An evolutionary view of the taxonomy and ecology of *Inocybe (Agaricales)*

with new perspectives gleaned from GenBank metadata

Martin Ryberg, 2009

UNIVERSITY OF GOTHENBURG

Faculty of Science, Department of Plant and Environmental Sciences

Akademisk avhandling för filosofie doktorsexamen i systematisk botanik, som enligt naturvetenskapliga fakultetens beslut kommer att offentligt försvaras torsdagen den 30:e april 2009, klockan 9.00 i föreläsningssalen, Institutionen för Växt- och Miljövetenskaper, Carl Skottbergsgata 22B, 413 19 Göteborg. Examinator: Bengt Oxelman. Fakultetsopponent: Dr. Ursula Peintner, Universität Innsbruck, Institute of Microbiology, Technikerstr. 25, A- 6020 Innsbruck, Austria.

ISBN: 978-91-85529-28-5

Dissertation Abstract

Martin Ryberg (2009). An evolutionary view of the taxonomy and ecology of *Inocybe* (*Agaricales*) with new perspectives gleaned from GenBank metadata.
University of Gothenburg, Department of Plant and Environmental Sciences, Box 461, 405 30 Göteborg, Sweden

Inocybe (*Inocybaceae*) is one of the most speciose genera among the gilled mushrooms (*Agaricales*) but large parts of its taxonomy and evolutionary history remain poorly explored. The present thesis shows that the traditional infrageneric classification of *Inocybe* does not fully reflect evolutionary relationships and that the two most commonly used taxonomic characters, spore shape and presence of a cortina, have not evolved in such a way as to define unique monophyletic groups. The section *Rimosae*, in its traditional circumscription, is divided into two separate clades by the present analyses, and *Rimosae* s.str. does not group with the subgenus *Inosperma* - where it is commonly placed - but forms a separate clade more closely related to the subgenus *Inocybe*. Also the sections of the subgenus *Inocybe*, the largest of the subgenera, should in general be interpreted in a strict sense if they are to reflect monophyletic groups. Such a view-point would leave many species with an uncertain placement. On the species level it is shown that many of the traditionally accepted taxa are insufficiently characterized in terms of circumscription and morphological variation.

This thesis also explores metadata associated with DNA sequences from molecular ecological studies. These sequences represent a large part of the fungal internal transcribed spacer (ITS) sequences in GenBank that are relatively well annotated regarding ecological and geographical data. They are, however, seldom used as sources of such information. One reason is that they typically lack precise taxonomic annotation and therefore are hard to search in a systematic context. To facilitate the exploration of such sequences, the software suite *emerencia* (www.emerencia.org) was developed and is presented as a component of this thesis. The *emerencia* software was employed to explore the distribution and ecology of *Inocybe* and was found to have the capacity to significantly expand on our knowledge of the world-wide distribution of the genus and of the ecology of its individual species. This new information was compiled together with information from other sources to explore whether host preference and three characters coding for preference for particular soil conditions are correlated to the phylogeny. This was done using ancestral state reconstruction methods. The results show that while soil moisture preference is not, host preference, preference for calcareous soils, and preference in soil nutritional status are indeed correlated to the phylogeny. This indicates that a well formulated taxonomy that is reflective of phylogenetic relationships can have predictive values for these ecological traits.

Key words: *Inocybe*, *Inocybaceae*, *emerencia*, GenBank, insufficiently identified sequences (IIS).

ISBN: 978-91-85529-28-5

To my nephews and the smiles they bring

道可道，非常道。名可名，非常名。

- 老子

An evolutionary view of the taxonomy and ecology of *Inocybe (Agaricales)* with new perspectives gleaned from GenBank metadata

List of papers

This thesis is based on the following papers. They are referred to in the text by their Roman numbers.

I. Nilsson RH, Kristiansson E, Ryberg M, Larsson K-H. 2005. Approaching the taxonomic affiliation of unidentified sequences in public databases - an example from the mycorrhizal fungi. *BMC Bioinformatics* **6:** 178.

II. Ryberg M, Kristiansson E, Sjökvist E, Nilsson RH. 2009. An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. *New Phytologist* **181:** 471-477.

III. Ryberg M, Nilsson RH, Kristiansson E, Töpel M, Jacobsson S, Larsson E. 2008. Mining metadata from unidentified ITS sequences in GenBank: A case study in *Inocybe* (*Basidiomycota*). *BMC Evolutionary Biology* **8:** 50.

IV. Larsson E, Ryberg M, Moreau P-A, Delacuse Mathiesen Å, Jacobsson S. Submitted. Taxonomy and evolutionary relationships within species of section *Rimosae* (*Inocybe*) based on ITS, LSU and mtSSU sequence data.

V. Ryberg M, Larsson E, Jacobsson S. Manuscript. An evolutionary perspective on morphological and ecological characters in the mushroom forming family *Inocybaceae* (*Agaricomycotina*, *Fungi*).

Contribution of the separate authors to the different papers.
In all papers the lead author had the main responsibility for compiling and writing the paper but all authors contributed in the writing process.

I. All authors contributed to the structure and functions of *emerencia*. RHN and EK wrote most parts of the *emerencia* core, the CGI scripts, and the database handlers. I was responsible for the mycological part, including literature comparisons, integrity testing, and data verification. KHL contributed with advice on fungal taxonomy and systematics.

II. I had the main responsibility for scripting the new search function and compiling and analyzing the data. EK contributed insights on scripting and statistics. ES compiled statistics on host data. RHN contributed to the database handling and scripting.

III. I had the main responsibility for the project and participated in all its parts. RHN was responsible for setting up the local installation of *emerencia*, assisted with database handling, and participated in the phylogenetic analysis. EK was responsible for the statistics and conducted the cluster analysis. MT was responsible for GIS applications and constructed the maps. SJ and EL had the main responsibility for taxonomic issues and EL did most of the molecular work.

IV. EL contributed taxonomic knowledge, lab work, and specimens, compiled and performed the final alignment of the data matrix, and performed the parsimony analysis. I performed the Bayesian analysis, helped in the alignment of the data matrix, contributed specimens, and participated in the lab work. ÅDM contributed lab work and preliminary analyses. PAM and SJ contributed specimens and taxonomic knowledge.

V. I aligned the DNA matrix, compiled the morphological and ecological character matrices, performed all analyses, contributed taxonomic knowledge, contributed specimens, and participated in the lab work. EL contributed taxonomic knowledge, participated in and had the main responsibility for some parts of the morphological and ecological character coding, performed lab work, and compiled the DNA data matrix. SJ contributed specimens and taxonomic knowledge.

Introduction

This thesis is centered around the systematics of *Inocybe* (Fr.) Fr. (*Inocybaceae* Jülich; Figure 1) and as Judd et al (2002) states "Systematics is the science of organismal diversity". This suggests that systematics aims at finding and describing all taxa and sorting out the relationships among them. Considering the human-induced mass extinction of biological life (Groom et al 2006) and that only approximately 97 000 (<7%; Kirk et al 2008) of the estimated 1.5 million (Hawksworth 2001) number of fungal species are described to date, it seems a daunting task to map and understand all species even when only considering the kingdom *Fungi*. The urgency of this task is exacerbated by the decreasing number of taxonomists (Lee 2000), and it is therefore important with projects that map this diversity. This thesis is part of a project financed by the Swedish Taxonomy Initiative (STI; Ronquist and Gärdenfors 2003). STI is an all-taxon biodiversity inventory that aims at finding and describing all multicellular species in Sweden. There is, however, more to diversity than just simple taxon counts. Here phylogenetic methods are used to investigate evolutionary relationships among taxa in *Inocybe* (**IV**; **V**). Furthermore, the infrageneric taxonomy is evaluated with respect to the estimated phylogenies (**IV**; **V**). For a deeper understanding of the discrepancies between morphological based taxonomy and molecular phylogenies, five morphological characters are investigated using ancestral state reconstruction methods (**V**). A new bioinformatics tool is introduced (**I**; **II**) and used to investigate the ecology and distribution of *Inocybe* with respect to DNA sequence metadata in GenBank (**III**) and four ecological characters are investigated to see if they are correlated with the phylogeny (**V**).

Taxonomy and phylogeny of *Inocybe* and their relation to morphology

An introduction to the taxonomy of Inocybe

Inocybe and the genus *Auritella* Matheny & Bougher are placed in the brown-spored family *Inocybaceae* (*Agariacales*; Matheny 2005; Matheny et al 2009). *Auritella* was not described until recently (Matheny and Bougher 2006) and the genus *Inocybe* has most frequently been placed in the family *Cortinariaceae* Pouzar. The use of *Inocybaceae* has, however, been justified by molecular phylogenetic studies that have shown it to be the sister clade of the family *Crepidotaceae* (Imai) Singer (Matheny 2005; Matheny et al 2006; **V**). Matheny et al (2006) also suggest that the ectomycorrhizal habit of *Inocybaceae* is one of several independent transformations to this nutritional habit and one that significantly separates it from *Crepidotaceae*.

The genus *Inocybe* contains more than 500 species (Kirk et al 2008) and is the larger of the two genera in the family and is one of the largest in *Agaricales* (Kirk et al 2008). The genus is recognized by its fibrillose to rimose cap texture. Most of the species are small and brown, there are, however, species that are relatively large and species with other colors e.g. white, liliac, or red. Most *Inocybe* species are poisonous, and even if they are small and brown and not particularly culinary appealing, the fact

Figure 1. A. *Inocybe dulcamara* var. *latispora* (subgenus *Mallocybe*). B. *I. rimosa* (section *Rimosae s.str.*). C. *I. relicina* (subgenus *Inocybe*). D. *I. salicis-herbaceae* (*praetervisa* group, subgenus *Inocybe*). E. *I. whitei* (*geophylla* group, subgenus *Inocybe*). F. *I. sindonia* (subgenus *Inocybe*).

that they commonly occur in parks and along paths increases the risks of consumption by children or pets (Gulden and Schumacher 1980). The most frequently occurring poisonous substance is muscarin (Kuyper 1986), a substance similar to acetylcholin that functions as a neurotransmitter in both the peripheral and the central nervous systems of many organisms, including humans (Gulden and Schumacher 1980). There are also species that have psilocybin, a substance with hallucinogenic effects (Kuyper 1986).

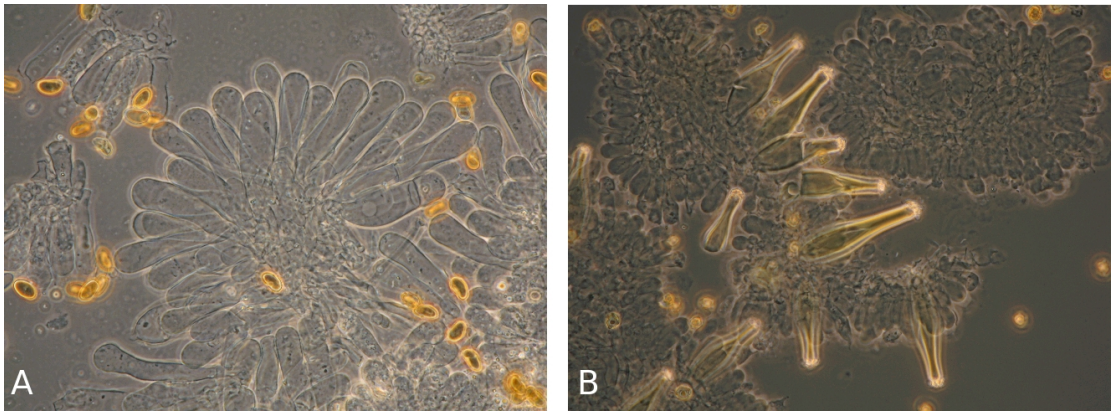

Figure 2. A. Cheilocystidia and phaseoliform spores of *I. maculata* (*maculata* clade, subgenus *Inosperma*). B. Yellow metuloid cystidia and nodulose spores of *I. petiginosa* (section *Petiginosae*, subgenus *Inocybe*).

The genus *Inocybe* is often further subdivided into three subgenera (*Inocybe*, *Inosperma* Kühner, and *Mallocybe* Kuyper; Kuyper 1986; Stangl 1989), which are sometimes informally treated as genera (e.g. *Mallocybe* in the GenBank taxonomy; Benson et al 2009). The subgenera *Mallocybe* and *Inosperma* are recognized by bean-shaped (phaseoliform) spores and subgenus *Mallocybe* is distinguished by necropigmented basidia and cheilocystidia originating from the hymenophoral hypha rather than from hymenial hypha as in subgenus *Inosperma* (Kuyper 1986; Jacobsson 2008; Figure 2A). The subgenus *Inocybe* is recognized by metuloid cystidia (also referred to as pleurocystidia) and amygdaloid or nodulose-angular spores (Figure 2B). Matheny (2005) showed that the subgenus *Inosperma* as circumscribed by Kuyper (1986) and Stangl (1989) is split into two separate clades. One of these clades is more closely related to the subgenus *Inocybe* than to the rest of the species in *Inosperma* s.str. This clade was called the "*Pseudosperma*" clade. With this exception the subgenus taxonomy seem to reflect the phylogeny rather well.

The subgenera *Inocybe* and *Inosperma* are subdivided into sections. *Inosperma* include the two sections *Cervicolores* Sing. and *Rimosae* (Fr.) Sacc. that can be distinguished based on the shape of the basidia, the often reddening context of most *Cervicolores* species, and the rimulose to rimose cap texture of many *Rimosae* species. The subgenus *Inocybe* (including subgenus *Inocibium* (Earle) Sing.) is divided into 11 sections according to Singer (1986) but Jacobsson (2008) recognizes only eight. These sections are separated mainly on spore shape and the presence and extent of cystidia on the stipe. This second character is often coupled with the presence of cortina, a veil between the stipe and the cap, that protects the lamella during fruiting body development. Cystidia are normally only present above the point where the cortina attaches to the stipe but this correlation is not perfect (e.g. in *I. sindonia* (Fr.) P. Karst.; Kuyper 1986; Figure 1F). Instead of sections Kuyper (1986) and Stangl (1989) use the two supersections *Cortinatae* (Kühner) Kuyper and *Marginatae* (Kühner) Kuyper. These supersections are based on the presence (supersection *Cortinatae*) or absence

(supersection *Marginatae*) of a cortina.

What this thesis adds to the story

When considering the phylogenetic estimates of *Inocybaceae* in **V**, the placement of the genus *Auritella* is somewhat uncertain and it is unclear if it belongs within or outside the genus *Inocybe* (also cf. Matheny 2005 and Matheny et al 2009). The section *Rimosae* (sensu Kuyper 1986) is divided into two groups of which one, *Rimosae* s.str., is more closely related to the subgenus *Inocybe* than to *Rimosae* p.p. (here referred to as the *maculata* clade; Figure 3; **IV**; **V**). The *Rimosae* s.str. clade is very likely to represent the "*Pseudosperma*" clade of Matheny (2005). The fact that *Rimosae* (sensu Kuyper 1986) is not monophyletic presents a taxonomic dilemma as there are few morphological characters to separate the *Rimosae* s.str. and the *maculata* clades (**IV**). Morphology is important in the everyday taxonomic work but phylogeny presents the framework in which many more characters can be interpreted, it is important to carefully weigh these two issues against each other. The taxonomic issues concerning section *Rimosae* are central to the question whether the subgenera of *Inocybe* should be elevated to genera. As this should be addressed in light of the overall taxonomy of *Inocybaceae,* no taxonomic revision is made here.

The taxonomic subdivision of the subgenus *Inocybe* is important as this is the largest subgenus of *Inocybe* and a single section of this taxa may hold as many species as any of the other subgenera (Jacobsson 2008). The supersections, *Marginatae* and *Cortinatae* as circumscribed by Kuyper (1986) and Stangl (1989), do not receive any support as monophyletic in the phylogenetic analysis (**V)** and the presence of a cortina is only moderately reflective of the phylogeny when investigated by ancestral state reconstruction methods (**V**). There are, however, indications that presence of a cortina can be used together with other characters to define smaller groups like *Lacerae* (Fr.) Kummer s.str. and the *lanuginosa* group (sensu **V**). Spore morphology is a more promising character for sorting species at this level (**V**) but the subgenus *Inocibium* does not receive any support as a separate taxon that include all the smooth-spored *Inocybe* with metuloid cystidia. The ancestral state reconstruction of **V** indicates that the character state of amygdaloid spores has arisen among species with nodulose spores. It is likely that this has rendered the nodulose-spored species paraphyletic and it is also probable that the amygdaloid spore shape has arisen on more than one occasion (**V**). It has been questioned if the presence of a bulb at the base of the stipe is a good taxonomic character (Kuyper 1986), and there is now evidence that it should be used with care as it is only moderately reflective of the phylogeny (**V**). Nevertheless, the character appears to be useful in describing some groups, e.g. the *napipes* group.

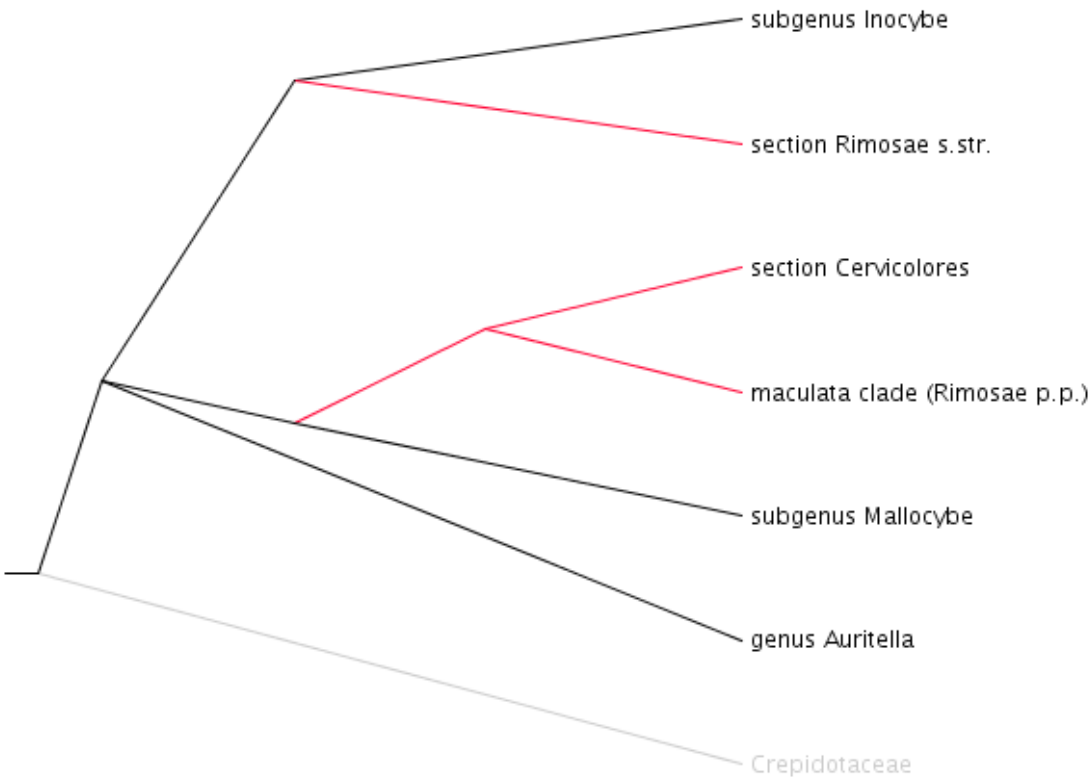

Figure 3. A schematic tree of *Inocybaceae*. The genus *Auritella*, subgenus *Inocybe* and subgenus *Mallocybe* are given at the tip of the branches. Branches leading to groups included in the subgenus *Inosperma* sensu Kuyper (1986) are colored red, and the section or clade names in this subgenus are given at the tips. *Crepidotaceae* is given as outgroup and is colored gray.

The hymenial metuloid cystidia of the subgenus *Inocybe* vary in wall thickness and color in KOH. Even if these characters also vary somewhat within each species, there are species-specific trends. In **V** it is shown that the metuloid cystidia color and wall thickness is moderately reflective of the phylogeny, and many of the clades defined in the paper have either weakly yellow to yellow or transparent cystidia. As most of the species of subgenus *Inocybe* have intermediate metuloid cystidia wall thickness, this character has limited application to separate groups, but there are clades dominated by thick-walled cystidia, e.g. the *inodora* group.

V identifies 11 well-supported clades, nine of which are relatively well defined in morphological terms. Among the cortinate nodulose-spored species the *napipes* group, *Lacearae* s.str., the *lanuginosa* group, and the *subcarpta* group are distinguishable. The *napipes* group does, however, also include *I. cf. asterospora,* a species that lacks cortina, as a sister species to the rest of the clade. This clade also forms a group with *I. calospora* Quél. and the *praetervisa* group that mainly contain species lacking cortina, although the underlying support is moderate. The *praetervisa* group includes *I. asterospora* Quél. which is the type species of the section *Marginatae* Kühner. Most of the species in the larger clade, that includes both *napipes* and *praetervisa,* have a bulb (marginate or not; Figure 1D) at the base of the stipe, but the

clade does not include all bulbous species. *Petiginosae* Heim is another well defined clade that includes small, nodulose-spored, species that lack cortina. The clade features the species of *Petiginosae* sensu Jacobsson (2008) that have yellow lamella when young. The *inodora* group is another well-defined clade that include nodulose-spored species that lack cortina and that have a somewhat bulbouse to marginate bulbouse stipe base. This group also forms the only well supported clade within subgenus *Inocybe* that includes both nodulose and smooth-spored species. *Inocybe inodora* Velen. has amygdaloid spores, the most common shape among the smooth-spored species of subgenus *Inocybe.* However, *Inocybe vulpinella* Bruylants and *I. pruinosa* R. Heim have spores that deviate from this shape. *Inocybe vulpinella* has somewhat ovoid spores and *I. pruinosa* has spores that sometimes are slightly irregular.

Among the smooth-spored taxa of the subgenus *Inocybe, Lactiferae* Heim is the most well defined group. *Lactiferae* comprises the species with a cortina, that often have reddening flesh, and a peculiar sweetish or unpleasant smell. In addition they commonly contain the poison psilocybin. There is another large well supported clade - clade XIII sensu **V** - that mainly, but not exclusively, contains species with a cortina. This clade contains several well supported subclades. One of these clades is the *geophylla* group which holds species that have a ground color that is white (Figure 1E) instead of the usual dull colors of *Inocybe* as well as one clade that includes many of the species of section *Tardae* Bon, including *I. tarda* Kühner (sensu Stangl).

The groups outlined above leave many species with an uncertain taxonomic placement. *Inocybe relicina* (Fr.) Quél. (Figure 1C) is one of them; this is the type species of *Inocybe* and also of the section (supersection) *Cortinatae* Kühner. There are also sections that have not been considered here and it is therefore premature to make large revisions to the section taxonomy of *Inocybe,* even if the current classification system reflects the phylogenetic estimates poorly.

At the start of this project less than 90 *Inocybe* species were known from Sweden (Stridvall et al 1989), a number that has risen to about 150 at the time of this writing (Jacobsson 2008). In addition, it is clear that many generally held taxon concepts in *Inocybe* circumscribe species complexes rather than individual species (**III; IV**). In **IV** the complex surrounding *I. rimosa* (Bull.:Fr.) P. Kumm. (Figure 1B) is investigated. There are many species descriptions associated with this complex (Kuyper 1986) but most authors have accepted a wide species concept (Kuyper 1986; Stangl 1989). In **IV** there were several separate clades in the molecular based phylogenetic estimate that could also be separated based on morphology and be matched with existing species descriptions. There were, however, specimens that did not group with any of these identified clades such that they probably represent different species. They could not be confidently matched to any species description and there was not enough data to warrant the description of new taxa. So even if **IV** sorts out some problems there are still reasons for further studies of this species group. **III** and **IV** point at several other species that should be investigated more throughly, including *I. cincinnata* (Fr.:

Fr.) Quél, *I. flavella* P. Karst, *I. geophylla* (Fr.:Fr.) P. Kumm., *I. hirtella* Bres., *I. maculata* Boud., *I. soluta* Velen., and *I. squamata* J.E. Lange.

Other recent advances

In addition to the five evolutionary lineages *Auritella*, *Mallocybe*, *Inosperma* s.str., section *Rimosae* s.str., and subgenus *Inocybe,* Matheny et al (2009) present two additional lineages, the "*Mallocybella"* clade that groups with *Inosperma* and *Mallocybe* and the "*Nothocybe"* clade that groups with the subgenus *Inocybe*. Considering *Inocybe* is a widely distributed genus, present in areas poorly explored by molecular methods (**II; III**) and that phylogenetically very distinct lineages may not be very easy to define morphologically (**V**), it is possible that even more evolutionary lineages equivalent to the subgenus level await discovery.

Sequences from molecular ecological studies as a data source

An introduction to the sequence data from molecular ecological studies

In the last decade it has become increasingly more common to investigate fungal ecology by applying molecular methods. This is often done by extracting DNA from soil samples (O'Brien 2005; Kjøller 2006) or from root tip samples in the case of ectomycorrhizal communities (Nara 2006; Mühlmann and Peintner 2008; Tedersoo et al 2008). The DNA is then used for amplification and sequencing of shorter regions of the genome. When attempting to determine the taxonomic affiliation of the sequence, the internal transcribed spacers (ITS) region of the nuclear ribosomal DNA are the most commonly used parts of the genome (Horton and Bruns 2001). The taxonomic affinity of such a sequence is inferred in relation to sequences with a known taxonomic identity (Nilsson 2007). As most fungal taxa have not been sequenced (**I**), many sequences from community studies cannot be identified to species or even genus level (**I; II; III**). As part of the process of scientific publication all sequences are deposited in public nucleotide sequence databases such as GenBank. A consequence of the lack of precise taxonomic assignment is that the unidentified sequences are hard to search in a systematic context, and the information tied to them is therefore largely inaccessible (**II**).

The introduction of a new bioinformatics tool

The web service *emerencia* (www.emerencia.org; Figure 4; **I**) was designed to facilitate the exploration of sequences lacking a precise taxonomic annotation. *emerencia* downloads all the fungal ITS sequences in the International Nucleotide Sequence Databases (INSD: GenBank, European Molecular Biology Laboratory (EMBL), and DNA Database of Japan (DDBJ); Benson et al 2009) from GenBank. It then sorts out the sequences lacking a full species name (insufficiently identified sequence(s) (IIS), in **III** referred to as unidentified sequences) from those that have one (fully identified sequence(s) (FIS)) and stores them in separate tables in a local MySQL database.

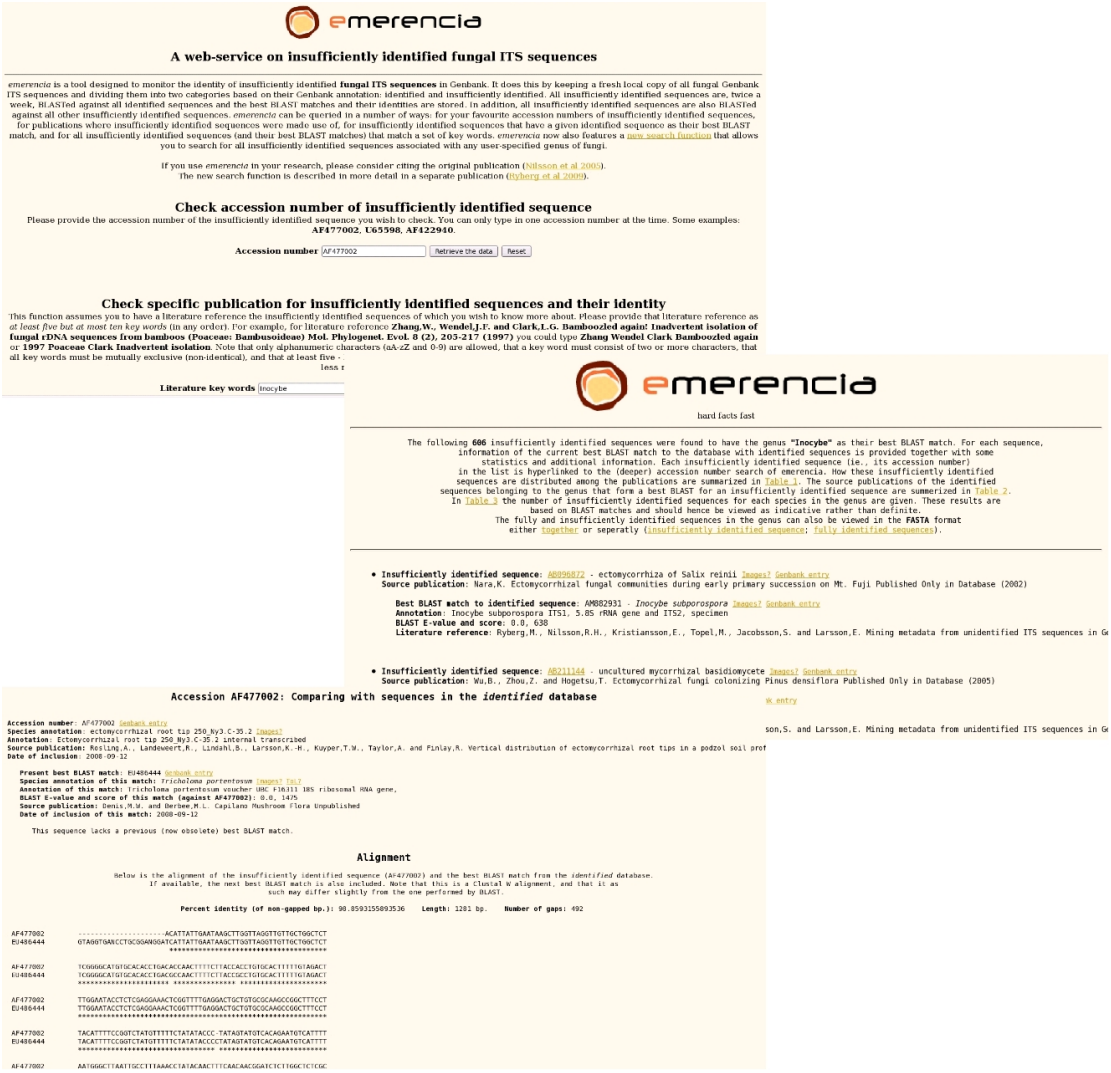

Figure 4. Screen shots from the *emerencia* web page (www.emerencia.org). Top picture is the start page where specific search functions can be selected. Middle picture is the top part of the output of a genus search. Bottom picture shows the top part of the output of a search on an individual insufficiently identified sequence.

BLAST (Altschul et al 1997) is used to find the FIS the most similar to each IIS and the result is stored in the database. Although BLAST results are not perfect substitutes for taxonomic proximity - especially as the taxonomic integrity of the FIS dataset is compromised by dubious annotations (Nilsson et al 2006) - it can serve as a rough estimate and gateway for further examination (Nilsson 2007). The resulting database opens up new possibilities to search the sequences in different ways compared to those offered through the interfaces associated with the INSD. One possibility is to search for IIS associated with a particular FIS (**I**) and as an extension of this it is now also possible to search for sequences associated witha particular genus (**II**).

What metadata is associated with the fungal ITS sequences?

To investigate the utility of the IIS and the FIS in providing ecological and geographical

data, the metadata associated with the sequences were compiled and analyzed (**II**). Both FIS and IIS are found on all continents, but North America and Western Europe stand out as more extensively sampled (**II**). In Asia, there are many ITS sequences from China and Japan but most other Asian countries have been poorly explored in this context. The IIS dataset forms a growing part of the fungal ITS sequences in INSD (**II**) covering a wide taxonomic scope. The highest number of IIS are available for the *Ascomycota* and the *Basidiomycota* with a bias towards parasitic and mycorrhizal genera (**II**; Table 1). In contrast to the taxonomic annotation, the IIS are often better annotated than the FIS regarding geographical and ecological metadata; 50% of the IIS have a country annotation while the same figure for the FIS is 37% (**II**). The IIS are also more often annotated with specific host (IIS: 38%, FIS: 18%) and isolation source (IIS: 55%, FIS: 10%; **II**). The IIS hold great potential in offering data on the diversity, geographical distribution, and the ecology of a genus (**II; III**). In **III,** *emerencia* is used to mine data from both the FIS and IIS datasets pertaining to the genus *Inocybe*.

Distribution and ecology of *Inocybe*: knowledge from GenBank metadata and correlation of ecological preferences with the phylogeny

Distribution

Matheny et al (2009) present support for the hypothesis that *Inocybaceae* originated in the Paleotropics during the Cretaceous era some 143 million years ago. Today the family is cosmopolitan and can be found on both hemispheres, in the old and new world, and from tropical to arctic/alpine climates (Kuyper 1986; Matheny et al 2009). These observations were expanded on in **III** where the distribution of IIS associated with *Inocybe* as seen through *emerencia* is correlated to the overall distribution of IIS (**II**). It was also shows that individual species can have a wide distribution (**III**).

Nutritional habit

Inocybe is generally considered to be ectomycorrhizal (Kuyper 1986). Mycorrhiza is an important symbiotic association between plants and fungi where the plants receive essential nutrients like nitrogen and phosphorous from the fungi and the fungi receive carbon compounds for its energy supply (Smith and Read 2008). As many of the stand forming trees in the temperate areas of the world are ectomycorrhizal this interaction constitutes a key element in the global carbon circle (Smith and Read 2008).

Although *Inocybe* species seldom dominate ectomycorrhizal communities (Horton and Bruns 2001), **II** and **III** indicate that they are often found in ecological studies based on root-tip sampling. **III** also suggests that *Inocybe* can form arbutoid and orchid mycorrhiza. Roy et al (2009) demonstrated that species of the genus *Inocybe* often are the dominating mycorrhizal partner of the mycoheterotrophic orchid *Epipogium aphyllum*. This rare orchid lacks chlorophyll and is therefore dependent on the mycorrhizal partner for its carbon supply.

Table 1. The 20 most common genera as represented by insufficiently identified sequences (IIS) in the International Nucleotide Sequence Databases as seen by the number of distinct studies they originate from and their generic affiliation as judged by BLAST. Data from http://www.emerencia.org/genuslist.html#bystudies, 2009-03-23.

Number of studies	Number of IIS	Genus
194	994	*Fusarium*
182	519	*Penicillium*
164	1216	*Tomentella*
147	1184	*Cortinarius*
142	886	*Russula*
119	761	*Phoma*
117	475	*Cladosporium*
112	499	*Sebacina*
111	606	*Inocybe*
110	1140	*Alternaria*
104	615	*Cryptococcus*
103	478	*Lactarius*
102	379	*Candida*
95	508	*Trichoderma*
92	238	*Aspergillus*
88	277	*Diaporthe*
88	397	*Hypocrea*
85	1607	*Mortierella*
84	195	*Leptodontidium*
80	226	*Tuber*

Host specificity

Species of *Inocybe* are typically associated with angiosperm hosts (Matheny et al 2009; **III**; **V**) and indeed **V** estimated this to be the most probable ancestral state of the family *Inocybaceae*. Known angiosperm host families for the genus *Inocybe* include the *Salicaceae*, *Betulaceae*, *Fagaceae*, *Rosaceae*, *Malvaceae*, *Cistaceae* (Kuyper 1986), *Myrtaceae*, *Phyllanthaceae, Fabaceae* (Matheny et al 2009), *Ericaceae,* and *Orchidaceae* (**III**). Even if association with angiosperm hosts is the dominating

condition, there are several species that associate with gymnosperms, mainly with the *Pinaceae* and *Cupressaceae* (Kuyper 1986). Host specificity can be observed at many different levels. For example a fungal species can be restricted to just one plant family or just one plant species or any other taxonomic level (Bruns et al 2002). Species within *Inocybe* are rarely restricted to one host species (Kuyper 1986; **III**) but associations at higher taxonomic levels have been suggested (Matheny et al 2009). At another level a species can have a preference for one host but may also associate with other hosts to a lesser extent. There are studies on the community level that show most ectomycorrhizal fungi to have such diffuse preferences (Ishida et al 2007; Tedersoo et al 2008).

V shows that when studied in a phylogenetic context in *Inocybaceae* it appears host association at the gymnosperm and angiosperm level is moderately evolutionary conserved. Host type is not fully reflective of the sectional classification of *Inocybe*, however, phylogenetically related species are more likely to be associated with the same type of hosts.

Soil preferences

There are many *Inocybe* species that are associated with rich soils and also many species associated with calcareous soils (**III**). The study of Tedersoo et al (2006), which was the study found to have most *Inocybe* species associated with it in **III**, was performed on a wooded meadow on calcareous soil. There are lineages such as the *lanuginosa* group that seem to be mainly associated with acidic and poor soil conditions. Both the preference for calcareous soil or not and the preference for rich or poor soils seem to be moderately evolutionarily conserved and the distribution of the traits among the species in the family is supported as being correlated with the phylogeny (**V**).

Not all traits are reflective of the phylogeny at the level explored in this thesis. When looking at the preference of soil moisture conditions, no support was found for a phylogenetic model when compared with a non-phylogenetic model (**V**). The species associated with dry soil conditions seem almost randomly distributed in the tree, and species associated with moist conditions seem to form a clade only rarely, one exception being the *lanuginosa* group.

Concluding discussion

As molecular studies of fungal diversity become increasingly common the mycological community is effectively building two largely separate sets of distributional and ecological information about fungi: one based on ITS sequence data and one based on morphological data. It is important to create and maintain links between these datasets to get an integrated picture of the biology of fungi.

In **III** data connected to sequences associated with *Inocybe* are explored and clustered using sequence similarity and phylogenetic methods. Although these clusters are not necessarily the same thing as species they are likely to be so in most cases. Even

if these putative species cannot be tied to a formal species description at the present, they are not without biologically relevant information, e.g. both *Inocybe* sp. 31 and *I.* sp. 2 (sensu **III**) seem to be species with a wide north temperate distribution. The fact that they are not directly linked to a species name may simply reflect the lack of FIS for many species (**I**) or it may be that they represent novel taxa. The ability to judge which of these two reasons is the most likely is largely dependent on the ability to place the putative species in a phylogenetic context and the completeness of the phylogenetic estimate regarding taxon sampling. If the taxonomy is in accordance with the evolutionary tree, phylogenetic placement will also help in giving an as exact as possible taxonomic annotation, which can facilitate scientific communication and comparison among studies. For this reason it is also alluring to assign standardized provisional names to such putative species clusters rather than using separate names in different papers (Thomas D. Bruns, MSA and FESIN talks, 2008, http://www.bio.utk.edu/fesin/FESIN2008/Talks_Sunday/Bruns2008.pdf).

As many morphological characters are reflective of the phylogeny (**V**), phylogenetic placement of an environmental sequence may be used to infer the character states for the specimen it originated from. This could help in finding matching fruiting bodies, which can then be used to generate sequences for further potential identification. As several ecological characters also are relatively conserved, and IIS are relatively well annotated regarding such characters (**II**), it is possible that IIS clusters can provide information for species where ecological characteristics are poorly known. It is likely that there are several other kinds of characters such as chemical content that can be linked between morphologically defined taxa and clusters circumscribed from sequence data. In these respects phylogenetic estimates represents much better frame-works for creation of informative links between organismal data based on morphology and sequences than e.g. plain sequence similarity. However, in order to assign any kind of probability to the predictions of character states we are reliant on explicit models of evolution (cf. Ronquist 2004). These models and their dependence on different parameters such as branch lengths need to be better explored (Ekman et al 2008; **V**). In *Inocybe* and several other taxa, there is also a problem linked to the use of ITS for phylogenetic placement as this region is difficult to align meaningfully over the entire family (**III**). Therefore the estimates of the phylogeny need to be better resolved to facilitate the formation of smaller groups for alignment, or possibly to facilitate the development of other procedures like supertree (Sanderson et al 1998) or supermatrix (de Queiroz and Gatesy 2007) methods. If connections between morphological data and sequence data cannot be built, valuable insights are sure to be missed and both the fields of molecular ecology and systematics are likely to suffer.

Acknowledgments

The road has been long and not entirely straight but I have not walked it alone. When I have needed to escape the university world I have always been able to go home to my parents. I have never needed to ask for their support as I always knew I have it (OK, it may be doubtful when it comes to moving to the US but I know I have it in that too). As my brother and sister let their families be also mine, I now, in addition to a sister and brother in law, also have three little boys that bring me joy. This has also made the most stressful times bearable.

When I started my academic career it was not with *Inocybe* in mind (I did not even know they existed) and it was not until Nils introduced me to mycology at "Växtvärdens mångfald och systematik" that I got a particular interest in fungi. Before that the inspired teachings of Thomas made me want to learn how to interpret nature and the history of a site based on what species and structures one finds there. The combination of these two interests lead me to do my "magister" work in a conservation biology project on wood-inhabiting fungi. Björn who supervised me during this work inspired and convinced me to try a career in science. So when the opportunity to pursue an PhD in a mycological project turned up I applied.

Ellen, from the beginning you have encouraged me to go my own way and never objected when I have been taking on side-projects. At the same time your have never declined me advice when I have asked for it. Thank you for introducing me to *Inocybe*, they are indeed the most interesting agarics. Stig you have been an invaluable source of knowledge on *Inocybe* and without your guidance I would never had been able to even start to grasp their diversity. Henrik you introduced me to bioinformatics, Linux, crazy bets, science citation indices (and all that comes with that), and you have never declined to read and correct my manuscripts. You also started the *emerencia* team and included Erik and I; the best science is done after midnight. I would also like to thank all others I have been working with these years, among these: Elisabet (thank you, also, for changing my curtains), Mats (the cluster buster), Ulf, Karl-Henrik (thanks also for editing **IV**), Åse, Nils, Urmas (and the rest of the Tartu crew), Björn, Frank, Mathias, Tobias, and Robert. Josephine thank you for being my native English speaker and all the Skyping. Tine, thank you for reading and commenting on this thesis. I gratefully acknowledge Bengt for being my "examinator" the second half of this work and for arranging interesting PhD courses. Thank also to the rest of the systematics group in Göteborg for science discussions and social gatherings and the rest of my colleges here at "Botan" for good times at "fika", lunches, and pubs. There are many more people that should been mentioned and for even more reasons, I hope you all forgive me for keeping this short. So last but not least I would like to thank _____.

Alex, Henrik, and Mats we made it through the dark years.

Literature

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25:** 3389-3402.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2009. GenBank. *Nucleic Acids Research* **36:** D26-D31.

Bruns TD, Bidartondo MI, Taylor DL. 2002. Host specificity in ectomycorrhizal communities: what do the exceptions tell us? *Integrative and Comparative Biology* **42:** 352–359.

Ekman S, Andersen H, Wedin M. 2008. The limitations of ancestral state reconstruction and the evolution of the ascus in the *Lecanorales* (lichenized *Ascomycota*). *Systematic Biology* **57:** 141-156.

de Queiroz A, Gatesy J. The supermatrix approach to systematics. *Trends in Ecology and Evolution* **22:** 34-41.

Groom MJ, Meffe GK, Carroll CR. 2006. *Principles of conservation biology,* 3ed edn. Sinauer Associates. Sunderland, USA.

Gulden G, Schumacher T. 1980. *Giftiga svampar och svampförgiftningar.* Lts förlag. Helsingborg, Sweden.

Hawksworth DL. 2001. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research* **105:** 1422-1432.

Horton TR, Bruns TD. 2001. The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. *Molecular Ecology* **10:** 1855-1871.

Ishida T.A., Nara K., Hogetsu T. 2007. Host effects on ectomycorrhizal fungal communities: insight from eight host species in mixed conifer–broadleaf forests. *New Phytologist* **174:** 430-440.

Jacobsson S. 2008. *Inocybe* (Fr.) Fr. pp. 868-906 in Knudsen H, Vesterholt J. eds. *Funga Nordica, Agaricoid, boletoid and cyphelloid genera*. Nordsvamp. Copenhagen, Denmark.

Judd WS, Cambell CS, Kellogg EA, Stevens PF, Donoghue MJ. 2002. *Plant systematics – a phylogenetic approach*, 2nd edn. Sinauer Associates, Inc. Massachusetts, USA.

Kirk PM, Cannon PF, Minter DW, Stalpers JA. 2008. *Dictionary of the Fungi.* 10th edn. CABI, Wallingford, UK.

Kjøller R. 2006. Disproportionate abundance between ectomycorrhizal root tips and their associated mycelia. *FEMS Microbiology Ecology* **58:** 214-224.

Kuyper TW. 1986. A revision of the genus *Inocybe* in Europe. I. Subgenus *Inosperma* and the smooth-spored species of subgenus *Inocybe*. *Persoonia* supplement volume **3:** 1-247.

Lee MSY. 2000. A worrying systematic decline. *Trends in Ecology and Evolution* **15:** 346.

Matheny PB. 2005. Improving phylogenetic inference of mushrooms with RPB1 and RPB2 nucleotide sequences (*Inocybe*; *Agaricales*). *Molecular Phylogenetics and Evolution* **35:** 1-20.

Matheny PB, Bougher NL. 2006. The new genus *Auritella* from Africa and Australia (*Inocybaceae*, *Agaricales*): molecular systematics, taxonomy and historical biogeography. *Mycological Progress* **5:** 2-17.

Matheny PB, Curtis JM, Hofstetter V, Aime MC, Moncalvo JM, Ge ZW, Yang ZL, Slot JC, Ammirati JF, Baroni TJ, Bougher NL, Hughes KW, Lodge DJ, Kerrigan RW, Seidl MT, Aanen DK, DeNitis M, Daniele GM, Desjardin DE, Kropp BR, Norvell LL, Parker A, Vellinga EC, Vilgalys R, Hibbett DS. 2006. Major clades of *Agaricales*: a multilocus phylogenetic overview. *Mycologia* **96:** 982-995.

Matheny PB, Aime MC, Bougher NL, Buyck B, Desjardin DE, Horak E, Kropp BR, Lodge DJ, Soytong K, Trappe JM, Hibbett DS. 2009. Out of Palaeotropics? Historical biogeography and diversification of the cosmopolitan ectomycorrhizal mushroom family *Inocybaceae*. *Journal of Biogeography* **36:** 577-592.

Mühlmann O, Peintner U. 2008. Ectomycorrhiza of *Kobresia myosuroides* at a primary successional glacier forefront. *Mycorrhiza* **18:** 355-362.

Nara K. 2006. Pioneer dwarf willow may facilitate tree succession by providing late colonizers with compatible ectomycorrhizal fungi in a primary successional volcanic desert. *New Phytologist* **171:** 187-198.

Nilsson RH. 2007. *Fungal taxonomy and systematics in the digital era, with an outlook on the cantharelloid clade (*Basidiomycota*).* PhD thesis University of Gothenburg, Department of Plant and Environmental Sciences.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Kõljalg U. 2006. Taxonomic reliability of DNA sequences in public databases: a fungal perspective. *PLoS ONE* **1:** e59.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R. 2005. Fungal community analysis by large scale sequencing of environmental samples. *Applied and Environmental Microbiology* **71:** 5544-5550.

Ronquist F. 2004. Bayesian inference of character evolution. *Trends in Ecology and Evolution* **19:** 475-481.

Ronquist F, Gärdenfors U. 2003. Taxonomy and biodiversity inventories: time to deliver. *Trends in Ecology and Evolution* **18:** 269-270.

Roy M, Yagame T, Yamato M, Iwase K, Heinz C, Faccio A, Bonfante P, Selosse M-A. 2009. Ectomycorrhizal *Inocybe* species associate with the mycoheterotrophic orchid *Epipogium aphyllum* but not its asexual propagules. *Annals of Botany* doi:10.1093/aob/mcn269.

Sanderson MJ, Purvis A, Henze C. 1998. Phylogenetic supertrees: Assembling the trees of life. *Trends in Ecology and Evolution* **13:** 105-109.

Singer R. 1986. *The Agaricales in modern taxonomy*, 4th edn. Koetlz Scientific Books

Koenigstein, Germany.

Smith SE, Read DJ. 2008. *Mycorrhizal Symbiosis,* 3ed edn. Academic Press. London, UK.

Stangl J. 1989. Die gattung *Inocybe* in Bayern. *Hopea* **46:** 5-388.

Stridvall L, Stridvall A, Jacobsson S. 1989. Släktet *Inocybe* i Sverige, en preliminär översikt. *Jordstjärnan* **10:** 29-76.

Tedersoo L, Suvi T, Larsson E, Kõljalg U. 2006. Diversity and community structure of ectomycorrhizal fungi in a wooded meadow. *Mycological Research* **110:** 734-748.

Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, Kõljalg U. 2008. Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. *New Phytologist* **180:** 479-490.

Svensk populärvetenskaplig sammanfattning.

Den här avhandlingen visar att man med hjälp av modärna bioinformatiska tekniker kan sortera och söka bland vetenskapligt insamlad data i internationella databaser på nya sätt för att få en inblick av en enskild svampgrupps ekologi och distribution. När dessa tekniker användes på svampsläket Inocybe *(trådingar) återfanns data från tropisk till alpin miljö och från så skilda växter som australiska eukalyptus till europeiska orkidéer. Det visas också att flera ekologiska karaktärer är relaterade till evolutionärt släktskap inom* Inocybe *men att den nuvarande namngivningen av grupper inom släktet inte fult ut stämmer med detta släktskap.*

Den här avhandlingen handlar om de evolutionära sambanden mellan olika arter i svampsläktet *Inocybe* (trådingar) och om hur väl namngivningen av grupper i släktet stämmer med det verkliga släktskapet. Avhandlingen tar också upp på vilket sätt olika karaktärer och ekologiska preferenser hänger samman med släktets evolutionära historia. För att undersöka ekologin inom släktet har moderna bioinformatiska metoder utvecklats och använts för att komplettera den bild man har från klassiska metoder som är baserade på fruktkroppsinventeringar.

Inocybe är ett stort släkte med fler än 150 arter i Sverige och med fler än 500 arter i hela världen. För att sortera upp och sätta in de olika arterna i ett sammanhang så grupperas de i så kallade taxa i ett hierarkiskt system. *Inocybe* är till exempel uppdelat i tre taxa på undersläktesnivån. Två av dessa är i sin tur uppdelade i fler taxa på sektionsnivå i det hierarkiska systemet. Det som oftast beskrivs med hjälp av dessa taxa är evolutionärt släktskap. Släktskapet har klassiskt sett härletts från morfologiska karaktärer, det vill säga formen av olika strukturer och vävnadstyper, och fokus har ofta varit på enstaka egenskaper.

De senaste decennierna har DNA-sekvenser kommit till allt större användning för att undersöka släktskap. DNA-sekvenser har fördelen att de kan ge stora mängder data relativt snabbt och lätt. Nackdelen är att de är relativt dyra och tekniskt komplicerade att ta fram. Det gör att mycket av det vardagliga taxonomiska (systematiserande och strukturerande) arbetet, som identifieringen av enskilda individer till art (eller annan taxonomisk rang), fortfarande i huvudsak görs för hand baserat på morfologin (utseende relaterade egenskaper). När man jämför en uppdelning av taxa gjord utifrån morfologi med ett släktskapsträd konstruerat från DNA-data så stämmer dessa inte alltid överens. Ofta måste de morfologiska karaktärerna omvärderas som ett led i uttolkandet av de grupper som DNA't visar på.

Inom *Inocybe* visar det sig att undersläktet *Inosperma,* såsom det hittills avgränsats, i själva verket innehåller arter från två evolutionärt sett separata linjer och att det därför bör delas upp. En sådan uppdelning kompliceras av att det är svårt att särskilja de två separata linjerna med hjälp av morfologi. Det blir därför svårt att till vardags - utan DNA-teknik - skilja de två nya grupperna åt.

Sektionerna inom undersläktet *Inocybe* stämmer också ofta dåligt överens med

släktskapsträdet och bör därför även de revideras. Här kompliceras situationen ytterligare av att det i många fall är svårt att exakt slå fast släktskapet mellan flera enskilda evolutionära linjer. Mer information behövs för att kunna skapa nya grupper som är både morfologiskt urskiljbara och som återspeglar den evolutionära historien.

Det vi ser av svamparna när vi är ute i naturen, fruktkropparna, utgör endast en liten del av den egentliga svampen och de visar sig bara tillfälligt när det är dags för sporspridning. Studier som bygger på fruktkroppar ger därför bara en begränsad bild av svampens liv och av vilka svampar som finns på en plats vid ett givet tillfälle. Man undersöker därför ofta svampsamhällen genom att extrahera DNA från jordprover eller växtmaterial som rötter. DNA-sekvenser från sådana prover jämförs sedan med sekvenser från fruktkroppar för att försöka koppla dem till ett artnamn. Men då endast en liten del av alla svampar finns vetenskapligt beskrivna och bara ett begränsat antal av dessa finns representerade som DNA-sekvenser så misslyckas man ofta med denna koppling och man får nöja sig med en mindre precis placering i det hierarkiska namngivningssystemet. Sekvenser som saknar artangivelse skickas ofta in till offentliga databaser som ett steg i dokumentationen av det vetenskapliga arbetet där de ingår. Tillsammans med sekvensen finns också ofta data om var de samlats och från vilken växt (eller annan värd eller substrat) de samlats. Denna information kan vara intressant när man söker kunskap om en speciell svampgrupp. Men eftersom sekvenserna ofta saknar ett namn som visar att de hör till den specifika gruppen är det svårt att ta tillvara informationen.

För att hjälpa till att hitta sekvenser med hjälp av sekvenslikhet så konstruerades webtjänsten *emerencia* (www.emerencia.org). Med hjälp av den går det att få fram sekvenser som är knutna till ett visst släkte. Dessa sekvenser kan sedan undersökas närmare. Denna metod har i arbetet med avhandlingen använts för att finna sekvenser av just *Inocybe*.

Arterna i *Inocybe* lever ett liv i symbios (nära association/beroende) med främst vedartade växter såsom gran, tall, björk, ek och bok. Men genom *emerencia* hittades även DNA-sekvenser som härstammar från örter som pyrola och olika orkidéer. DNA-sekvenser hittades även från så vitt skilda värdväxter och platser som granar i Sverige, ekar i Kalifornien, eukalyptus träd i Australien, vide i Japan och tropiska träd i Thailand.

Till sist undersöktes det även om det går att se någon korrelation mellan släktskap och vilken ekologisk preferens en art har. Det visade sig att det gick i tre av de fyra egenskaper som undersöktes nämligen, om arten föredrar kalkrik mark eller inte, om den föredrar fattig eller rik jordmån och med vilken typ av värdväxter den föredrar att bilda symbios. Detta betyder att det genom att placera en art i ett släktskapsträd går att uppskatta vilka levnadsförhållanden den troligast föredrar. Det innebär i sin tur också att om det går att förena den taxonomiska uppdelningen av *Inocybe* med de evolutionära släktskapförhållandena så kan taxonomin också användas för sådana förutsägelser.