für Microbiologie* 102, pp. 95–101. (Cit. on p. 74).
Ricciardi, A. (2005). "Facilitation and synergistic interactions between introduced aquatic species". In: *Invasive Alien Species*. Ed. by H. A. Mooney, R. N. Mack, J. A. McNeely, L. E. Neville, P. J. Schei, and J. K. Waage. London UK: Island press, pp. 162–178. (Cit. on p. 154).
Rice, P., I. Longden, and A. Bleasby (2000). "EMBOSS: the European molecular biology open software suite". In: *Trends in Genetics* 16.6, pp. 276–277. (Cit. on p. 11).
Richardson, D. M., N. Allsopp, C. M. d'Antonio, S. J. Milton, and M. Rejmanek (2000). "Plant invasions - the role of mutualisms". In: *Biological Reviews* 75.1, pp. 65–93. (Cit. on p. 154).
Rico-Gray, V. and P. S. Oliveira (2007). *The ecology and evolution of ant-plant interactions*. IL USA: The University of Chicago Press. (Cit. on p. 154).
Ridley, M. (1989). "The cladistic solution to the species problem". In: *Biology and Philosophy* 4.1, pp. 1–16. (Cit. on p. 3).
Rieth, A. (1952). "Über die vegetative Vermehrung bei *Sphaeropha wilmani* Fritsch et Rich." In: *Flora*, p. 28. (Cit. on p. 60).
– (1953). "Zur Kenntnis der Gattung *Sphaeroplea*, *Sphaeroplea caubrica* Fritsch". In: *Flora* 140, pp. 130–139. (Cit. on p. 60).

Riffell, J. A., H. Lei, T. A. Christensen, and J. G. Hildebrand (2009). "Characterization and Coding of Behaviorally Significant Odor Mixtures". In: *Current Biology* 19.4, pp. 335–340. (Cit. on p. 154).

Rino, J. A. (1972). "Contribucion para o conhecimento das algas de agua doce de Mocambique". In: *Revista de Sciencis Biologica, ser. A* 5, pp. 121–264. (Cit. on p. 74).

Rissler, L. J. and J. J. Apodaca (2007). "Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*)". In: *Systematic Biology* 56.6, p. 924. (Cit. on p. 3).

Rodriguez, F., J. L. Oliver, A. Marin, and J. R. Medina (1990). "The general stochastic model of nucleotide substitution". In: *Journal of Theoretical Biology* 142.4, pp. 485–501. (Cit. on pp. 5, 60).

Rokas, A. and S. B. Carroll (2005). "More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy". In: *Molecular Biology and Evolution* 22.5, p. 1337. (Cit. on p. 46).

Ronquist, F. (1997). "Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography". In: *Systematic Biology* 46.1, p. 195. (Cit. on pp. 11, 124).

Ruhl, M. W., M. Wolf, and T. M. Jenkins (2010). "Compensatory base changes illuminate morphologically difficult taxonomy". In: *Molecular Phylogenetics and Evolution* 54.2, pp. 664 –669. (Cit. on p. 241).

Saar, D. E., N. O. Polans, and P. D. Sørensen (2003). "A phylogenetic analysis of the genus *Dahlia* (Asteraceae) based on internal and external transcribed spacer regions of nuclear ribosomal DNA". In: *Systematic Botany*, pp. 627–639. (Cit. on p. 38).

Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees". In: *Molecular Biology and Evolution* 4.4, p. 406. (Cit. on pp. 5, 17, 60).

Sanders, E. R., K. G. Karol, and R. M. McCourt (2003). "Occurrence of matK in a trnK group II intron in charophyte green algae and phylogeny of the Characeae". In: *American Journal of Botany* 90.4, p. 628. (Cit. on p. 86).

Savolainen, V., R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane (2005). "Towards writing the encyclopaedia of life: an introduction to DNA barcoding". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 360.1462, p. 1805. (Cit. on p. 86).

Sayers, E. W., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, et al. (2009). "Database resources of the National Center for Biotechnology Information". In: *Nucleic Acids Research* 37, pp. D5–D15. (Cit. on p. 154).

Schagerl, M., D. G. Angeler, and A. W. Coleman (1999). "Infraspecific phylogeny of *Pandorina morum* (Volvocales, Chlorophyta) inferred from molecular, biochemical and traditional data". In: *European Journal of Phycology* 34.1, pp. 87–93. (Cit. on p. 86).

Schill, R., F. Förster, T. Dandekar, and M. Wolf (2010). "Using compensatory base change analysis of internal transcribed spacer 2 secondary structures to identify three new species in *Paramacrobiotus* (Tardigrada)". In: *Organisms Diversity & Evolution*, pp. 1–10. (Cit. on p. 241).

Schlötterer, C., M. T. Hauser, A. von Haeseler, and D. Tautz (1994). "Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*". In: *Molecular Biology and Evolution* 11.3, p. 513. (Cit. on pp. 6, 7).

Schnepf, E. and E. Hegewald (1993). "*Didymogenes palatina* Schmidle and *Didymogenes anomala* (GM Smith) Hind. (Chlorococcales): taxonomy, ultrastructure, autosporogenesis and autospore wall assembly". In: *Archiv für Protistenkunde* 143.1-3, pp. 41–53. (Cit. on p. 74).

Schoch, C. L., G. H. Sung, F. Lopez-Giraldez, J. P. Townsend, J. Miadlikowska, V. Hofstetter, B. Robbertse, P. B. Matheny, F. Kauff, Z. Wang, et al. (2009). "The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits". In: *Systematic Biology* 58.2, p. 224. (Cit. on p. 6).

Schöniger, M. and A. von Haeseler (1994). "A stochastic model for the evolution of autocorrelated DNA sequences." In: *Molecular Phylogenetics and Evolution* 3.3, p. 240. (Cit. on pp. 6, 46, 238).

Schultz, J. and M. Wolf (2009). "ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics". In: *Molecular Phylogenetics and Evolution* 52.2, pp. 520–523. (Cit. on pp. 8, 38, 46, 74, 86, 124, 154, 230, 237, 238).

Schultz, J., S. Maisel, D. Gerlach, T. Müller, and M. Wolf (2005). "A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota." In: *RNA* 11.4, p. 361. (Cit. on pp. 6, 28, 38, 60, 74, 86, 124, 239).

Schultz, J., T. Müller, M. Achtziger, P. N. Seibel, T. Dandekar, and M. Wolf (2006). "The internal transcribed spacer 2 database–a web server

for (not only) low level phylogenetic analyses". In: *Nucleic Acids Research* 34.Web Server issue, W704. (Cit. on pp. 6, 28, 46, 60, 74, 86, 124, 237).
Schulz, S. and J. S. Dickschat (2007). "Bacterial volatiles: the smell of small organisms". In: *Natural Product Reports* 24.4, pp. 814–842. (Cit. on p. 210).
Schurian, K. G. (1989). "Bemerkungen zu *Lysandra cormion* Nabokov, 1941 (Lepidoptera: Lycaenidae)". In: *Nachrichten Entomologischer Verein Apollo* 10.2, pp. 183–192. (Cit. on p. 124).
– (1991). "Nachtrag zu den "Bemerkungen zu *Lysandra cormion*" (Lepidoptera: Lycaenidae)". In: *Nachrichten Entomologischer Verein Apollo* 12.3, pp. 193–195. (Cit. on p. 124).
– (1997). "Freilandexemplare des *Hybriden cormion*(= *Polyommatus* (*Meleageria*) *coridon* x *P.* (*M.*) *daphnis*) (Lepidoptera: Lycaenidae)". In: *Nachrichten Entomologischer Verein Apollo* 18.2/3, pp. 227–230. (Cit. on p. 124).
Seberg, O. and G. Petersen (2009). "How many loci does it take to DNA barcode a Crocus?" In: *PLOS one* 4.2. (Cit. on p. 86).
Seehausen, O. (2004). "Hybridization and adaptive radiation". In: *Trends in Ecology and Evolution* 19.4, pp. 198–207. (Cit. on p. 124).
Seibel, P. N., T. Müller, T. Dandekar, J. Schultz, and M. Wolf (2006). "4 SALE – A tool for synchronous RNA sequence and secondary structure alignment and editing". In: *BMC Bioinformatics* 7.1, p. 498. (Cit. on pp. 6, 11, 17, 23, 28, 46, 60, 74, 86, 124, 237).
Seibel, P. N., T. Müller, T. Dandekar, and M. Wolf (2008). "Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4 SALE". In: *BMC Research Notes* 1.1, p. 91. (Cit. on pp. 11, 17, 23, 28, 38, 46, 74, 86, 124, 154, 210, 237).
Seifert, K. A. (2009). "Progress towards DNA barcoding of fungi". In: *Molecular Ecology Resources* 9.1, pp. 83–89. (Cit. on pp. 124, 240).
Selig, C., M. Wolf, T. Müller, T. Dandekar, and J. Schultz (2008). "The ITS2 Database II: homology modelling RNA structure for molecular systematics." In: *Nucleic Acids Research* 36.Database issue, p. D377. (Cit. on pp. 6, 28, 46, 60, 74, 86, 124, 237).
Senn, G. (1899). "Über einige coloniebildende einzellige Algen". PhD thesis. Universität Basel Switzerland. (Cit. on p. 74).
Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks". In: *Genome Research* 13.11, p. 2498. (Cit. on pp. 11, 285).
Shapiro, B. A., Y. G. Yingling, W. Kasprzak, and E. Bindewald (2007). "Bridging the gap in RNA structure prediction". In: *Current Opinion in Structural Biology* 17.2, pp. 157–165. (Cit. on pp. 230, 242).
Shoup, S. and L. A. Lewis (2003). "Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta)". In: *Journal of Phycology* 39.4, pp. 789–796. (Cit. on p. 60).
Simberloff, D. and B. von Holle (1999). "Positive interactions of nonindigenous species: invasional meltdown?" In: *Biological Invasions* 1, pp. 21–32. (Cit. on p. 154).
Simpson, G. (1951). "The species concept". In: *Evolution* 5.4, pp. 285–298. (Cit. on p. 3).
Slowinski, J. B. and R. Lawson (2002). "Snake phylogeny: evidence from nuclear and mitochondrial genes". In: *Molecular Phylogenetics and Evolution* 24.2, pp. 194–202. (Cit. on p. 46).
Sluiman, H. J., C. Guihal, and O. Mudimu (2008). "Assessing phylogenetic affinities and species delimitations in Klebsormidiales (Streptophyta): nuclear-encoded rDNA phylogenies and its secondary structure models in *Klebsormidium*, *Hormidiella*, and *Entransia*". In: *Journal of Phycology* 44.1, pp. 183–195. (Cit. on p. 86).
Small, R. L., J. A. Ryburn, R. C. Cronn, T. Seelanan, and J. F. Wendel (1998). "The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in a recently diverged plant group". In: *American Journal of Botany* 85.9, p. 1301. (Cit. on p. 6).
Smith, G. M. (1916). *A monograph of the algal genus* Scenedesmus *based upon pure culture studies*. WI USA: Academy of Sciences, Arts and Letters. (Cit. on p. 74).
– (1920). *Phytoplankton of the inland lakes of Wisconsin*. MI USA: State Michigan. (Cit. on p. 74).
Smith, V. S. (2005). "DNA barcoding: perspectives from a "Partnerships for Enhancing Expertise in Taxonomy" (PEET) debate". In: *Systematic Biology* 54.5, pp. 841–844. (Cit. on pp. 86, 124).
Sodomkova, M. (1970). "Taxonomische Übersicht der Gattung *Coelastrum* Nägeli". In: *Acta Universitatis carolinae. Biologica*. (Cit. on p. 74).
Soltis, D. E., P. S. Soltis, and J. J. Doyle (1998). *Molecular systematics of plants* 2. Köln Germany: Kluwer Academic Publishing. (Cit. on pp. 6, 7, 237).
Sorhannus, U., J. D. Ortiz, M. Wolf, and M. G. Fox (2009). "Microevolution and Speciation in

Thalassiosira weissflogii (Bacillariophyta)". In: *Protist* 161.2, pp. 237–249. (Cit. on pp. 7, 241).

Stamatakis, A., P. Hoover, and J. Rougemont (2008). "A rapid bootstrap algorithm for the RAxML web servers". In: *Systematic Biology* 57.5, pp. 758–771. (Cit. on pp. 17, 18, 60).

Stang, M., P. G. L. Klinkhamer, and E. van der Meijden (2006). "Size constraints and flower abundance determine the number of interactions in a plant-flower visitor web". In: *Oikos* 112, pp. 111 –121. (Cit. on p. 154).

– (2007). "Asymmetric specialization and extinction risk in plant–flower visitor webs: a matter of morphology or abundance?" In: *Oecologia* 151, pp. 442 –453. (Cit. on p. 154).

Stewart, K. D. and K. R. Mattox (1975). "Comparative cytology, evolution and classification of the green algae with some consideration of the origin of other organisms with chlorophylls a and b". In: *The Botanical Review* 41.1, pp. 104–135. (Cit. on p. 60).

Stockwell, V. O. (2005). "Flowers: diverse and mutable microbial habitats". In: *Phytopathology* 95.6. (Cit. on pp. 210, 242).

Stocsits, R., H. Letsch, J. Hertel, B. Misof, and P. F. Stadler (2009). "Accurate and efficient reconstruction of deep phylogenies from structured RNAs". In: *Nucleic Acids Research* 37.18, p. 6184. (Cit. on p. 238).

Stone, C. P., C. W. Smith, and J. T. Tunison (1992). *Alien plant invasions in native ecosystems of Hawaii: Management and research*. HI USA: University of Hawaii Press. (Cit. on p. 154).

Stringer, L. D., A. M. El-Sayed, L. M. Cole, L. A. M. Manning, and D. M. Suckling (2008). "Floral attractants for the female soybean looper, Thysanoplusia orichalcea (Lepidoptera: Noctuidae)". In: *Pest Management Science* 64.12, pp. 1218–1221. (Cit. on p. 154).

Suárez-Díaz, E. and V. H. Anaya-Muñoz (2008). "History, objectivity, and the construction of molecular phylogenies". In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 39.4, pp. 451–468. (Cit. on p. 4).

Sundin, G. W. and J. L. Jacobs (1999). "Ultraviolet radiation (UVR) sensitivity analysis and UVR survival strategies of a bacterial community from the phyllosphere of field-grown peanut (*Arachis hypogeae* L.)" In: *Microbial Ecology* 38.1, pp. 27–38. (Cit. on p. 210).

Svirenko, D. O. (1924). "Al'gologiceskie nabljudenija". In: *Russkij Arkhiv Protistologii* 3, pp. 175–182. (Cit. on p. 74).

Swofford, D. L. (2002). *PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0 b10*. MA USA: Sinauer Associates. (Cit. on pp. 11, 17, 60).

Symstad, A. J. (2000). "A test of the effects of functional group richness and composition on grassland invasibility". In: *Ecology* 81.1, pp. 99–109. (Cit. on p. 154).

Tamura, K., J. Dudley, M. Nei, and S. Kumar (2007). "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0". In: *Molecular Biology and Evolution* 24, pp. 1596–1599. (Cit. on p. 124).

Tavaré, S. (1986). "Some probabilistic and statistical problems in the analysis of DNA sequences". In: *Some mathematical questions in biology—DNA sequence analysis* 17, pp. 57–86. (Cit. on pp. 5, 60).

Taylor, F. (1977). "Foraging behavior of ants - experiments with 2 species of Myrmecine ants". In: *Behavioral Ecology and Sociobiology* 2.2, pp. 147–167. (Cit. on p. 154).

Teiling, E. (1957). "Some little known Swedish phytoplankters". In: *Svensk Botanisk Tidskrift* 51.1. (Cit. on p. 74).

Tell, G. and A. Couté (1979). "Ultrastructure de la paroi cellulaire de *Coelastrum sphaericum* var. *rugulosum* (Thom.) Sodomkova en microscopie electronique a balayage". In: *Revue algologique*, p. 163. (Cit. on p. 74).

Templeton, A. R. (1992). "The meaning of species and speciation: a genetic perspective". In: *The Units of Evolution. Essays on the Nature of Species, MIT Press, Cambridge, MA, USA*, pp. 159–183. (Cit. on p. 3).

Thompson, I. P., M. J. Bailey, J. S. Fenlon, T. R. Fermor, A. K. Lilley, J. M. Lynch, P. J. McCormack, M. P. McQuilken, K. J. Purdy, and P. B. Rainey (1993). "Quantitative and qualitative seasonal changes in the microbial community from the phyllosphere of sugar beet (*Beta vulgaris*)". In: *Plant and Soil* 150.2, pp. 177–191. (Cit. on p. 210).

Thompson, J. D., D. G. Higgins, and T. J. Gibson (1994). "Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". In: *Nucleic Acids Research* 22.22, p. 4673. (Cit. on pp. 46, 60, 237).

Thomson, E. and D. Tollervey (2005). "Nop53p is required for late 60S ribosome subunit maturation and nuclear export in yeast". In: *RNA* 11.8, p. 1215. (Cit. on pp. 5, 28).

Tillier, E. R. M. and R. A. Collins (1998). "High apparent rate of simultaneous compensatory base-pair substitutions in ribosomal RNA". In: *Genetics* 148.4, p. 1993. (Cit. on p. 46).

Tilman, D. (1997). "Community invasibility, recruitment limitation, and grassland biodiversity". In: *Ecology* 78.1, pp. 81–92. (Cit. on p. 154).

Tomczykowa, M., M. Tomczyk, P. Jakoniuk, and E. Tryniszewska (2008). "Antimicrobial and antifungal activities of the extracts and essential oils of *Bidens tripartita*". In: *Folia Histochemica et Cytobiologica* 46.3, pp. 389–393. (Cit. on p. 210).

Torres, R. A., M. Ganal, and V. Hemleben (1990). "GC balance in the internal transcribed spacers ITS 1 and ITS 2 of nuclear ribosomal RNA genes". In: *Journal of Molecular Evolution* 30.2, pp. 170–181. (Cit. on pp. 6, 7).

Traut, W., K. Sahara, and F. Marec (2007). "Sex chromosomes and sex determination in Lepidoptera". In: *Sexual Development* 1.6, pp. 332–346. (Cit. on p. 124).

Tsarenko, P. M. and O. A. Petlevanny (2001). "Addition to the diversity of algae of Ukraine". In: *Algologia* Supplement, pp. 1–130. (Cit. on p. 74).

Tshikolovets, V. V. (1997). *The butterflies of Pamir*. Bratislava, Slowakia: Bratislava. (Cit. on p. 124).

– (1998). *The butterflies of Turkmenistan*. Ukraine: Nacional Na Akademija Nank Ukrainy. (Cit. on p. 124).

– (2000). *The butterflies of Uzbekistan*. Ukraine: Nacional Na Akademija Nank Ukrainy. (Cit. on p. 124).

– (2004). *The butterflies of Tajikistan*. Author's edition. (Cit. on p. 124).

– (2005a). *The butterflies of Kyrgyzstan*. Author's edition. (Cit. on p. 124).

– (2005b). *The butterflies of Ladakh (N.-W. India)*. Author's edition. (Cit. on p. 124).

Tshikolovets, V. V., A. Bidzilya, and M. Golovushkin (2002). *The butterflies of Transbaikal Siberia*. Author's edition. (Cit. on p. 124).

Tsuji, K., A. Hasyim, H. Nakamura, and K Nakamura (2004). "Asian weaver ants, Oecophylla smaragdina, and their repelling of pollinators". In: *Ecological Research* 19, pp. 669 –673. (Cit. on p. 154).

Turmel, M., C. Otis, J. C. de Cambiaire, J. F. Pombert, and C. Lemieux (2002). "The chloroplast genome sequence of *Chlorokybus atmophyticus*: evidence that charophycean green algae from an early diverging lineage adapted to terrestrial life". In: *Journal of Phycology* 38.s1, pp. 35–36. (Cit. on p. 86).

Turmel, M., J. S. Brouard, C. Gagnon, C. Otis, and C. Lemieux (2008). "Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses". In: *Journal of Phycology* 44.3, pp. 739–750. (Cit. on p. 60).

United Nations (2010). *International Year of Biodiversity*. URL: http://www.unep.org/iyb. (Cit. on p. 3).

van de Peer, Y. and R. de Wachter (1993). "TREECON: a software package for the construction and drawing of evolutionary trees". In: *Bioinformatics* 9.2, p. 177. (Cit. on p. 74).

van der Pijl, L. (1955). "Some remarks on myrmecophytes". In: *Phytomorphology* 5, pp. 190–200. (Cit. on p. 154).

van der Sande, C. A. F. M., M. Kwa, R. W. van Nues, H. van Heerikhuizen, H. A. Raué, and R. J. Planta (1992). "Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA". In: *Journal of molecular biology* 223.4, pp. 899–910. (Cit. on p. 60).

van Hannen, E. J., M. Lürling, and E. van Donk (2000). "Sequence analysis of the ITS-2 region: a tool to identify strains of *Scenedesmus* (Chlorophyceae)". In: *Journal of Phycology* 36.3, pp. 605–607. (Cit. on p. 86).

van Hannen, E. J., P. Fink-Godhe, and M. Lurling (2002). "A revised secondary structure model for the internal transcribed spacer 2 of the green algae *Scenedesmus* and *Desmodesmus* and its implication for the phylogeny of these algae". In: *European Journal of Phycology* 37.2, pp. 203–208. (Cit. on pp. 28, 60, 74, 86, 239).

van Nues, R. W., J. M. Rientjes, S. A. Morré, E. Mollee, R. J. Planta, J. Venema, and H. A. Raué (1995). "Evolutionarily conserved structural elements are critical for processing of Internal Transcribed Spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA". In: *Journal of Molecular Biology* 250.1, pp. 24–36. (Cit. on pp. 28, 60).

van Oppen, M. J. H., B. J. McDonald, B. Willis, and D. J. Miller (2001). "The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: reticulation, incomplete lineage sorting, or morphological convergence?" In: *Molecular Biology and Evolution* 18.7, p. 1315. (Cit. on p. 46).

van Valen, L. (1976). "Ecological species, multispecies, and oaks". In: *Taxon*, pp. 233–239. (Cit. on p. 3).

Vanormelingen, P., E. Hegewald, A. Braband, M. Kitschke, T. Friedl, K. Sabbe, and W. Vyverman (2007). "The systematics of a small spineless *Desmodesmus* species, *D. costat-granulatus* (Sphaeropleales, Chlorophyceae), based on ITS2 rDNA sequence analyses and cell wall

morphology". In: *Journal of Phycology* 43.2, pp. 378–396. (Cit. on p. 74).

Velickovic, D. T., N. V. Ransjelovic, M. S. Ristic, A. S. Velicovic, and A. A. Simelcerovic (2003). "Chemical constituents and antimicrobial activity of the ethanol extracts obtained from the flower, leaf and stem of *Salvia officinalis* L." In: *Journal of the Serbian Chemical Society* 68.1, pp. 17–24. (Cit. on p. 210).

Venema, J. and D. Tollervey (1999). "Ribosome Synthesis in *Saccharomyces cerevisiae*". In: *Annual review of genetics* 33.1, pp. 261–311. (Cit. on pp. 5, 28, 230).

– (2004). "Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*". In: *Yeast* 11.16, pp. 1629–1650. (Cit. on pp. 5, 6, 28).

Verbruggen, H., O. D. Clerck, T. Schils, W. H. C. F. Kooistra, and E. Coppejans (2005). "Evolution and phylogeography of *Halimeda* section *Halimeda* (Bryopsidales, Chlorophyta)". In: *Molecular Phylogenetics and Evolution* 37.3, pp. 789–803. (Cit. on p. 86).

Vet, L. E., J. Van Lenteren, M. Heymans, and E. Meelis (1983). "An airflow olfactometer for measuring olfactory responses of hymenopterous parasitoids an other small insects". In: *Physiological Entomology* 8, pp. 97–106. (Cit. on p. 154).

Vila, M., I. Bartomeus, A. C. Dietzsch, T. Petanidou, I. Steffan-Dewenter, J. C. Stout, and T. Tscheulin (2009). "Invasive plant integration into native plant-pollinator networks across Europe". In: *Proceedings of the Royal Society B-Biological Sciences* 276.1674, pp. 3887–3893. (Cit. on p. 154).

Vodolazhsky, D. I., M. Wiemers, and B. V. Stradomsky (2009). "A comparative analysis of mitochondrial and nuclear DNA sequences in blue butterflies of the subgenus *Polyommatus* (s. str.) Latreille, 1804 (Lepidoptera: Lycaenidae: *Polyommatus*)". In: *Kavkazskij entomologiceskij bjulleten* 5.1, pp. 115–120. (Cit. on p. 124).

Vollmer, S. V. and S. R. Palumbi (2004). "Testing the utility of internally transcribed spacer sequences in coral phylogenetics". In: *Molecular Ecology* 13, pp. 2763–2772. (Cit. on p. 124).

Waelti, M. O., J. K. Muhlemann, A. Widmer, and F. P. Schiestl (2008). "Floral odour and reproductive isolation in two species of *Silene*". In: *Journal of Evolutionary Biology* 21.1, pp. 111–121. (Cit. on p. 210).

Wagner, W. L., D. R. Herbst, and S. H. Sohmer (1990). *Manual of the flowering plants of Hawai'i*. HI USA: Bishop Museum. (Cit. on pp. 154, 241).

Wahlberg, N. and C. W. Wheat (2008). "Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera". In: *Systematic Biology* 57.2, pp. 231–242. (Cit. on p. 124).

Wakeham-Dawson, A. and P. Spurdens (1994). "Anomalous blue butterflies of the genus *Agrodiaetus* Hubner (Lepidoptera: Lycaenidae) in southern Greece". In: *Entomologist's Gazette* 45.1, pp. 13–20. (Cit. on p. 124).

West, L. and F. E. Fritsch (1927). *A treatise on the british freshwater algae*. Cambridge UK: Cambridge University Press. (Cit. on p. 60).

West, W. and G. S. West (1896). "Algae from central Africa". In: *Journal of Botany British and Foreign* 34, pp. 377–384. (Cit. on p. 74).

Wetterer, J. K. (1998). "Nonindigenous ants associated with geothermal and human disturbance in Hawai'i Volcanoes National Park". In: *Pacific Science* 52, pp. 40–50. (Cit. on pp. 154, 241).

Wheeler, Q. D. (2005). "Losing the plot: DNA "barcodes" and taxonomy". In: *Cladistics* 21, pp. 7405–407. (Cit. on p. 86).

Wheeler, W. C. and R. L. Honeycutt (1988). "Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications". In: *Molecular Biology and Evolution* 5.1, p. 90. (Cit. on pp. 6, 7, 237).

Wheeler, W. M. (1934). "Revised list of Hawaiian ants". In: *Bishop Museum Occasional Papers* 10, pp. 1–21. (Cit. on p. 154).

Whelan, S., P. Lin, and N. Goldman (2001). "Molecular phylogenetics: state-of-the-art methods for looking into the past". In: *Trends in Genetics* 17.5, pp. 262–272. (Cit. on p. 46).

Whinnett, A., A. V. Z. Brower, M. M. Lee, K. R. Willmott, and J. Mallet (2005). "Phylogenetic utility of tektin , a novel region for inferring systematic relationships among Lepidoptera". In: *Annals of the Entomological Society of America* 98.6, pp. 873–886. (Cit. on p. 124).

White, T. J., T. Bruns, S. Lee, and J. Taylor (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics". In: *PCR Protocols: A Guide to Methods and Applications*. Ed. by M. A. Innis, D. H. Gelfand, J. J. Shinsky, and T. J. White. CA USA: Academic Press, pp. 315–322. (Cit. on pp. 7, 28, 60, 86, 124, 240).

Whitfield, J. B. and K. M. Kjer (2008). "Ancient rapid radiations of insects: challenges for phylogenetic analysis". In: *Annual Reviews of Entomology* 53, pp. 449–472. (Cit. on p. 124).

Wiemers, M. (2003). "Chromosome differentiation and the radiation of the butterfly sub-

genus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*) – a molecular phylogenetic approach". PhD thesis. University of Bonn, Germany. URL: http://hss.ulb.uni-bonn.de/diss_online/math_nat_fak/2003/wiemers_martin/index.htm. (Cit. on pp. 124, 240).

Wiemers, M. and J. de Prins (2004). "*Polyommatus* (*Agrodiaetus*) *paulae* sp. nov. (Lepidoptera: Lycaenidae) from northwest Iran, discovered by means of molecular, karyological and morphological methods". In: *Entomologische Zeitschrift* 114.4, pp. 155–162. (Cit. on p. 124).

Wiemers, M. and K. Fiedler (2007). "Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae)". In: *Frontiers in Zoology* 4, p. 8. (Cit. on p. 124).

Wiemers, M., A. Keller, and M. Wolf (2009). "ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*)". In: *BMC Evolutionary Biology* 9.1, p. 300. (Cit. on pp. 15, 124, 240).

Wikimedia Foundation Inc. (2010). *Wikipedia®*. URL: http://wikipedia.org. (Cit. on p. 11).

Wilcox, L. W. and G. L. Floyd (1998). "Ultrastructure of the gamete of *Pediastrum duplex* (Chlorophyceae)". In: *Journal of Phycology* 24.2, pp. 140–146. (Cit. on p. 60).

Will, K. W., B. D. Mishler, and Q. D. Wheeler (2005). "The perils of DNA barcoding and the need for integrative taxonomy". In: *Systematic Biology* 54.5, pp. 844–851. (Cit. on pp. 86, 124).

Wille, N. (1909). "Conjugatae und Chlorophyceae". In: *Die natürlichen Pflanzenfamilien. Nachträge zu I. Teil, Abt* 2, pp. 1–96. (Cit. on p. 74).

Willmer, P. G. and G. N Stone (1997). "How aggressive ant-guards assist seed-set in Acacia flowers". In: *Nature* 388.6638, pp. 165–167. (Cit. on p. 154).

Willmer, P. G., C. V. Nuttman, N. E. Raine, G. N. Stone, J. G. Pattrick, K. Henson, P. Stillman, L. McIlroy, S. G. Potts, and J. T. Knudsen (2009). "Floral volatiles controlling ant behaviour". In: *Functional Ecology* 23, pp. 888–900. (Cit. on p. 154).

Wilson, P., M. C. Castellanos, J. N. Hogue, J. D. Thomson, and W. S. Armbruster (2004). "A multivariate search for pollination syndromes among penstemons". In: *Oikos* 104, pp. 345–361. (Cit. on p. 154).

Windels, C. E. (2000). "Economic and social impacts of *Fusarium* head blight: changing farms and rural communities in the Northern Great Plains". In: *Phytopathology* 90.1, pp. 17–21. (Cit. on p. 210).

Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms". In: *Proceedings of the National Academy of Sciences of the USA* 74.11, p. 5088. (Cit. on pp. 5, 28).

Woese, C. R., O. Kandler, and M. L. Wheelis (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya". In: *Proceedings of the National Academy of Sciences of the USA* 87.12, p. 4576. (Cit. on pp. 5, 46).

Wolf, M. and J. Schultz (2009). "Barcoding of plants and fungi: ITS better than its reputation". In: *Science Online E-letters* 325.5941, p. 682. (Cit. on pp. 86, 240).

Wolf, M., L. Krienitz, and D. Hepperle (2002a). "Phylogenetic position of *Actinastrum hantzschii* Lagerheim 1882 (Chlorophyta, Trebouxiophyceae)". In: *Archiv für Hydrobiologie* 104, pp. 59–67. (Cit. on pp. 74, 86).

Wolf, M., M. A. Buchheim, E. Hegewald, L. Krienitz, and D. Hepperle (2002b). "Phylogenetic position of the Sphaeropleaceae (Chlorophyta)". In: *Plant Systematics and Evolution* 230.3, pp. 161–171. (Cit. on pp. 60, 86, 239).

Wolf, M., D. Hepperle, and L. Krienitz (2003a). "On the phylogeny of *Radiococcus*, *Planktosphaeria* and *Schizochlamydella* (Radiococcaceae, Chlorophyta)". In: *Biologia* 58.4, pp. 759–766. (Cit. on pp. 74, 86).

Wolf, M., E. Hegewald, D. Hepperle, and L. Krienitz (2003b). "Phylogenetic position of the Golenkiniaceae (Chlorophyta) as inferred from 18S rDNA sequence data". In: *Biologia* 58.4, pp. 433–436. (Cit. on p. 86).

Wolf, M., J. Friedrich, T. Dandekar, and T. Müller (2005a). "CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures". In: *In silico Biology* 5.3, pp. 291–294. (Cit. on pp. 11, 18, 38, 74, 124).

Wolf, M., M. Achtziger, J. Schultz, T. Dandekar, and T. Müller (2005b). "Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures". In: *RNA* 11.11, p. 1616. (Cit. on pp. 6, 15, 23, 28, 46, 60, 74, 86, 237).

Wolf, M., B. Ruderisch, T. Dandekar, J. Schultz, and T. Müller (2008). "ProfDistS: (profile-) distance based phylogeny on sequence–structure alignments". In: *Bioinformatics* 24.20, p. 2401. (Cit. on pp. 6, 11, 17, 18, 22, 28, 38, 46, 74, 86, 124, 154, 210, 230, 237).

Wuyts, J., G. Perriere, and Y. van de Peer (2004). "The European ribosomal RNA database". In: *Nucleic Acids Research* 32.Database Issue, p. D101. (Cit. on pp. 12, 13, 28).

Yamada, T. K., K. Miyaji, and H. Nozaki (2008). "A taxonomic study of *Eudorina unicocca* (Volvocaceae, Chlorophyceae) and related species, based on morphology and molecular phylogeny". In: *European Journal of Phycology* 43.3, pp. 317–326. (Cit. on p. 86).

Yamagishi, T. and M. Akiyama (1996). "Photomicrographs of the freshwater algae". In: *Uchida Rokakuho* 17, pp. 5–24. (Cit. on p. 74).

Yang, C. H., D. E. Crowley, J. Borneman, and N. T. Keen (2001). "Microbial phyllosphere populations are more complex than previously realized". In: *Proceedings of the National Academy of Sciences of the USA* 98.7, p. 3889. (Cit. on p. 210).

Yang, Z. (1998). "On the best evolutionary rate for phylogenetic analysis". In: *Systematic Biology* 47.1, pp. 125–133. (Cit. on pp. 46, 238).

Young, I. and A. W. Coleman (2004). "The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example". In: *Molecular Phylogenetics and Evolution* 30.1, pp. 236–242. (Cit. on pp. 46, 60).

Zakharov, E. V., N. F. Lobo, C. Nowak, and J. J. Hellmann (2009). "Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperiidae)". In: *Heredity* 102, pp. 590–599. (Cit. on p. 124).

Zechman, F. W. (2003). "Phylogeny of the Dasycladales (Chlorophyta, Ulvophyceae) based on analyses of Rubisco large subunit (rbcL) gene sequences". In: *Journal of Phycology* 39.4, pp. 819–827. (Cit. on pp. 86, 240).

Zimmer, E. A., S. L. Martin, S. M. Beverley, Y. W. Kan, and A. C. Wilson (1980). "Rapid duplication and loss of genes coding for the alpha chains of hemoglobin". In: *Proceedings of the National Academy of Sciences of the USA* 77.4, p. 2158. (Cit. on p. 86).

ACRONYMS

A Adenine.
AIC Akaike information criterion.

C Cytosine.
CBC compensatory base change.
CO Cytochrome c oxidase.
CPU computational processing unit.

DDR double data rate.
DNA deoxyribonucleic acid.

G Guanine.
GPL general public license.
GTR general time reversible.

H Hydrogen.
hCBC hemi compensatory base change.
HMM hidden Markov model.
HPC high performance computing.
HTU hypothetical taxonomic unit.

ITS internal transcribed spacers.
ITS2 internal transcribed spacer 2.

LSU large subunit.

ML maximum likelihood.
MP maximum parsimony.

NJ neighbor joining.
NMR nuclear magnetic resonance.
NNI nearest neighbour interchange.

OTU operational taxonomic unit.

PDF portable document format.
PNJ profile neighbor joining.

RNA ribonucleic acid.
rRNA ribosomal RNA.

SDRAM synchronous dynamic random access memory.
SSU small subunit.
SVG scalable vector graphics.

T Thymine.

U Uracil.

LIST OF FIGURES

LIST OF TABLES

CURRICULUM VITAE

Personal Information

Name	Alexander Keller
Gender	Male
Date of Birth	31.05.1981
Place of Birth	Heidelberg, Germany
Nationality	German

Contact Information

Private Adress	Jakob-Riedinger Straße 8, 97074 Würzburg
Office Adress	Dept. of Bioinformatics, Biocenter, Am Hubland, 97074 Würzburg
Private Phone	+49 178 5512971
Office Phone	+49 931 3184427
Email	a.keller@biozentrum.uni-wuerzburg.de

Education

			14.07	2010	Biocenter Science Award Winner, University of Würzburg
01.04	2008	–	today		Elite Network Bavaria Graduate School Student
01.11	2007	–	today		University of Würzburg: PhD studies in Biology
01.06	2007	–	31.10	2007	University of Würzburg: Research Assistant
			16.05	2007	**Diplom (Diploma) in Biology**
15.04	2004	–	16.05	2007	University of Würzburg: Graduate studies in Biology
			20.01	2004	**Vordiplom (Pre-Diploma) in Biology**
15.10	2001	–	14.04	2004	University of Karlsruhe: Undergraduate studies in Biology
03.07	2000	–	31.05	2001	Civil Service at the Red Cross, Pforzheim
			27.06	2000	**Abitur**
14.09	1998	–	26.06	2000	Hebel Gymnasium Pforzheim
26.08	1991	–	29.07	1998	Lise Meitner Gymnasium Königsbach
17.08	1987	–	10.07	1991	Grundschule Ispringen

LIST OF PUBLICATIONS

Journal Publications associated with this Thesis

Buchheim, M. A., A. Keller, C. Koetschan, F. Förster, B. Merget, and M. Wolf (2010b). "Internal transcribed spacer 2 (nu ITS2 rRNA) sequence-structure phylogenetics: towards an automated reconstruction of the green algal tree of life". In: submitted. (Cit. on pp. 86, 239).

Hegewald, E., M. Wolf, A. Keller, T. Friedl, and L. Krienitz (2010). "ITS2 sequence-structure phylogeny in the Scenedesmaceae with special reference to *Coelastrum* (Chlorophyta, Chlorophyceae), including the new genera *Comasiella* and *Pectinodesmus*". In: *Phycologia* 49, pp. 325–335. (Cit. on pp. 74, 86, 239).

Junker, R. R., C. C. Daehler, S. Dötterl, A. Keller, and N. Blüthgen (2010a). "Ant-flower networks in Hawaii: native plants are exploited, introduced plants defended". In: *Ecological Monographs*, in press. (Cit. on pp. 154, 241).

Junker, R. R., C. Loewel, R. Gross, S. Dötterl, A. Keller, and N. Blüthgen (2010b). "Flower- and leaf-specificity of epiphytic bacterial communities – implications for pollination ecology". In: to be submitted. (Cit. on p. 242).

Keller, A., T. Schleicher, F. Förster, B. Ruderisch, T. Dandekar, T. Müller, and M. Wolf (2008c). "ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales)". In: *BMC Evolutionary Biology* 8.1, p. 218. (Cit. on pp. 18, 28, 46, 60, 86, 124, 154, 230, 239, 241).

Keller, A., T. Schleicher, J. Schultz, T. Müller, T. Dandekar, and M. Wolf (2009a). "5.8S-28S rRNA interaction and HMM-based ITS2 annotation". In: *Gene* 430.1-2, pp. 50–57. (Cit. on pp. 15, 28, 38, 46, 74, 124, 154, 237).

Keller, A., F. Förster, T. Müller, T. Dandekar, J. Schultz, and M. Wolf (2010a). "Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees." In: *Biology Direct* 5.1, p. 4. (Cit. on pp. 46, 86, 154, 210, 230, 237).

Keller, A., M. Wolf, and T. Dandekar (2010b). "Ribosomal RNA phylogenetics: the third dimension". In: *Biologia* 65.3, pp. 388–391. (Cit. on pp. 230, 242, 243).

Koetschan, C., F. Förster, A. Keller, T. Schleicher, B. Ruderisch, R. Schwarz, T. Müller, M. Wolf, and J. Schultz (2010). "The ITS2 Database III–sequences and structures for phylogeny". In: *Nucleic Acids Research* 38.Database issue, p. D275. (Cit. on pp. 12, 15, 17, 38, 86, 230, 237, 239).

Wiemers, M., A. Keller, and M. Wolf (2009). "ITS2 secondary structure improves phylogeny estimation in a radiation of blue butterflies of the subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*)". In: *BMC Evolutionary Biology* 9.1, p. 300. (Cit. on pp. 15, 124, 240).

Other Journal Publications

Ernst, R., G. Landburg, A. Keller, and F. Dziock (2010). "Convergent evolution of trait-habitat relations and universal habitat templets: A cross-continental comparison of trait-environment relationships and environmental filters in tropical anuran amphibian assemblages". In: *Global Ecology and Biogeography*, submitted (24.06.2010).

Grafe, T. U. and A. Keller (2009). "A Bornean amphibian hotspot: the lowland mixed dipterocarp rainforest at Ulu Temburong National Park, Brunei Darussalam". In: *Salamandra* 45.1, pp. 25–38.

Keller, A. (2009). "Die Artenvielfalt der Amphibien in einem Tieflandregenwald auf Borneo [*english title*: Amphibian Diversity of a Bornean lowland rainforest]". In: *Terraria* 18, pp. 80–83.
Keller, A., M. Siegle, and T. U. Grafe (2008a). "Geographic distribution: *Parias sumatranus*". In: *Herpetological Review* 39, p. 373.
– (2008b). "Geographic distribution: *Xenodermus javanicus*". In: *Herpetological Review* 39, p. 373.
Keller, A., M.-O. Rödel, K. E. Linsenmair, and T. U. Grafe (2009b). "The importance of environmental heterogeneity for species diversity and assemblage structure in Bornean stream frogs". In: *Journal of Animal Ecology* 78.2, pp. 305 –314.

Poster Presentations

Achtziger, M., T. Dandekar, F. Förster, D. Gerlach, B. Hammesfahr, et al. (04.-06.03.2009). "ITS2 - it's 2 in 1 - Sequence-Structure Analyses". In: *Celebrating Darwin: From The Origin of Species to Deep Metazoan Phylogeny*. Berlin, Germany.
– (21.04.2009). "ITS2 - it's 2 in 1 - Sequence-Structure Analyses". In: *(R)evolution Research - Life and Sciences: A journey through time*. Würzburg, Germany.
Förster, F., A. Keller, R. O. Schill, T. Dandekar, and M. Wolf (03.-06.08.2009). "Distinguishing species in *Paramacrobiotus* (Tardigrada, Macrobiotidae)". In: *11th International Symposium on Tardigrada*. Tübingen, Germany.
– (04.-06.03.2009). "Distinguishing species in *Paramacrobiotus* (Tardigrada, Macrobiotidae) via compensatory base change analysis of internal transcribed spacer 2 secondary structures". In: *Celebrating Darwin: From The Origin of Species to Deep Metazoan Phylogeny*. Berlin, Germany.
– (21.04.2009). "Distinguishing species in *Paramacrobiotus* (Tardigrada, Macrobiotidae) via compensatory base change analysis of internal transcribed spacer 2 secondary structures". In: *(R)evolution Research - Life and Sciences: A journey through time*. Würzburg, Germany.
Keller, A. (22.-24.07.2009a). "ITS2 phylogenetics with secondary structures". In: *Structure Days*. Bayreuth, Germany.
Schlegelmilch, K., A. Keller, V. Monz, F. Jakob, and N. Schütze (04.-06.03.2010). "Funktionsaufklärung von WISP-Proteinen in mesenchymalen Stammzellen (MSC) und Chondrozyten [*english title*: Identification of WISP protein functions in mesenchymal stem cells (MSC) and chondrocytes]". In: *Osteology Congress*. Berlin, Germany.

Oral Presentations

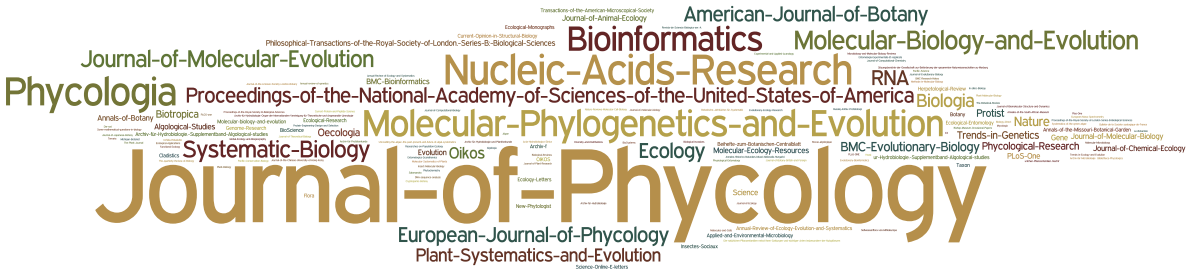
Grafe, T. U. and A. Keller (11.-13.06.2007). "Monitoring amphibian diversity in a pristine lowland tropical rainforest: baseline data for conservation". In: *Biodiversity Crisis on Tropical Islands*. Bandar Seri Begawan, Brunei Darussalam.
Keller, A. (21.-25.09.2008). "ITS2 structure and evolution". In: *BIGSS & BaCaTec Summer School*. Bayreuth, Germany.
– (22.-24.07.2009b). "Phylogenetics with RNA secondary structures". In: *Structure Days*. Bayreuth, Germany.
– (23.04.2010). "Structure in Phylogenetics!" In: *Tri-Beta Honors Society Induction Ceremony*. Edinboro University, PA, USA.
Keller, A., M.-O. Rödel, and T. U. Grafe (03.-07.10.2007). "Amphibian communities of the Ulu Temburong National Park". In: *DGHT/ÖGH Annual Symposium*. Hallein, Austria.

THESIS STATISTICS

To get a nice overview about authors, topics and scientific journals relevant to this thesis, I decided to scan for, cluster and visualize (Feinberg 2009; Shannon et al. 2003) such information from the bibliographical library used for this thesis. In my opinion, with a little background knowledge about the journals and the referred authors, this gives a nice impression and abstracts the fields of research most relevant to this thesis. Further, the title words characterise the most important topics of this work stunningly good. As this has been a totally automatic procedure performed with a self written perl script without my manual corrections, it presents a very objective and summarizing view on this thesis.

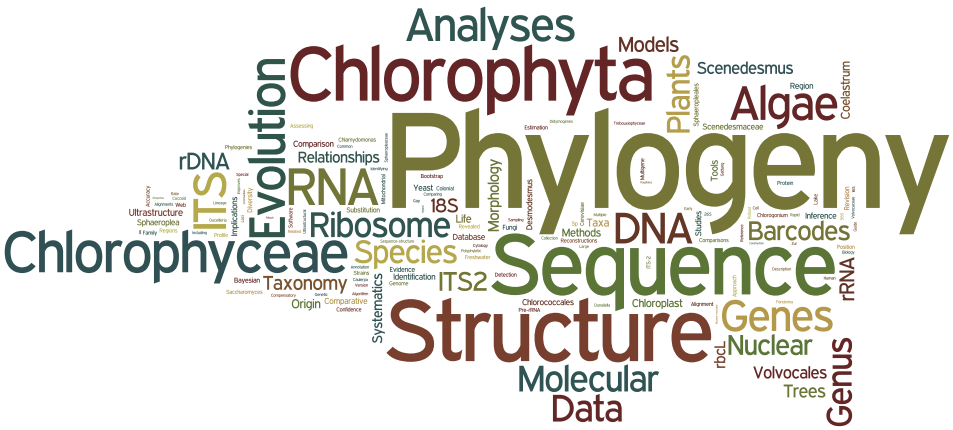
Journals with most References

1. Journal of Phycology 39x
2. Nucleic Acids Research 18x
3. Phycologia 15x
4. Molecular Phylogenetics and Evolution 15x
5. Bioinformatics 13x
6. Molecular Biology and Evolution 12x
7. Systematic Biology 10x
8. Ecology 9x
9. RNA 9x
10. Proceedings of the National Academy of Sciences of the USA 9x
11. Journal of Molecular Evolution 9x
12. European Journal of Phycology 8x
13. Plant Systematics and Evolution 8x
14. Oikos 8x
15. American Journal of Botany 8x
16. Biologia 7x
17. Science 5x
18. Nature 5x
19. BMC Evolutionary Biology 5x
20. Oecologia 5x
21. Trends in Genetics 4x
22. Applied and Environmental Microbiology 4x
23. Molecular Ecology Resources 4x
24. Biotropica 4x
25. Evolution 4x



Most prominent Title Words of cited Articles

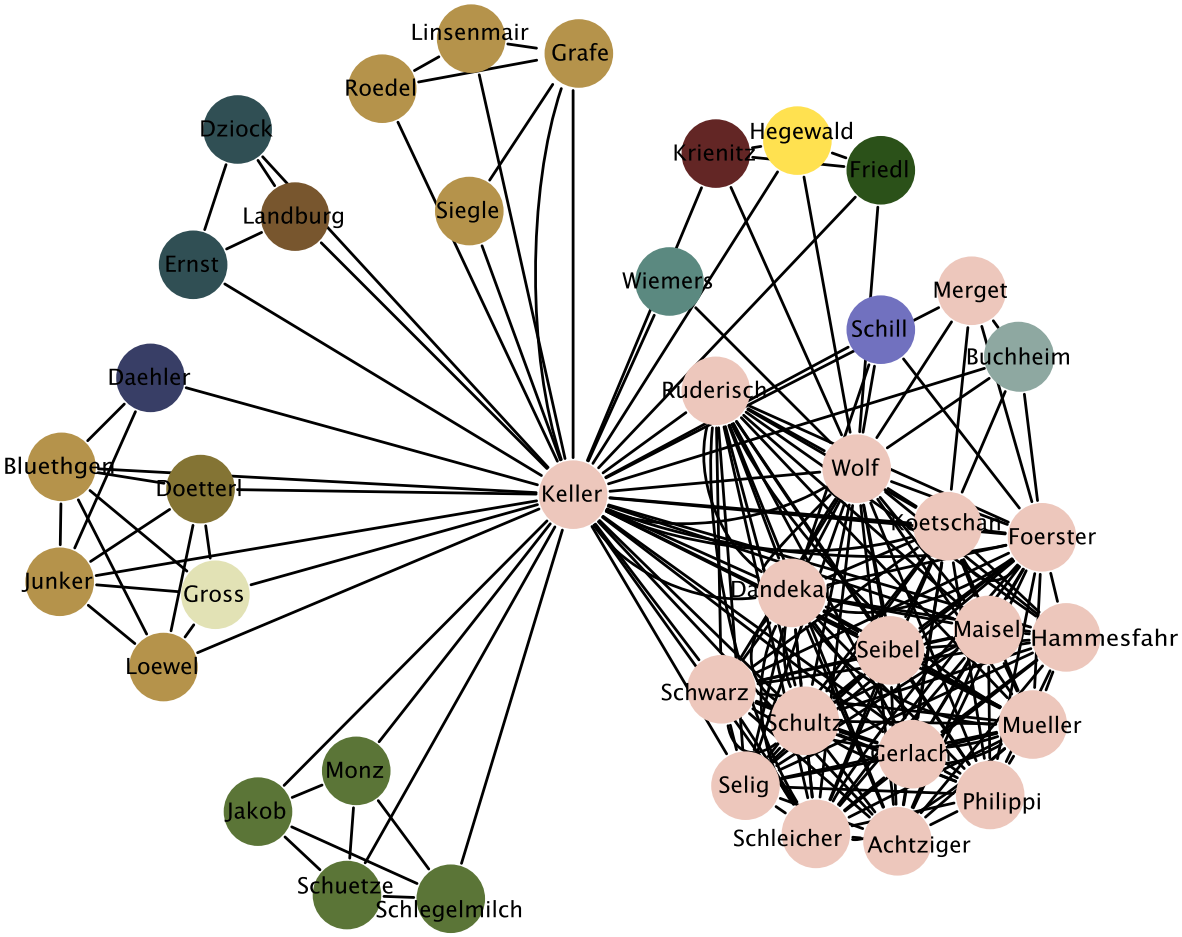
1.	Phylogeny	98x	26.	18S	14x
2.	Structure	67x	27.	Taxonomy	14x
3.	Sequence	62x	28.	Volvocales	12x
4.	Chlorophyta	54x	29.	Morphology	12x
5.	Evolution	41x	30.	Relationships	12x
6.	RNA	41x	31.	Systematics	11x
7.	Chlorophyceae	41x	32.	rbcL	11x
8.	Species	39x	33.	Origin	10x
9.	Analyses	37x	34.	*Scenedesmus*	10x
10.	Algae	34x	35.	Life	9x
11.	DNA	33x	36.	Microbial	9x
12.	Plants	32x	37.	Methods	9x
13.	ITS	32x	38.	Chloroplast	8x
14.	Genes	30x	39.	*Coelastrum*	8x
15.	Data	28x	40.	Diversity	8x
16.	Molecular	28x	41.	Ecology	7x
17.	Genus	27x	42.	Phylogenies	7x
18.	ITS2	25x	43.	Tools	7x
19.	Ribosome	24x	44.	Identification	7x
20.	Barcodes	22x	45.	Bacterial	7x
21.	Nuclear	20x	46.	Studies	7x
22.	rDNA	17x	47.	Ultrastructure	7x
23.	Models	16x	48.	Implications	6x
24.	rRNA	15x	49.	Taxa	6x
25.	Trees	15x	50.	Database	6x



Authors with most References (including Posters and Talks)

1. Wolf, M. 37x
2. Keller, A. 28x
3. Dandekar, T. 22x
4. Hegewald, E. 21x
5. Müller, T. 20x
6. Coleman, A. W. 16x
7. Schultz, J. 15x
8. Nozaki, H. 15x
9. Krienitz, L. 13x
10. Buchheim, M. A. 12x
11. Blüthgen, N. 11x
12. Hepperle, D. 10x
13. Förster, F. 10x
14. Lewis, L. A. 8x
15. Junker, R. R. 7x
16. Friedl, T. 7x
17. Chase, M. W. 7x
18. Pröschold, T. 6x
19. Lach, L. 6x
20. Watanabe, M. M. 6x
21. Grafe, T. U. 6x
22. Raguso, R. A. 5x
23. Tollervey, D. 5x
24. Chapman, R. L. 5x
25. Dötterl, S. 5x
26. Seibel, P. N. 5x
27. Turmel, M. 5x
28. Schleicher, T. 5x
29. Felsenstein, J. 5x
30. Ruderisch, B. 5x
31. Hillis, D. M. 4x
32. Olsen, J. L. 4x
33. Galen, C. 4x
34. Floyd, G. L. 4x
35. Hanagata, N. 4x
36. Otis, C. 4x
37. Buchheim, J. A. 4x
38. Huelsenbeck, J. P. 4x
39. Jossinet, F. 4x
40. Cowan, R. S. 4x
41. Westhof, E. 4x
42. Kuroiwa, T. 4x
43. Koetschan, C. 4x
44. Lipman, D. J. 4x
45. Achtziger, M. 4x
46. Peculis, B. A. 3x
47. Selig, C. 3x
48. Jürgens, A. 3x
49. Beattie, A. J. 3x
50. Uchida, H. 3x

Interaction Network of Co-Authorships (including Posters and Talks)



- Department of Bioinformatics, University of Würzburg, Germany
- Department of Biodiversity Dynamics, TU Berlin, Germany
- Department of Orthopedics, University of Würzburg, Germany
- Department of Animal Ecology and Tropical Biology, University of Würzburg, Germany
- Environmental Department, Anton de Kom University of Suriname, Suriname
- Department of Population Ecology, University of Vienna, Austria
- Institute of Chemistry and Dynamics of the Geosphere, Research Centre Jülich, Germany
- Experimental Phycology and Culture Collection of Algae (SAG), University of Göttingen, Germany
- Leibniz-Institute of Freshwater Ecology & Inland Fisheries, Germany
- Department of Zoology, University of Stuttgart, Germany
- Department of Biological Science, University of Tulsa, OK, USA
- Department of Microbiology, University of Würzburg, Germany
- Department of Plant Systematics, University of Bayreuth, Germany
- Botany Department, University of Hawai'i at Manoa, Hawaii

ERKLÄRUNG

Hiermit erkläre ich ehrenwörtlich, dass ich die vorliegende Dissertation selbständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die Dissertation wurde bisher weder in gleicher noch ähnlicher Form in einem anderen Prüfungsverfahren vorgelegt. Außer dem Diplom in Biologie von der Universität Würzburg habe ich bisher keine weiteren akademischen Grade erworben oder versucht zu erwerben.

Würzburg, den 19. 10. 2010

_____

Alexander Keller