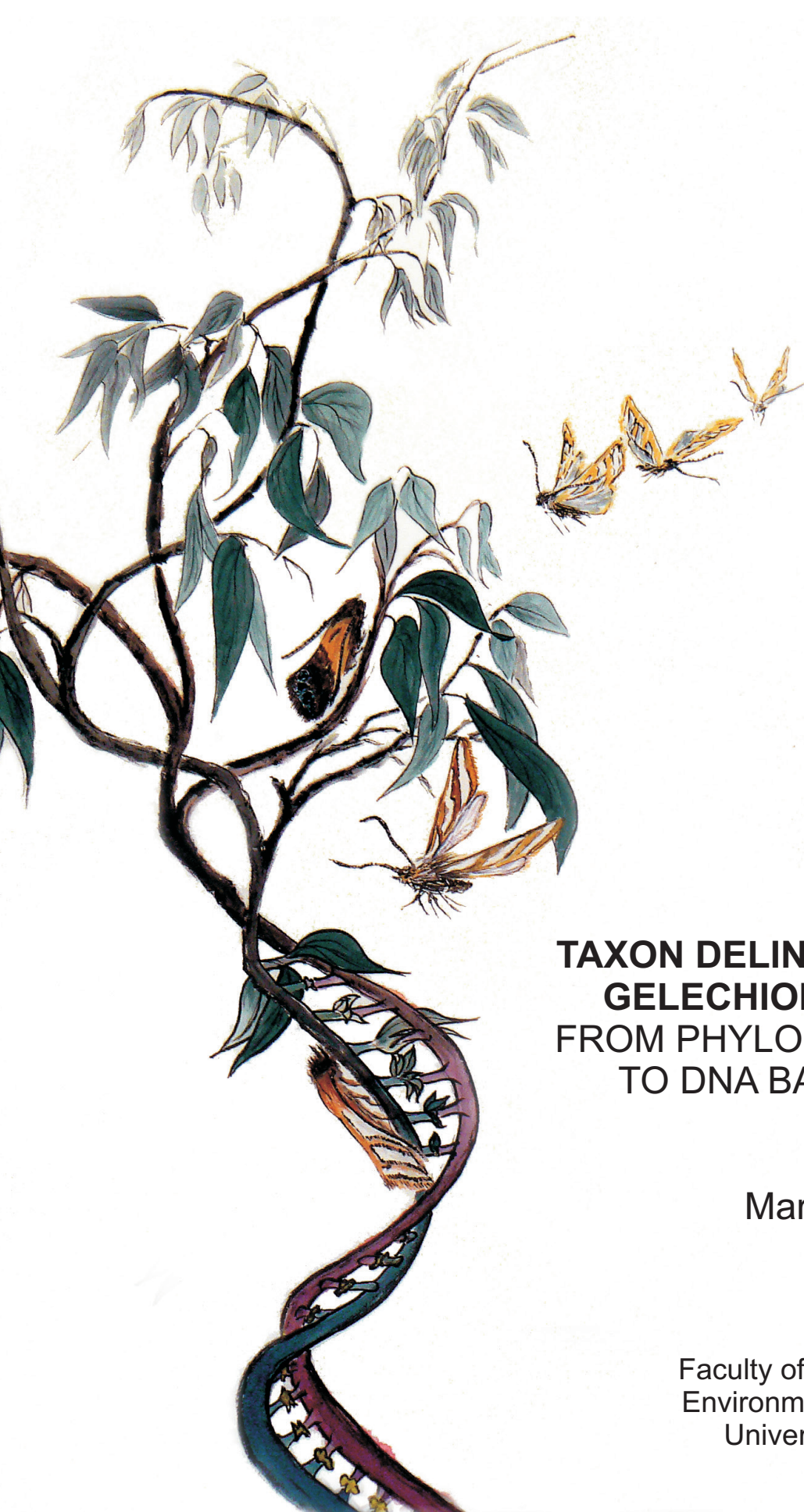




**TAXON DELINEATION IN
GELECHIOID MOTHS:
FROM PHYLOGENETICS
TO DNA BARCODING**

Mari Kekkonen

Faculty of Biological and
Environmental Sciences
University of Helsinki

TAXON DELINEATION IN GELECHIOID MOTHS:

FROM PHYLOGENETICS TO DNA BARCODING

Mari Kekkonen

Finnish Museum of Natural History
Zoology Unit
University of Helsinki
Finland

Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Finland

ACADEMIC DISSERTATION

To be presented for public examination
with the permission of the Faculty of Biological and Environmental Sciences of
the University of Helsinki in Auditorium (C205), Economicum (Arkadiankatu 7),
on May 16th 2014 at 12 noon.

Helsinki 2014

SUPERVISED BY:

Dr Lauri Kaila
Finnish Museum of Natural History, Finland

Dr Marko Mutanen
University of Oulu, Finland

REVIEWED BY:

Dr Michael Balke
Bavarian State Collection of Zoology, Germany

Dr Tommi Nyman
University of Eastern Finland, Finland

EXAMINED BY:

Dr Carlos Lopez-Vaamonde
French National Institute for Agricultural
Research, France

CUSTOS:

Prof. Jouni Laakso
University of Helsinki, Finland

MEMBERS OF THESIS ADVISORY COMMITTEE:

Prof. Jaakko Hyvönen
University of Helsinki, Finland

Prof. Jyrki Muona
Finnish Museum of Natural History, Finland

Author's address:

Finnish Museum of Natural History, Zoology Unit
P.O.Box 17, FI-00014, University of Helsinki,
Finland
e-mail: mari.kekkonen@helsinki.fi, m.e.kekkonen@gmail.com

ISBN 978-952-10-9894-9 (paperback)

ISBN 978-952-10-9895-6 (PDF)

<http://ethesis.helsinki.fi>

Painosalama Oy

Turku 2014

*Nothing in life is unrealistic
– just varying levels of challenge!*

CONTENTS

Abstract.....	6
1 Introduction.....	7
1.1 It is all about species	7
Box 1 Species concept	8
1.2 Fifty shades of species delineation	10
Box 2 DNA barcoding.....	11
1.3 Systematics – the discipline of evolution-based classification	13
1.4 Taxonomy and the need for efficient tools	14
1.5 The flowchart of taxonomy – right pieces in the right order	15
1.6 Gelechioidea – a world-wide variety of small moths	17
2 Outline of the thesis.....	19
3 Material & Methods	20
3.1 Sampling and the use of outgroup comparison.....	20
3.2 Character data 1: morphology, molecules or both?.....	21
3.3 Character data 2: single- or multi-locus DNA?	22
3.4 Tools for tree reconstruction	23
3.5 OTU delineation methods.....	25
3.5.1 General Mixed Yule-coalescent (GMYC)	26
3.5.2 Statistical parsimony networks (TCS)	28
3.5.3 Barcode Index Number (BIN)	28
3.5.4 Automatic Barcode Gap Discovery (ABGD)	29
3.6 Evaluating the results of OTU delineation	30
3.7 Associating short sequences from type specimens with OTUs	30
4 Results & Discussion	31
4.1 Combined data provides improved support in Gelechioidea.....	31
4.2 DNA barcode-based delineation of putative species.....	32
4.3 Congruence in resulting OTUs	33
4.4 Performance of delineation methods	37
4.5 Criteria for conflicting results.....	39
4.6 Associating types with OTUs – the challenge of short sequences	40
4.7 Contribution to taxonomical research.....	41
5 Conclusions & Prospects.....	42
Acknowledgements.....	44
References	46

This thesis is based on the following articles, which are referred to in the text by their Roman numerals.

- I. Heikkilä, M., Mutanen, M., **Kekkonen, M.** & Kaila, L. 2013. Morphology reinforces proposed molecular phylogenetic affinities: a revised classification for Gelechioidea (Lepidoptera). – *Cladistics*, early view. DOI: <http://dx.doi.org/10.1111/cla.12064>
- II. **Kekkonen, M.**, Mutanen, M., Kaila, L., Nieminen, M. & Hebert, P.D.N., Taxon-specific performance of DNA barcode-based species delineation methods. – *Manuscript*.
- III. **Kekkonen, M.** & Hebert, P.D.N., DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. – *Molecular Ecology Resources*, early view. DOI: <http://dx.doi.org/10.1111/1755-0998.12233>
- IV. Mutanen, M., **Kekkonen, M.**, Prosser, S.W.J., Hebert, P.D.N. & Kaila, L., One species in eight: DNA barcodes from type specimens resolve a taxonomic quagmire. – *Submitted manuscript*.

CONTRIBUTION TABLE

	I	II	III	IV
Original idea	LK, MM	MK	MK	MM, LK
Study design	LK, MH	MK	MK	MK , MM, LK
Data sampling	LK, MM, MH	LK, MM	MK	LK, MM, MK
Molecular data	MM	PDNH	PDNH	SWJP, PDNH
Morphology	LK, MH, MK	MN	-	LK
Analyses	MH, LK	MK , MN	MK	MK , MM, LK
Manuscript preparation	MH, LK, MM, MK	MK , MM, LK, MN, PDNH	MK , PDNH	MM, MK , LK, SWJP, PDNH

Mari Kekkonen (MK), Paul D.N. Hebert (PDNH), Maria Heikkilä (MH), Lauri Kaila (LK), Marko Mutanen (MM), Marko Nieminen (MN), Sean W.J. Prosser (SWJP)

Summary © Mari Kekkonen
 Chapter I © The Willi Hennig Society
 Chapters II, III & IV © The Authors
 Cover picture © Hannu Kekkonen a.k.a. Lào Yà

DISCLAIMER

This publication is not available for purposes of zoological nomenclature in accordance with the International Code of Zoological Nomenclature, fourth edition, article 8.3.

ABSTRACT

Systematics, phylogenetics and taxonomy are the scientific fields of species discovery, delimitation, description, classification and evolutionary history. The major task of these fields is to form meaningful groups, such as species and higher taxa (i.e. genera, families, orders), based on certain rules and characters. Species and higher taxa create the units of the Linnean hierarchic classification system, which is needed as the basis of all knowledge on biodiversity. Indeed, there is a great need for a complete classification, covering every species on earth, because all subsequent studies and applications are hindered as long as species remain undescribed. However, the estimated count of nine million species on earth, of which ca. 86% are currently unknown to science, together with the numerous threats to biodiversity pose a real challenge to taxonomy, and efficient tools and procedures are strongly needed.

In order to be both effective and high-quality, the taxonomic workflow needs to be divided into different steps in the correct sequence. A sampling scheme, and the choice of characters and analytical tools, for instance, are dependent on the phase performed along the workflow. The main aim of this doctoral thesis is to study and conduct different steps along this ‘taxonomic flowchart’ by using various gelechioid moths as focal species. Specifically, the thesis includes four parts: a phylogenetic examination of the superfamily Gelechioidea based on combined data of multi-locus DNA and morphology (**I**), and three studies focusing on delineation of putative species (operational taxonomic units, OTUs) based on single-locus DNA barcodes (i.e. the standardized 648bp region of the mitochondrial *COI* gene) (**II-IV**). The delineation studies cover various topics, from testing different OTU delineation methods (BIN, TCS, ABGD, GMYC) with reference species (**II**), developing criteria for discordant results (**III**) and a protocol for associating short sequences from type specimens to delineated groups (**IV**), to employing the methods, criteria and protocols in practice (**III, IV**).

Chapter **I** presents a phylogenetic hypothesis for the Gelechioidea with the best support for families to date. This was achieved mainly due to the use of both molecular and morphological data. We also provide a new family-level classification with redefined 16 families. The three OTU delineation studies revealed taxon-dependent performance within the four datasets (Finnish Gelechiinae and Australian Elachistinae, **II**; Australian Hypertrophinae, **III**; *Elachista dispunctella* complex, **IV**), but otherwise rather congruent results between the methods (with some exceptions regarding GMYC and ABGD; **II**). Chapter **III** provides 120 putative species for poorly-known hypertrophine moths and criteria for evaluating discordant delineation results. Finally, in **IV**, the *Elachista dispunctella* group with a high number of poorly-defined species is re-examined based on DNA barcodes from both fresh non-type and old type specimens.

This doctoral thesis is a part of a long-term study of Australian Hypertrophinae, aiming to provide well-supported phylogeny and species boundaries for the group. The results here not only enhance the study on hypertrophines, but also offer tools to benefit the taxonomic research in general. As a result of the great challenge of taxonomy to describe all species, every new innovation to speed up the workflows without compromising the quality is of crucial importance.

1 INTRODUCTION

Grouping is the simplest definition of this thesis' main idea: to form taxonomic groups called families (**I**) or putative species (**II-IV**), which are constructed by certain units, and delineated based on certain rules. Here, the units are subfamilies represented by, ideally several, species (**I**), or individual specimens (**II-IV**), respectively. But what makes such groups so important that one whole doctoral thesis is dedicated to their delineation?

1.1 IT IS ALL ABOUT SPECIES

A species is the basic unit in biology. As a consequence, the initial point for all biological research is the definition of species; their description based on diagnostic characters and naming with a unique binomial. A species' diagnosis is crucial for its identification, and the name serves as an identifier. Hence, taxonomy provides the foundation for all accumulating knowledge related to a particular species. Although many ecological and evolutionary studies are based on populations, every species' identity needs to be outlined prior to these examinations (Bortolus 2008). As a matter of fact, if studies are focused on species with faulty taxonomy, it might lead to a situation where none of the published papers can be connected with any particular species. Because species can be considered as the main units of evolution, studies lacking precise focal species are only telling the story of a group of species, and thus may be of limited significance for examining many evolutionary questions. Locke and Coates (2008) have estimated that one such a situation has led to losses of nearly four million dollars in the case of the coral *Madracis mirabilis*. Due to the importance of species description, undescribed species pose a considerable problem as they are excluded from most research and conservation actions (Mace 2004). Furthermore, unknown species can provide crucial benefits for humanity, such as ecosystem services (Ehrlich & Mooney 1983) and novel medicines (e.g. Chanda *et al.* 2013), but the use and management of these resources is hindered as long as species remain without specified boundaries and description. This is because all knowledge needs to be bound to something known, and a species' name and description forms an ideal basis for this.

Despite its importance, 'the species' is a difficult topic starting with a question: Are species real, or are they simply artificial groups formed for the sake of convenience? It has been said that even Darwin, the author of "The Origin of Species" (Darwin 1859), did not believe that the species exist (Wheeler 2012), although others claim that this argument is based on a misinterpretation (Coyne & Orr 2004, Barberousse & Samadi 2010). Nevertheless, the great number of varying species concepts and the sheer

Box 1. Species concept

The species concept is essentially a theoretical concept, aiming to define what a species is. As the literature surrounding different species concepts is both massive and diverse, it is reasonable to present only a small fraction of it here.

The most well-known species concept in biology is probably the Biological Species Concept (BSC) (Dobzhansky 1937, Mayr 1940, Wright 1940), which is based on potential for interbreeding and reproductive barriers between species. Although widely applied, the BSC has many drawbacks, which are especially clear with the strict interpretation of the concept. If no hybridization between sister species is allowed, a large fraction of species (e.g. many birds) cannot be considered as species anymore (Coyne & Orr 2004). By contrast, if some amount of hybridization is allowed, the next obvious question is “How much?”, and thus the lack of precise conditions would make the concept very subjective (Wheeler 2012). Due to its reliance on reproductive isolation, the concept is inapplicable for asexual organisms. Furthermore, relying on the BSC in the operational sense requires that sister species exist in sympatry (Coyne & Orr 2004) and thus the concept is unsuitable for allopatric and allochronic populations alike. Finally, being a process-based species concept, the BSC has little relevance to pattern-based systematics. The related Genotypic Cluster Species Concept (Mallet 1995) would offer a more suitable option for pattern-based fields as it defines a species as a “distinguishable group [...] that has few or no intermediates when in contact with other such clusters” (Mallet 1995). Genotypic clusters can be inferred from different traits or genetic data, but it shares the problem of allopatric populations with the BSC.

Lineage-based species concepts, the Evolutionary Species Concept (ESC; Simpson 1961, Wiley 1978, Wiley & Mayden 2000) and the Hennigian Species Concept (Hennig 1966, Meier & Willmann 2000), take the temporal aspect into consideration. Species are defined as “individuals with origins, existences, and ends” (Wiley & Mayden 2000), and especially the ESC has been claimed to be more universally applicable than the BSC as it has no restrictions for the mode of reproduction (Wiley 1978). The main difference between the Hennigian Species Concept and the ESC is that in the former, species have specific beginning and end points in splitting events, whereas the latter lacks such definition (Wheeler 2012).

The two main phylogenetic species concepts, Autapomorphic or Monophyletic Species Concept (Rosen 1979, Mishler & Donoghue 1982, Donoghue 1985, Mishler 1985) and the Diagnostic Species Concept (Nelson & Platnick 1981, Cracraft 1983, Nixon & Wheeler 1990), have both originated within the systematics community. The Autapomorphic Species Concept focuses on monophyletic groups defined by autapomorphic characters (i.e. unique to only one group within a phylogenetic analysis). Two issues are especially problematic with regard

to using monophyly as the definition of a species. According to some authors (Davis & Nixon 1992, Goldstein & DeSalle 2000, Goldstein *et al.* 2000), phylogenies are inappropriate for studying processes below species-level (i.e. tokogenetic relationships; Hennig 1966) as these are considered to be reticular instead of hierarchical (although maternally inherited mitochondrial DNA is an exception (Davis & Nixon 1992)), and the term monophyly should not be applied for species themselves (Nixon and Wheeler 1990). The second drawback of the concept is that gene trees and species trees are not necessarily identical and thus the same species can appear as monophyletic in one tree and paraphyletic in another depending on the loci used (Avice & Ball 1990). These topics are further discussed in sections 3.1. and 3.3.

In the Diagnostic Species Concept, species are defined on the presence of a diagnostic (i.e. unique) character or a set of characters shared by all members of a particular species and lacking from others. A species is “a smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent” (Cracraft 1983). This concept is more theoretically sound with the principles of cladistics than the Autapomorphic Species Concept, and it also allows species to be paraphyletic. However, the Diagnostic Species Concept suffers from being typological, as do most other concepts to some extent, because any trait can in principle serve as a diagnostic one (either apomorphic or not).

The most inclusive concept is probably the General Lineage Concept (de Queiroz 1998, 1999). It aims to unify all previous species concepts under one general concept, and defines species as “separately evolving (segments of) metapopulation lineages” (de Queiroz 2005a). The requirements of other concepts are regarded as contingent properties. This follows from the notion that most species concepts define species of various ages, and they become congruent given enough time (de Queiroz 1998, 2005b). For instance, the complete reproductive isolation and monophyly are evolving rather slowly (Avice & Ball 1990, Coyne & Orr 2004) and thus species expressing these traits have diverged a long time ago.

existence of biology indicate that many do believe that species are real, but the problem is how to define them. Argumentation surrounding this topic has been going on for decades (e.g. Wheeler & Meier 2000), producing a wide variety of species concepts (Box 1). Achieving one commonly accepted concept seems to be too great a challenge for the academic community, and it might simply be impossible to fit the vast variation of different patterns existing in nature into one species concept. Therefore, some authors suggest keeping ‘species’ as a theoretical concept, and concentrating on testing different components of this concept (Barraclough *et al.* 2009b). Yet, with

the lack of a universal species concept, the comparison of different studies can be difficult if the used species concept varies without explicit mention of the concept applied (Wheeler 2012).

All other taxa than species (e.g. subspecies, genera, families, orders) are arbitrary and thus incomparable between studies unless they are sister-groups, which are coeval by definition (Wheeler 2012). Because the total number of species is massive (one of the latest estimates resulting in ca. 8.7 million (Mora *et al.* 2011), although most other estimates range between 10 and 100 millions of species (Blaxter 2004, Savolainen *et al.* 2005)), a Linnean hierarchical system including higher taxa is needed to understand the big picture of biodiversity. To do so, we need to know the basics: which species are related to each other, in what way, and how closely. Also, taxa below the species rank provide useful information on diversity and they are sometimes elevated to species after additional investigations (e.g. Mallet 2009).

1.2 FIFTY SHADES OF SPECIES DELINEATION

Species delineation is currently a hot topic with many novel tools published in recent years (Sites & Marshall 2003, 2004, Pons *et al.* 2006, Knowles & Carstens 2007, Monaghan *et al.* 2009, Yang & Rannala 2010, Hao *et al.* 2011, Jones *et al.* 2011, Puillandre *et al.* 2012, Ratnasingham & Hebert 2013, Zhang *et al.* 2013). Despite its presumable simplicity, the topic is most complex. As noted above, the species concept has been a source of ongoing arguments, and this has inevitable consequences for the operational side of the same issue – species delineation. Most delineation studies focus on a small number of species, usually belonging to a challenging species complex in a larger group that is sufficiently well known (Roe & Sperling 2007, Silva-Brandão *et al.* 2008, Dinca *et al.* 2011, Reilly *et al.* 2012, Edwards & Knowles 2014). These studies usually employ data from several sources, such as multi-locus DNA from both mitochondrial and nuclear genomes, and morphometrics. In addition, the number of specimens per species is generally high as these examinations aim to maximize intraspecific variation in the sample, although such maximizing might raise other problems as noted by Hey (2009)

In poorly-known groups with a high number of undescribed species, such detailed studies are unfeasible, and rougher methods are needed for discovering boundaries between putative species. DNA barcodes (a standardized 648bp region of the mitochondrial *COI* gene; Box 2) are suitable for species delineation for groups including tens of species (II-IV), due to their suitable rate of change for delineation of most species (Hebert 2003a, b), and easier amplification than in many other loci resulting from the high number of copies in each cell. However, as explained below (section 3.3), relying on mitochondrial single-locus DNA poses certain drawbacks,

Box 2. DNA barcoding

DNA barcoding (Hebert *et al.* 2003a, b) is based on a standardized region of mitochondrial genome (648bp of cytochrome *c* oxidase subunit one gene, *COI*). Although using DNA for the identification of species was not a novel innovation back in 2003 (Meier 2008), the idea of employing the same, relatively short fragment for all species on Earth was unique (Stoeckle & Hebert 2008). The main purpose of barcoding is the identification of previously described species by comparing a barcode sequence from a specimen to the extensive reference database. Of the two largest web-based databases, the Barcode of Life Data System (BOLD; Ratnasingham & Hebert 2007) is exclusively focused on the barcode regions, and includes more than three million sequences to date. In order to make this identification tool as efficient and well-known as possible, a major global consortium (International Barcode of Life, iBOL) is promoting the construction of barcode libraries (e.g. aphids: Footitt *et al.* 2008; North American Lepidoptera: Hebert *et al.* 2010; Bavarian Lepidoptera: Hausmann *et al.* 2011). The barcode region includes enough information to provide a correct match even when only a small proportion of the whole barcode is available (Hajibabaei *et al.* 2006, Meusnier *et al.* 2008, Shokralla *et al.* 2011). For instance, a 93bp sequence from an old type specimen provides a 100% match with its corresponding species when searched against over two million barcodes in the BOLD Identification System (unpublished data).

In the distribution of pairwise distances between all sequences of a given dataset, a gap is usually present between intra- and interspecific diversity. This gap has been named as the ‘barcode gap’ and it is a crucial factor for the successful use of DNA barcoding (Hebert *et al.* 2003b). The barcode gap has been in the center of a heated debate for several years (Hebert *et al.* 2004b, Meyer & Paulay 2005, Astrin *et al.* 2006, Meier *et al.* 2006, Dalebout *et al.* 2007, Wiemers & Fiedler 2007, Lahaye *et al.* 2008, Kerr *et al.* 2009, Astrin *et al.* 2012). While some studies have resulted with clear gaps for the majority of the species (Hogg & Hebert 2004, Barrett & Hebert 2005, Monaghan *et al.* 2005, Vences *et al.* 2005, Astrin *et al.* 2006, Hajibabaei *et al.* 2006, Mikkelsen *et al.* 2007, Hebert *et al.* 2010), some authors have reported weak differences between intra- and interspecies distances or even a shortage of the whole gap (Meyer & Paulay 2005, Kaila & Ståhls 2006, Meier *et al.* 2006, 2008, Elias *et al.* 2007, Wiemers & Fiedler 2007, Astrin *et al.* 2012).

As an identification tool, DNA barcoding has a vast number of potential applications. It has already been used to discover food frauds in various commercial products (e.g. herbal teas (Stoeckle *et al.* 2011) and fish (Hanner *et al.* 2011)). In addition, barcoding is a valuable tool for customs officers tracing illegal trade of endangered species (Luo *et al.* 2014), and the police for criminal forensics (Meiklejohn *et al.* 2011). Because DNA barcoding is a simple technique, it is very suitable for educational

purposes in various science projects (e.g. School Malaise Trap Project (<http://malaiseprogram.ca/>), high school student project in Coastal Marine Biolabs (Santschi *et al.* 2013)). DNA barcoding has also been used as a part of ecological studies. For example, the high discrimination power has provided insights into food webs (Kaartinen *et al.* 2010, Wirta *et al.* 2014). Next generation sequencing-based metabarcoding leads ‘traditional’ barcoding to the next level with its increased effectiveness in data production, and enables exploration of many new areas, such as gut contents (Soininen *et al.* 2013), aquatic and soil ecosystems (Thomsen *et al.* 2012, Yoccoz *et al.* 2012, Schmidt *et al.* 2013), ancient DNA (Parducci *et al.* 2013) and global scale biodiversity assessments.

The relationship between DNA barcoding and taxonomy has certainly been two-sided. Barcoding has become a very efficient tool for taxonomists due to its multiple uses. For example, it has been used for discovering hidden diversity (i.e. cryptic species) (Hebert *et al.* 2004a), and clarifying challenging species complexes (Huemer & Mutanen 2012, Mutanen *et al.* 2013). It also enables identification of broken specimens, tissue samples, and various life stages where appropriate morphological characters are lacking (Ahrens *et al.* 2007), and provides an additional tool for examining type specimens (Puillandre *et al.* 2011, **IV**). Despite these benefits, the strongest opponents of DNA barcoding can probably be found among taxonomists.

The vast majority of the critique can be divided into two categories: (i) drawbacks of barcoding, and (ii) shared problems between barcoding and taxonomy. The following drawbacks are mainly concentrated on the barcoding of animals. The work is more complicated with fungi and especially plants where *COI* is not evolving at a sufficient speed, and thus other standard regions have been applied (CBOL Plant Working Group 2009, Hollingsworth 2011, Schoch *et al.* 2012). The first category includes issues related to the use of single-locus data (Dupuis *et al.* 2012), mitochondrial DNA (Moritz & Cicero 2004), Neighbor-Joining trees (Will & Ruinoff 2004, Prendini 2005), the K2P model (Srivathsan & Meier 2012), and strict thresholds for separating intra- and interspecific variation (DeSalle *et al.* 2005, Mallet *et al.* 2005, Meier *et al.* 2008). The drawbacks of single-locus mitochondrial DNA data are indisputable, and they limit the capacity of DNA barcoding with very young focal species, although the scale of this problem is controversial (e.g. Meier 2008, Stoeckle & Hebert 2008). The potential pitfalls of Neighbor-Joining trees and the K2P model are easily avoidable with the use of alternative methods (e.g. Bayesian or maximum likelihood trees) and substitution models (e.g. GTR), although the drawbacks of NJ trees are rather minor on species-level studies (see further discussion in section 3.4) and the effect of the K2P model is controversial (Collins *et al.* 2012). Furthermore, novel species delineation methods do not rely on strict thresholds, removing the problems related to their use. These issues are further discussed in section 2, and **II-IV**.

The second category of critique includes themes shared between DNA barcoding and taxonomy. Barcoding is claimed to produce false identifications due to references based on misidentified specimens in museum collections and on the large number of undescribed species (Meier & Dikow 2004, Quicke 2004, Seberg 2004). It is evident that a large proportion of the reference library is based on undescribed species. For instance, the BOLD database currently includes 323 891 BIN clusters for animals, but only 141 320 formally described species (site <http://www.boldsystems.org> visited on 4 February 2014). However, it is unreasonable to blame barcoding for this, because the task to provide descriptions belongs to the taxonomic community. In general, the relationship between DNA barcoding and taxonomy should be considered as mutually beneficial, instead of parasitic, because both have the capacity to enhance the other.

and thus delineated entities should be treated as putative species (referred here as operational taxonomic units, OTUs). Instead of ‘final’ species ready to be described, they serve better as a framework for future studies where they will be validated, or optionally, redefined.

1.3 SYSTEMATICS – THE DISCIPLINE OF EVOLUTION-BASED CLASSIFICATION

Systematics, phylogenetics and taxonomy, the scientific fields of species discovery, delimitation, description, classification and evolutionary histories, have a great responsibility to provide the basic level of knowledge of all organisms on earth. Systematics can be considered as a research of historical events, because it aims to elucidate past species genealogy based on present-day species. Due to these temporal constraints, systematics is a comparative rather than an experimental field of science.

Especially after the introduction of cladistics, developed by Hennig (1966), the reconstruction of the evolutionary history of species (i.e. phylogeny) has been the main focus of systematics (Farris 1983). Phylogeny is usually reconstructed based on discrete characters, which can be morphological, molecular or any other biological, like behavioral, traits. According to the concept of ‘total-evidence’, data from various sources should be pooled and analyzed simultaneously as this approach reveals the global agreement among data (Kluge 1989, Kluge & Wolf 1993). The crucial issue is to separate characters derived from the same origin (i.e. homologies) from

characters developed by convergent evolution (i.e. homoplasies). In essence, the different status between the characters at different levels can be revealed only a posteriori as they are defined based on their positions in a resulting tree. In cladistics, relationships between species are genealogical, and they are expressed as nested sister-groups in phylogenetic trees (Wheeler 2012). A special emphasis is placed on the shared derived attributes of characters (i.e. synapomorphies) for establishing sister-groups. Another cladistics-based criterion, following from the previous criteria, is to form ‘natural’ groups, that is to delineate higher taxa based on monophyly (Farris 1974, Wheeler 2012). A monophyletic group, characterized by synapomorphies, can be defined as including all descendants of a common ancestor and the common ancestor itself, whereas a paraphyletic group contains a common ancestor and some, but not all, of its descendants and is defined by ancestral characters (i.e. plesiomorphies) (Farris 1974). Finally, polyphyly is the inclusion of some descendants of a common ancestor, but the ancestor is excluded. These groups are characterized by homoplasies. The information content of a phylogeny can be transformed into a formal classification, aiding its subsequent use (Lipscomb 1998).

1.4 TAXONOMY AND THE NEED FOR EFFICIENT TOOLS

According to Mora *et al.* (2011), 86% of terrestrial species are awaiting description. Although the need for high quality taxonomic research is obvious, the reality has been different. The ‘taxonomic impediment’ has received a lot of attention, and the related discussion was especially heated ten years ago, partially boosted by novel ideas of DNA barcoding (Hebert *et al.* 2003a, b; Box 2) and DNA taxonomy (Tautz *et al.* 2002, 2003) (e.g. Mallet & Willmott 2003, Godfray & Knapp 2004, Janzen 2004, Lee 2004, Lipscomb *et al.* 2003, Wheeler *et al.* 2004, Will & Rubinoff 2004, Ebach & Holdrege 2005, Gregory 2005, Hebert & Gregory 2005, Schindel & Miller 2005, Will *et al.* 2005). The main difference between DNA barcoding and DNA taxonomy is that the former was originally developed as an identification tool, whereas the latter emphasized the role of DNA as a basis of taxonomy. The key concerns of the taxonomic impediment related to the shortage of appreciation for taxonomy, leading to the scarcity of both funding and young people entering the field. The opponents of DNA barcoding were concerned that a ‘fashionable’ new technique would scoop all available funding from already restricted resources (Will *et al.* 2005). As a whole, taxonomy has suffered from the lack of novel innovations, which could enhance the workflows and turn the field into a flourishing one.

Since then, the arguments have gradually faded despite some occasional critiques (e.g. Ebach 2011), and especially DNA barcoding seems to have found its own niche as a versatile tool for varying uses (Box 2). But what

about taxonomy? Has it grown into its new glory? Although much improvement has taken place due to many global and national research programs (e.g. United Nations' Global Taxonomy Initiative (GTI; www.cbd.int/gti/), American Partnerships for Enhancing Expertise in Taxonomy (PEET; Rodman & Cody 2003), Swedish Taxonomy Initiative (STI; Ronquist & Gärdenfors 2003), Finnish Research Program of Deficiently Known and Threatened Forest Species (PUTTE; Juslén *et al.* 2008), the pace of change is too slow in the face of strong global threats to biodiversity, such as climate change and habitat destruction (Ehrlich & Wilson 1998, Lane 2009, Li 2009).

1.5 THE FLOWCHART OF TAXONOMY – RIGHT PIECES IN THE RIGHT ORDER

Taxonomical work has its own steps and a reasonable order for them. Consider a poorly-known subfamily of moths including probably tens of undescribed species. The mission would be to form an evolution-based classification for the group, and study all species, redefine already described ones and provide descriptions for newly discovered species. Naturally, the first question is where to start from? To be able to sample a dataset, one must know which specimens belong to the focal subfamily. This knowledge is only available from superfamily- or family-level phylogenies, i.e. from the phylogenetic hypothesis of a higher taxonomic level than the group under study (Fig. 1.a).

After clarifying the boundaries of the focal subfamily, another phylogenetic analysis should take place for creating a genus-level classification following the corresponding genealogy (Fig. 1.c). However, in order to draw samples from species belonging to the particular subfamily, (putative) species boundaries need to be known. Traditional grouping based on morphology (i.e. 'morpho grouping') has its use, but when the focal subfamily includes tens of species, the task becomes exhaustive. Such large-scale sampling based on morphology has been conducted (Tänzler *et al.* 2012), but previous knowledge on the group under study is essential in these cases. Especially di- and polymorphic species and species lacking clear external characters are extremely challenging (Krell 2004, **II**). The above-mentioned DNA barcode-based delineation of putative species (OTUs) provides a well-suited option here (Fig. 1.b).

The last phase includes the hard core of taxonomy where OTUs are examined for distinguishing (i.e. diagnostic) characters and redefined if needed (Fig. 1.d). The examination should be based on all available data (e.g. morphology, DNA barcodes, other DNA loci, phenology, ranges), following the concept of 'integrative taxonomy' (Dayrat 2005). After the validation of putative species to either well-supported species worthy of description or ambiguous entities in need of further investigation, each new species must be

named. The major question here is whether a given species already possesses a taxonomic binomial. Every described species is defined based on a type specimen (or several), and thus the species name is also bound with types. As a consequence, linking types to the corresponding species is the only way to know whether the species already has a name. There are many ways to do this. The dissection of genitalia is the most common approach with small lepidopterans, but occasionally unfeasible (IV). DNA barcodes provide a novel way to associate types with species even when the type specimens are in poor condition. In addition to the actual type specimens, original descriptions provide crucial information such as diagnostic characters and type localities where type specimens have been collected.

Although the flowchart presents only four steps along the taxonomic work process, it should be emphasized that delineated, described and named species are not final facts (Sites & Crandall 1997, Barberousse & Samadi 2010). Instead, they are hypotheses, which may be either supported or rejected in further investigations with new data and/or criteria (Hey *et al.* 2003), and thus the work does not end after the four steps presented here.

Similar flowcharts of taxonomic work have been presented in earlier publications. Puillandre *et al.* (2012b) presented a very inclusive protocol for marine molluscs, where they first defined primary species hypotheses based on DNA barcodes, and then evaluated them using various tools (Klee diagrams; Sirovich *et al.* 2009) and sources of data (morphology, an additional locus, bathymetry, gene flow and monophyly). The approach of Puillandre *et al.* (2012b) can be considered to cover the second and the

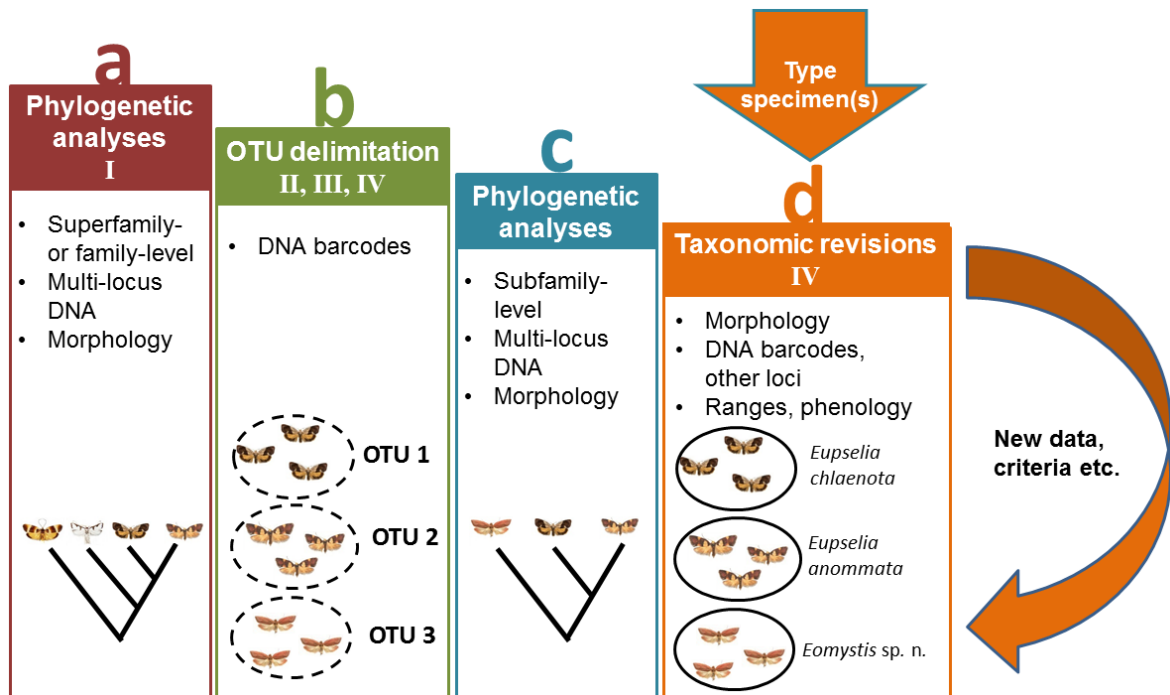


Fig. 1 The flowchart of taxonomy. Roman numerals refer to the four chapters.

fourth steps presented in this section (Fig. 1b,d). A different structure of the flowchart is presented here, because it is reasonable to keep initial OTU delineation and subsequent validation as separate steps requiring different sampling schemes and analyses. In addition, keeping each step disconnected aids their repetitiveness as only one dataset is used at a time. A corresponding workflow was conducted with weevils (Riedel *et al.* 2013b), covering species delineation based on morphology and DNA barcodes (Tänzler *et al.* 2012), and descriptions of 101 species (Riedel *et al.* 2013a) in the ‘turbotaxonomic’ framework (Butcher *et al.* 2012). The main difference between the approach of Riedel *et al.* (2013b) and the one presented here (III) is the use of morphology: in the former, morphology is employed as a null hypothesis of putative species, whereas here, initial OTUs are based on DNA barcodes (although external morphology and previous taxonomic assignments for museum specimens affected sampling to some extent).

1.6 GELECHIOIDEA – A WORLD-WIDE VARIETY OF SMALL MOTHS

The Gelechioidea (I) are one of the largest lepidopteran radiations globally, but also one of the most poorly-known groups (Hodges 1998, Kaila 2004, Kaila *et al.* 2011). The superfamily Gelechioidea includes ca. 18 000 species, but Hodges (1998) has estimated that the current species count could cover only one quarter of the actual number. In addition to its great diversity, the superfamily is a significant group in many senses. It includes several serious pest species, such as the tomato leafminer (*Tuta absoluta*) and the pink bollworm (*Pectinophora gossypiella*), and, contrastingly, many threatened species. For instance, one third of endangered lepidopterans in Finland consist of gelechioid moths (Rassi *et al.* 2010). Furthermore, gelechioids are an important part of many ecosystems. In Australia, lepidopteran fauna is dominated by gelechioids, which include one third of all Australian moths and butterflies (Nielsen *et al.* 1996), and oecophorids are the largest family acting as the main decomposers of *Eucalyptus* leaf litter (Common 1990).

The lack of consensus of family-level classification has hindered both basic and applied studies of Gelechioidea. Although the Gelechioidea have previously been included in the primitive the most basally diverged Ditrysian Lepidoptera close to the Yponomeutoidea (Minet 1991, Hodges 1998) recent evidence suggests that the superfamily belongs to the Apoditrysia (Kaila 2004, Mutanen *et al.* 2010, Cho *et al.*, 2011, Regier *et al.* 2012, 2013) or even to the Obectomera together with pyralids, butterflies and ‘macromoths’ (Regier *et al.* 2013). The monophyly of the Gelechioidea appears to be well-supported (Kaila 2004, Mutanen *et al.* 2010, Cho *et al.*, 2011, Regier *et al.* 2012, 2013).

Gelechioids are rather well-known in Northern Europe (Huemer & Karsholt 1999, 2010; **II**) due to long traditions and the high number of professionals and amateurs alike, but the situation is the opposite in the Southern Hemisphere. Apart from the Oecophoridae (Hodges (1998) optimistically estimated ca. 50% as described), the other Australian gelechioid species are very poorly known. Kaila's recent work (2011) on the Elachistinae (**II**) has offered some progress, but most other families still remain virtually unstudied. Hypertrophinae (Gelechioidea: Depressariidae) (**III**) is an endemic Australian subfamily with 51 described and numerous undescribed species (Common 1996). Not much is known about these exceptionally brightly colored moths, but many are known to live in arid and semi-arid environments and the majority of their recognized food plants belong to the genus *Eucalyptus* (Common 1990).

As many gelechioid species are small and drably colored, their classification can be challenging. A good example of this is the European *Elachista dispunctella-triseriatella* complex (Gelechioidea: Elachistidae; from here on referred to as the *Elachista dispunctella* complex) (**IV**) where poorly-justified splitting based on minute differences in wing venation has resulted in numerous named species lacking appropriate diagnostic characters. The complex included 11 species until the revisionary works of Traugott-Olsen (1988, 1992), when the count rose to 64, causing problems for subsequent studies.

2 OUTLINE OF THE THESIS

The main aims of this doctoral thesis can be divided into two parts: the first is to produce a phylogeny-based classification for the superfamily Gelechioidea and the second is to test, develop and use different approaches for the delineation of putative species. The four studies represent different parts of the above-mentioned flowchart (see Roman numerals in Fig. 1) and thus can be seen as procedures complementary to each other. These approaches also serve as useful tools for the long-standing taxonomic work on Australian hypertrophines. In detail, the aims can be specified as follows:

- (i) To reconstruct a total-evidence phylogeny for the Gelechioidea based on thorough taxon sampling and of both morphological and molecular data (**I**).
- (ii) To provide a revised family-level classification for the Gelechioidea (**I**).
- (iii) To compare the performance of five OTU delineation methods (one based on external morphology and four based on DNA barcodes) using two reference groups of gelechioid moths (Finnish Gelechiinae, Australian Elachistinae) (**II**).
- (iv) To develop an approach for OTU delineation in poorly-known groups and to use it for Australian hypertrophines (**III**).
- (v) To develop and use a procedure for linking type specimens to OTUs in the *Elachista dispunctella* complex (**IV**).

3 MATERIAL & METHODS

3.1 SAMPLING AND THE USE OF OUTGROUP COMPARISON

A sampling scheme is different for taxon delimitation at family- and species-levels. Phylogeny reconstruction relies on exemplars (Yeates 1995), i.e. one species represents each terminal taxon. A species is represented by one specimen (or a few if characters are coded from different sexes and/or developmental stages) instead of pooling character states over a larger amount of individuals due to difficulties in species definitions (Vrana & Wheeler 1992). By contrast, the delimitation of (putative) species necessitates multiple specimens per species to be included in the analyses (Hey 2009).

Another difference between studies on higher taxa and species is the need for an outgroup. The outgroup is a taxon or several taxa sampled in the same way as the ingroup (i.e. the taxa of interest), but does not belong to the focal group (Nixon & Carpenter 1993). Ideally, it represents the closest relative of the group under study. Using outgroup taxa too distantly related to the ingroup may lead to topological errors due to ‘long branch attraction’ (Bergsten 2005) or saturation of informative sites in DNA sequences (Vandamme 2009). The crucial point is to choose a taxon outside the focal group, but the uncertainty of the closest relative usually leads to the use of several outgroup taxa. An outgroup is used for rooting a phylogenetic tree, and as the root indicates the direction of the evolutionary process, it has the central role in defining which character states are ancestral (plesiomorphic) and which are derived (apomorphic) (Nixon & Carpenter 1993). As noted above (section 1.3), synapomorphies define monophyletic groups. However, in the delineation of (putative) species, the role of the outgroup is less relevant. Some authors consider phylogenetic methods inappropriate for studying species (see Box 1), but many species delineation studies rely on the tree-based methods (section 3.5). The key point here is to understand the difference between phylogenetic and species delineation analyses (e.g. Velasco 2009). The same tools are used for both purposes, but the interpretation of the result is different. The tree-based methods delineate clusters, which correspond to monophyletic groups, although they are not characterized by synapomorphies. Whether the used tree is rooted or not, makes no difference. As a consequence, some authors advocate the use of another term instead of monophyletic, such as ‘exclusive’ (Wiens & Penkrot 2002). In **II** and **III**, the term ‘monophyly’ has been adopted in this ‘non-phylogenetic’ sense.

Our phylogenetic study of the Gelechioidea (**I**) included 184 species from 35 of the 37 subfamilies defined by Hodges (1998). As the identity of the

closest relatives of gelechioids are uncertain, three outgroup taxa were used (*Cossus cossus* (Cossioidea), *Thyris fenestrella* (Thyridoidea), and *Epermenia illigerella* (Epermenioidea)), but due to the instability of the root, analyses were also executed without outgroups.

In the species delineation studies (II–IV), a varying number of specimens per species were sampled. The lowest sampling effort was in II in which approximately three to five specimens per species were included. This low a number was a consequence of a sampling scheme aiming to maximize the species coverage of the DNA barcode library being constructed. By contrast, the highest specimens per species count was achieved in IV, where 215 non-type specimens and 42 type specimens were sampled and sequenced corresponding to either 11 or 64 species depending on the source (see section 1.6). The highest total number of included specimens was in III (originally 864 specimens), but due to the poorly-known taxonomy of the focal group, the sampling was inevitably uneven among the included species.

3.2 CHARACTER DATA 1: MORPHOLOGY, MOLECULES OR BOTH?

Morphological characters represent the traditional type of data in phylogenetic and taxonomic studies, but they have been strongly challenged by DNA since the 1980s. As long as specimens are available, the information contained in their morphological structures is also available, whereas this is not necessarily the case for molecular data as many factors influence the availability of DNA. For instance, the age of the specimen and how it has been preserved have a significant effect on the integrity of its DNA (Hebert *et al.* 2013). Morphological character data is the only type of information which is available from fossils. Coding of morphological characters is a time-consuming process, requiring a sufficient level of expertise, whereas the extraction, amplification and sequencing of DNA form a faster process as a whole, when sufficiently good quality DNA is available. As a consequence, the acquisition of a large amount of data is by far easier with DNA than with morphology. The recognition of truly homological characters remains an issue as explained above (section 1.3). Both types of data share the problem. For non-coding DNA and gene regions with gaps, the alignment of sequences can pose difficulties in regard to homology (Wheeler 2012). Despite the different strengths and weaknesses of the data types, or rather because of them, both morphology and molecules contain valuable information on the evolutionary history of the subject and thus are equally useful for phylogeny reconstruction. In taxonomy, the use of DNA data has remained in the minority of studies longer than in phylogenetics, but its popularity has also increased with the adoption of DNA barcoding.

The data of chapter I comprised both morphological and molecular characters. Morphological data covered 167 species and altogether 253

characters (139 adult including both sexes, 49 pupal, and 65 larval characters). Molecular data consisted of eight genes (*COI*, *EF-1 α* , *RpS5*, *CAD*, *MDH*, *IDH*, *GAPDH*; 6 127bp in total). The same set of genes has been shown to successfully reflect relationships in lepidopterans (Wahlberg & Wheat 2008) and have been employed in several previous studies (Mutanen *et al.* 2010, Kaila *et al.* 2011, Zahiri *et al.* 2011, 2012).

In the chapters related to species delineation (**II**, **IV**), morphological characters were mainly studied to allow for comparison between the information content of morphology and the DNA barcode region. **II** included a test where external characters (mainly wing patterns) were used for sorting dry and pinned specimens. The test was conducted by an experienced lepidopterist with no previous taxonomic knowledge on the focal groups as in Tänzler *et al.* (2012). The rationale behind the test was to allow the comparison of DNA-based methods and ‘parataxonomic sorting’ (Janzen 2004) commonly used in biodiversity assessments in the tropics (e.g., Basset *et al.* 2000, Janzen & Hallwachs 2011). In **IV**, the characters of male genitalia were studied due to their likely close correspondence with species boundaries (e.g. Kaila 2011).

3.3 CHARACTER DATA 2: SINGLE- OR MULTI-LOCUS DNA?

When a tree is constructed based on more than one gene representing at least two genomes (i.e. multi-locus DNA), the result can be called a species tree and it is assumed to be an accurate estimation of the true phylogeny. When a tree is constructed based on one gene only (i.e. single-locus DNA), it is no longer a species tree, but instead, a gene tree (Pamilo & Nei 1988). Why has such a change in definitions been adopted? Different genes and genomes have different rates of change and types of inheritance, and thus a gene tree based on a mitochondrial locus may be incongruent with another gene tree based on a locus of a nuclear genome. Several phenomena might cause this pattern (e.g. genetic polymorphism in the ancestral species, incomplete lineage sorting, introgression, horizontal gene transfer), and they are especially common among closely related species (Dupuis *et al.* 2012). The same reasons can also be the underlying cause for non-monophyletic species on trees based on a mitochondrial locus or loci (Funk & Omland 2003). The rationale behind using multi-locus DNA from different genomes for phylogeny reconstruction is to avoid possible problems of gene trees and thus to produce a phylogenetic hypothesis with improved support (Ballard & Whitlock 2004).

Despite potential risks, single-locus data have their uses. They are widely used in population-level studies (e.g. Bromilow & Sperling 2011), and DNA barcoding of animals and fungi is based on a single gene (Box 2). Benefits of the single-locus approach are especially obvious for barcoding due to its

ambitious goal of eventually covering all species on Earth. Minimizing the amount of data per species enables maximizing the coverage. In addition, a common agreement to use only one region of a certain gene (or two in plants), i.e. standardization, makes different studies comparable and allows extensive meta-analyses. In Lepidoptera, the use of DNA barcode data removes problems with alignment as gaps are absent (only one tineid subfamily (Scardiinae) possessing a gap is currently known (P.D.N. Hebert, pers. comm.)). Finally, single-locus data simplify analyses as there is always only one dataset (**III**).

As indicated above, chapter **I** was based on multi-locus data from eight genes (one mitochondrial, seven nuclear), whereas in **II-IV**, we relied on single-locus DNA barcodes. All sequences in **II** (altogether 562 barcodes divided into two datasets: Finnish Gelechiinae: 92 species, 307 sequences, 187 haplotypes; Australian Elachistinae: 103 species, 255 sequences, 178 haplotypes) and **III** (Australian Hypertrophinae: 51 previously described species, 502 barcode sequences, 294 haplotypes) were of equal length (654bp as the BOLD aligner implemented in BOLD workbench (Ratnasingham & Hebert 2007, <http://www.boldsystems.org/>) reduces the original length of 658bp by omitting the first and the last three bases). This was done to remove possible problems introduced by missing bases. In **IV**, sequences were divided into two groups: (i) >400bp (191 sequences including one type, 78 haplotypes) and (ii) >400 and <325bp (223 sequences including the same 191 sequences, and 32 short barcodes (56-325bp) from type specimens).

3.4 TOOLS FOR TREE RECONSTRUCTION

Methods for reconstructing trees for phylogeny estimation and other uses can be divided into four categories *sensu* Wheeler (2012) based on their optimality criterion (i.e. the criterion defining the best tree): (i) distance methods, (ii) cladistic methods with parsimony as the optimality criterion, (iii) maximum likelihood preferring the solution which maximizes the likelihood function, and (iv) Bayesian inference, a very similar approach to maximum likelihood, relying on posterior probabilities.

The first category includes methods mainly lacking an optimality criterion such as Neighbor-Joining (NJ; Saitou & Nei 1987) and Unweighted Pair Group Method using arithmetic Averages (UPGMA; Michener & Sokal 1957). Distance-based approaches are user-friendly in the sense that they are very fast and return with only one solution. Their downside is that they discard a lot of information as all characters are treated equally for counting pairwise distances. As a consequence, the terminals are connected based on their overall similarity, which was an especially popular idea among phenetic systematists. Due to their drawbacks, distance-based methods are usually absent from studies aiming for phylogeny reconstruction. However, they, especially NJ, are commonly adopted in DNA barcoding (Box 2). A thing to

bear in mind is that phylogenetics and DNA barcoding have one crucial difference: while the former poses hypotheses on species relationships (all nodes on the output tree make a difference), the latter focuses on forming groups imitating species (only the closest nodes to the terminals are noted). As the order of species-like groups on the tree is irrelevant, the deep divergences on barcode gene trees make no difference. As a consequence, all problems that relying on overall similarity might cause for tree topology are irrelevant as long as the most similar individuals are placed together.

Cladistic methods relying on the principle of parsimony (i.e. the solution requiring the smallest number of changes should be preferred) have long remained in the nucleus of phylogenetic analysis (Farris 1983). The methods are equally suitable for morphological, molecular and combined data (Schuh 2000). When compared to the model-based approaches (maximum likelihood, Bayesian inference), parsimony is much simpler as it requires neither models of character evolution nor parameters of branch lengths. A cladogram is the plainest reconstruction of phylogenetic relationships, only visualizing nested sister-groups without providing statements about the evolutionary process (Wheeler 2012).

Model-based approaches including both maximum likelihood (ML) and Bayesian inference (BI) are by far more complex than previous methods. They require models of character evolution for each included partition, and if partitions are based on characters that evolve differently, they should be used with corresponding models. Two types of models are employed: (i) substitution models to estimate character transformation, and (ii) distributional models of character change rates. The former type includes several options varying in complexity (i.e. in the number of parameters) in base frequencies and substitution types. JC69 (Jukes & Cantor 1969) and MK (Lewis 2001) represent the simplest models, whereas General-Time-Reversible (GTR) is among the most complex ones. Gamma-distributed rates (Yang 1994) and the fraction of invariant sites (Hasegawa *et al.* 1985) are the most commonly used models of character change rates (Strimmer & von Haeseler 2009, Wheeler 2012). In addition, branch length parameters have an important role in the likelihood function, and they are also expressed in the output trees. The major difference between maximum likelihood and Bayesian inference is the use of priors (i.e. prior distributions of parameters) and employing Markov Chain Monte Carlo (MCMC) in the latter. Priors allow the implementation of previous knowledge on parameters, while MCMC enables the determination of posterior probabilities. The main priors to be defined by the user in phylogenetic analyses are tree priors and substitution models (and priors related to the rate estimation when relaxed molecular clock models are used).

Although cladistic and model-based methods are mainly different, they are commonly used in concert (e.g. Zahiri *et al.* 2012; I). Such a procedure enables the evaluation of congruence among resulting trees, and thus can be viewed as improving confidence in results. In chapter I, phylogenetic

analyses were performed with (i) the combined dataset of morphological and molecular data (with and without outgroups), (ii) both types of data separately, and (iii) each gene separately. Molecular and combined data were analyzed with parsimony and model-based maximum likelihood and Bayesian inference, but morphological data only with parsimony. The GTR model was used with molecular data and the MK model with morphological data. Bayesian inference gene trees were employed with tree-based delineation approaches (**II-IV**; see section 3.5), and maximum likelihood, Bayesian inference and Neighbor-Joining trees for associating short sequences with OTUs (**IV**). Additional maximum likelihood gene trees were used to enable comparison between ML and BI topologies (**III, IV**).

3.5 OTU DELINEATION METHODS

As noted above (section 1.2), several novel methods for species delineation have been published in recent years. Due to the sole use of DNA barcode data (**II-IV**), only methods suitable for single-locus DNA are considered here. Available approaches can be divided into three categories: character-, tree- and distance-based methods. In character-based methods, all individuals belonging to the same species share the same fixed diagnostic bases, which are lacking from other closely related species. Depending on the procedure used, diagnostics can be either just single bases or unique combinations. Tree-based methods, such as General Mixed Yule-coalescent (GMYC; Pons *et al.* 2006, Monaghan *et al.* 2009, Fujisawa & Barraclough 2013), require a certain kind of phylogenetic tree as an input and, thus, are more or less dependent on the tree-constructing method used to acquire the starting tree. In distance-based methods (e.g. Statistical parsimony networks (Templeton *et al.* 1992), Barcode Index Number System (Ratnasingham & Hebert 2013), Automatic Barcode Gap Discovery (Puillandre *et al.* 2012a)), individuals are grouped by their genetic similarity according to different clustering algorithms. Pairwise genetic distances are calculated from the sequences without making any differences between diagnostic and non-diagnostic bases. OTU (operational taxonomic unit) delineation studies (**II-IV**) utilized both tree- and distance-based methods, which are presented below. No character-based delineation method was employed here as two currently available options, Population Aggregation Analysis (PAA; Davis & Nixon 1992) and Character Attribute Organization System (CAOS; Sarkar *et al.* 2002, 2008, Bergmann *et al.* 2009), were inappropriate (PAA requires a priori definition of populations and CAOS apparently suffers from some malfunctions).

3.5.1 GENERAL MIXED YULE-COALESCENT (GMYC)

General Mixed Yule-coalescent (GMYC; Pons *et al.* 2006, Monaghan *et al.* 2009, Fujisawa & Barraclough 2013) is a tree-based likelihood method for species delimitation where a threshold is approximated based on the interplay between speciation and coalescent processes. According to Pons *et al.* (2006), this method aims to locate the transition point (i.e. the threshold) where branching rate of the tree changes from speciation to coalescent events. This change can be seen as a sudden increase on the lineage-through-time plot (where node heights are plotted against the logarithm of the number of nodes (Nee *et al.* 1992)). The amount of putative species equals to the number of lineages crossing the threshold line. Assuming the threshold a particular node on the tree, this method uses the internode intervals (i.e. waiting times between bifurcations) to find the maximum likelihood solution for the threshold (Pons *et al.* 2006). Speciation events are modeled with Yule stochastic birth-only model (Yule 1924) and intraspecific coalescent events are based on the neutral coalescent model of Kingman (1982). GMYC assumes that the most recent speciation event happened before the oldest intra-specific coalescent event within the pool of species under study (Pons *et al.* 2006). This makes the method conservative, because it is unable to detect recently diverged species (Papadopoulou *et al.* 2008). GMYC applies a null model, which consists only of coalescent process, and thus implies the presence of one species in the data (Pons *et al.* 2006). The method described above is called a single-threshold model. Monaghan *et al.* (2009) published another version of GMYC, which computes several thresholds allowing different speciation rates throughout the tree.

GMYC requires a fully resolved (i.e. fully dichotomous) ultrametric tree as a starting tree and it should always be used with haplotype data (including only one unique haplotype) (Pons *et al.* 2006, Monaghan *et al.* 2009). An ultrametric tree (clock-constrained phylogram) is a tree with equal root-to-tip lengths for all lineages, i.e. branch lengths equal to time units (Wheeler 2012). It can be constructed using the data from taxa evolved under a strict molecular clock or nearly so (Felsenstein 2004). However, given the rarity of true ultrametric data, the majority of trees have been constructed with relaxed molecular clock models, which allow variation in evolution rates over branches (Wheeler 2012).

There are multiple ways to construct a suitable starting tree for GMYC. The most common way is to use the software BEAST, which employs Bayesian inference and clock models (Drummond *et al.* 2006, Drummond & Rambaut 2007). The uncorrelated relaxed lognormal clock model, preferred by Monaghan *et al.* (2009), is by far the most popular of all clock models used with BEAST (Leliaert *et al.* 2009, Boykin *et al.* 2011, Pons *et al.* 2011, Vuataz *et al.* 2011, Brown *et al.* 2012, Esselstyn *et al.* 2012, Fernandez *et al.* 2012, Puillandre *et al.* 2012b, Tänzler *et al.* 2012, Williams *et al.* 2012) although others have been used as well (e.g. strict clock model: Bode *et al.* 2010). Ceccarelli *et al.* (2012) tested the impact of different relaxed clock

models for the results of GMYC. Their findings suggest that the relaxed log-normal clock model recovers more accurate clock estimates and thus produces more reliable species delimitations (measured by the congruence between morphology and molecular markers) than other clock models.

Since BEAST is a true Bayesian program, priors need to be defined. As a tree prior, both Yule and coalescent priors have been implemented as GMYC starting trees with the former usually for species-level and the latter for population-level data (Drummond *et al.* 2006). Both tree priors have been applied for GMYC studies (Yule: Bode *et al.* 2010, Pons *et al.* 2011, Williams *et al.* 2012; Coalescent: Boykin *et al.* 2011, Vuataz *et al.* 2011, Fernandez *et al.* 2012, Tänzler *et al.* 2012), sometimes even in the same study allowing the comparison of the tree priors (Monaghan *et al.* 2009, Yassin *et al.* 2009, Ceccarelli *et al.* 2012, Sauer & Hausdorf 2012; II). Monaghan *et al.* (2009) recommended the use of the coalescent prior, whereas Ceccarelli *et al.* (2012) noted that the effect of a prior depends on the marker used: with *cytb* they preferred Yule and with COI coalescent (both are mitochondrial genes), although the differences among the tree priors used were minor.

As a UPGMA tree is much simpler and faster to construct than a Bayesian tree, it provides an appealing option for GMYC. Still only a few studies have employed UPGMA starting trees (Lahaye *et al.* 2008, Barraclough *et al.* 2009a), possibly due to the strong critique for the method in the history of phylogenetics (e.g. Farris 1981).

Like its name implicates, General Mixed Yule-coalescent model is not specified to any particular type of evolution; instead, it can be used with many different species (Barraclough *et al.* 2009a) and even with asexuals (e.g. Fontaneto *et al.* 2007, Bode *et al.* 2010, Xiang *et al.* 2011, Obertegger *et al.* 2012), bacteria (Barraclough *et al.* 2009a) and fungi (Powell *et al.* 2011). The threshold value is always optimized for the dataset at hand rather than assuming a global cut-off value (Barraclough *et al.* 2009a). When GMYC clusters have been compared to other delineation methods, the conclusions have been somewhat contradicting: in earlier studies, results have showed congruence with morphology (Pons *et al.* 2006, Ahrens *et al.* 2007, Monaghan *et al.* 2009) and ABGD (Puillandre *et al.* 2012b, Ratnasingham & Hebert 2013), whereas in some cases, GMYC seems to deliver higher species counts than morphology and other methods (Esselstyn *et al.* 2012, Paz & Crawford 2012, Sauer & Hausdorf 2012, Miralles & Vences 2013, Talavera *et al.* 2013).

Multiple cases of a failure to reject the null model, where many species were sampled, have been reported (Powell *et al.* 2011, Vuataz *et al.* 2011, Ceccarelli *et al.* 2012, Sauer & Hausdorf 2012, Williams *et al.* 2012). One explanation for this result is a uniform branching rate through the tree, which leads to the exclusion of a sudden increase on a lineage-through-time plot, and complicates the function of GMYC (Williams *et al.* 2012). Ahrens *et al.* (2007) noted the effect of sampling on the change in branching rate. They stated that low intraspecific sampling could lead to the loss of a clear

transition point. In general, the phenomenon seems to be relevant for several groups of species like bumblebees (Williams *et al.* 2012), land snails (Sauer & Hausdorf 2012) and fungi (Powell *et al.* 2011). In the case of Ceccarelli *et al.* (2012), the null model was accepted with the nuclear marker *Wingless*, but rejected with all mitochondrial markers. A similar case was observed by Vuataz *et al.* (2011) with nuclear protein coding *PEPCK*.

3.5.2 STATISTICAL PARSIMONY NETWORKS (TCS)

The method originally introduced by Templeton *et al.* (1992; commonly used with software TCS (Clement *et al.* 2000)) is based on using networks instead of trees. Although having a slightly misleading name, the method is purely distance-based (Clement *et al.* 2000). The procedure begins with constructing a p-distance matrix from pairwise genetic distances. Pairs sharing the shortest genetic distances are combined to the same network. The procedure continues until an a priori stated cut-off value is crossed. The higher the cut-off value, the smaller the amount of steps needed and the higher the count of unconnected networks formed. In other words, setting a high cut-off value produces a high species count and *vice versa*. Hence, the method attempts to discover the most parsimonious solution for the given cut-off value.

Statistical parsimony networks have been used in several studies with varying focuses: phylogeography (Barrett & Freudenstein 2011), species delimitation (Astrin *et al.* 2012) and with both nuclear (Sota & Sasabe 2006) and mitochondrial sequence data (Paquin & Hedin 2004, Cardoso & Vogler 2005, Gompert *et al.* 2006).

3.5.3 BARCODE INDEX NUMBER SYSTEM (BIN)

Although all species delimitation methods used in this thesis can be employed with DNA barcodes, Barcode Index Numbers (BIN; Ratnasingham & Hebert 2013) and Automatic Barcode Gap Discovery (ABGD; Puillandre *et al.* 2012a) are especially designed for barcodes. As indicated in Box 2, the existence of the ‘barcode gap’ is a crucial factor for successful barcoding. Due to the case-specificity of the gap, strict thresholds generally perform poorly (Ferguson 2002, Holland *et al.* 2004, Bichain *et al.* 2007, Gómez *et al.* 2007, Meier *et al.* 2008). This issue is particularly important when barcode data are used for species discovery, and both BIN and ABGD methods have been developed to deal with this challenge.

In the BIN system, the delineation is conducted with RESL algorithm as a two-phased procedure, which starts with single linkage clustering and is subsequently followed by Markov clustering using pairwise p-distances (Ratnasingham & Hebert 2013). As linkage clustering requires a priori definition of a threshold level for maximum intraspecific divergence, the BIN system employs 2.2%, which was found optimal in several datasets covering a

wide variety of taxa (Ratnasingham & Hebert 2013). Single linkage clustering is followed by Markov clustering in order to verify and, where needed, redefine groupings gained from the first step.

The interface of BOLD is easy to use and the RESL algorithm requires very little time, which make the BIN system a useful method. This is further supported by the fact that the BIN system output consists of only one result. This can, however, also be seen as a drawback since the user has basically no opportunity to affect the parameters of the analysis.

3.5.4 AUTOMATIC BARCODE GAP DISCOVERY (ABGD)

The main aim of ABGD is to locate the barcode gap from the dataset under study and use this location for producing putative species boundaries (Puillandre *et al.* 2012a). More specifically, the method first splits data into groups based on the statistically inferred barcode gap, and then recursively applies the same procedure to the groups obtained in the former step. This two-phased procedure enables the use of several thresholds in the data (Puillandre *et al.* 2012a).

According to Puillandre *et al.* (2012a), the protocol starts with computing and ranking pairwise distances. Substitution models currently available are JC69 and K2P, but using p-distance is also possible. Distances are then used for computing a local slope function. The method detects peaks of slope values and reports the distance performing the local maximum as the exact value of the gap. However, the chosen slope must fulfill two conditions: (i) to be the first local maximum slope after $\text{dist}_{\text{limit}}$ (i.e. a threshold value under which distances are more likely to be intraspecific) and (ii) to be X (i.e. an estimate of relative gap width chosen by the user) times larger than any gap within the intraspecific part of all distances (Puillandre *et al.* 2012a).

Besides the value of X , ABGD also requires minimum and maximum values of prior intraspecific divergence (P) as an input (Puillandre *et al.* 2012a). The method uses P -values as limits for computing an average pairwise distance value to be used as an estimator of population mutation rate θ . Population mutation rate is an essential part of a coalescent model (Tajima 1983), and it is used by ABGD for detection of the barcode gap, but its estimation is very problematic when the number of species in a dataset is unknown (Puillandre *et al.* 2012a).

Only a handful of studies have utilized ABGD and BIN so far (Jörger *et al.* 2012, Pantaleoni & Badano 2012, Paz & Crawford 2012, Puillandre *et al.* 2012b, Rosso *et al.* 2012, Hausmann *et al.* 2013, Hendrixson *et al.* 2013, Mutanen *et al.* 2013, Prévot *et al.* 2013, Ratnasingham & Hebert 2013, Weigand *et al.* 2013).

3.6 EVALUATING THE RESULTS OF OTU DELINEATION

We used two different protocols for comparing the composition of OTUs: (i) the direct examination method of Ratnasingham and Hebert (2013) where OTUs are divided into four categories (MATCH, SPLIT, MERGE, MIXTURE) based on their similarity with previously studied reference species (**II**), and (ii) a comparison of resulting OTUs based on their congruence (**III**). In the latter, OTUs were divided into three categories (MATCH, PARTIAL MATCH, DISCORDANT) similarly as in the direct examination, but instead of comparing OTUs with reference species, the sets of OTUs produced by three delineation methods were compared among themselves. The OTUs included in PARTIAL MATCH and DISCORDANT categories were subsequently evaluated against criteria (monophyly, diagnostic characters, sympatry) derived from different species concepts (Box 1).

3.7 ASSOCIATING SHORT SEQUENCES FROM TYPE SPECIMENS WITH OTUS

Short sequences (<300bp) are problematic with OTU delineation methods as they are either automatically excluded from analyses (BIN) or may cause false results (GMYC, ABGD; **IV**). Yet their inclusion may be crucial in particular cases, like linking type barcodes with OTUs. Type specimens are usually old and their DNA degraded, yielding only short fragments of DNA (Hajibabaei *et al.* 2006). To overcome these problems, we applied another procedure (**IV**): three tree-based methods and one relying on pairwise distances were used in concert. The types were associated with OTUs based on the tree topology (types were linked with their nearest neighbors) or according to the closest match in distances.

4 RESULTS & DISCUSSION

4.1 COMBINED DATA PROVIDES IMPROVED SUPPORT IN GELECHIOIDEA

The phylogenetic study of the Gelechioidea (**I**) provides the best support to date for family-level classification, due to the comprehensive taxon sampling and the use of combined data including both morphological and molecular characters. All analyses of combined data offered better resolution and improved support than analyses based on molecular or morphological characters alone, although support values remained generally low, especially in deep divergences. The topologies in both model-based phylogenies reconstructed with combined data, and in the maximum likelihood tree with molecular data were very similar. Bayesian inference for molecular characters was unsuccessful as MCMC chains failed to converge if outgroup taxa were included. Parsimony analysis of morphological and combined data produced well-resolved strict consensus trees with rather weak support values, whereas the tree based on molecular data alone was less well resolved. In addition to improved support values, the use of combined data seems to reduce the negative effect of ‘rogue taxa’ (i.e. taxa which are unstable across various phylogenetic analyses, causing weakly supported clades and poorly resolved trees).

A revised classification of the Gelechioidea includes 16 families of which 11 are well-supported in all model-based analyses (narrowly restricted Oecophoridae, Gelechiidae, Cosmopterigidae, Momphidae, Blastobasidae, Stathmopodidae, Scythrididae, Batrachedridae, Coleophoridae, Autostichidae, and Lecithoceridae). Newly formed Depressariidae (including the following ten subfamilies: Acriinae, Aeolanthinae, Cryptolechiinae, Depressariinae, Ethmiinae, Hypercalliinae, Hypertrophinae, Oditinae, Peleopodinae, and Stenomatinae) is based on a large, yet weakly supported clade consisting taxa from several former families such as the Elachistidae *s.l.* (Kaila 2004, van Nieukerken *et al.* 2011). The family Elachistidae in its earlier composition was mainly based on pupal characters (the presence of lateral condyles), but the current phylogenetic hypothesis suggests that these characters have resulted from convergent evolution and thus are referred to as homoplasies. In general, morphological homoplasies are numerous within the Gelechioidea. This has undoubtedly hindered the discovery of appropriate morphological characters, which would act as synapomorphies at family-level.

The Depressariidae formed, for the first time, a monophyletic clade and possessed synapomorphic characters derived from the male genitalia and pupae (**I**), although they also exist in the Elachistinae and the Coleophoridae. A spinose, lobe-like gnathos in the male genitalia is present in many, but not

all depressariids. A paired spur group ventrally on abdominal segment 9 forms pupal-based synapomorphies together with the above-mentioned lateral condyles. As the sampling in this study was not sufficient enough for delineating intrafamily taxa, we did not propose a classification for subfamilies. The monophyly of the Hypertrophinae and its placement in the revised Depressariidae are supported by the results.

4.2 DNA BARCODE-BASED DELINEATION OF PUTATIVE SPECIES

This thesis includes three studies focusing on delineation of putative species (OTUs) with DNA barcode data (four datasets; Fig. 2), and with an emphasis on either method-testing (**II**) or employing various methods for different taxonomic purposes (**III**, **IV**). In the method-testing study, we present a comparison of four DNA-based methods (BIN, TCS, ABGD, GMYC) with two datasets representing well-known reference species (Finnish Gelechiinae and Australian Elachistinae datasets are presented in section 3.3) (**II**). Of these two datasets, Australian elachistines exemplify a challenging group known to possess discordances with barcoding (Kaila & Ståhl 2006). The OTU-delineation study uses three of four tested methods (BIN, ABGD, GMYC; unpublished results from TCS and GMYC with UPGMA are also included in sections 4.3 and 4.4) to delimit OTUs in the poorly-known Australian subfamily Hypertrophinae (see section 3.3) and demonstrates a novel approach to evaluate conflicting OTUs (**III**). The last study repeats the same

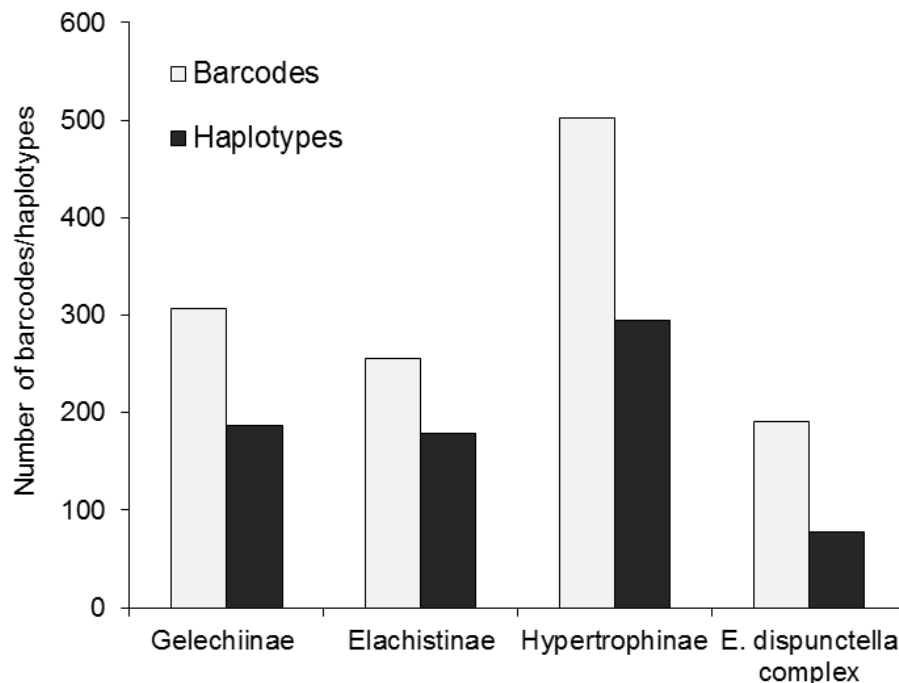


Fig. 2 Datasets used for OTU delineation studies (**II-IV**).

delineation procedure as in **III** with barcodes from fresh specimens (>400bp), but adds a new combination of well-established methods (ML, BI, NJ, pairwise distance comparison) for linking short DNA barcodes (most <200bp) from type specimens with OTUs (**IV**). This two-phased procedure is employed for *Elachista dispunctella* complex (see section 3.3) to re-evaluate species boundaries and estimate the number of synonymies (**IV**).

Genetic distance provides an estimate on how closely related two species are (Vandamme 2009). Fig. 3 shows K2P pairwise distances of all datasets used in **II-IV**, illustrating the differences in barcode gaps (Box 2) between studied groups. The narrowest gap is possessed by the Elachistinae dataset, whereas the Gelechiinae and the Hypertrophinae show much wider gaps. The *Elachista dispunctella* complex is characterized by steep edges on both sides of the gap, making it especially favorable for these delineation methods.

4.3 CONGRUENCE IN RESULTING OTUS

OTU delineation methods provide results of two kinds: OTU count and OTU composition. The OTU count means the number of delimited groups and offers an estimate of the species count, whereas the OTU composition refers to the individuals included in each group, which is particularly important to subsequent taxonomical work. The evaluation of delineation results based on an OTU count alone can be misleading as close (and even same) OTU counts can be produced by different combinations of OTU composition if splits and merges are cancelling each other out.

The range of OTU counts produced by different methods depends on the dataset used. Both the Gelechiinae and the *E. dispunctella* complex resulted in rather stable OTU counts, but the Hypertrophinae and the Elachistinae showed more diverse outcomes (Figs 4 & 5). Reference species count was available only for the Gelechiinae and the Elachistinae, enabling the comparison between previously described species and resulting OTUs. The OTU counts were generally higher than the reference in the Gelechiinae, but lower in the Elachistinae (**II**).

We compared the OTU composition among delineation methods in two studies (**III**, **IV**), but contrasting OTU composition with reference species was possible only in the Gelechiinae and the Elachistinae (**II**). With regard to the reference species, the results indicate an evident difference between the two datasets: ca. 90% of delineated OTUs were congruent with references in the Gelechiinae, whereas only 60-70% of OTUs matched with references in the Elachistinae.

To be able to compare the OTU composition among the methods used in all four datasets, an additional analysis was conducted. The OTU composition was evaluated as in **III** by using one result from three methods (BIN, GMYC, ABGD) where OTU counts were most congruent (Gelechiinae:

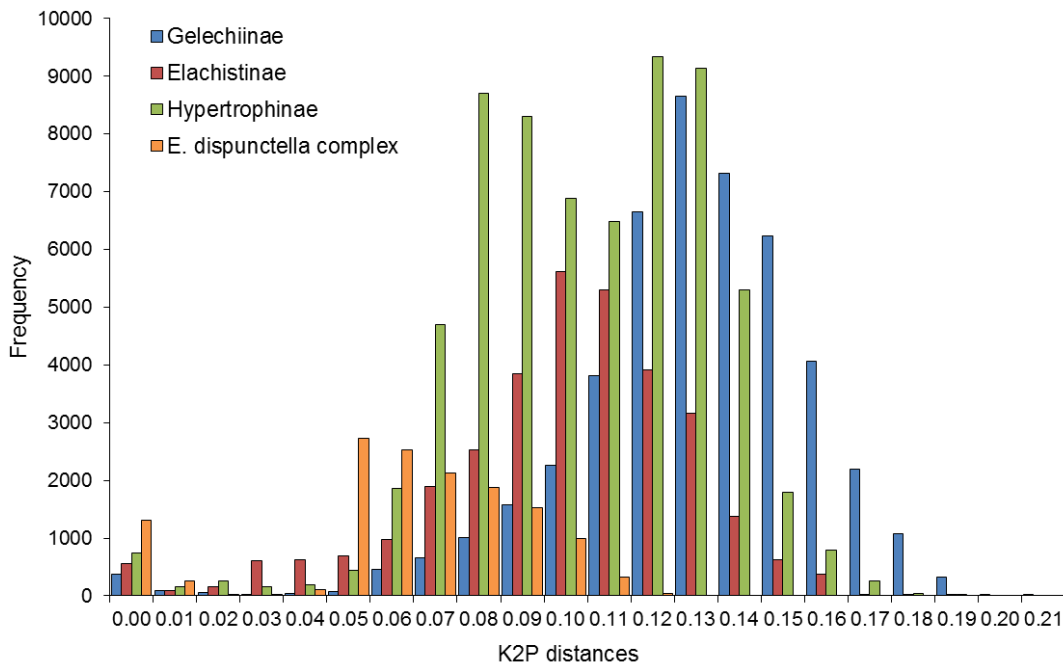


Fig. 3 Pairwise K2P distances without a priori grouping for Gelechiinae and Elachistinae (II), Hypertrophinae (III) and *E. dispunctella* complex (IV).

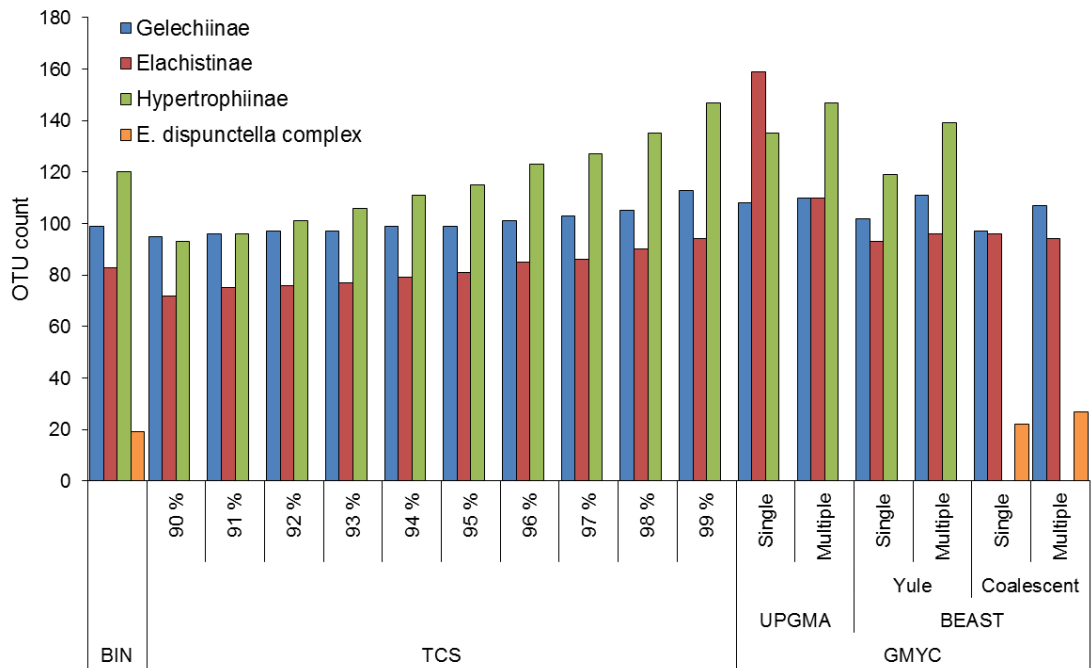


Fig. 4 OTU counts resulting from BIN, TCS and GMYC for Gelechiinae and Elachistinae (II), Hypertrophinae (III; results from TCS and GMYC with UPGMA: unpublished data) and *E. dispunctella* complex (IV).

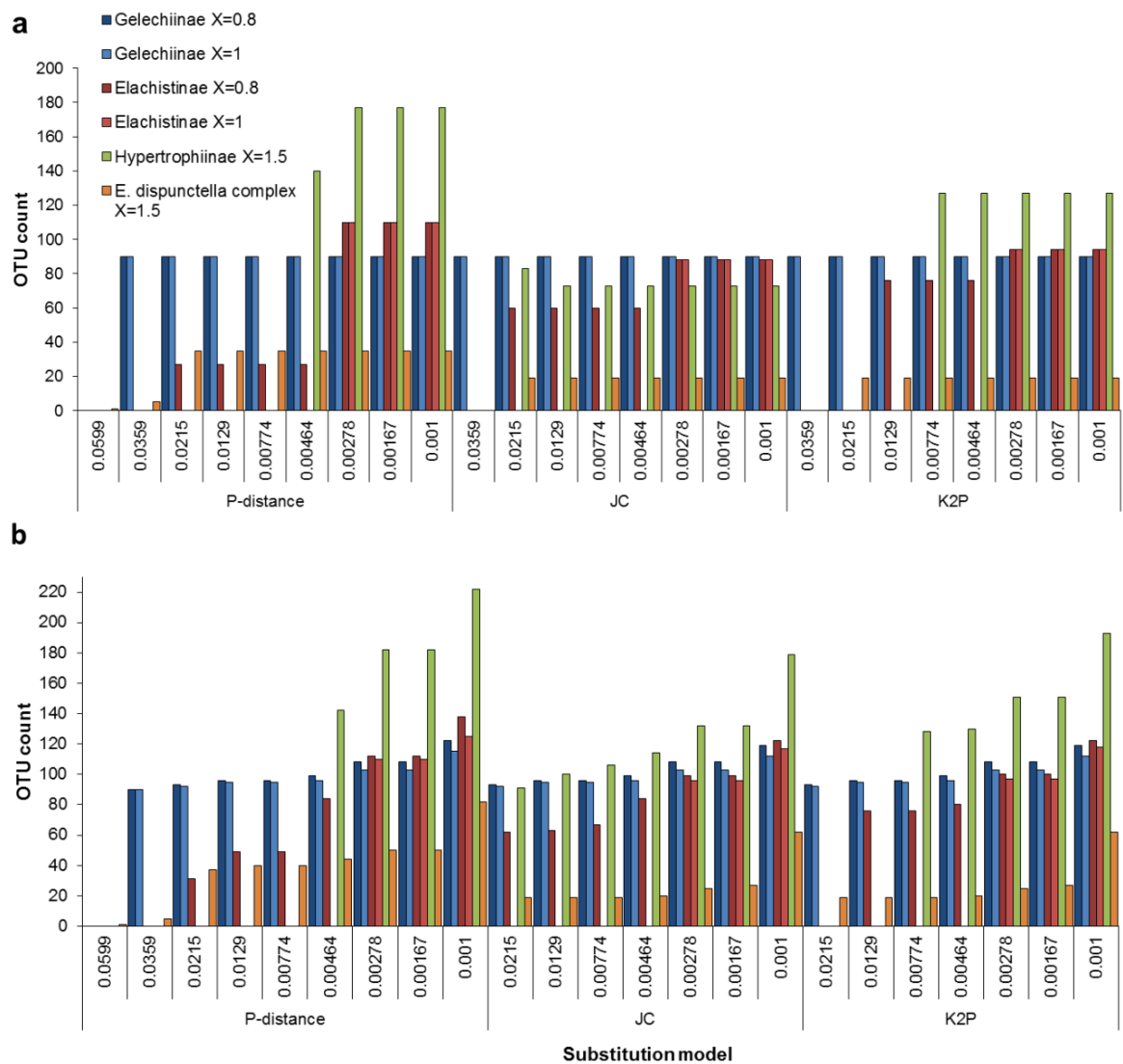


Fig. 5 OTU counts resulting from ABGD (a) initial partitions, (b) recursive partitions for Gelechiinae and Elachistinae (II), Hypertrophinae (III) and *E. dispunctella* complex (IV). Figures below the results indicate prior intraspecific divergence (*P*) values (in reverse order by substitution model).

GMYC with single threshold model and Yule tree prior, ABGD with K2P, $X = 0.8$, $P = 0.00464$; Elachistinae: GMYC with single-threshold model and Yule tree prior, ABGD with K2P, $X = 0.8$, $P = 0.00278$ initial partition; Hypertrophinae: see **III**; *E. dispunctella* complex: OTU11 was excluded from GMYC and ABGD as the corresponding sequence was omitted by the BIN system, otherwise see **IV**). Results from these three methods were divided into three categories: FULL MATCH, PARTIAL MATCH and DISCORDANT (for definitions of the categories see **III**). As illustrated in Fig. 6, the percentage of results in the FULL MATCH category is rather stable (80-90%) in all datasets, being the highest in the Gelechiinae and the lowest in the Hypertrophinae. The amount of OTUs in the FULL MATCH category seems to correlate to some extent with the size of the dataset instead of the width of the barcode gap, although the sample size ($n = 4$) is very small, making this comparison unreliable. All conflicting results, except two cases in the Hypertrophinae, are PARTIAL MATCHES, denoting that one of three results disagrees with the others. Altogether, the OTU compositions resulting from different methods seem to be mainly congruent, supporting the findings in an earlier study (Ratnasingham & Hebert 2013). It should be noted, however, that several factors can affect these results. For instance, the datasets include varying numbers of sequences and (putative) species, and the ABGD result with the OTU count closest to the other methods was chosen for the comparison (see further discussion in **III**).

Using several methods with different analytical approaches increases the reliability of results (Carstens *et al.* 2013), although systematic errors are possible even then (e.g. resulting from inappropriate sampling or the use of mitochondrial DNA). Of the three OTU delimitation studies presented in this thesis, none included completely congruent results (Figs 4, 5 & 6). There are

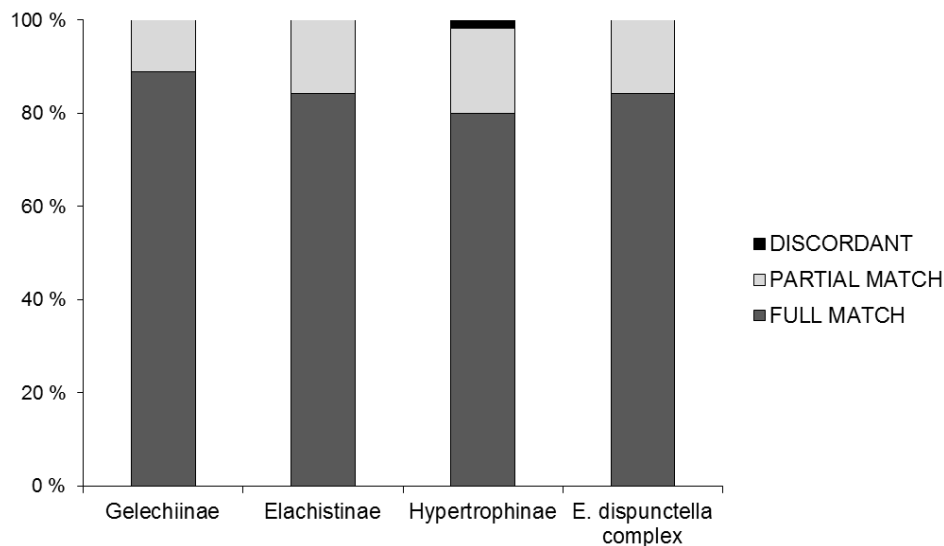


Fig. 6 OTU composition for Gelechiinae and Elachistinae (**II**), Hypertrophinae (**III**) and *E. dispunctella* complex (**IV**). Description of FULL MATCH, PARTIAL MATCH, and DISCORDANT categories are provided in **III**.

several possible explanations to this. First, a method or all methods may return false results. False is interpreted here as being strongly discordant with other available evidence, or when other types of evidence are lacking due to a conflict with different methods. Strongly discordant results have been reported, although effects of methodological errors are usually unknown (e.g. Sauer & Hausdorff 2012). Second, a given dataset may be challenging for the methods used, leaving a lot of room for different outcomes. This is especially evident in **II** where the resulting OTU counts of the ‘easier’ dataset (Gelechiinae) are more congruent than those of the more challenging Elachistinae, although the difference is minor when the closest results of OTU composition are compared (Fig. 6). A possible explanation to this difference is the narrow barcode gap of Elachistinae (Fig. 2). Third, all employed methods (GMYC, BIN, ABGD) may have a tendency to split ‘outliers’, i.e. specimens which have been collected far away from the other specimens belonging to the same OTU. This explanation is especially evident in the large Hypertrophinae dataset (**III**). Although these singleton OTUs might reflect the true boundary between two species, they should be interpreted with caution as the difference may only result from the geographic distance. Many PARTIAL MATCHES are caused by these ‘outlier splits’ (**III**). Fourth, discordant results may be due to ABGD returning a range of results if it is used with the default settings (Fig. 5). Parsimony networks (TCS) can also produce several results when employed with a range of cut-off values (Fig. 4), although that method is generally used with only one value (95%). Both methods are designed to produce many results starting from more inclusive grouping with lower number of OTUs, and increasing splitting gradually until the OTU count is very high. Obviously, this feature needs to be considered when comparing the results between methods.

4.4 PERFORMANCE OF DELINEATION METHODS

The performance of four methods was evaluated in **II** with two datasets (Gelechiinae and Elachistinae). These two datasets vary with regard to genetic distance (Fig. 2, **II**), intra- and interspecific distances (**II**), and the number of singletons and monophyletic and non-monophyletic species (monophyly was evaluated based on the species’ structure on a mtDNA gene tree) (**II**). In all these, the Elachistinae is more challenging than the Gelechiinae, having narrow barcode gap and a higher proportion of non-monophyletic species. The Elachistinae dataset also includes more singletons, but it can have both positive and negative effects on the results. Delineation methods can deliver incorrect results if intraspecific sampling is very low (Puillandre *et al.* 2012a), but by contrast, a higher number of

singletons may result in more matching OTUs as splitting is not an option for OTUs including one specimen.

The performance was evaluated with the direct examination method (Ratnasingham & Hebert 2013) by comparing the composition of resulting OTUs to the composition of reference species, and dividing results subsequently into four categories: MATCH, SPLIT, MERGE and MIXTURE (for definitions of categories see **II**). The performance was also evaluated separately for singletons and monophyletic and non-monophyletic species (**II**). As the direct examination relies on the underlying taxonomy, all falsely identified species and specimens affect the results. Because Australian elachistines seem to include fairly young species (Kaila & Ståhls 2006, **II**), the group might be also be challenging for traditional taxonomic approaches, hindering the establishment of a stable classification. A contrasting pattern is also possible with the Gelechiinae dataset. It includes only species occurring in Finland, but many of these species have ranges reaching other countries as well. Our sampling scheme thus includes only a restricted amount of intraspecific diversity, which might raise the proportion of MATCHES as genetic distances between specimens can be falsely long (Bergsten *et al.* 2012). These topics are further discussed in **II**.

The comparison pointed out methods with apparent drawbacks. Of the two available models in GMYC, the multiple-threshold model produced a higher OTU count with all starting trees used in three out of four datasets (Fig. 4). A similar result was also discovered by Fujisawa and Barraclough (2013) with simulated data. Also, the use of a UPGMA starting tree generally returned a higher OTU count than Bayesian trees (Fig. 4). The same pattern was obvious when the OTU composition was compared to reference species (**II**): the multiple-threshold model and UPGMA starting tree produced the lowest amount of MATCHES in the Gelechiinae and were also among the lowest in the Elachistinae. This indicates that both methods should be either used with caution or avoided. By contrast, Bayesian starting trees with the single threshold model produced a very congruent outcome with other methods, especially with BIN and TCS 95%, although GMYC slightly outperformed them with the challenging Elachistinae dataset. Used tree prior (Yule or coalescent) seemed to have a small effect only (**II**).

As expected, the OTU count rises with the rising cut-off values (Fig. 4) in TCS. When the OTU composition was compared with references, an increase in the cut-off value improved the match, but only in the Elachistinae (**II**). In the Gelechiinae, the best match was produced by 93% cut-off and decreased when the cut-off value rose. ABGD seems to be a very controversial method. It produced the highest number of MATCHES in both Gelechiinae and Elachistinae, but these were based on the use of contrasting prior intraspecific (*P*) values. As emphasized in **II** and **III**, ABGD poses problems with poorly-known groups. Choosing one result from several options is difficult when the same *P*-value can produce contrasting results with different datasets. Furthermore, ABGD shows discordant results between

substitution models (p-distance, JC and K2P) (Fig. 5), and especially p-distance seems to be producing much higher OTU counts than the other two distance metric options.

In general, none of the tested methods produced MATCHES with non-monophyletic species. By contrast, singletons did not seem to cause discordances, but the problematic methods, such as multiple-threshold GMYC and GMYC with UPGMA, performed poorly with singletons as well. The above-mentioned slight improvement of GMYC with the Elachistinae compared with BIN and TCS seems to be caused by the better performance with singletons (**II**).

Sufficient sampling effort is crucial for the delineation of species, but is an important factor also for studies focusing on OTUs (e.g. Hey 2009, Lohse 2009). However, covering an adequate amount of intraspecific variation can be challenging as the number of samples (i.e. sequences) and haplotypes (i.e. unique sequences) are commonly unequal, and increasing the sampling effort may only raise the number of sequences, but not of haplotypes. In the four datasets used here, all have different amounts of barcode sequences and haplotypes (Fig. 2). Another potential problem of sampling is the commonness of rarity in nature (Lim *et al.* 2012). The focal group in **III** currently includes 51 described species, and as the sample covers 502 sequences, it is supposed to reach ten times the minimum coverage. However, the results revealed that a large proportion (46%) of delineated OTUs includes only one or two specimens indicating that the sampling has been uneven. As such difficulties in accomplishing sufficient sampling might be common among studies on poorly-known groups, methods employed for the OTU delineation should be robust towards intraspecific sampling. Still, for instance ABGD has been reported to result in an unreliable outcome when OTUs include less than three specimens (Puillandre *et al.* 2012a, b). This mismatch emphasizes the need for studies evaluating the performance of the method, and discourages its use as a sole method for the OTU delineation in poorly-known groups.

4.5 CRITERIA FOR CONFLICTING RESULTS

As noted above, the results from delineation methods are rarely fully congruent, especially with tens of putative species. Then, the following question is how to deal with discordant results. Previously adopted approaches, conservative (Weigand *et al.* 2013) and minimum consensus (Jörger *et al.* 2012), discard much information. We present criteria for evaluating OTUs belonging to PARTIAL MATCH and DISCORDANT categories (**III**). In order to be useful for evaluating conflicting OTUs, the criterion should point out differences among OTUs. Altogether three criteria derived from different species concepts were tested, but both monophyly and diagnostic characters proved unfeasible as nearly all OTUs were

monophyletic and possessed diagnostics. This is a likely consequence of the use of OTU delineation methods as they mainly form clusters, which are monophyletic initially. Diagnostic bases were discovered for all conflicting OTUs, possibly due to the low number of specimens within each OTU. By contrast, the sympatry criterion proved useful (Fig. 4 in **III**). According to that criterion, allopatric OTUs form one entity, because their status is impossible to evaluate. Instead, the sympatric ranges of sister OTUs provide sufficient evidence to keep them divided. Although this criterion is based on the Biological Species Concept and the Genotypic Cluster Species Concept (Box 1), it should be noted that mitochondrial DNA cannot provide direct evidence of reproductive isolation, so nuclear DNA data are needed for its confirmation. However, if two sister OTUs have overlapping ranges, they need to be ‘independently limited’, meaning that individuals of both OTUs mainly contribute genes to subsequent generations within their own OTU (Barracough *et al.* 2009b). In sympatry, this is usually caused by ecological differences between the groups.

4.6 ASSOCIATING TYPES WITH OTUS – THE CHALLENGE OF SHORT SEQUENCES

Type specimens are crucially important for taxonomic studies, but as many types are old, their DNA can be degraded and therefore provide only short fragments (Hajibabaei *et al.* 2006, but see Strutzenberger *et al.* 2012). Even short sequences have been shown to include sufficient information to link them with their conspecific specimens (Hajibabaei *et al.* 2006, Meusnier *et al.* 2008, Shokralla *et al.* 2011). Although OTU delineation methods offer a valuable tool for examining species boundaries, short sequences cannot be included into these analyses (**IV**, Ratnasingham & Hebert 2013). In **IV**, we employed a two-phased procedure where OTUs were first delineated with three methods (GMYC, BIN, ABGD) based on longer sequences (>400bp) and subsequently linked with short type sequences with three tree-based methods (ML, BI, NJ) and one approach relying on pairwise distances. The use of several approaches for association of type barcodes is required with very short sequences due to the different drawbacks of the methods. A small trial (see **IV**) revealed that both maximum likelihood and Neighbor-Joining tend to place any short sequence on a distinct node even when it was identical with several clusters. This was the case with two identical sequences from different types (164bp and 65bp), which were placed in different clusters in ML and NJ trees (although located correctly in BI). However, two clusters represented the same OTU based on initial delineation and, thus, two types were assigned to one OTU. This example emphasizes the importance of conducting both phases: initial OTU delineation and subsequent linking of types. The use of pairwise distance comparison can reveal false placements by tree-based methods, but it also has its own pitfalls.

Those might appear when an overlap in the target sequence and the reference sequences is partial, usually caused by very short sequence lengths.

4.7 CONTRIBUTION TO TAXONOMICAL RESEARCH

In **III** and **IV**, the methods were used for solving taxonomical questions related to different gelechioid moths. The OTU delineation of the poorly-known Australian subfamily Hypertrophinae (**III**) revealed a more than two-fold increase from the currently recognized species number (from 51 to 124) and provided an excellent framework for subsequent taxonomic work (Kekkonen, in prep.). In the problematic *Elachista dispunctella* complex, both the initial OTU delineation and the type association offered evidence on the presence of 19 OTUs and revealed a high number of likely synonymies (two to eight types were linked to seven OTUs). In addition, both phases showed close correspondence with morphology, and thus provided a useful scaffold for following revisionary work (Kaila & Tabell, in prep.).

5 CONCLUSIONS & PROSPECTS

As emphasized in the Introduction, the taxonomic workflow includes several steps to be conducted in a certain order (Fig. 1). In this doctoral thesis, I present four studies concentrating on different steps along this ‘taxonomic flowchart’. This doctoral thesis forms a part of a long-term study of Australian Hypertrophinae, aiming to provide a well-supported phylogeny and species boundaries for the group. The results of this thesis not only enhance the study of hypertrophines, but also offer tools to benefit the taxonomic research in general. The first chapter (**I**) offers evidence on the synapomorphies and monophyly of 16 families in the Gelechioidea. The newly formed family Depressariidae is of particular interest here as the subfamily Hypertrophinae belongs to it. The phylogeny of the Gelechioidea also supports the monophyly of the Hypertrophinae and reveals its close relatives, although the sampling was insufficient for consistent conclusions.

We studied various methods for the DNA barcode-based delineation of OTUs to reveal their strengths and weaknesses (**II**), and then employed some of these methods on hypertrophines (**III**). I will continue to study these hypertrophine OTUs in the near future based on varying sources of evidence, and either validate them via diagnostic characters or discard them (i.e. divide them into two or more species or merge with some other species). Due to the current phase of the taxonomic work, the congruence between OTU and species boundaries of hypertrophines cannot be evaluated here. Nevertheless, the initial delineation based on DNA barcodes has provided an efficient start for the work and an essential introduction to the diversity of the group, which is especially desirable for a subfamily lacking an expert in the present-day.

In the last step of the taxonomical workflow, I will describe and name new species (Fig. 1.d). As the type specimens provide the main link between newly studied species and existing names, various means are used for revealing this linkage. In addition to the external morphology and genitalia, the procedure for associating short DNA barcodes with OTUs (**IV**) will be most beneficial for this purpose.

The superfamily Gelechioidea is a good example of the lack of knowledge of biodiversity. The Lepidoptera are generally the best-known insect group, yet one of their largest radiations remains poorly-studied. Nevertheless, much progress has taken place in recent years in the form of phylogenetic studies and taxonomic revisions (e.g. Kaila 2004, 2011, Bucheli & Wenzel 2005, Lee & Brown 2008, Bucheli 2009, Huemer & Hebert 2011, Kaila *et al.* 2011, Bauer *et al.* 2012, Karsholt *et al.* 2013), and the newly-formed newsletter I.N.G.A. aims to further strengthen the global collaboration (http://mississippiantomologicalmuseum.org.msstate.edu/Researchtaxapages/Lepidoptera/Gelechioidea/INGA_newsletter.html).

As emphasized in the Introduction, taxonomy is in dire need of new innovations. Even the need for binomial names has been questioned in discussions, but most novel ideas concentrate on more modest issues. Research on DNA-based taxonomic methods is evolving rapidly, and these tools speed up the workflow considerably without sacrificing the high quality of the outcome. As noted by Riedel *et al.* (2013a), providing DNA barcodes of species to be described, including type specimens, adds to the information content of the description substantially and enables the production of more compact descriptions, which in turn increases efficiency. DNA-based methods are by no means the only innovations in taxonomy. New journals provide not only Open Access publication and data storage, but also the opportunity to utilize additional web-based features, such as writing tools (e.g. <http://biodiversitydatajournal.com>) and ‘wikis’ for each newly-described species. Especially wikis have great potential as they facilitate the updating of knowledge whenever new information arises (Riedel *et al.* 2013a, b).

In general, applying the optimal combination of different methods, either traditional or novel, along the ‘taxonomic flowchart’ most likely leads to increased productivity in both quantity and quality of the work. Furthermore, as DNA barcoding has proved, ‘big science’ funding is also possible for taxonomy-related fields of research, but it seems to require sufficiently bold ideas (Meier 2008). The grand question is: Will the taxonomic community have the courage to broadmindedly discuss new bold ideas?

ACKNOWLEDGEMENTS

In 2007, an open position of museum technician was announced in the FMNH. Senior curator, Dr Lauri Kaila and Jaakko 'Jaska' Kullberg were brave enough to hire two girls with limited amount of experience on lepidopteran species (me and Katja Rönkä). Seven years later, I am still working in the FMNH, and it is evident that this thesis would have never seen daylight without the courage and vision of Lauri and Jaska.

I am very grateful to Lauri for being my supervisor. I thank greatly for your faith in me, sharing your knowledge on morphology, gelechioids and other topics with me and all your support towards my work. There have been ups and downs, but we definitely survived as winners!

I am also very grateful to my other supervisor Dr Marko Mutanen for helping and supporting me with varying molecular and other works, and acting as a wonderful host during my visit to Oulu. You also introduced me to the fascinating world of DNA barcoding. I was very lucky to have two supervisors complementing each other so well.

Although I have two official supervisors, I have many times felt that I would have had three instead of two. Canadian professor Paul Hebert visited Finland in May 2011. Since then, Paul's impact on this thesis has been considerable. Paul, you never had any official obligations towards me, but you still chose to help me tremendously. I am so grateful to you for your advice, support and help in various topics, and your hospitality in Australia and Guelph. You have been my unofficial mentor, a role model and a source of inspiration. Thank you so much for believing in me.

I am most grateful to Dr Carlos Lopez-Vaamonde for acting as my opponent. Dr Michael Balke and Dr Tommi Nyman are acknowledged for their valuable contribution towards this thesis during the pre-examination. Professor Jouni Laakso kindly agreed to act as a custos.

I am grateful to the members of my thesis advisory group, Professor Jaakko Hyvönen and Professor Jyrki Muona, for all inspiring discussions, useful knowledge and pieces of advice. The coordinator of LUOVA graduate school Anni Tonteri has offered invaluable support during my doctoral studies, which I am very grateful. Professor Veijo Kaitala provided assistance in many issues throughout the years.

Dr Marianne Horak and Dr Ted Edwards were extremely hospitable during my visit to ANIC, and shared their information and knowledge with me. I am honored to be able to continue the work of past Dr Ian Common on hypertrophines. I thank You Ning Su, Beth Mantle and other kind people in ANIC for your help. I am also grateful to Dr John LaSalle for hospitality and assistance.

I have been lucky to be able to work at the zoology unit of the Finnish Museum of Natural History (LUOMUS) with many great people, but I am especially grateful to Maria Heikkilä. You have been such a great friend during these years and provided tons of peer support. Thank you so much! Marianna Teräväinen kindly corrected the language of the summary chapter. All other PhD students in the FMNH (Aino,

Andrea (currently in MRG), Annina, Diego, Elina, Eeva-Maria, Hanna La., Hanna Li., Heidi, Juho, Kalle, Maria Hä., Sanna, Sara) have provided many good moments and memories. I am also grateful to the director of the FMNH Professor Leif Schulman and the director of the zoology unit Dr Aino Juslén for your support towards my doctoral studies. Jaska, Hasse, Jere, Jaakko M. and others belonging to the staff of LUOMUS are acknowledged for many great memories, interesting discussions during lunches, coffee breaks and Christmas parties. Also, many thanks to librarian Marjatta Mikkonen who has always offered her kind assistance.

In addition to Paul, I had the privilege to meet many great people at the Biodiversity Institute of Ontario (BIO), but I am especially indebted to Sujeevan Ratnasingham for your support and advice, and Sean Prosser for all of your efforts towards my work and being such a great fellow worker. In addition, Megan Milton provided valuable help in various cases, Dr Bob Hanner kindly invited me to join his lab meetings, Dr Evgeny Zakharov shared his insights during several interesting discussions and Dr Jeremy deWaard provided many pieces of advice. I am so grateful to Sean and Jayme Sones (and Delgado) for your kindness and hospitality (I never forget poutine!! :). I am also very grateful to Dr Reza Zahiri for sharing the BIO house with me and for your friendship. Paul, Jer, Jaclyn McKeown, Jayme, Patrick Strutzenberger and Bradley Zlotnick made my journey to Australia a memorable one. Finally, thanks to Dr Dirk Steinke for your great blog! Keep up the good work!

I am grateful to Dr Panu Välimäki and Marianne Välimäki for your hospitality during my trip to Oulu. Lastly, I would like to acknowledge following people for helping me in different ways: Graeme Cocks, Dr Doug Hilton, Dr Axel Kallies, Ian McMillan, Dr Scott Miller, Dr Andrew Mitchell, Dr Jadranka Rota, Andy Young and Dr Niklas Wahlberg.

This work was supported by Research Foundation of the University of Helsinki, Finnish Concordia Fund, and the Ella and Georg Ehrnrooth Foundation. Sequence analysis of DNA barcodes was enabled by funds from the government of Canada through Genome Canada and the Ontario Genomics Institute in support of the International Barcode of Life project.

Finally, my deepest gratitude goes to my family: my son Heikki Nieminen, common-law husband Dr Marko Nieminen, mother Mirja Kekkonen, father Hannu Kekkonen and Marko's daughters Neea and Ninette Metsäkylä. Thank you to all of you (and all four-legged furry friends) for your support and for reminding me that there is life outside of work as well. Kiitos kaikille tuestanne ja työn ulkopuolisen elämän olemassaolon muistuttamisesta! Hannu Kekkonen (a.k.a. Lào Yà) created the amazing artwork for the cover picture. Marko, you are my colleague, co-author, fellow worker, best friend and partner. Could I possibly ask for more? Thank you so much for everything!

REFERENCES

- Ahrens, D., Monaghan, M.T. & Vogler, A.P. 2007, DNA-based taxonomy for associating adults and larvae in multi-species assemblages of chafers (Coleoptera: Scarabaeidae), *Molecular phylogenetics and evolution*, 44, 436-449.
- Astrin, J.J., Huber, B.A., Bernhard, M. & Kluetsch, C.F.C. 2006, Molecular taxonomy in pholcid spiders (Pholcidae, Araneae): evaluation of species identification methods using CO1 and 16S rRNA, *Zoologica Scripta*, 35, 441-457.
- Astrin, J.J., Stueben, P.E., Misof, B., Waagele, J.W., Gimnich, F., Raupach, M.J. & Ahrens, D. 2012, Exploring diversity in cryptorhynchine weevils (Coleoptera) using distance-, character- and tree-based species delineation, *Molecular Phylogenetics and Evolution*, 63, 1-14.
- Avise, J.C. & Ball, R.M. Jr. 1990, Principles of genealogical concordance in species concepts and biological taxonomy, *Oxford Surveys in Evolutionary Biology*, 7, 45-67.
- Ballard, J.W.O. & Whitlock, M.C. 2004, The incomplete natural history of mitochondria, *Molecular Ecology*, 13, 729-744.
- Barberousse, A. & Samadi, S. 2010, Species from Darwin onward, *Integrative Zoology*, 5, 187-197.
- Barracough, T.G., Hughes, M., Ashford-Hodges, N. & Fujisawa, T. 2009a, Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data, *Biology Letters*, 5, 425-428.
- Barracough, T.G., Fontaneto, D., Herniou, E.A. & Ricci, C. 2009b, The evolutionary nature of diversification in sexuals and asexuals, 29-45. In R. Butlin, J. Bridle, D. Schluter (eds.), *Speciation and Patterns of Diversity*. Cambridge University Press, Cambridge.
- Barrett, C.F. & Freudenstein, J.V. 2011, An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid, *Molecular Ecology*, 20, 2771-2786.
- Barrett, R. & Hebert, P.D.N. 2005, Identifying spiders through DNA barcodes, *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 83, 481-491.
- Basset, Y., Novotny, V., Miller, S.E. & Pyle, R. 2000, Quantifying Biodiversity: Experience with Parataxonomists and Digital Photography in Papua New Guinea and Guyana, *BioScience*, 50, 899-908.
- Bauer, F., Stübner, A., Neinhuis, C. & Nuss, M. 2012, Molecular phylogeny, larval case architecture, host-plant associations and classification of European Coleophoridae (Lepidoptera), *Zoologica Scripta*, 41, 248-265.
- Bergmann, T., Hadrys, H., Breves, G. & Schierwater, B. 2009, Character-based DNA barcoding: a superior tool for species classification, *Berliner und Munchener tierärztliche Wochenschrift*, 122, 446-450.
- Bergsten, J. 2005, A review of long-branch attraction, *Cladistics*, 21, 163-193.
- Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barracough, T.G. & Vogler, A.P. 2012, The effect of geographical scale of sampling on DNA barcoding, *Systematic Biology*, 61, 851-869.
- Bichain, J., Boisselier-Dubayle, M., Bouchet, P. & Samadi, S. 2007, Species delimitation in the genus *Bythinella* (Mollusca : Caenogastropoda : Rissooidea): a first attempt combining molecular and morphometrical data, *Malacologia*, 49, 2, 293-311.

- Blaxter, M.L. 2004, The promise of a DNA taxonomy, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 359, 669–679.
- Bode, S.N.S., Adolfsson, S., Lamatsch, D.K., Martins, M.J.F., Schmit, O., Vandekerckhove, J., Mezquita, F., Namiotko, T., Rossetti, G., Schoen, I., Butlin, R.K. & Martens, K. 2010, Exceptional cryptic diversity and multiple origins of parthenogenesis in a freshwater ostracod, *Molecular Phylogenetics and Evolution*, 54, 542–552.
- Bortolus, A. 2008, Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology, *Ambio*, 37, 114–118.
- Boykin, L., Armstrong, K.L., Kubatko, L. & De Barro, P. 2012, Species delimitation and global biosecurity, *Evolutionary Bioinformatics*, 8, 1–37.
- Bromilow, S.M. & Sperling, F.A.H. 2011, Phylogeographic signal variation in mitochondrial DNA among geographically isolated grassland butterflies, *Journal of Biogeography*, 38, 299–310.
- Brown, S.D.J., Armstrong, K.F. & Cruickshank, R.H. 2012a, Molecular phylogenetics of a South Pacific sap beetle species complex (*Carpophilus* s, Coleoptera: Nitidulidae), *Molecular Phylogenetics and Evolution*, 64, 428–440.
- Bucheli, S.R. & Wenzel, J., 2005. Gelechioidea (Insecta: Lepidoptera) systematics: a reexamination using combined morphology and mitochondrial DNA data. *Molecular Phylogenetics and Evolution*, 35, 380–394.
- Bucheli, S.R. 2009, Annotated review and discussion of phylogenetically important characters for families and subfamilies of Gelechioidea (Insecta: Lepidoptera), *Zootaxa*, 1, 22.
- Butcher, B.A., Smith, M.A., Sharkey, M.J. & Quicke, D.L.J. 2012, A turbo-taxonomic study of Thai aleiodes (aleiodes) and aleiodes (arcaleiodes) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species, *Zootaxa*, 3457, 1–232.
- Cardoso, A. & Vogler, A. 2005, DNA taxonomy, phylogeny and Pleistocene diversification of the *Cicindela hybrida* species group (Coleoptera : Cicindelidae), *Molecular Ecology*, 14, 3531–3546.
- Carstens, B.C., Pelletier, T.A., Reid, N.M. & Satler, J.D. 2013, How to fail at species delimitation, *Molecular Ecology*, 22, 4369–4383.
- Ceccarelli, S.F., Sharkey, M.J. & Zaldivar-Riveron, A. 2012, Species identification in the taxonomically neglected, highly diverse, neotropical parasitoid wasp genus *Notiospathius* (Braconidae: Doryctinae) based on an integrative molecular and morphological approach, *Molecular Phylogenetics and Evolution*, 62, 485–495.
- Chanda, S., Sarethy, I.P., De, B. & Singh, K. 2013, *Paederia foetida* - a promising ethno-medicinal tribal plant of northeastern India, *Journal of Forestry Research*, 24, 801–808.
- Cho, S., Zwick, A., Regier, J.C., Mitter, C., Cummings, M.P., Yao, J., Du, Z., Zhao, H., Kawahara, A.Y., Weller, S., Davis, D.R., Baixeras, J., Brown, J.W., Parr, C., 2011, Can deliberately incomplete gene sample augmentation improve a phylogeny estimate for the advanced moths and butterflies (Hexapoda: Lepidoptera)? *Systematic Biology*, 60, 782–796.
- Clement, M., Posada, D. & Crandall, K. 2000, TCS: a computer program to estimate gene genealogies, *Molecular Ecology*, 9, 1657–1659.
- Collins, R.A., Boykin, L.M., Cruickshank, R.H. & Armstrong, K.F. 2012, Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification, *Methods in Ecology and Evolution*, 3, 457–465.
- Common, I.F.B. 1990, *Moths of Australia*. Melbourne University Press, Victoria.

REFERENCES

- Common, I.F.B. 1996, Hypertrophidae, 91–92. In E.S. Nielsen, E.D. Edwards, T.V. Rangsi (eds.), *Checklist of the Lepidoptera of Australia*. Monographs on Australian Lepidoptera. CSIRO, Melbourne.
- Coyne, J.A. & Orr, H.A. 2004, *Speciation*. Sinauer Associates, Sunderland, MA, U.S.A.
- Cracraft, J., 1983, Species concepts and speciation analysis, *Current Ornithology*, 159–187.
- Dalebout, M.L., Baker, C.S., Steel, D., Robertson, K.M., Chivers, S.J., Perrin, W.F., Mead, J.G., Grace, R.V. & Schofield, T.D., Jr. 2007, A divergent mtDNA lineage among Mesoplodon beaked whales: Molecular evidence for a new species in the tropical Pacific?, *Marine Mammal Science*, 23, 954–966.
- Darwin, C.R. 1859, *On the Origin of Species by Means of natural selection, or the preservation of favoured races in the struggle for life*, John Murray, London.
- Davis, J. & Nixon, K. 1992, Populations, Genetic-Variation, and the Delimitation of Phylogenetic Species, *Systematic Biology*, 41, 421–435.
- Dayrat, B. 2005, Towards integrative taxonomy, *Biological Journal of the Linnean Society*, 85, 407–415.
- de Queiroz, K. 1998, The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations 57–75. In D.J. Howard, S.H. Berlocher, (eds.), *Endless Forms. Species and Speciation*. Oxford University Press, Oxford.
- de Queiroz, K. 1999, The general lineage concept of species and the defining properties of the species category. 49–89. In R.A. Wilson (ed.), *Species. New Interdisciplinary Essays*. MIT Press, Cambridge, Massachusetts.
- de Queiroz, K. 2005a, Different species problems and their resolution, *BioEssays*, 27, 1263–1269.
- de Queiroz, K. 2005b, A Unified Concept of Species and Its Consequences for the Future of Taxonomy, *Proceedings of the California Academy of Sciences*, 56, 196–215.
- DeSalle, R., Egan, M.G. & Siddall, M. 2005, The unholy trinity: taxonomy, species delimitation and DNA, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 1905–1916
- Dinca, V., Lukhtanov, V.A., Talavera, G. & Vila, R. 2011, Unexpected layers of cryptic diversity in wood white Leptidea butterflies, *Nature Communications*, 2, 324.
- Dobzhansky, T. 1937, *Genetics and the Origin of Species*, Columbia University Press, New York.
- Donoghue, M. 1985, A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative, *Bryologist*, 88, 172–181.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J. & Rambaut, A. 2006, Relaxed phylogenetics and dating with confidence, *PloS Biology*, 4, 699–710.
- Drummond, A.J. & Rambaut, A. 2007, BEAST: Bayesian evolutionary analysis by sampling trees, *BMC Evolutionary Biology*, 7, 214.
- Dupuis, J.R., Roe, A.D. & Sperling, F.A.H. 2012, Multi-locus species delimitation in closely related animals and fungi: one marker is not enough, *Molecular Ecology*, 21, 4422–4436.
- Ebach, M.C. & Holdrege, C. 2005, DNA barcoding is no substitute for taxonomy, *Nature*, 434, 697.
- Ebach, M.C. 2011, Taxonomy and the DNA barcoding enterprise, *Zootaxa*, 2742, 67–68.

- Edwards, D.L. & Knowles, L.L. 2014, Species detection and individual assignment in species delimitation: can integrative data increase efficacy? *Proceedings of the Royal Society B-Biological Sciences*, 281, 20132765.
- Ehrlich, P.R. & Mooney, H.A. 1983, Extinction, substitution, and ecosystem services, *Bioscience*, 33, 248-254.
- Ehrlich, P.R. & Wilson, E.O. 1991, Biodiversity studies – science and policy, *Science*, 253, 758-762.
- Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J. & Jiggins, C.D. 2007, Limited performance of DNA barcoding in a diverse community of tropical butterflies, *Proceedings of the Royal Society B-Biological Sciences*, 274, 2881-2889.
- Esselstyn, J.A., Evans, B.J., Sedlock, J.L., Khan, F.A.A. & Heaney, L.R. 2012, Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats, *Proceedings of the Royal Society B-Biological Sciences*, 279, 3678-3686.
- Farris, J.S. 1974, Formal definitions of paraphyly and polyphyly, *Systematic Zoology*, 23, 548-554.
- Farris, J.S. 1981, Distance data in phylogenetic analysis, 3-23. In V.A. Funk, D.R. Brooks (eds.), *Advances in Cladistics*, 1, New York Botanical Garden, New York.
- Farris, J.S. 1983, The logical basis of phylogenetic analysis, 7-36. In N.I. Platnic, V.A. Funk (eds.), *Advances in Cladistics, Proceedings of the second meeting of the Willi Hennig Society*, volume 2, Columbia University Press, New York.
- Felsenstein, J. 2004, *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA, U.S.A.
- Ferguson, J. 2002, On the use of genetic divergence for identifying species, *Biological Journal of the Linnean Society*, 75, 509-516.
- Fernandez, R., Almodovar, A., Novo, M., Simancas, B. & Diaz Cosin, D.J. 2012, Adding complexity to the complex: New insights into the phylogeny, diversification and origin of parthenogenesis in the *Aporrectodea caliginosa* species complex (Oligochaeta, Lumbricidae), *Molecular Phylogenetics and Evolution*, 64, 368-379.
- Fontaneto, D., Herniou, E.A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C. & Barraclough, T.G. 2007, Independently evolving species in asexual bdelloid rotifers, *PloS Biology*, 5, 914-921.
- Footitt, R.G., Maw, H.E.L., Von Dohlen, C.D. & Hebert, P.D.N. 2008, Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes, *Molecular Ecology Resources*, 8, 1189-1201.
- Fujisawa, T. & Barraclough, T.G. 2013, Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evolution on simulated data sets, *Systematic Biology*, 62, 707-724.
- Funk, D.J. & Omland, K.E. 2003, Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA, *Annual Review of Ecology, Evolution, and Systematics*, 34, 397-423.
- Godfray, H.C.J. & Knapp, S. 2004, Introduction, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 359, 559-569.
- Goldstein, P. & DeSalle, R. 2000, Phylogenetic species, nested hierarchies, and character fixation, *Cladistics*, 16, 364-384.
- Goldstein, P., DeSalle, R., Amato, G. & Vogler, A. 2000, Conservation genetics at the species boundary, *Conservation Biology*, 14, 120-131.
- Gómez, A., Wright, P.J., Lunt, D.H., Cancino, J.M., Carvalho, G.R. & Hughes, R.N. 2007, Mating trials validate the use of DNA barcoding to reveal cryptic

REFERENCES

- speciation of a marine bryozoan taxon, *Proceedings of the Royal Society B-Biological Sciences*, 274, 199-207.
- Gompert, Z., Nice, C., Fordyce, J., Forister, M. & Shapiro, A. 2006, Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly, *Molecular Ecology*, 15, 1759-1768.
- Gregory, T.R. 2005, DNA barcoding does not compete with taxonomy, *Nature*, 434, 1067.
- Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B. & Hebert, P.D.N. 2006, A minimalist barcode can identify a specimen whose DNA is degraded, *Molecular Ecology Notes*, 6, 959-964.
- Hanner, R., Becker, S., Ivanova, N.V. & Steinke, D. 2011, FISH-BOL and seafood identification: Geographically dispersed case studies reveal systemic market substitution across Canada, *Mitochondrial DNA*, 22, 106-122.
- Hao, X., Jiang, R. & Chen, T. 2011, Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering, *Bioinformatics*, 27, 611-618.
- Hasegawa, M., Kishino, H., & Yano, T.A. 1985, Dating of the human ape splitting by a molecular clock of mitochondrial-DNA, *Journal of Molecular Evolution*, 22, 160-174.
- Hausmann, A., Haszprunar, G. & Hebert, P.D.N. 2011, DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions, *PLoS ONE*, 6, e17134.
- Hausmann, A., Godfray, H.C.J., Huemer, P., Mutanen, M., Rougerie, R., van Nieuwerkerken, E.J., Ratnasingham, S. & Hebert, P.D.N. 2013, Genetic patterns in European geometrid moths revealed by the Barcode Index Number (BIN) System, *PLoS ONE*, 8, e84518.
- Hebert, P.D.N., Cywinska, A., Ball, S. & DeWaard, J. 2003a, Biological identifications through DNA barcodes, *Proceedings of the Royal Society B-Biological Sciences*, 270, 313-321.
- Hebert, P.D.N., Ratnasingham, S. & deWaard, J. 2003b, Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species, *Proceedings of the Royal Society B-Biological Sciences*, 270, S96-S99.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. 2004a, Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*, *Proceedings of the National Academy of Sciences of the United States of America*, 101, 14812-14817.
- Hebert, P.D.N., Stoeckle, M., Zemlak, T. & Francis, C. 2004b, Identification of birds through DNA barcodes, *PloS Biology*, 2, 1657-1663.
- Hebert, P.D.N. & Gregory, T.R. 2005, The promise of DNA barcoding for taxonomy, *Systematic Biology*, 54, 852-859.
- Hebert, P.D.N., deWaard, J.R. & Landry, J. 2010, DNA barcodes for 1/1000 of the animal kingdom, *Biology Letters*, 6, 359-362.
- Hebert, P.D.N., deWaard, J.R., Zakharov, E.V., Prosser, S.W.J., Sones, J.E., McKeown, J.T.A., Mantle, B. & Le Salle, J. 2013, A DNA 'Barcode Blitz': rapid digitization and sequencing of a natural history collection, *PLoS ONE*, 8, e68535.
- Hendrixson, B.E., DeRussy, B.M., Hamilton, C.A. & Bond, J.E. 2013, An exploration of species boundaries in turret-building tarantulas of the Mojave Desert (Araneae, Mygalomorphae, Theraphosidae, Aphonopelma), *Molecular Phylogenetics and Evolution*, 66, 327-340.
- Hennig, W. 1966, *Phylogenetic Systematics*, University of Illinois Press, Urbana, U.S.A.

- Hey 2009, On the arbitrary identification of real species, 15-28. In R. Butlin, J. Bridle, D. Schluter (eds.), *Speciation and Patterns of Diversity*. Cambridge University Press, Cambridge.
- Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K. & Harrison, R.G. 2003, Understanding and confronting species uncertainty in biology and conservation, *Trends in Ecology and Evolution*, 18, 597-603.
- Hodges, R.W. 1998, The Gelechioidea, 131–158. In N.P. Kristensen (ed.), *Handbook of Zoology, Vol. IV, Arthropoda: Insecta, part 35, Lepidoptera, Moths and Butterflies*. Springer, Berlin.
- Hogg, I. & Hebert, P.D.N. 2004, Biological identification of springtails (Hexapoda : Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes, *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 82, 5, 749-754.
- Holland, B., Dawson, M., Crow, G. & Hofmann, D. 2004, Global phylogeography of *Cassiopea* (Scyphozoa : Rhizostomeae): molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands, *Marine Biology*, 145, 1119-1128.
- Hollingsworth, P.M. 2011, Refining the DNA barcode for land plants, *Proceedings of the National Academy of Sciences of the United States of America*, 108, 19451–19452.
- Huemer, P. & Karsholt, O 1999, *Gelechiidae I (Gelechiinae: Teleiodini, Gelechiini)*, 1-356. In P. Huemer, O. Karsholt, L. Lyneborg (eds.), *Microlepidoptera of Europe 3*. Apollo Books, Stenstrup.
- Huemer, P. & Karsholt, O. 2010, *Gelechiidae II (Gelechiinae: Gnorimoschemini)*, 1-586. In P. Huemer, O. Karsholt, M. Nuss (eds.), *Microlepidoptera of Europe 6*. Apollo Books, Stenstrup.
- Huemer, P. & Hebert, P.D.N. 2011, Cryptic diversity and phylogeography of high alpine *Sattleria*—a case study combining DNA barcodes and morphology (Lepidoptera: Gelechiidae), *Zootaxa*, 2981: 1–22.
- Huemer, P. & Mutanen, M. 2012, Taxonomy of spatially disjunct alpine *Teleiopsis albifemorella* sensu auctt. revealed by molecular and morphological data - how many species are there? *Zootaxa*, 3580, 1–23.
- Janzen, D.H. 2004, Setting up tropical biodiversity for conservation through non-damaging use: participation by parataxonomists, *Journal of Applied Ecology*, 41, 181-187.
- Janzen, D.H. & Hallwachs, W. 2011, Joining inventory by parataxonomists with DNA barcoding of a large complex tropical conserved wildland in northwestern Costa Rica, *PLoS ONE*, 6, e18123.
- Jones, M., Ghoorah, A. & Blaxter, M. 2011, jMOTU and Taxonator: Turning DNA barcode sequences into annotated operational taxonomic units, *PLoS ONE*, 6, e19259.
- Jukes, T.H. & Cantor, C.R. 1969, Evolution of protein molecules, 21-132. In N.H. Munro (ed.), *Mammalian Protein Metabolism*, Academic Press, New York.
- Juslén, A, Kuusinen, M., Muona, J., Siitonen, J. & Toivonen, H. (eds.) 2008, *Research program of deficiently known and threatened forest species—the final report*, Finnish Environment, 1, 1–146.
- Jörger, K.M., Norenburg, J.L., Wilson, N.G. & Schrödl, M. 2012, Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs, *BMC Evolutionary Biology*, 12, 245.
- Kaartinen, R., Stone, G. N., Hearn, J., Lohse, K. & Roslin, T. 2010, Revealing secret liaisons: DNA barcoding changes our understanding of food webs, *Ecological Entomology*, 35, 623-638.
- Kaila, L. 2004, Phylogeny of the superfamily Gelechioidea (Lepidoptera: Ditrysia): an exemplar approach, *Cladistics*, 20, 303–340.

REFERENCES

- Kaila, L. 2011, *Elachistine Moths of Australia: (Lepidoptera:Gelechioidea: Elachistidae)*. In Monographs on Australian Lepidoptera. CSIRO Publishing, Collingwood, 1–443.
- Kaila, L. & Stahls, G. 2006, DNA barcodes: Evaluating the potential of COI to differentiate closely related species of *Elachista* (Lepidoptera : Gelechioidea : Elachistidae) from Australia, *Zootaxa*, 1170, 1-26.
- Kaila, L., Mutanen, M. & Nyman, T., 2011. Phylogeny of the megadiverse Gelechioidea (Lepidoptera): adaptations and determinants of success, *Molecular Phylogenetics and Evolution*, 61, 801–809.
- Karsholt, O., Mutanen, M., Lee, S. & Kaila, L., 2013, A molecular analysis of the Gelechiidae (Lepidoptera, Gelechioidea) with an interpretative grouping of its taxa, *Systematic Entomology*, 38, 334–348.
- Kerr, K.C.R., Birks, S.M., Kalyakin, M.V., Redkin, Y.A., Koblik, E.A. & Hebert, P.D.N. 2009, Filling the gap - COI barcode resolution in eastern Palearctic birds, *Frontiers in Zoology*, 6, 29.
- Kingman, J.F.C. 1982, The coalescent, *Stochastic Processes and their Applications*, 13, 235–248.
- Kluge, A.G. & Wolf, A.J. 1993, Cladistics – Whats in a word? *Cladistics*, 9, 183–199.
- Kluge, A.G. 1989, A concern for evidence and a phylogenetic hypothesis of relationships among epicrates (Boidae, Serpentes), *Systematic Zoology*, 38, 7–25.
- Krell, F.-T. 2004, Parataxonomy vs. taxonomy in biodiversity studies – pitfalls and applicability of ‘morphospecies’ sorting, *Biodiversity and Conservation*, 13, 795–812.
- Knowles, L.L. & Carstens, B.C. 2007, Delimiting species without monophyletic gene trees, *Systematic Biology*, 56, 887–895.
- Lahaye, R., Van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G. & Savolainen, V. 2008, DNA barcoding the floras of biodiversity hotspots, *Proceedings of the National Academy of Sciences of the United States of America*, 105, 2923–2928.
- Lane R. 2009, Taxonomy comes of age, *Nature*, 461, 597.
- Lee, M.S.Y. 2004, The molecularisation of taxonomy, *Invertebrate Systematics*, 18, 1–6.
- Lee, S. & Brown, R.L. 2008, Phylogenetic relationships of Holarctic Teleiodini (Lepidoptera: Gelechiidae) based on analysis of morphological and molecular data, *Systematic Entomology*, 33, 595–612.
- Leliaert, F., Verbruggen, H., Wysor, B. & De Clerck, O. 2009, DNA taxonomy in morphologically plastic taxa: Algorithmic species delimitation in the Boodlea complex (Chlorophyta: Cladophorales), *Molecular Phylogenetics and Evolution*, 53, 122–133.
- Lewis, P.O. 2001, A likelihood approach to estimating phylogeny from discrete morphological character data, *Systematic Biology*, 50, 913–925.
- Li J. 2009, China searches for an 11th-hour lifesaver for a dying discipline, *Science*, 325, 31.
- Lim, G.S., Balke, M. & Meier, R. 2012, Determining species boundaries in a world full of rarity: Singletons, species delimitation methods, *Systematic Biology*, 61, 165–169.
- Lipscomb, D. 1998. *Basics of cladistics analysis*. George Washington University, Washington D.C.
- Lipscomb, D., Platnick, N. & Wheeler, Q. 2003, The intellectual content of taxonomy: a comment on DNA taxonomy, *Trends in Ecology and Evolution*, 18, 64–66.

- Locke, J.M. & Coates, K.A. 2008, What are the costs of bad taxonomic practices: and what is *Madracis mirabilis*? *Proceedings of the 11th International Coral Reef Symposium*, 1348-1351.
- Lohse, K. 2009, Can mtDNA barcodes be used to delimit species? A response to Pons *et al.* (2006), *Systematic Biology*, 58, 439-441.
- Luo, J.-y., Yan, D., Song, J.-y., Zhang, D., Xing, X.-y., Han, Y.-m., Yang, M.-h., Dong, X.-p., Peng, C., Chen, S.-l. & Xiao, X.-h. 2014, A strategy for trade monitoring and substitution of the organs of threatened animals, *Scientific Reports*, 3, 3108.
- Mace, G.M. 2004, The role of taxonomy in species conservation, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 359, 711-719.
- Mallet, J. 1995, A species definition for the Modern Synthesis, *Trends in Ecology and Evolution*, 10, 294-299.
- Mallet 2009, Rapid speciation, hybridization and adaptive radiation in the *Heliconius melpomene* group, 177-194. In R. Butlin, J. Bridle & D. Schluter (eds), *Speciation and Patterns of Diversity*, Cambridge University Press, Cambridge.
- Mallet, J. & Willmott, K. 2003, Taxonomy: renaissance or Tower of Babel? *Trends in Ecology and Evolution*, 18, 57-59.
- Mallet, J., Isaac, N.J.B. & Mace, G.M. 2005, Response to Harris and Froufe, and Knapp *et al.*: Taxonomic inflation, *Trends in Ecology and Evolution*, 20, 8-9.
- Mayr, E. 1940, Speciation phenomena in birds, *American Naturalist*, 74, 249-278.
- Meier, R. 2008, DNA Sequences in Taxonomy: Opportunities and Challenges, 95-126. In Q.D. Wheeler (ed.) *The New Taxonomy*. CRC Press, Boca Raton.
- Meier, R. & Dikow, T. 2004, Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data, *Conservation Biology*, 18, 478-488.
- Meier, R. & Willman, R. 2000, The Hennigian species concept 30-43. In Q.D. Wheeler & R. Meier (eds.), *Species Concepts and Phylogenetic Theory*, Columbia University Press, New York.
- Meier, R., Shiyang, K., Vaidya, G. & Ng, P.K.L. 2006, DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success, *Systematic Biology*, 55, 715-728.
- Meier, R., Zhang, G. & Ali, F. 2008, The use of mean instead of smallest interspecific distances exaggerates the size of the barcoding gap and leads to misidentification, *Systematic Biology*, 57, 809-813.
- Meiklejohn, K.A., Wallman, J.F. & Dowton, M. 2011, DNA-based identification of forensically important Australian Sarcophagidae (Diptera), *International Journal of Legal Medicine*, 125, 27-32.
- Meusnier, I., Singer, G.A., Landry, J.-F., Hickey, D.A., Hebert, P.D.N. & Hajibabaei, M. 2008, A universal DNA mini-barcode for biodiversity analysis, *BMC Genomics*, 9, 214.
- Meyer, C. & Paulay, G. 2005, DNA barcoding: Error rates based on comprehensive sampling, *PloS Biology*, 3, 2229-2238.
- Michener, C. & Sokal, R. 1957, A quantitative approach to a problem in classification, *Evolution*, 11, 130-162.
- Mikkelsen, N.T., Schander, C. & Willassen, E. 2007, Local scale DNA barcoding of bivalves (Mollusca): a case study, *Zoologica Scripta*, 36, 455-463.
- Minet, J. 1991, Tentative reconstruction of the ditrysian phylogeny (Lepidoptera, Glossata), *Entomologica Scandinavica*, 22, 69-95.

REFERENCES

- Miralles, A. & Vences, M. 2013, New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards, *PLoS ONE*, 8, e68242.
- Mishler, B. 1985, The morphological, developmental, and phylogenetic basis of species concepts in Bryophytes, *Bryologist*, 88, 207-214.
- Mishler, B. & Donoghue, M. 1982, Species concepts - a case for pluralism, *Systematic zoology*, 31, 491-503.
- Monaghan, M., Balke, M., Gregory, T. & Vogler, A. 2005, DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 1925-1933.
- Monaghan, M.T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D.J.G., Lees, D.C., Ranaivosolo, R., Eggleton, P., Barraclough, T.G. & Vogler, A.P. 2009, Accelerated species inventory on Madagascar using coalescent-based models of species delineation, *Systematic Biology*, 58, 298-311.
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. 2011, How many species are there on Earth and in the Ocean? *PLoS ONE*, 9, e1001127.
- Moritz, C. & Cicero, C. 2004, DNA barcoding: promise and pitfalls, *PloS Biology*, 2, 1529-1531.
- Mutanen, M., Wahlberg, N. & Kaila, L. 2010, Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies, *Proceedings of the Royal Society B-Biological Sciences*, 277, 2839-2848.
- Mutanen, M., Kaila, L. & Tabell, J. 2013, Wide-ranging barcoding aids discovery of one-third increase of species richness in presumably well-investigated moths, *Scientific Reports*, 3, 2901.
- Nee, S., Mooers, A. & Harvey, P. 1992, Tempo and mode of evolution revealed from molecular phylogenies, *Proceedings of the National Academy of Sciences of the United States of America*, 89, 8322-8326.
- Nelson, G.J. & Platnick, N.I. 1981, *Systematics and biogeography: Cladistics and vicariance*, Columbia University Press, New York.
- Nielsen, E.S., Edwards, D.H. & Rangsi, T.V. 1996. *Checklist of the Lepidoptera of Australia*. In Monographs on Australian Lepidoptera. CSIRO Publishing, Collingwood.
- Nixon, K.C. & Carpenter, J.M. 1993, On outgroups, *Cladistics*, 9, 413-426.
- Nixon, K. & Wheeler, Q. 1990, An amplification of the phylogenetic species concept, *Cladistics*, 6, 211-223.
- Obertegger, U., Fontaneto, D. & Flaim, G. 2012, Using DNA taxonomy to investigate the ecological determinants of plankton diversity: explaining the occurrence of *Synchaeta* s (Rotifera, Monogononta) in mountain lakes, *Freshwater Biology*, 57, 1545-1553.
- Pamilo, P. & Nei, M. 1988, Relationships between gene trees and species trees, *Molecular Biology and Evolution*, 5, 568-583.
- Pantaleoni, R.A. & Badano, D. 2012, *Myrmeleon punicanus* n. sp., a new pit-building antlion (Neuroptera Myrmeleontidae) from Sicily and Pantelleria, *Bulletin of Insectology*, 65, 1, 139-148.
- Papadopoulou, A., Bergsten, J., Fujisawa, T., Monaghan, M.T., Barraclough, T.G. & Vogler, A.P. 2008, Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 363, 2987-2996.
- Paquin, P. & Hedin, M. 2004, The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Cicurina* (Araneae : Dictynidae) from Texas caves, *Molecular Ecology*, 13, 3239-3255.

- Parducci, L., Matetovici, I., Fontana, S.L., Bennett, K.D., Suyama, Y., Haile, J., Kjaer, K.H., Larsen, N.K., Drouzas, A.D. & Willerslev, E. 2013, Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia, *Molecular Ecology*, 22, 3511-3524.
- Paz, A. & Crawford, A.J. 2012, Molecular-based rapid inventories of sympatric diversity: A comparison of DNA barcode clustering methods applied to geography-based vs clade-based sampling of amphibians, *Journal of Biosciences*, 37, 887-896.
- Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. 2006, Sequence-based species delimitation for the DNA taxonomy of undescribed insects, *Systematic Biology*, 55, 4, 595-609.
- Pons, J., Fujisawa, T., Claridge, E.M., Savill, R.A., Barraclough, T.G. & Vogler, A.P. 2011, Deep mtDNA subdivision within Linnean species in an endemic radiation of tiger beetles from New Zealand (genus *Neocicindela*), *Molecular Phylogenetics and Evolution*, 59, 251-262.
- Powell, J.R., Monaghan, M.T., Oepik, M. & Rillig, M.C. 2011, Evolutionary criteria outperform operational approaches in producing ecologically relevant fungal species inventories, *Molecular Ecology*, 20, 655-666.
- Prendini, L. 2005, Comment on "Identifying spiders through DNA barcodes", *Canadian Journal of Zoology – Revue Canadienne de Zoologie*, 83, 498-504.
- Prévot, V., Jordaens, K., Sonet, G. & Backeljau, T. 2013, Exploring species level taxonomy and species delimitation methods in the facultatively self-fertilizing land snail genus *Rumina* (Gastropoda: Pulmonata), *PLoS ONE*, 8, e60736.
- Puillandre, N., Macpherson, E., Lambourdière, J., Cruaud, C., Boisselier-Dubayle, M. & Samadi, S. 2011, Barcoding type specimens helps to identify synonyms and an unnamed new species in *Eumunida* Smith, 1883 (Decapoda: Eumunididae), *Invertebrate Systematics*, 25, 322-333.
- Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. 2012a, ABGD, Automatic Barcode Gap Discovery for primary species delimitation, *Molecular Ecology*, 21, 1864-1877.
- Puillandre, N., Modica, M.V., Zhang, Y., Sirovich, L., Boisselier, M.-C., Cruaud, C., Holford, M. & Samadi, S. 2012b, Large-scale species delimitation method for hyperdiverse groups, *Molecular Ecology*, 21, 2671-2691.
- Quicke, D. 2004, The world of DNA barcoding and morphology – collision or synergism and what of the future? *The Systematist*, 23, 8-11.
- Rassi, P., Hyvärinen, E., Juslén, A. & Mannerkoski, I. (eds.) 2010, *Suomen lajien uhanalaisuus 2010*, Ympäristöministeriö ja Suomen ympäristökeskus.
- Ratnasingham, S. & Hebert, P.D.N. 2007, BOLD: The Barcode of Life Data System (www.barcodinglife.org), *Molecular Ecology Notes*, 7, 355-364.
- Ratnasingham, S. & Hebert, P.D.N. 2013, A DNA-based registry for all animal species: the Barcode Index Number (BIN) system, *PLoS ONE*, 8, e66213.
- Regier, J.C., Mitter, C., Solis, M.A., Hayden, J.E., Landry, B., Nuss, M., Simonsen, T.J., Yen, S.-H., Zwick, A. & Cummings, M.P. 2012 A molecular phylogeny for the pyraloid moths (Lepidoptera: Pyraloidea) and its implications for higher-level classification, *Systematic Entomology*, 37, 635-656.
- Regier, J.C., Mitter, C., Zwick, A., Bazinet, A.L., Cummings, M.P., Kawahara, A.Y., Sohn, J.C., Zwickl, D.J., Cho, S., Davis, D.R., Baixeras, J., Brown, J., Parr, C., Weller, S., Lees, D.C. & Mitter, K.T. 2013, A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (Moths and Butterflies), *PLoS ONE*, 8, e58568.

REFERENCES

- Reilly, S.B., Marks, S.B. & Jennings, W.B. 2012, Defining evolutionary boundaries across parapatric ecomorphs of Black Salamanders (*Aneides flavipunctatus*) with conservation implications, *Molecular Ecology*, 21, 5745–5761.
- Riedel, A., Sagata, K., Surbakti, S., Tänzler, R. & Balke, M. 2013a, One hundred and one new species of trigonopterus weevils from New guinea, *Zookeys*, 280, 1–150.
- Riedel, A., Sagata, K., Suhardjono, Y.R., Tänzler, R. & Balke, M. 2013b, Integrative taxonomy on the fast track – towards more sustainability in biodiversity research, *Frontiers in Zoology*, 10, 15.
- Rodman, J.E. & Cody, J.H. 2003, The taxonomic impediment overcome: NSF's partnerships for enhancing expertise in taxonomy (PEET) as a model, *Systematic Biology*, 52, 428–435.
- Roe, A.D. & Sperling, F.A.H. 2007, Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach, *Molecular Ecology*, 16, 3617–3633.
- Ronquist, F. & Gärdenfors, U. 2003, Taxonomy and biodiversity inventories: time to deliver, *Trends in Ecology and Evolution*, 18:269–270.
- Rosen, D.E. 1979, Fishes from the uplands and intermontane basins of Guatemala : revisionary studies and comparative geography, *Bulletin of the American Museum of Natural History*, 162, article 5.
- Rosso, J.J., Mabragaña, E., Gonzáles Castro, M. & Díaz de Astarloa, J.M. 2012, DNA barcoding Neotropical fishes: recent advances from the Pampa Plain, Argentina, *Molecular Ecology Resources*, 12, 999–1011.
- Saitou, N. & Nei, M. 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution*, 4, 406–425.
- Santschi, L., Hanner, R.H., Ratnasingham, S., Riconscente, M. & Imondi, R 2013, Barcoding life's matrix: translating biodiversity genomics into high school settings to enhance life science education, *PloS Biology*, 11, e1001471.
- Sarkar, I., Thornton, J., Planet, P., Figurski, D., Schierwater, B. & DeSalle, R. 2002, An automated phylogenetic key for classifying homeoboxes, *Molecular Phylogenetics and Evolution*, 24, 3, 388–399.
- Sarkar, I.N., Planet, P.J. & Desalle, R. 2008, CAOS software for use in character-based DNA barcoding, *Molecular Ecology Resources*, 8, 6, 1256–1259.
- Sauer, J. & Hausdorf, B. 2012, A comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy, *Cladistics*, 28, 300–316.
- Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K. & Lane, R. 2005, Towards writing the encyclopaedia of life: an introduction to DNA barcoding, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 1805–1811.
- Schindel, D.E. & Miller, S.E. 2005, DNA barcoding a useful tool for taxonomists, *Nature*, 435, 17.
- Schmidt, P.-A., Bálint, M., Greshake, B., Bandowa, C., Römbke, J. & Schmitt, I. 2013, Illumina metabarcoding of a soil fungal community, *Soil Biology & Biochemistry*, 65, 128–132.
- Schoch, C.L. *et al.* 2012, Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi, *Proceedings of the National Academy of Sciences of the United States of America*, 6241–6246.
- Schuh, R.T. 2000, *Biological Systematics: principles and applications*, 1–236. Cornell University Press, New York.
- Seberg, O. 2004, The future of systematics: assembling the tree of life, *The Systematist, Newsletter of the Systematics Association*, 2–8, 23.

- Shokralla, S., Zhou, X., Janzen, D.H., Hallwachs, W., Landry, J.-F., Jacobus, L.M., & Hajibabaei, M. 2011, Pyrosequencing for mini-barcoding of fresh and old museum specimens, *PLoS ONE*, 7, e21252.
- Silva-Brandão, K.L., Azeredo-Espin, A.M.L. & Freitas, A.V.L. 2008, New evidence on the systematic and phylogenetic position of *Parides burchellanus* (Lepidoptera: Papilionidae), *Molecular Ecology Resources*, 8, 502–511.
- Simpson, G.G. 1961, *Principles of animal taxonomy*, Columbia University Press, New York.
- Sirovich, L., Stoeckle, M. & Zhang, Y. 2009, A scalable method for analysis and display of DNA sequences, *PLoS ONE*, 4, e7051.
- Sites, J.W. jr. & Crandall, K.A. 1997, Testing species boundaries in biodiversity studies, *Conservation Biology*, 11, 1289–1297.
- Sites, J. & Marshall, J. 2003, Delimiting species: a renaissance issue in systematic biology, *Trends in Ecology & Evolution*, 18, 462–470.
- Sites, J. & Marshall, J. 2004, Operational criteria for delimiting species, *Annual Review of Ecology Evolution and Systematics*, 35, 199–227.
- Soininen, E.M., Zinger, L., Gielly, L., Bellermain, E., Bråthen, K.A., Brochmann, C., Epp, L.S., Gussarova, G., Hassel, K., Henden, J.-A., Killengreen, S.T., Rämä, T., Stenøien, H.K., Yoccoz, N.G. & Ims, R.A. 2013, Shedding new light on the diet of Norwegian lemmings: DNA metabarcoding of stomach content, *Polar Biology*, 36, 1069–1076.
- Sota, T. & Sasabe, M. 2006, Utility of nuclear allele networks for the analysis of closely related species in the genus *Carabus*, subgenus *Ohomopterus*, *Systematic Biology*, 55, 329–344.
- Srivathsan, A. & Meier, R. 2012, On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature, *Cladistics*, 28, 190–194.
- Stoeckle, M.Y. & Hebert, P.D.N. 2008, Barcode of life, *Scientific American*, 299.
- Stoeckle, M.Y., Gamble, C.C., Kirpekar, R., Young, G., Ahmed, S. & Little, D.P. 2011, Commercial teas highlight plant DNA barcode identification successes and obstacles, *Scientific Reports*, 1, 42.
- Strimmer, K. & von Haeseler, A. 2009, Genetic distances and nucleotide substitution models, 111–125. In P. Lemey, M. Salemi, A.-M. Vandamme (eds.), *The Phylogenetic Handbook*. Cambridge University Press, Cambridge.
- Strutzenberger, P., Brehm, G. & Fiedler, K. 2012, DNA barcode sequencing from old type specimens as a tool in taxonomy: a case study in the diverse genus *Eois* (Lepidoptera: geometridae), *PLoS ONE*, 7, e49710.
- Tajima, F. 1983, Evolutionary relationship of Dna-sequences in finite populations, *Genetics*, 105, 437–460.
- Talavera, G., Dinca, V. & Vila, R. 2013, Factors affecting species delimitations with the GMYC model: insights from a butterfly survey, *Methods in Ecology and Evolution*, 4, 1101–1110.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. 2002, DNA points the way ahead in taxonomy, *Nature*, 418, 479.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. 2003, A plea for DNA taxonomy, *Trends in Ecology and Evolution*, 18, 71–74.
- Templeton, A., Crandall, K. & Sing, C. 1992, A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and Dna-sequence data.3. Cladogram estimation, *Genetics*, 132, 619–633.
- Thomsen, P., Kielgast, J., Iversen, L., Møller, P., Rasmussen, M. & Willerslev, E. 2012, Detection of a diverse marine fish fauna using environmental DNA from seawater samples, *PLoS ONE*, 7, e41732.

REFERENCES

- Traugott-Olsen, E. 1988, The *Elachista triseriatella* Stainton complex, with descriptions of eight new species (Lepidoptera: Elachistidae), *Entomologist's Gazette*, 39, 293–312.
- Traugott-Olsen, E. 1992, The *Elachista dispunctella* (Duponchel, 1843) complex with descriptions of new taxa (Lepidoptera: Elachistidae), *SHILAP Revista de lepidopterologica*, 20, 197–316.
- Tänzler, R., Sagata, K., Surbakti, S., Balke, M. & Riedel, A. 2012, DNA Barcoding for community ecology - how to tackle a hyperdiverse, mostly undescribed Melanesian fauna, *PLoS ONE*, 7, e28832.
- Wahlberg, N. & Wheat, C.W. 2008, Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera, *Systematic Biology*, 57, 231–242.
- van Nieuwerkerken, E.J., Kaila, L., Kitching, I.J., Kristensen, N.P., Lees, D.C., Minet, J., Mitter, C., Mutanen, M., Regier, J.C., Simonsen, T.J., Wahlberg, N., Yen, S.-H., Zahiri, R., Adamski, D., Baixeras, J., Bartsch, D., Bengtsson, B.Å., Brown, J.W., Bucheli, S.R., Davis, D.R., De Prins, J., De Prins, W., Epstein, M.E., Gentili-Poole, P., Gielis, C., Hattenschwiler, P., Hausmann, A., Holloway, J.D., Kallies, A., Karsholt, O., Kawahara, A.Y., Koster, J.C., Kozlov, M.V., Lafontaine, J.D., Lamas, G., Landry, J.-F., Lee, S., Nuss, M., Park, K.-T., Penz, C., Rota, J., Schmidt, B.C., Schintlmeister, A., Sohn, J.-S., Solis, M.A., Tarmann, G.M., Warren, A.D., Weller, S., Yakovlev, R.V., Zolotuhin, V.V. & Zwick, A. 2011, Order lepidoptera. In Zhang, Z.-Q. (Ed.), *Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness*, *Zootaxa*, 3148, 212–221.
- Vandamme, A.-M. 2009, Basic concepts of molecular evolution, 3–30. In P. Lemey, M. Salemi, A.-M. Vandamme (eds.), *The Phylogenetic Handbook*. Cambridge University Press, Cambridge.
- Weigand, A.M., Jochum, A., Slapnik, R., Schnitzler, J., Zarza, E. & Klussmann-Kolb, A. 2013, Evolution of microgastropods (Ellobioidea, Carychiidae): integrating taxonomic, phylogenetic and evolutionary hypotheses, *BMC Evolutionary Biology*, 13, 18.
- Velasco, J.D. 2009, When monophyly is not enough: exclusivity as the key to defining a phylogenetic species concept, *Biology and Philosophy*, 24, 473–486.
- Vences, M., Thomas, M., Bonett, R. & Vieites, D. 2005, Deciphering amphibian diversity through DNA barcoding: chances and challenges, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 360, 1859–1868.
- Vrana, P. & Wheeler, W. 1992, Individual organisms as terminal entities - Laying the species problem to rest, *Cladistics*, 8, 67–72.
- Vuataz, L., Sartori, M., Wagner, A. & Monaghan, M.T. 2011, Toward a DNA taxonomy of alpine *Rhithrogena* (Ephemeroptera: Heptageniidae) using a mixed Yule-coalescent analysis of mitochondrial and nuclear DNA, *PLoS ONE*, 6, e19728.
- Wheeler, Q.D. & Meier, R. (eds.) 2004, *Species Concepts and Phylogenetic Theory*, Columbia University Press, New York.
- Wheeler, W.C. 2012, *Systematics. A Course of Lectures*, Wiley-Blackwell, West Sussex, U.K.
- Will, K.W. & Rubinoff, D. 2004, Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification, *Cladistics*, 20, 47–55.
- Will, K.W., Mishler, B.D. & Wheeler, Q. D. 2005, The perils of DNA barcoding and the need for integrative taxonomy, *Systematic Biology*, 54, 844–851.
- Wirta, H.K., Hebert, P.D.N., Kaartinen, R., Prosser, S.W., Várkonyi, G. & Roslin, T. 2014, Complementary molecular information changes our perception of food web structure, *Proceedings of the National Academy of Sciences of the United*

- States of America*, Early edition,
www.pnas.org/cgi/doi/10.1073/pnas.1316990111.
- Wiemers, M. & Fiedler, K. 2007, Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae), *Frontiers in Zoology*, 4, article 8.
- Wiens, J.J. & Penkrot, T.A. 2002, Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*), *Systematic Biology*, 51, 69-91. .
- Wiley, E.O. 1978, The evolutionary species concept reconsidered, *Systematic Zoology*, 27, 88-92.
- Wiley, E.O. & Mayden, R.L. 2000, The evolutionary species concept, 70-89. In Q.D. Wheeler, R. Meier (eds.), *Species Concepts and Phylogenetic Theory*, Columbia University Press, New York.
- Williams, P.H., Brown, M.J.F., Carolan, J.C., An, J., Goulson, D., Aytekin, A.M., Best, L.R., Byvaltsev, A.M., Cederberg, B., Dawson, R., Huang, J., Ito, M., Monfared, A., Raina, R.H., Schmid-Hempel, P., Sheffield, C.S., Sima, P. & Xie, Z. 2012, Unveiling cryptic species of the bumblebee subgenus *Bombus* s. str. worldwide with COI barcodes (Hymenoptera: Apidae), *Systematics and Biodiversity*, 10, 21-56.
- Wright, S. 1940, The statistical consequences of Mendelian heredity in relation to speciation, 161-184. In J. Huxley (ed.), *The new systematics*, Oxford University Press, London.
- Xiang, X., Xi, Y., Wen, X., Zhang, G., Wang, J. & Hu, K. 2011, Patterns and processes in the genetic differentiation of the *Brachionus calyciflorus* complex, a passively dispersing freshwater zooplankton, *Molecular Phylogenetics and Evolution*, 59, 386-398.
- Yang, Z. 1994, Estimating the pattern of nucleotide substitution, *Journal of Molecular Evolution*, 39, 105-111.
- Yang, Z. & Rannala, B. 2010, Bayesian species delimitation using multilocus sequence data, *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9264-9269.
- Yassin, A., Amedegnato, C., Cruaud, C. & Veuille, M. 2009, Molecular taxonomy and species delimitation in Andean *Schistocerca* (Orthoptera: Acrididae), *Molecular Phylogenetics and Evolution*, 53, 404-411.
- Yeates, D.K. 1995, Groundplans and exemplars: Paths to the tree of life, *Cladistics*, 11, 343-357.
- Yoccoz, N.G., Brathen, K.A., Gielly L., Haile, J., Edwards, M.E., Goslar, T., Von Stedingk, H., Brysting, A.K., Coissac, E., Pompanon, F., Sønstebo, J.H., Miquel, C., Valantini, A., De Bello, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory, F., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Brochmann, C., Willerslev, E. & Taberlet, P. 2012, DNA from soil mirrors plant taxonomic and growth form diversity, *Molecular Ecology*, 21, 3647-3655.
- Yule, G.U. 1924, A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, FRS, *Philosophical Transactions of the Royal Society B-Biological Sciences*, 213, 21-87.
- Zahiri, R., Kitching, I.J., Lafontaine, J.D., Mutanen, M., Kaila, L., Holloway, J.D. & Wahlberg, N. 2011, A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera), *Zoologica Scripta*, 40, 158-173.
- Zahiri, R., Holloway, J.D., Kitching, I.J., Lafontaine, J.D., Mutanen, M. & Wahlberg, N. 2012, Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea), *Systematic Entomology*, 37, 102-124.

REFERENCES

- Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. 2013, A general species delimitation method with applications to phylogenetic placements, *Bioinformatics*, 29, 2869-2876.